Genome expression dynamics reveals parasitism regulatory landscape of the root-knot nematode *Meloidogyne incognita* and a promoter motif associated with effector genes

Running title:
Concerted expression and a putative cis regulatory element associated with effector gene expression in *Meloidogyne spp*.

Authors

Martine Da Rocha*[1], Caroline Bournaud*[2], Julie Dazenière[1], Peter Thorpe[3], Clement Pellegrin[2], Marc Bailly-Bechet[1], Arthur Péré[1], Priscila Grynberg[4], Laetitia Perfus-Barbeoch[1], Sebastian Eves-van den Akker*[2] and Etienne GJ Danchin*[1].

*: equal contribution

Affiliations

[1] INRAE, Université Côte d'Azur, CNRS, Institut Sophia Agrobiotech, Sophia Antipolis, France
[2] Department of Plant Sciences, University of Cambridge, UK, CB2 3EA
[3] University of St. Andrews, School of Medicine, Medical & Biological Sciences, North Haugh, St Andrews, United Kingdom
[4] Embrapa Genetic Resources and Biotechnology, Brasília DF 70770-917, Brazil

Abstract

Root-knot nematodes are the major contributor to the crop losses caused by nematodes. Root-knot nematodes secrete effectors into the plant, derived from two sets of pharyngeal gland cells, to manipulate host physiology and immunity. Successful completion of the life cycle, involving successive molts from egg to adult, covers morphologically and functionally distinct stages and will require precise control of gene expression, including effectors. The details of how root-knot nematodes regulate transcription remain sparse.

Here, we report a life stage-specific transcriptome of *Meloidogyne incognita*. Combined with an available annotated genome, we explore the spatio-temporal regulation of gene expression. We reveal gene expression clusters and predicted functions that accompany the major developmental transitions. Focusing on effectors, we identify a putative cis-regulatory motif associated with expression in the dorsal glands: providing an insight into effector regulation. We combine the presence of this motif with several other criteria to predict a novel set of putative dorsal gland effectors. Finally, we show this motif, and thereby its utility, is broadly conserved across the *Meloidogyne* genus and termed it Mel-DOG.

Taken together, we provide the first genome-wide analysis of spatio-temporal gene expression in a root-knot nematode, and identify a new set of candidate effector genes that will guide future functional analyses.

Keywords

RNAseq, transcriptome, spatio-temporal expression patterns, effector prediction, promoter motif.

Introduction

Plant parasitic nematodes are microscopic worms that threaten the security of most major agricultural crops such as soybean, cotton, peanut, banana, coffee, and potato [1,2]. Damage caused by nematodes therefore represent an important constraint on global food security in the developed and developing world, and are estimated to cost world agriculture 173 billion US dollar per year [3]. The majority of these loses are attributed to the obligate biotrophic sedentary endoparasites: the root-knot nematodes (RKN - *Meloidogyne spp.)* and the cyst nematodes (CN, *Heterodera spp*. and *Globodera spp.*).

Root-knot nematodes are considered obligate biotrophic endoparasites, as they spend the majority of their life cycle within host roots and feed from living tissues. The typical RKN life cycle is complex and usually completed in 4-6 weeks, although heavily dependent on the host plant and nematode species as well as environmental conditions such as temperature and moisture. Regardless of the timing, a number of key transitions are required for successful completion of the life cycle and consequently of plant parasitism [4]. Generally ~500 eggs are deposited in a glyco-proteic gelatinous matrix, secreted by the rectal glands of adult females at the root surface, to protect against biotic and abiotic stress [5]. Within the eggs, embryonic development leads to a first stage vermiform and coiled juvenile (J1) which then moults into a second stage juvenile (J2). Favorable moisture and temperature conditions trigger hatching of vermiform J2s from the egg, which constitutes the only infective stage, able to penetrate root tissue and thus also called pre-parasitic J2 (ppJ2). Through the perception of root diffusates and other factors, the ppJ2 migrates towards the host root system, to subsequently enter the plant tissues, often close to a root tip. Root penetration is made possible by a combination of mechanical and enzymatic disruption of the plant cell wall. At this stage, cell wall degrading/modifying enzymes (referred to as CWMEs) are secreted through a syringe-like stylet connected to secretory gland cells. Once inside the plant, the nematode migrates intracellularly between the cortical cells towards the root tip [6]. Upon reaching the host vasculature, the parasitic nematode induces the formation of five to seven hypertrophied and multinucleated “giant cells” from parenchymal root cells, via secretion of effector proteins and other molecules manipulating cell divisions and other functions. These metabolically hyperactive giant cells will remain the sole source of nutrition for the developing endoparasite. Once the permanent feeding site is initiated, the nematode uptakes plant nutrients through its stylet and rapidly becomes a swollen fusiform parasitic J2. In a compatible interaction, the parasitic J2 undergoes two successive molts to reach the J3 and then J4 juvenile stages. These two stages are characterized by the absence of a functional stylet and hence the nematode does not feed at these stages [5,7]. Late J2, J3 and J4 stages are morphologically difficult to differentiate and this can only be achieved by a meticulous examination to identify the presence of a functional stylet or to count the number of superimposed outer cuticles, resulting from the successive molts [7]. The last transition leads to adult individuals with a marked dimorphism between males and females. Males return to a vermiform morphology, and no evidence for feeding at this stage has been shown. Males are extremely rare and do not participate in the generation of offspring in obligate parthenogenetic Meloidogyne species such as *M. incognita*. The proportion of males can be higher in conditions unfavorable to the development of adult females and production of offspring. Adult females are morphologically distinct; being pear-shaped and having a functional stylet, they resume feeding and produce hundreds of eggs that will eventually form the next generation of parasitic juveniles and thus complete the life cycle. The parasitic life cycle of root-knot nematodes is thus characterized by successive transitions punctuated by molts and profound morphological and functional changes.

Accompanying these morphological changes, successful manipulation of host physiology, structure, development, and immunity must be achieved. It is generally accepted that this is achieved by delivery of a repertoire of “effectors” into and around host cells. Effectors in plant-parasitic

nematodes are primarily produced in two sets of pharyngeal glands: the dorsal gland (DG) and subventral gland (SvG) cells [8,9]. The SvGs are primarily, although not exclusively, active at the earlier stages of parasitism, and therefore thought to contribute to J2 penetration and migration. The DGs are primarily, although not exclusively, active at the later stages of infection, and therefore thought to contribute to the development and maintenance of the feeding site [10]. Hence, PPNs must control the transcription of effectors and other genes in space and time to successfully parasitise their host, while undergoing profound morphological changes of their own.

Advances in genomics and transcriptomics provide a means to explore the nature and regulation of parasitic nematode genomes during infection. Notably, large scale transcriptome studies of various different plant-parasitic nematode groups have revealed global trends in gene expression at different stages of the life cycle [11], and important insights in the transcriptional regulation of effectors. An emerging theme in many different plant-parasitic nematodes is the presence of conserved putative cis regulatory elements that are associated with effectors expressed in specific gland cells [12]. For example, the DOG box is conserved in cyst nematodes [11,13], enriched in the promoters of DG effectors [11,13,14], and can be used to predict novel effectors [11,13–15]. These analyses have also been expanded to the migratory endoparasites (e.g. *Bursaphelenchus xylophilus* STATAWAARS motif [14]).

Despite that the RKNs are the major contributors to the crop losses due to nematodes, no study to date has comprehensively described the transcriptional changes across its life cycle, nor explored the cis-regulatory elements associated with effector expression in this group. Here, we report a comprehensive transcriptomic survey of *M. incognita* using life stage-specific RNA-seq data to understand the global expression changes and to profile parasitism genes throughout a complete pathogen reproductive cycle. Availability of an annotated genome sequence comprehensively representing the set of protein-coding genes [16] coupled with this RNA-seq time course, provides an opportunity to explore the spatio-temporal regulation of gene expression in *M. incognita*, with a focus on effectors. Our study reveals gene expression variations that accompany the major developmental transitions. We identify the first putative cis-regulatory motif associated with tissue specific expression in the root knot nematodes; providing a novel insight into effector regulation in the DG cell. We combine the presence of this motif with several other criteria to predict a novel set of putative DG effectors in this species. Finally, we show this motif, and thereby its utility, is broadly conserved across the Meloidogyne genus and termed it Mel-DOG.

Material and Methods

Gene expression levels during four developmental life stages of *M. incognita*

To determine expression levels of protein-coding genes, we used RNA-seq data of four life stages of *M. incognita*, strain Morelos (ppJ2s, parasitic J3-J4, adult females and eggs) produced in triplicates in a previous analysis [16] and that have been recently re-mapped to the genome with STAR [17] using more stringent end-to-end parameters [18].

Read counts and expected read counts were calculated on the predicted genes from the *M. incognita* GFF annotation [16] as FPKM and TPM values using two different methods:

- With RSEM [19] to take into account the multi-mapped reads via expectation maximization as explained in [18].
- With htseq-count [20], where only uniquely mapped reads are counted.

Identification of differentially expressed genes between life stages

We used three independent methods to identify genes differentially expressed between the four life stages. RSEM expected read counts were used to identify differential gene expression using EBseq [21]. Read counts from htseq_count were used to identify differential gene expression using DEseq2 [22] and EdgeR exact tests [23]. We considered genes differentially expressed if they returned a Log2 fold change value >2 and a false discovery rate (adjusted p-value) <0.05, consistently, from all the three above-mentioned methods.

Clustering of differentially expressed genes

To classify differentially expressed genes in expression clusters, we used the log10 of the median FPKM value over the 3 replicates (+1) as determined by RSEM and as previously published in [18] with data publicly available at https://doi.org/10.15454/YM2DHE. For gene expression data, the goal of a clustering approach is to minimize the divergence within a cluster yet maximize the divergence between clusters. We used the Pearson correlation (r) to estimate this divergence between gene expression patterns. This metric is appropriate for temporal variations of expression values such as a developmental life cycle. Many differential gene expression analyses employ an arbitrarily predetermined number of clusters in which genes have to be assigned according to their expression pattern. This kind of approach is not ideal but determining an optimal number of clusters is far from evident. Here, because our goal was to obtain a low number of clusters, we used a divisive clustering. Divisive approaches first divide the dataset in two most dissimilar clusters and progressively until the desired number of clusters is reached. For the divisive analysis we used the Diana method in the cluster package of R. To determine the most appropriate number of clusters, we used the ClusterSim packages in R. We used three indicator metrics (CH, Silhouette and C-index). We allowed the number of clusters to vary from min. 2 to max. 25.

Identification of overrepresented Gene Ontology terms

Gene Ontology (GO) terms have been assigned to the *M. incognita* protein-coding genes based using InterProScan [24] in a previous analysis [18] with all the results publicly available at https://doi.org/10.15454/9BFFKG.
To identify overrepresented GO terms, we used a hypergeometric test as implemented in func [25] within the R package GOfuncR. Overrepresented GO terms were searched at each transition between the four life stages and within each of the expression clusters. A family-wise error rate (FWER) threshold value of 0.05/(number of comparisons) was set to consider a GO term as significantly overrepresented in the dataset.

Mapping of genes known to be specifically expressed in secretory gland cells on Meloidogyne genomes

Based on the literature, we established a manually curated list of *M. incognita* genes specifically expressed in the secretory Sub-ventral Gland (SvG) or Dorsal Gland (DG) cells. We started from the list published in [26] and manually checked the figures associated with the cited references to only keep genes with evidence for specific expression via in-situ hybridization. We completed this list with the Mi-PNF3 gene published more recently [27] and eliminated redundancy between the sequences at the protein level. We retrieved the corresponding CDS sequences and aligned them to the *M. incognita* genome sequence [16] using the splice aware aligner SPALN [28] using the following parameters.
'spaln -Q7 -O4 -M -d Mi.mfa test-CDS > result.gff
-M = report multiple hits and not only the best
-Q4= DP mode, 5-7 = maximum number of reported hsp
-O4= megablast-like output format
To investigate the conservation of effectors in the Meloidogyne genus, we also aligned these same CDS to the genomes of the evolutionary close relatives *M. arenaria, M. javanica* and *M. enterolobii* [16,29] as well as the more distant *M. hapla* [30] using the same parameters.

Identification of specific motifs in the upstream regions of genes expressed in secretory gland cells

In order to identify putative regulatory motifs associated with effector genes, a custom python program was written, with unit tests, to extract intergenic regions in the *M. incognita* genome (https://github.com/peterthorpe5/intergenic_regions). Various length regions (up to 10,000 bp) 5' of the coding start site (as defined in the gff3 annotations) were extracted from the genome for all genes. If another gene was located within this region, the script will return the sequence between the two coding regions. The resulting intergenic regions were divided into positive and negative test sets, and taken forward for further motif analysis.

As a positive test set, we retained *M. incognita* genomic loci corresponding to alignments of the known SvG and DG genes as described above encompassing the start (ATG) position and covering at least ⅔ of the effector length with a minimum of 99% identity. When several effectors mapped on an overlapping locus we kept the best scoring alignment. When one effector mapped to multiple loci, we kept up to the three best-scoring alignments matching the above-mentioned criteria.

As a negative set, we selected 167 *M. incognita* genes from an Orthofinder [31] analysis previously performed on 64 genomes, including 62 nematodes and two outgroup tardigrade species [18]. The selection criteria were as follows: (i) the vast majority (163 genes) came from orthogroups conserved in at least 80% of the species and 85% of the Tylenchida and did not form multigene families. Being

widely conserved across nematodes, including in many non phytoparasitic species and not forming multigene families, these genes are more likely involved in core housekeeping functions than coding for effector proteins. (ii) four genes specific to root-knot nematodes but bearing no signal peptide for secretion, including one gene specific to *M. incognita*. The annotated list is available at https://doi.org/10.15454/KYPEN0.

Promoter regions (n bp upstream of the start codon) were analysed for the presence of enriched motifs as previously described [11], with the following modifications. Enrichment of motifs between categories (DG versus non-effectors, DG versus SvG, etc.) was calculated using HOMER [32] with default parameters. Instances of the motif were identified in FASTA sequences of promoter regions using the FIMO web server [33]. Correlations between the number of motifs in a given promoter and the presence or absence of signal peptides and transmembrane domains encoded by the corresponding gene (as determined above) were calculated. To identify variants of the motif that are able to similarly predict whether the corresponding gene encodes a putatively secreted protein, a custom python script was written. The script identifies the occurrence of all 1bp mismatch variants of the motif in the promoter regions of all genes, and calculates the proportion of genes that encode putatively secreted proteins for 1, 2, or 3 copies of each motif variant (https://github.com/sebastianevda/Mel_DOG_scripts). The same initial motif identification procedure (HOMER) was repeated on the genomes of the four other Meloidogyne species introduced in the above section using 2kb promoter regions.

Identification of putative secreted proteins and effectors

We used the list of 2,811 *M. incognita* proteins bearing a signal peptide for secretion, and no transmembrane region, as putative secreted proteins (PSP) as described in [18]. The list is publicly available at https://doi.org/10.15454/JCYZDI. To further differentiate candidate effectors from the rest of PSP, we only retained PSP bearing at least one of the four motifs enriched in known *M. incognita* effectors in their 100 first amino-acids as explained in [18,34]. The list of 2,146 PSP bearing an effector protein motif is publicly available at https://doi.org/10.15454/CSTXU2.

Evolutionary origin of *M. incognita* effectors

Based on the mapping of non-redundant known *M. incognita* SvG and DG effectors on the genomes of five *Meloidogyne* species ancestral numbers were deduced across their phylogeny using parsimony inference with Mesquite [35]. In the genomes of *M. incognita*, *M. arenaria*, *M. javanica* and *M. entetolobii*, only genes mapping with >90% identity on >66% of their length were considered unambiguously present (character state '1' in Mesquite). Genes with a significant spaln mapping but that did not meet the % identity and % coverage thresholds were considered as unsure to be present (character state '?' in Mesquite). Given that *M. hapla* is phylogenetically much more distant, only a length coverage threshold was used to infer unambiguous presence following a significant spaln mapping from the ATG position. Genes that returned no significant spaln mapping were considered absent from the genomes under consideration (character state '0' in Mesquite). The tree topology of the five *Meloidogyne* species used for ancestral states reconstruction was taken from [36]. We used the Mesquite analysis 'Trace all Characters' to reconstruct characters (presence / absence of each effector) at each ancestral branch. To infer the minimal unambiguous number of effectors at each ancestral branch the sum of 1's at this branch was calculated while equi-parsimonious 0/1's ancestral state were considered an absence in a conservative scenario. All the data used to produce the ancestral reconstruction as well as the raw results are available at https://doi.org/10.15454/OJMRDD.

Results

Predicted functions of differentially-expressed genes are consistent with transitions between developmental stages

During the course of the *M. incognita* life cycle, the nematode undergoes several major transitions, from a sedentary egg to a migratory second stage juvenile in the soil, culminating in a sedentary endoparasite within the living host root system. To understand the changes in gene expression during these transitions, and progression through the *M. incognita* life cycle in general, the present study deployed RNA sequencing (RNA-seq) of four time points: eggs, ppJ2s, a mixture of larval J3/J4s and adult females [16]. Using these RNA-seq data, differential gene expression was analyzed to capture biological signatures between these different developmental transitions.

Approximately 30 % of the protein-coding genes (12,461) are significantly differentially expressed between at least two developmental life stages, according to three independent statistical methods (https://doi.org/10.15454/VLN8UC). Interestingly, all three methods converge on similar log2 (fold change) values, despite using different counting strategies. Three of the four transition stages are rich (>4,500) in differentially expressed genes (DEGs), with the exception of the transition larval J3/J4 to adult females, with the smallest number of DE genes (555), nearly 10-fold lower (Figure 1A and https://doi.org/10.15454/VLN8UC). The largest number of DEGs (6,967) was observed in the transition between the pre-parasitic J2s and the mix of J3-J4 sedentary stages. Interestingly, this transition is generally characterised by large scale gene repression: the vast majority of DEGs (5,196) were significantly downregulated (log2FC down to <-16) and only 1,771 were significantly up-regulated (log2FC up to >14). Moreover, the largest upregulation in gene expression (3,948 DEG at log2FC up to 10.9) is observed at the developmental transition from adult females to eggs. A total of 9,145 DEGs (70%) were differentially expressed between multiple comparisons, while 3,271 were differentially expressed specifically at one given transition. The highest number of specifically differentially expressed genes (607) was, from pre-parasitic J2 to mixed J3-J4. Together, our filtered dataset provides a first comprehensive view of transcription during the entire *M. incognita* life cycle.

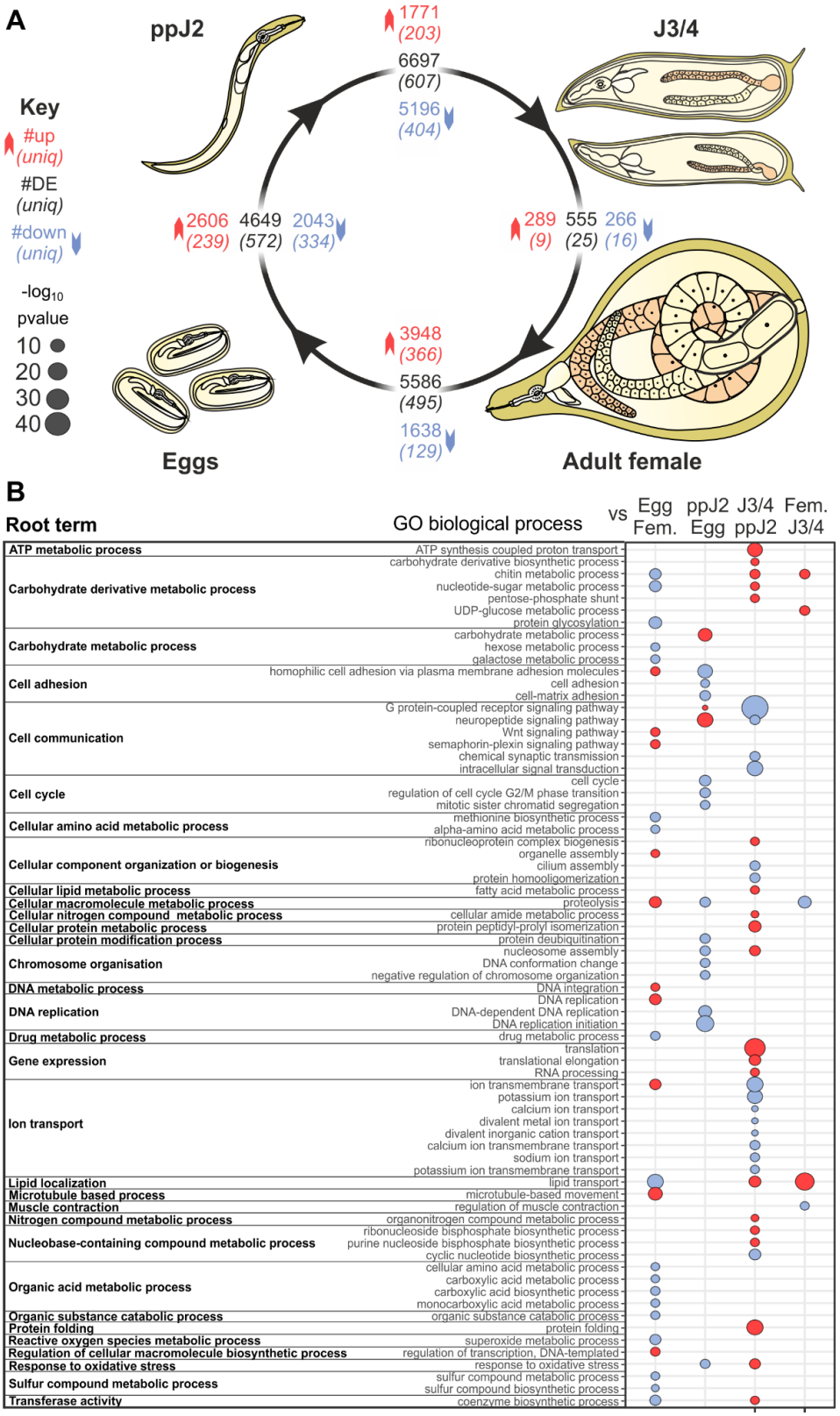

Figure 1. Distinct patterns of 12,461 differentially expressed genes (DEGs) and Gene ontology (GO) term enrichment between the life stages of *M. incognita.* A) Representation of the nematode life cycle. At each transition between stages (black arrows), the total number of DEGs is indicated in black in the central box. Genes significantly upregulated and downregulated are indicated in red and in blue, respectively. The number DEGs only at this transition is indicated in italics and parentheses.

B) The "biological processes" of Gene Ontology (GO) terms significantly enriched in each transition are shown, grouped by their root terms. Bubble plots illustrate the most significant GO terms that were overrepresented (in red) and underrepresented (in blue) for each term. If no buble is present, this term was not significantly different at that transition. Bubble size is calculated as -log10 (refined p_value), resulting in a p-value FWER threshold at 0.0125.

To gain insight into the biological changes occurring during the time-course of the *M. incognita* life cycle, we identified and analysed significantly overrepresented GO terms in genes differentially regulated between each pair of adjacent life stages (methods, Figure 1 and https://doi.org/10.15454/VLN8UC). At the transition "eggs to ppJ2s", a large proportion of upregulated DEGs have GO terms related to "sensory perception" (e.g. Cell communication, GPCR signaling, neuropeptide signaling, response to stimulus etc. https://doi.org/10.15454/NOT2LH). The GO terms related to "locomotion/movement" (e.g myofibril, contractile fiber, striated muscle thin filament etc.) were also enriched. These functional terms make sense with the particular time point of the pre-parasitic life stages of the animal, in which the triggering of these signaling pathways accompanied hatched ppJ2s that actively sense and move towards plant root exudates. Significantly enriched GO terms in this transition were also related to "Carbohydrate metabolism" including those related to plant cell-wall degradation (e.g. Carbohydrate degradation, hydrolases active on glycosyl bond, polygalacturonase activity, pectate lyase activity). The parasitic J2s are known to secrete a range of carbohydrate-active enzymes (CAZymes, e.g. cellulases, pectate lyases, polygalacturonases), many of which were acquired via horizontal gene transfer, for the degradation of plant cell wall to facilitate the penetration and moving and also supply nutrients to the animal [37,38]. Another notable feature was an enrichment of GO terms including iron binding, heme binding and oxidoreductase activity, which may indicate the beginning of an "response to oxidative stress" at this transition. The cellular component GO term "extracellular region" was also overrepresented in this transition, consistent with the secretion of many enzymes and other proteins at this stage.

At the transition from "ppJ2s to sedentary larval J3/J4s", terms related to a biotrophic parasitism were overrepresented. GO terms related to "response to stressors" were upregulated (e.g. response to oxidative stress; oxidation-reduction process; oxidoreductase activity, peroxidase, antioxidant, https://doi.org/10.15454/VE2MLH). Other overrepresented terms were related to host colonization and/or survival such as "serine-type peptidase activity, exopeptidase activity". This may suggest the nematode activates important processes of parasitism, such as invasion and migration or responses to oxidative stress imposed by the plant as a defence against the parasite. Furthermore, we noted a down-regulation of genes coding for sensory perception (i.e G protein-coupled receptor signaling pathway, neuropeptide signaling pathway, chemical synaptic transmission), indicating that perception of environmental cues may be less important once the nematode has established its feeding site. Other interesting enriched GO terms in upregulated genes encompassed the following main functional categories related to "lipid biosynthesis, transport and storage" (e.g. fatty-acid synthase, lipid transporter activity, lipid transport, lipid localization), reflecting a need to store nutrients and energy to sustain the important metamorphosis to the next stage (adult). As these transitions involve metamorphoses and important morphological / structural changes, it is also consistent with other overrepresented terms such as "structural constituents of the cuticle" and "chitin binding". This transition reveals a marked shift in the nematode lifestyle with differential regulation of various genes promoting growth and pathogenic success.

At the transition "sedentary J3/J4 larvae to adult females", the nematode resumes feeding phase and develops hundreds of eggs. We observed that genes coding for locomotion-related functions were down-regulated (i.e regulation of muscle contraction) confirming most nematode movements have stopped. The overrepresented GO terms associated with upregulated genes (https://doi.org/10.15454/5UJNHS) were similar to the ones identified at the previous transition

(from ppJ2 to sedentary J3-J4 stages). These GO terms included those related to lipid transport and localization (e.g. lipid transport, lipid localization, lipid transporter activity, macromolecule localization) and related to defense against oxidative stress (e.g. response to oxidative stress), or possible remodeling of the cuticle (e.g. chitin metabolic process, chitin binding). These GO terms support that the processes and functions triggered at the previous transition are further amplified at this transition.

By contrast, comparing the 'eggs' to the 'adult females' stages (https://doi.org/10.15454/6EYGIX) showed most of induced distinctive DEGs were associated with biological processes such as ion transport, ion transmembrane transport, metal / potassium ion transport, related to DNA dynamics (i.e DNA metabolic process, DNA replication, DNA integration) and also involved in "movement of cell". Important changes in gene expression were observed with overrepresented GO terms associated with molecular function such as "ion channel activity, channel activity, transmitter-gated ion channel activity, potassium channel activity, calcium / metal ion binding, DNA-directed DNA polymerase, DNA binding". In the cellular component category, there were a significantly high number of DEGs classified as "plasma membrane, membrane, synaptic membrane, cation channel complex". The distribution of DEGs may be correlated with an overall transcriptional change in gene expression related to embryonic development. In addition, we observed a significant down-regulation of genes encoding stress response (i.e superoxide dismutase, oxido-reductase activity, L-ascorbic acid binding), digestion enzymes (serine-type endopeptidase inhibitor activity, metalloendopeptidase inhibitor activity, hydrolase activity) and general metabolism (i.e carbohydrate binding, lipid transporter activity, galactosyltransferase activity), underlying that the nematode has stopped feeding and by extension completes its parasitic lifestyle.

Most of the known *M. incognita* effectors are conserved in multiple Meloidogyne genomes and were probably inherited from a common ancestor

To explore the transcriptional regulation of parasitism, we first identified genome loci that encode known effectors with experimental evidence of expression in the two sets of pharyngeal glands: sub-ventral (SvG) or dorsal pharyngeal gland (DG) cells. From the literature, we identified 48 SvG and 34 DG non-redundant *M. incognita* effector genes and studied their conservation in the *M. incognita* and other root-knot nematode genomes (https://doi.org/10.15454/P5YIGX). Two anomalies are noted. Firstly, the effector 35A02 (msp25) described as specifically expressed in the SvG in [39] is 97% identical to Minc02097, conversely detected in the DG in [40]. Both genes map to the same genomic location in *M. incognita* and to the same CDS (Minc3s00202g07465). Secondly, Minc18033 described as SvG-specific in [40] is 99% identical to 16E05 (=msp17) described as DG-specific in [39]. Both Minc18033 and 16E05 map the same genomic locations and the same two CDSs (Minc3s02105g28312 and Minc3s03024g32468). In both cases, we eliminated redundancy between the two sets, and assigned the most recent traceable gland cell location.

Of the 48 known and non-redundant SvG effector genes, all mapped from the ATG position to the *M. incognita* genome assembly (https://doi.org/10.15454/P5YIGX). These gene candidates mapped to 80 different loci in the genome which is consistent with the allopolyploid genome structure [16]. A total of 45 SvG effector CDS mapped with at least 90% identity and across >66% of their length, the three other effectors, 31H06, Mi-Pel1 and Minc03866, did not meet the 90% identity threshold (% identities range 85-89%). The 45 SvG effectors mapping with high confidence corresponded to 74 different loci and 64 predicted gene models in the *M. incognita* annotated reference genome. (https://doi.org/10.15454/P5YIGX).

The 48 non redundant *M. incognita* SvG effectors were mapped to the genomes of three other polyploid and parthenogenetic species (*M. arenaria*, *M. javanica* and *M. enterolobii*) belonging to the same *Meloidogyne* clade (I) than *M. incognita* [41] to determine their evolutionary conservation. All (48/48) mapped from the ATG with >90% identity and across >66% of their length to at least one of the above-mentioned genomes. Most of these effectors mapped to multiple loci in each of these genomes, which is again consistent with their polyploid structures. Only 8, 5, and 2 effectors did not meet the 90% identity and 66% coverage thresholds in the genomes of *M. arenaria*, *M. javanica* and *M. enterolobii*, respectively (of which 3, 1, and 0 did not map at all). Reconstruction of ancestral states from the phylogenetic distribution of SvG effector genes across these *Meloidogyne* species indicated that at least 42/48 effectors were present in a common ancestor of clade I root-knot nematodes (Figure 2). To assess whether these SvG effector genes were evolutionarily more anciently conserved, we also mapped them to the genome of *M. hapla*, a facultative parthenogenetic and diploid species belonging to the more distantly related clade II [41]. Only 25 /48 effectors mapped from the ATG on at least 66% of their length. However, due to the higher evolutionary distance of *M. hapla*, no alignment reached 90% identity (range=63.8% - 89.2%). Except Mi-VAP2 all the mapping candidate effectors matched one single locus in the *M. hapla* genome, consistent with its diploid homozygous structure. Overall, at least 52% (25/48) of the SvG effectors present in *M. incognita* were probably inherited from a common ancestor of Clade I and Clade II Meloidogyne, the rest either being actually absent from that ancestor, too distantly diverged to map the *M. hapla* genome or actually missing from this genome assembly (Figure 2).

Of the 34 non redundant effector genes with DG-specific expression, 33 map from the ATG position to the *M. incognita* genome assembly. One known DG effector, 28B04, did not map on the *M. incognita* genome, hence it will not be possible to study its expression. As previously observed with SvG effectors, the 33 DG effectors map to 57 different loci in the genome which is consistent with the allopolyploid genome structure. A total of 29 DG effector CDS map with >90% identity and across >66% of their length from the ATG position (https://doi.org/10.15454/P5YIGX). This corresponds to 51 different loci and 30 predicted gene models in the *M. incognita* annotated reference genome. The four effectors not meeting these criteria were 5C03B (86% id), 1C05B (87.7% id); 4F05B (67.5% id) and 17H02 (57.8% cov).
Except 7A01, which was absent from all the other Meloidogyne genomes investigated and might represent a *M. incognita*-specific gene, all the rest of DG effectors mapped from their ATG position in at least another Clade I *Meloidogyne* species. Although 28B04 was not found in the *M. incognita* genome assembly, this gene mapped to the genome of *M. javanica*, suggesting it might be missing in the *M. incognita* assembly. Overall, 27, 27, and 28 DG effector genes mapped from the ATG position with >90% identity and across >66% of their length in the genomes of *M. arenaria*, *M. javanica,* and *M. enterolobii*, respectively and the majority mapped to multiple loci, consistent with the polyploid genome structures. Reconstruction of ancestral states based on the DG effector genes phylogenetic distribution suggests that at least ~82% (28/34) were inherited from a common ancestor of Clade I Meloidogyne species (Figure 2). In *M. hapla* only 8 DG effectors genes map to the genome and only one (14-3-3) passed the 90% identity threshold, which is consistent with the higher evolutionary distance. Hence only ~23% of the known *M. incognita* DG effectors could be traced back to a common ancestor of Clade I and Clade II Meloidogyne species (Figure 2 and https://doi.org/10.15454/P5YIGX).

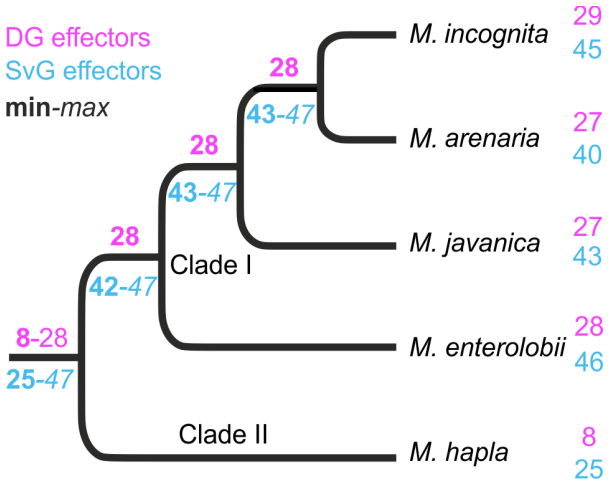

Figure 2. Reconstruction of ancestral numbers of homologues of *M. incognita* effectors across the Meloidogyne phylogeny. Based on the mapping of known *M. incognita* SvG (cyan) and DG (magenta) effectors on the genomes of five *Meloidogyne* species, ancestral numbers were reconstructed using parsimony. Minimal effector numbers are given in bold at the corresponding branches and maximal numbers in italics, depending on whether equi-parsimonious absence / presence of effectors are considered actually absent / present.

Hierarchical clustering analysis highlights changes in expression of genes related to parasitism as a whole

To provide a more holistic view of gene expression, we clustered and classified the 12,461 differentially expressed genes based on their expression profile in all life stages and studied the enriched functions and the distribution of effector genes in these clusters. Eight distinct clusters (named thereafter A to H) were generated, as indicated by the C-index and measured relative to the median expression level throughout the entire life cycle (https://doi.org/10.15454/2XJCJQ). The eight clusters varied in number of genes across an order of magnitude (from 611 to 5,441 genes, Figure 3). Three distinct and sequential gene expression patterns were identified according to the timing in developmental stages of *M. incognita*. We noted that the clusters A and B contain 17% of all DEGs and exhibit a peak of expression at the egg stage. These genes were enriched in GO terms related to embryogenesis, such as cuticle formation, cell adhesion and migration (https://doi.org/10.15454/V3SCRC). We also identified an enrichment of hydrolytic enzymes (i.e metalloendopeptidase activity) and pattern-receptors (i.e scavenger receptor activity) which may act as well in physiological and pathological processes. Interestingly, the majority of DEGs (6,934) were grouped into just two clusters (C and D) characterized by a peak of expression at the ppJ2 stage. These two clusters were significantly enriched in genes associated with pathogenesis (https://doi.org/10.15454/V3SCRC).

Interestingly, cluster C was overrepresented in DEGs encoding a mixture of potential secreted plant cell wall degrading enzymes including CAZymes (glycoside hydrolases, pectate lyases) and phosphatases (serine/threonine protein phosphatase), that are known to be involved in host colonization process but also in survival [42,43]. Consistent with secretion of many enzymes and other proteins, the only enriched GO term in the cellular component ontology is “extracellular region”. We also noted in cluster C a significant enrichment of terms related to transcription factor activity that might be related to genome-wide response to plant invasion. In cluster D, various different GO terms related to ion-channel and transport activities to DNA replication and binding

were significantly enriched, probably reflecting the preparation for metamorphosis from the late J2 to J3 stage.

The next two clusters E and F only encompass 695 and 611 DEGs respectively, and seem to be characterized by a peak of expression in eggs, a trough in the ppJ2, and a recovery of expression in the mix of sedentary J3-J4 larvae stage. Enriched terms at clusters E and F correspond to DNA replication, DNA binding, RNA translation / protein biosynthesis and folding activities, probably reflecting the profound changes morphological remodeling and physiological changes at these stages (https://doi.org/10.15454/V3SCRC). As the third key time point in the nematode's life cycle, the clusters G and H encompassed ~17% of DEGs and were characterized by a marked up-regulation during the sedentary J3-J4 larval stages. Both clusters showed an enrichment of gene categories related to nutrient processing / energy acquisition (metabolism of carbohydrates and lipids), and detoxification process (oxidation-reduction processes). Intriguingly, the clusters G and H differed by their content of GO terms related to hydrolytic and detoxification enzymes according to the type of regulation and substrates. For example, Glutathione hydrolase, serine-type carboxypeptidase were prevalent in G while metalloendopeptidase inhibitor activity was identified in H. Overall, the result is consistent with the idea there is a shift in the expression of a suite of genes at this given developmental stage driving the preparation and the progression in the parasitism of *M. incognita*.

Expression Cluster	egg ppJ2 J3/J4 female	Total N° of DEG in Cluster	N° of SvG (Percent per Cluster)	N° of DG (Percent per Cluster)	Top3 enriched GO terms (Molecular Function Category)
A		1345	1 (2)	0 (0)	Metalloendopeptidase activity Scavenger receptor activity Structural constituent of cuticle
B		807	0 (0)	0 (0)	Semaphorin receptor activity Thiol-dependent ubiquitinyl hydrolase activity Histone acetyltransferase activity
C	★	1493	19 (36)	4 (18)	Hydrolase activity DNA-binding transcription factor activity Transferase activity
D		5441	15 (29)	3 (14)	G protein-coupled receptor activity Ion channel activity DNA binding
E		695	3 (6)	2 (9)	RNA binding Structural constituent of ribosome Translation elongation factor activity
F		611	0 (0)	0 (0)	DNA biding Protein heterodimerization activity DNA helicase activity
G		675	11 (21)	7 (32)	Oxidoreductase activity Glutathione hydrolase activity Serine-type carboxxypeptidase activity
H		1394	3 (6)	6 (27)	Superoxide dismutase activity Lipid transporter activity Metalloendopeptidase inhibitor activity

Figure 3. Temporal gene expression profiles throughout the *M. incognita* infection. Hierarchical clustering identified 8 distinct temporal patterns of differentially expressed genes and known effector genes (SvG and DG effectors in cyan and magenta lines respectively). The *x* axis represents four developmental stages. Black star indicates those clusters significantly enriched in the cellular component term "extracellular region". Black lines represent the average profile for each cluster.

While the hierarchical clustering reveals global transcriptional changes during the time course of *M. incognita* life cycle, we also more specifically focused on expression patterns of effector genes over time. The 30 and 62 known *M. incognita* DG and SvG effectors mapped to 94 different genome loci in which a gene model was predicted in the genome and allowed studying their expression. Respectively 22 (73.3%) DG and 52 (81.3%) SvG effector gene models showed significantly different gene expression between at least two different life stages during the *M. incognita* parasitic life cycle. Using a X^2 test, neither the DG nor the SvG genes were randomly distributed in the eight expression

clusters (p-value <2.2E-16). We found that cluster C was significantly enriched in known effector genes (adjusted FDR 1.91E-07). Cluster C presents a pattern of higher expression in the ppJ2 stage. SvG genes are mostly expressed (approx. 71%) at the early stage of infection represented by the clusters C, D and E and positively upregulated at 65% during the pre-parasitic J2 stage. In contrast, only 1.9% of differentially expressed SvG genes were found in the clusters A and B, characterized by a peak of expression at egg stage. The rest of differentially expressed SvG effector genes (27%) were grouped in the two last clusters G (21.2%) and H (5.8%). These clusters are characterized by a peak of expression at J3-J4 stages.

Interestingly, and consistent with the previous description, differentially expressed DG effector genes were mostly (~59%) distributed in these latest G and H clusters. These DG effector genes were significantly up-regulated (64%) at J3-J4 stage. SvG and DG effector genes show a pattern of expression that was positively and temporally correlated with the infection course as previously described in [44]. In the current observations, we noted that SvG and DG effectors were significantly enriched in clusters C, G and H containing core parasitism genes, which are specific members of gene families encoding plant cell wall degrading enzymes and detoxification enzymes known to be involved in the parasite invasion process.

A promoter motif is associated with Dorsal Gland effectors and secreted proteins

An emerging theme in plant-parasitic nematodes is the *in silico* discovery of effector regulatory motifs in several phylogenetically distant species, the pine-wilt *B. xylophilus*, the cyst nematode *Globodera pallida* and the root lesion nematode *Pratylenchus penetrans* [11,14,45]. These studies provide powerful insights in the understanding of adaptation to plant parasitism in nematodes. Given that effectors in *M. incognita* are specifically regulated in space and time, similarly to other phytoparasites, we extended the search of cis-regulatory sequence motifs enriched in the promoters of genes related to RKN parasitism. From the coordinated of loci corresponding to the comprehensive set of known DG and SvG effectors mapped to the *M. incognita* genome; a series of putative promoter regions were extracted (500, 1000, and 1500 bp 5' of the start codon) using custom Python software (https://github.com/peterthorpe5/intergenic_regions). The 5' regions of DG effectors and SvG effectors were compared to one another, as well as to a set of 5' regions from 167 genes not annotated as effectors but widely conserved in nematodes besides *Meloidogyne* species (i.e. likely non-effectors). Gene predictions at the ends of scaffolds were omitted due to possibly truncated data. After this filtering step, 92% of all annotated protein-coding genes (40,312/43,718) were kept, which encompassed 90 % of mapped DG CDS (46/51), and 83% of mapped SvG CDS (62/74).

Comparing DG effector genes upstream regions with putative non-effector genes upstream regions using a differential motif discovery algorithm (HOMER, [32]) identified a series of similar motifs (of approximate consensus TGCACTT) in all three length categories (500 bp - TGCACT (32% on target, i.e. of the positive set that contain the motif), 1000 bp - TGCACTT (32% on target) and 1500 bp - TTGCACTT (50% on target), Figure 4A). A similar motif was enriched in the DG effector genes upstream regions when compared to upstream regions of SvG effector genes, suggesting this motif is specifically enriched in regulatory regions of DG effector genes. In contrast, no consistent motifs were enriched in the set of SvG upstream regions compared to either DG, or putative non-effector, upstream regions.

On a whole-genome scale, 17.3 % of all genes that contain one or more copies of TGCACTT motif in their 1000bp upstream region encode putatively secreted proteins. The proportion of putatively secreted proteins rises to over a quarter (26.9 %) for genes having two or more copies of the motif, t

in their 1000bp upstream region. This represents a~4-fold enrichment over the remainder of the genome (c.f. 6.4% putatively secreted proteins, FDR 1.8E-26). To determine whether variants of this motif provide similar enrichment, or enhance the overall enrichment, a custom python script (https://github.com/sebastianevda/Mel_DOG_scripts) was used to search for all 1 base-pair mismatch variants of the motif in the 1000 bp upstream of all genes, and calculate the proportion of putative secreted proteins. Two variants consistently enriched for putatively secreted proteins (TGCACTT and TGCCCTT, Figure 4B). We next searched in the 1000 bp proximal promoter sequences of the known DG effector gene loci to further explore the occurrence of the two variants. The distribution of the variants revealed a clear preference for one variant over the other in the upstream regions of known effectors, comprising 43,5% with only TGCACTT (20/46), while 3/46 have only TGCCCTT, and 4/46 have both TGCACTT and TGCCCTT. On a genome wide-scale, 20% of all genes that encode a putative secreted protein have 1 or more copy of TGCMCTT in the 1000 bp 5' region (524/2737), nearly half (48%; 251/524) have the variant TGCACTT, followed by 37% (194/524) with TGCCCTT, and 15% (79/524) with both variants. For a random set of genes, 4% (3/74) have TGCACTT, 4% (3/74) have TGCCCTT, and 1.3% have both (1/74). Consequently, TGCMCTT is enriched approximately 10-fold over random in the promoters of genes encoding putatively secreted proteins, and hereafter considered as a core motif for all further analyses.

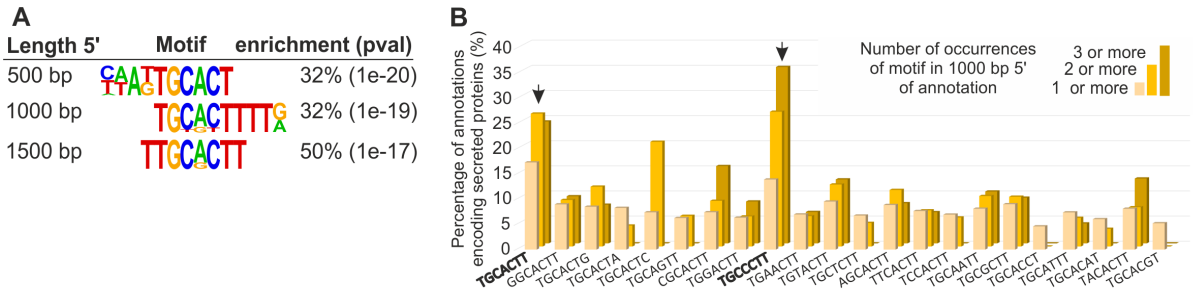

Figure 4. Identification of a promoter motif associated with DG effector genes in *M. incognita.* A) Selected similar motifs enriched in 500, 1000, or 1500 bp promoter regions of *M. incognita* DG effector genes compared to a set of non-effectors. A core motif derived from these predictions is TGCACTT. B) The proportion of annotations that encode secreted proteins is shown for those that contain one (yellow), two (orange), or three (dark orange) copies of each 1bp mismatch variant of TGCACTT.

Generally, the more copies of the motif in the upstream region of a gene, the more likely the corresponding gene encodes a putative secreted protein (Figure 4B). A consistent and positive correlation was observed between the proportion of putatively secreted proteins and the number of copies of the motif in the upstream region, at all lengths categories from 300 to 10,000 bp (Figure 5A). The number of DG effector genes identified using 10,000 bp upstream regions was used to set an upper limit for the DGs that are "identifiable" using this motif (60.87% (28/46)). Of these "motif-identifiable" DG effector genes, 95% were recovered by searching for one or more motifs within the first 1,000 bp upstream region (Figure 5B). To determine whether the predictive power of this motif is restricted to the non-coding upstream region of a gene, the enrichment analysis was repeated to include 200 bp coding sequence. The inclusion of 200 bp of coding sequence reduced the proportion of genes encoding putatively secreted proteins, and is therefore detrimental to prediction (e.g. Figure 5C). This supports the motif as a putative cis-regulatory elements.
Finally, we determined the positional enrichment of the motif within the upstream regions of DG effectors, and of all genes that encode putatively secreted proteins. Plotting the frequency distribution of TGCMCTT in known DG effectors from 200 bp downstream of the start codon to 1000 bp upstream reveals a peak between -350 to -150 bp (Figure 5D). The location of this peak is consistent with both the lack of enrichment for secretory proteins using 100 bp upstream of the start codon (Figure 5A), and the worse enrichment when including coding sequence (Figure 5C), because

the peak is non-coding and occurs further upstream. For those genes that contain at least one copy of the motif in the 1000 bp upstream of the start codon and encode a putative secretory protein but are not already classified as a known DG effectors, a similar frequency distribution of the motif is found (Figure 5E), complementary to the known DG set, with peaks at approximately -200, and -500 bp. Combining these distributions provides a broad peak from -150 to -500 bp for "motif-identifiable" putatively secreted proteins.

Taken together, these data describe a possible cis-regulatory promoter motif present in the first 1000 bp upstream of the majority of previously known DG effector loci (approx. 60%), and highlights a substantial but reasonable subset (approx. 20%) of the genes that encode putatively secreted proteins that have not, yet, been characterised as actual DG effectors. Moreover, "motif-identifiable" putatively secreted proteins are also enriched in the differential expression cluster C (peak at J2, FDR = 1.91E-07), consistent with a role in parasitism. Similar to previous dorsal gland cell motifs in other plant-parasitic nematodes, this motif may provide a way to predict root-knot nematode DG effectors, narrow down the candidate effector repertoire, and by extension explore the regulation of effectors in root-knot nematodes.

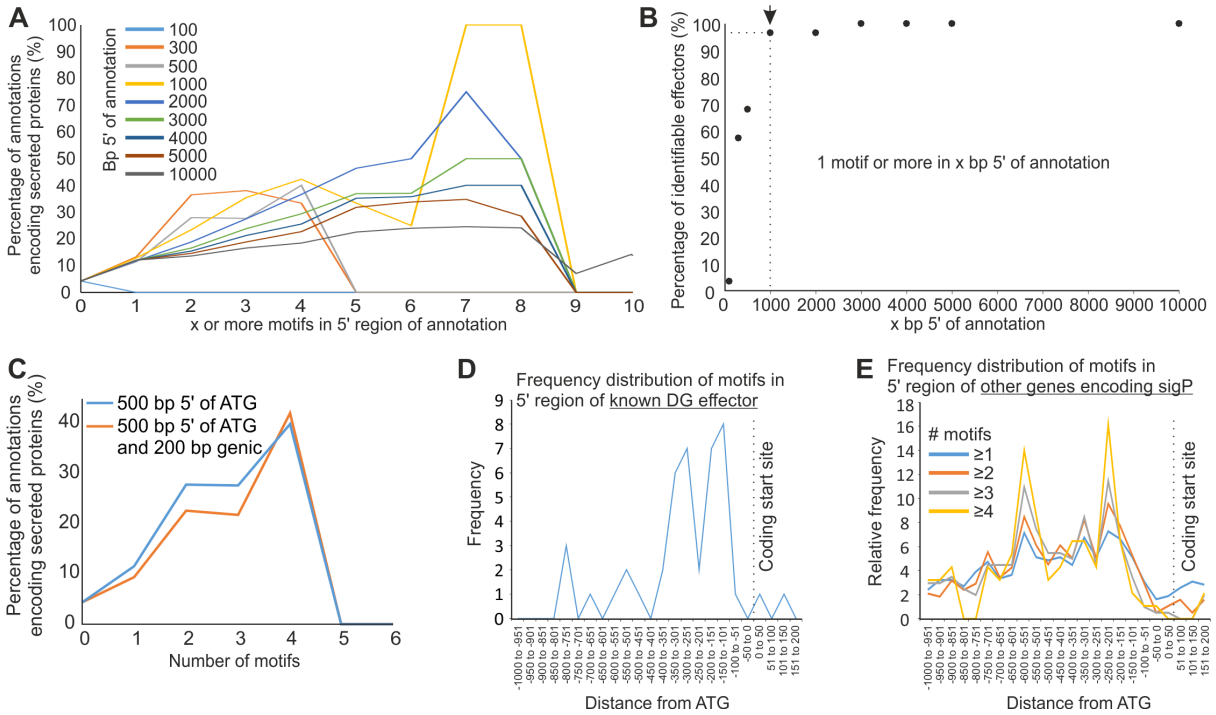

Figure 5. Location of a promoter motif associated with Dorsal Gland effector genes in *M. incognita*. A) For those promoters that contain x or more copies of the motif (where x varies from 0 to 10) the proportion of annotations that encode putatively secreted proteins is shown. Colours represent repeating the analysis with different promoter lengths (ranging from 100bp to 10,000bp). B) For those effectors that contain one or more motifs in the 10,000bp upstream of the coding start site, what proportion are identified with an x bp promoter (where x varies from 100 to 10,000). C) Using either 500 bp upstream of the ATG (blue), or 500 bp upstream and 200bp downstream of the ATG (orange), the relationship between the number of promoter motifs and whether the corresponding genes encode putatively secreted proteins is shown. D) Frequency distribution of motifs in the promoters of known DG effector genes. E) Frequency distribution of motifs in the promoters of other genes (i.e. excluding known effectors).

To arrive at a refined list of putative DG effector proteins for future studies, we combined the predictive power of a variety of attributes including; the presence of a signal peptide for secretion;

the absence of transmembrane domain/s; and the presence of 1 or more cis-regulatory motifs within the first 1000 bp upstream of the start codon. A total of 457 candidates fulfilled these criteria. Of note, 77% also have a MERCI motif in the encoded protein sequence [18,34]. Putative DG effector candidates identified in this way were functionally annotated using Gene Ontologies. This revealed a significant enrichment in two biological processes (Proteolysis (GO:0006508) and Carbohydrate metabolic process (GO:0005975)), two molecular functions (Hydrolase activity, hydrolysing O-glycosyl compounds (GO:0004553), Cysteine-type peptidase activity (GO:0008234)), and two cellular components (Extracellular region (GO:0005576) and acetylcholine-gated channel complex (GO:0005892), Figure 6). Interestingly, and consistent with the previous descriptions, most novel predicted DG effector genes were associated with clusters C and G, respectively characterized by an up-regulation at ppJ2 and J3/J4 stages. Despite only ~26% of putative DG effectors (118/457) are functionally annotated with at least one GO term, these data are consistent with what is known about the functions of previously characterized effectors. The list of 457 predicted DG effectors, their functional annotation and full results of enriched GO terms are available at https://doi.org/10.15454/2O77EF.

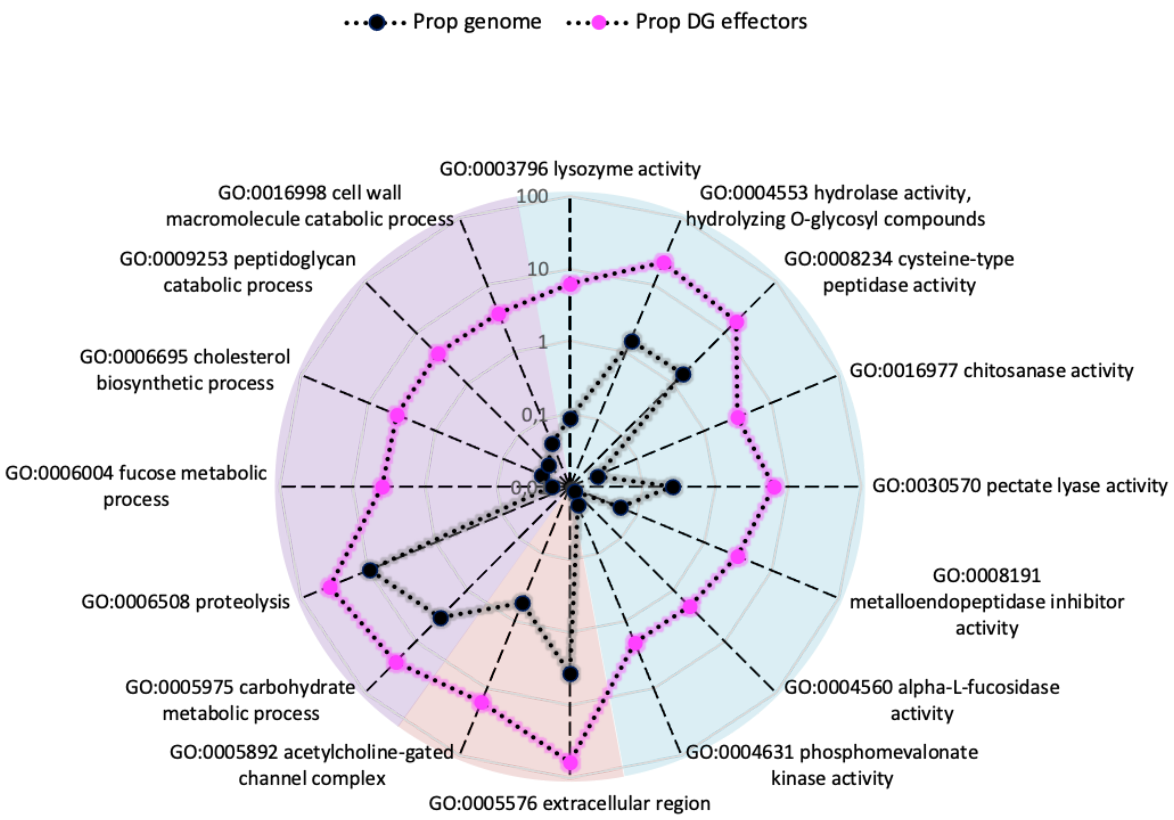

Figure 6. GO annotation enrichment of DG effectors in *M. incognita*. Radar plot with the log10FC of overrepresented GO terms related to all genes (Prop. Genome, black dotted line) and DG effectors (Prop. DG effectors; magenta dotted line) among three categories: biological process (purple pie chart), molecular function (blue pie chart) and cellular component (red pie chart).

Finally, to broaden the utility of this motif to identify novel putative effectors, the DG promoter enrichment analyses were repeated with four other species across the genus (Figure 7). Extremely similar motifs were identified as significantly enriched in 2kb promoter regions of *M. incognita* DG effectors mapped to the *M. arenaria, M. javanica*, and *M. enterolobii* genomes when compared to a

control set of non-effector promoters in each respective genome. The analysis could not be reliably reported for the promoters of *M. hapla* DG effectors because the number of *M. incognita* DG effectors unambiguously mapping to the genome (8) is too low to draw meaningful conclusions. Taken together, we conclude this promoter motif (termed Mel-DOG for 'Meloidogyne DOrsal Gland'), and thereby its utility to predict effectors, is broadly conserved across at least the Clade I root-knot nematodes.

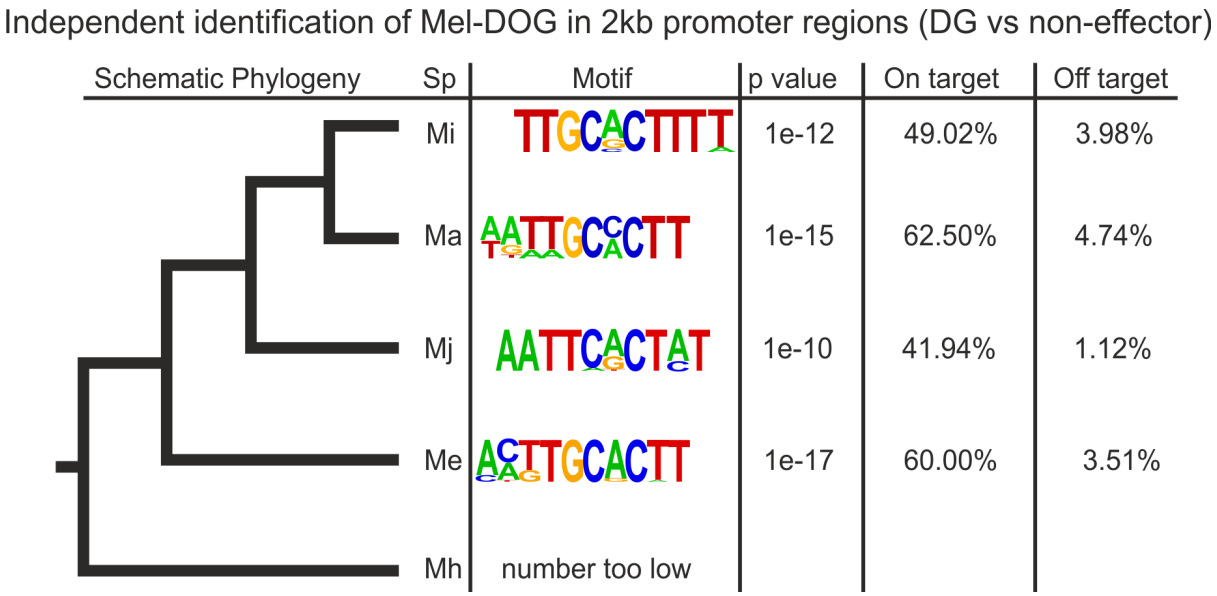

Figure 7. Independent identification of DG promoter motifs in the Meloidogyne genus. Left, a schematic representation of the phylogenetic relationship between *M incognita* (Mi), *M. arenaria* (Ma), *M. javanica* (Mj), *M. enterolobii* (Me), and *M. hapla* (Mh). Right, selected motifs identified as enriched in the promoter regions of DG effector genes when compared to a set of non-effectors. On target represents the proportion of DG effector genes that possess this motif in their 2kb upstream region. Off target represents the proportion of non-effectors that contain this motif in their 2kb upstream region.

Discussion

We performed a genome-wide survey of *M. incognita* transcription to identify trends in/causes of gene expression that likely contribute to nematode parasitism. These analyses enabled us to identify biological signatures of key developmental transitions during the lifecycle. We approach the question of parasitism gene regulation in spatio-temporal manner. Together, these results provide a step forward in our understanding of root-knot nematode transcriptional regulation by the identification of a putative cis-regulatory element "Mel-DOG" associated with DG effector gene expression in Meloidogyne spp.. By extension, these discoveries reveal an additional criterion on which to predict DG effectors in the most economically important plant-parasitic nematode genus.

Despite an increased amount of information regarding the mechanistic aspects of plant/nematode interaction, there is a paucity of molecular information about the parasite itself, i.e. the life stage specific gene expression in *M. incognita,* with a particular focus on parasitism genes [46,47]. We analyzed the transcriptome time-course of the RKN comprising four crucial stages of development: two pre-parasitic stages (eggs and ppJ2) and two parasitic stages (a mixture of J3-J4 larvae and adult females). Our data highlight gene functions that are consistent with the abilities of the nematode to complete critical transitions and ultimately become an enduring and successful parasite [4]. We thus followed the nematode's lifecycle from migratory to sedentary through and describe the key functional classes of genes employed.

At early stages (transition from eggs to ppJ2), *M. incognita* expresses genes involved in chemosensation and movement, supporting the host recognition and attraction. Induction of these groups of genes correlated with the preparation of parasitism have been reported in other PPNs, for example *Meloidogyne graminicola* [48], *Globodera pallida* [49] and the migratory endoparasite *Radopholus similis* [50]. RNAi mediated down-regulation of such chemosensory (e.g *odr* genes) or neuropeptide genes (e.g *npl-3* or *npl-12* genes) displayed a significant alteration of attraction and penetration of *M. incognita* ppJ2 to host roots, supporting their importance at early events of plant parasitism [51,52]. Similarly, we observed up-regulation of genes responsible for carbohydrate degradation as a signature of this transition. It is known that the motile J2 nematode produces and secretes a myriad of enzymes for plant cell wall softening/degradation during the invasion of root tissues [38,53,54]. Silencing of certain cell-wall degrading enzymes reduced root invasion by *M. incognita* [55,56] and other PPN species such as the migratory endoparasite *Pratylenchus coffeae* [57]*, Globodera rostochiensis* [58].

A second transcriptional step was observed from the early migratory to sedentary phases (ppJ2 to J3/J4). During this transition the juvenile undergoes physiological and developmental changes, highlighting that this transition is a turning point for the parasite. The downregulation of genes involved in chemosensory behavior suggests that once inside the root, the parasite no longer needs to follow attractive stimuli. The success of infection mostly relies on its ability to cope with the first line of plant defense and immunity and to move towards the host tissues [49,59–61]. At the J3/J4 stage, significantly up-regulated genes included putatively secreted proteins and enzymes such as superoxide dismutase, serine-type carboxypeptidase that may be involved in stress and defense responses. We hypothesise they may allow the protection of the endoparasite from the toxic effects of damaging ROS produced by the surrounding host plant cells, as supported by RNAi-mediated down-regulation of peroxiredoxins (PRXs) in *M. incognita* negatively impacting on its parasitic ability [62]. At this stage of infection, the juvenile probably does not acquire nutrient or energy sources from the plant host, but may instead utilize its lipid reserves [46,47,52] and also undergoes morphological changes (molting and reabsorption of the cuticle). Consistent with this, significantly enriched GO terms related to lipid metabolism were identified.

In comparison with the previous developmental transition, the transition from J3/J4 to adult female is marked a return to feeding and production of hundreds of eggs. A particularity of this transition was the down-regulation of genes associated with locomotion, supporting the idea that the parasitic worm does not need to move anymore once the feeding site is established. Conversely, we observed an upregulation of genes predicted to encode hydrolytic enzymes (serine-type endopeptidase) at this transition. While some of which may be involved in the evasion or suppression of plant immune responses [43,55,63] their roles are manifold, and require functional characterization to better understand their importance in nematode biology as well as in pathogenicity.

Finally, to complete *M. incognita* life cycle, comparison of the 'eggs' to the 'adult females' stages showed a diverse array of gene expression changes necessary for embryogenesis and, in a sense, marks the achievement of nematode parasitism with production of offspring. As reported in a previous study, our data analyses also suggest the differential regulation of genes involved in various basic biological processes (e.g. carbohydrate, amino acid, and lipid metabolism [64]), the semaphorin/plexin pathway, and cell adhesion and cuticle formation). At this step of lifecycle, a dampening of expression of genes related to stress responses occurs (e.g superoxide dismutase activity, oxidoreductase activity, L-ascorbic acid binding), suggesting that the host is no longer a threat for the parasitic animal, or the threat is no longer of concern since the adult female basically dies while extruding eggs in a gelatinous protective matrix on the root surface.

In addition to the DEG analysis at the transition between the four life stages, we also performed hierarchical clustering of DEGs to define and study groups of genes that share similar expression patterns during the parasite life cycle. DEG clusters showing peaks of expression at the pre-parasitic J2 and parasitic J3-J4 stages were associated with genes encoding for members of the *M. incognita* degradome [63]. Genes encoding a subset of enzymes, hydrolases (GO:0004553), pectate lyases (GO:0030570) and protein tyrosine/serine/threonine phosphatases (GO:0008138) were over-expressed at ppJ2 while the subset of glutathione hydrolases (GO:003674) and serine-type carboxypeptidases (GO:0004185) showed a peak of expression at J3/J4 stages. In agreement with previous reports, our data suggest that genes encoding metalloendopeptidases are likely to be involved in early events of *M. incognita'*s life cycle, with a significant upregulation at egg stage (cluster A) and conversely, a downregulation at J3-J4 stages (cluster G). Metallopeptidases are known to be mainly expressed at egg and ppJ2 in *M. incognita* [54], while others, such as cysteine peptidase, are highly expressed at later stages, both subsets of proteases are largely represented in the *M. incognita* genome [18,63,65,66]. The paramount importance of these enzymes has been previously proposed in nematode parasitism, for example a transmembrane metalloprotease Neprilysin is positively differentially expressed in *Heterodera avenae* during exposure to root exudates [67] and may be involved in the regulation of pathogenicity [68]. Known to be physiologically essential for many organisms including PPNs, but globally restricted to Metazoans, they are referred to as core parasitism genes [69] because they are associated with a role in tissue maceration/penetration and common to both parasitic nematodes of plants and animals [70–73]. In addition, these transcriptomic profiles associated with GO term analysis described in *M. incognita* mirrored similar studies in other PPNs, supporting the general commonalities that define the parasitic lifestyle [63,70,74], but the proportion and type of these enzymes vary according to the lifestyle [11,18,50]. Peptidases and CAZymes are both the result of a convergent adaptation in parasitic nematodes and an apparent ancient microbial origin acquired by multiple independent lateral gene transfers from different bacterial and fungal sources [37,38].

Like many other PPN species, most *M. incognita* known effectors are produced from the two main sets of secretory gland cells, DG and SvG, for which secretions are regulated at specific-stage during the lifecycle of the nematode. Previous works showed that SvG are more active during the nematode penetration, while DG is mostly involved in the formation and maintenance of nematode feeding

[75–77]. For the root-knot nematodes, it is not yet clear how evolutionarily conserved known effectors are across the genus, nor how the expression of effectors is distributed across the lifecycle of a species. To investigate evolutionary conservation of known *M. incognita* SvG and DG effectors, we mapped the corresponding CDS sequences to the genomes of *M. incognita* [16] and four other Meloidogyne species. Three species; *M. arenaria*, *M. javanica* and *M. enterolobii* [16,29] were from the same RKN Clade I [41], while *M. hapla* [30] was from Clade II and thus more distantly related. More than 87% (42/48) and 82% (28/34) of the known *M. incognita* SvG and DG effectors, respectively, could be traced back as inherited from a common ancestor of all the Clade I Meloidogyne investigated here. Alignment of effector CDS to the Clade II *M. hapla* genome allowed tracing back an even more ancient origin in a common ancestor of Clade I et II Meloidogyne for 52% (25/48) of SvG effectors but only above 23% (8/34) of DG effectors known in *M. incognita*. Only a more comprehensive sampling of Meloidogyne species in the different clades for genome sequencing will allow refining their precise evolutionary origin and path in the future.

The generally lower conservation of DG effectors compared to SvG effectors has several plausible explanations that are not necessarily mutually exclusive. Firstly, SvG effectors are generally deployed at the earlier stages of parasitism, are involved in migration through host tissue, and are conserved with the migratory endoparasites (i.e. Pratylenchus) that predate the genus [15]. These effectors are thus probably involved in core parasitic function common to multiple plant-parasitic nematodes and have probably evolved under purifying selection limiting divergence at the sequence level over evolutionary time. The DG effectors, in contrast, are generally accepted to be responsible for the precise manipulation of host physiology and development. They are therefore likely more recent, and under positive selection imposed by the different defence systems of the host. This might explain their lower degree of sequence-level conservation in the Meloidogyne genus and lack of identifiable orthologs in other species.

We subsequently cross referenced this comprehensive known effector catalogue with the gene expression data to describe patterns of transcription throughout the lifecycle of *M. incognita*. This revealed two main waves of effector expression in pre-parasitic and parasitic stages occur. Genes encoding known SvG effectors were mostly differentially expressed (81,3%) and up-regulated during the ppJ2 stage (clusters C and D), while 73,3% of DE DG effector genes were associated at the sedentary J3/J4 stage (clusters G and H), most of which were upregulated at this specific time point. Similar transcriptome analyses of the cyst nematode *Globodera pallida* showed three waves of effector expression, pre-parasitic (J2), early (7/14 dpi), and late (21-35 dpi), with a fourth category expressed throughout the biotrophic stages [11]. We observed that DG and SvG gene expression patterns coincide with those of certain subsets of degrading/modifying enzymes and those of stress responses which, intuitively, may act in concert to orchestrate the successful parasitic process [9]. In a previous study [56], transcriptional change of another set of CDWE effectors through has been demonstrated through RNAi silencing of pioneer MSP effectors, supporting a link between the expression of pioneer effectors and activities of such CDWEs in the RKN *M. incognita*. Overall, our analysis of the global gene expression coupled with GO annotation gives a more detailed view on the deployment of these effectors over time.

In addition to temporal control, the transcription of effectors is also precisely spatially regulated in the DG and SvGs. Recent exciting discoveries of cis-regulatory elements in PPNs associated with spatial regulation strengthen the interest to identify a putative regulatory promoter motif in the RKN *M. incognita* [11,14,45]. Using a combination of *in silico* and experimental transcriptomic analyses, we have thus identified for the first time a putative non-coding DNA motif: TGCMCTT (termed Mel-DOG) specifically enriched in the promoter regions of effector genes experimentally validated for dorsal expression in *M. incognita* (references in https://doi.org/10.15454/P5YIGX). Mel-DOG is conceptually similar, but sequence unrelated, to other cis-regulatory sequence motif among PPNs.

Cis-regulatory sequence motifs associated with gland cells do not appear to be conserved between the major lineages. For example, the DOG box, first identified in Globodera, is descriptive and predictive of effectors in cyst nematodes but not RKN nor pine wilt nematodes [11,13–15]) and the STATAWAARS motif first identified in *B. xylophilus* is descriptive and predictive of effectors in this species, but not cyst nematodes nor RKN [11,14,45]. Similarly, Mel-DOG appears to be associated with effectors in the RKN but not the cyst nor the pine wilt. Finally, Mel-DOG follows that trend that no motifs specifically associated with the SvG gland cells have been identified so far.

Following the identification of Mel-DOG, we used this motif as an additional criterion to predict effectors *in silico*. Consistently, Mel-DOG enrichment in the RKN genome was identified in promoters of genes encoding secreted proteins and comprising certain with specific functions in pathogenicity, such as terms related to CAZymes. For instance, a-L-Fucosidase (GH29) is the part of top 20 most represented CAZyme families from the genome-wide analysis of Excretory/Secretory proteins *M. incognita* [54] and from the secretome of the pine wilt nematode *B. xylophilus* [78,79]. Using hierarchical clustering, Mel-DOG was functionally enriched with DEGs from the cluster C, the main cluster comprising a major transcriptional shift in gene expression with a sharp peak at ppJ2, and although not statistically supported, Mel-DOG is also likely associated at a late stage of development (larval J3/J4 - cluster H). This data support the involvement of Mel-DOG in the transcriptional regulation of degrading/modifying enzymes, pointing to a role in tissue maceration/penetration and precising specific temporal window of Mel-DOG gene expression during *M. incognita*'s life cycle. Of importance, Mel-DOG is also associated with genes encoding putative secreted proteins with unknown function, in keeping with previous studies on other PPNs. Overall, the existence of non-coding DNA motifs associated with effector genes adds an important piece to the puzzle of parasitism gene regulation in PPNs. It remains undoubtedly essential to better understand the Mel-DOG function at a specific stage to further unveil its regulatory role during nematode parasitism and to extend this discovery by dissecting gene regulation networks that link with promoter motif. With respect to insights, one ongoing hypothesis is that a small number of regulators act in concert for effector expression, and thus, the gene disruption of these few regulators can simultaneously disrupt the function of hundreds of putative effectors and have a major effect on parasitism. Progress in understanding in effector biogenesis and biology is limited by a lack of tools for functional genetics on plant-parasitic nematodes, mostly due to their biology and gonad accessibility, but although this is hitherto challenging recent improvements on mRNA delivery appear realistic and promising [80]. The current approaches for their study are mostly based on the analyses of a single effector at a time through its overexpression in plants, target gene silencing via RNAi and the identification of target host proteins [9]. However, silencing of a single effector gene can be difficult to interpret because RNAi is never done at 100%, sometimes it can have a redundant role, and by consequence, its functional role may be not enough to completely alter the nematode parasitism [9,81–83]. Thus, targeting the regulators associated with Mel-DOG might represent an attractive and accurate target to develop novel nematode control strategies. For example, a well conserved Zn2Cy6 transcription factor (PnPf2) is involved in the expression of genes associated with nutrient assimilation and numerous effectors in the phyto-necrotrophic fungus *Parastagonospora nodorum*. Disrupting the transcription factor expression dramatically alters the plant cell wall degradation, axesual reproduction as well as virulence [84].

In addition to the potential for translation, the discovery of Mel-DOG provides additional criterion to predict and annotate effector repertoires in Meloidogyne genomes. Prior to this study, the prediction of nematode effector was based on identification of a putative N-terminal secretory peptide, no transmembrane region, and/or MERCI motif in their 100 first amino acids [18,34]. Thus the promoter motif complements the "roadmap" for the discovery of the PPN effectors in general and, for the first time it might be extended in the RKN *M. incognita* [85,86]. Here we combine our understanding of effectors to provide a high-confidence list of putative effectors for future validation and study.

Conclusion

The life stage specific transcriptome of *M. incognita* provides a global overview of the biological functions, and by extension may highlight the challenges associated with parasitism. We identify both a core set of effectors, and an associated promoter motif, that are broadly conserved across Clade I Meloidogyne spp.. The identification of the Mel-DOG box motif opens up a new way to study effector biogenesis in the root-knot nematodes, and provides an additional important criterion for the *in silico* prediction of effectors.

Authors Contributions

EGJD, CB, SEVDA: wrote the manuscript, designed the analysis, analyzed the results.
MDR: processed RNA-seq data and performed the differential gene expression analysis, analyzed the results.
MDR, JD, PG, PT, LPB: contributed to further versions of the manuscript.
JD, MBB: performed the clustering of genes according to gene expression data and statistical tests.
PG, EGJD: constructed the inventory of known DG and SvG genes, mapped them to RKN genomes and provided a curated list of corresponding gene models.
EGJD: performed the ancestral state reconstruction
SEVDA, PT, CP: performed the DG motif identification and enrichment analyses.
AP: performed the GO enrichment analysis.

Acknowledgement

Work on plant-parasitic nematodes at the University of Cambridge is supported by DEFRA licence 125034/359149/3, and funded by BBSRC grants BB/R011311/1, BB/N021908/1, and BB/S006397/1. CP and CB are supported by Marie Skłodowska-Curie Actions Individual Fellowships.

References

1. Decraemer, W.; Hunt, D.J. Structure and classification. In *Plant nematology*; CABI, 2013; pp. 3–39 ISBN 978-1-78064-151-5.
2. Jones, J.T.; Haegeman, A.; Danchin, E.G.J.; Gaur, H.S.; Helder, J.; Jones, M.G.K.; Kikuchi, T.; Manzanilla-López, R.; Palomares-Rius, J.E.; Wesemael, W.M.L.; et al. Top 10 Plant-Parasitic Nematodes in Molecular Plant Pathology. *Mol. Plant Pathol.* **2013**, *14*, 946–961, doi:10.1111/mpp.12057.
3. Elling, A.A. Major Emerging Problems with Minor Meloidogyne Species. *Phytopathology* **2013**, *103*, 1092–1102, doi:10.1094/PHYTO-01-13-0019-RVW.
4. Goverse, A.; Smant, G. The Activation and Suppression of Plant Innate Immunity by Parasitic Nematodes. *Annu. Rev. Phytopathol.* **2014**, *52*, 243–265, doi:10.1146/annurev-phyto-102313-050118.
5. Moens, M.; Perry, R.N.; Starr, J.L. Meloidogyne species - a diverse group of novel and important

plant parasites. In *Root-knot nematodes*; Perry, R.N., Moens, M., Starr, J.L., Eds.; CABI: Wallingford, 2009; pp. 1–17 ISBN 978-1-84593-492-7.
6. Abad, P.; Castagnone-Sereno, P.; Rosso, M.N.; Engler, J. de A.; Favery, B. Invasion, feeding and development. In *Root-knot nematodes*; Perry, R.N., Moens, M., Starr, J.L., Eds.; CABI: Wallingford, 2009; pp. 163–181 ISBN 978-1-84593-492-7.
7. Triantaphyllou, A.C.; Hirschmann, H. Post infection development of Meloidogyne incognita Chitwood 1949 (Nematoda-Heteroderidae). *Ann. Inst. Phytopathol. Benaki* **1960**, *3*, 1–11.
8. Baum, T.J.; Hussey, R.S.; Davis, E.L. Root-Knot and Cyst Nematode Parasitism Genes: The Molecular Basis of Plant Parasitism. *Genet. Eng. (N. Y.)* **2007**, *28*, 17–43, doi:10.1007/978-0-387-34504-8_2.
9. Mitchum, M.G.; Hussey, R.S.; Baum, T.J.; Wang, X.; Elling, A.A.; Wubben, M.; Davis, E.L. Nematode Effector Proteins: An Emerging Paradigm of Parasitism. *New Phytol.* **2013**, *199*, 879–894, doi:10.1111/nph.12323.
10. Davis, E.L.; Hussey, R.S.; Baum, T.J.; Bakker, J.; Schots, and A.; Rosso, M.-N.; Abad, and P. Nematode Parasitism Genes. *Annu. Rev. Phytopathol.* **2000**, *38*, 365–396, doi:10.1146/annurev.phyto.38.1.365.
11. Eves-van den Akker, S.; Laetsch, D.R.; Thorpe, P.; Lilley, C.J.; Danchin, E.G.J.; Da Rocha, M.; Rancurel, C.; Holroyd, N.E.; Cotton, J.A.; Szitenberg, A.; et al. The Genome of the Yellow Potato Cyst Nematode, Globodera Rostochiensis, Reveals Insights into the Basis of Parasitism and Virulence. *Genome Biol.* **2016**, *17*, 124, doi:10.1186/s13059-016-0985-1.
12. Eves-van den Akker, S. Plant–Nematode Interactions. *Curr. Opin. Plant Biol.* **2021**, *62*, 102035, doi:10.1016/j.pbi.2021.102035.
13. Masonbrink, R.; Maier, T.R.; Muppirala, U.; Seetharam, A.S.; Lord, E.; Juvale, P.S.; Schmutz, J.; Johnson, N.T.; Korkin, D.; Mitchum, M.G.; et al. The Genome of the Soybean Cyst Nematode (Heterodera Glycines) Reveals Complex Patterns of Duplications Involved in the Evolution of Parasitism Genes. *BMC Genomics* **2019**, *20*, 119, doi:10.1186/s12864-019-5485-8.
14. Espada, M.; Eves-van den Akker, S.; Maier, T.; Vijayapalani, P.; Baum, T.; Mota, M.; Jones, J.T. STATAWAARS: A Promoter Motif Associated with Spatial Expression in the Major Effector-Producing Tissues of the Plant-Parasitic Nematode Bursaphelenchus Xylophilus. *BMC Genomics* **2018**, *19*, 553, doi:10.1186/s12864-018-4908-2.
15. Vieira, P.; Maier, T.R.; Eves-van den Akker, S.; Howe, D.K.; Zasada, I.; Baum, T.J.; Eisenback, J.D.; Kamo, K. Identification of Candidate Effector Genes of Pratylenchus Penetrans. *Mol. Plant Pathol.* **2018**, doi:10.1111/mpp.12666.
16. Blanc-Mathieu, R.; Perfus-Barbeoch, L.; Aury, J.-M.; Rocha, M.D.; Gouzy, J.; Sallet, E.; Martin-Jimenez, C.; Bailly-Bechet, M.; Castagnone-Sereno, P.; Flot, J.-F.; et al. Hybridization and Polyploidy Enable Genomic Plasticity without Sex in the Most Devastating Plant-Parasitic Nematodes. *PLOS Genet.* **2017**, *13*, e1006777, doi:10.1371/journal.pgen.1006777.
17. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21, doi:10.1093/bioinformatics/bts635.
18. Grynberg, P.; Coiti Togawa, R.; Dias de Freitas, L.; Antonino, J.D.; Rancurel, C.; Mota do Carmo Costa, M.; Grossi-de-Sa, M.F.; Miller, R.N.G.; Brasileiro, A.C.M.; Messenberg Guimaraes, P.; et al. Comparative Genomics Reveals Novel Target Genes towards Specific Control of Plant-Parasitic Nematodes. *Genes* **2020**, *11*, 1347, doi:10.3390/genes11111347.
19. Li, B.; Dewey, C.N. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics* **2011**, *12*, 323, doi:10.1186/1471-2105-12-323.
20. Anders, S.; Pyl, P.T.; Huber, W. HTSeq—a Python Framework to Work with High-Throughput Sequencing Data. *Bioinformatics* **2015**, *31*, 166–169, doi:10.1093/bioinformatics/btu638.
21. Leng, N.; Dawson, J.A.; Thomson, J.A.; Ruotti, V.; Rissman, A.I.; Smits, B.M.G.; Haag, J.D.; Gould, M.N.; Stewart, R.M.; Kendziorski, C. EBSeq: An Empirical Bayes Hierarchical Model for Inference

in RNA-Seq Experiments. *Bioinformatics* **2013**, *29*, 1035–1043, doi:10.1093/bioinformatics/btt087.
22. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, doi:10.1186/s13059-014-0550-8.
23. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26*, 139–140, doi:10.1093/bioinformatics/btp616.
24. Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. InterProScan: Protein Domains Identifier. *Nucleic Acids Res* **2005**, *33*, W116-20.
25. Prüfer, K.; Muetzel, B.; Do, H.-H.; Weiss, G.; Khaitovich, P.; Rahm, E.; Pääbo, S.; Lachmann, M.; Enard, W. FUNC: A Package for Detecting Significant Associations between Gene Sets and Ontological Annotations. *BMC Bioinformatics* **2007**, *8*, 41.
26. Truong, N.M.; Nguyen, C.-N.; Abad, P.; Quentin, M.; Favery, B. Function of Root-Knot Nematode Effectors and Their Targets in Plant Parasitism. In *Advances in Botanical Research*; Elsevier, 2015; Vol. 73, pp. 293–324 ISBN 978-0-12-417161-9.
27. Leelarasamee, N.; Zhang, L.; Gleason, C. The Root-Knot Nematode Effector MiPFN3 Disrupts Plant Actin Filaments and Promotes Parasitism. *PLOS Pathog.* **2018**, *14*, e1006947, doi:10.1371/journal.ppat.1006947.
28. Iwata, H.; Gotoh, O. Benchmarking Spliced Alignment Programs Including Spaln2, an Extended Version of Spaln That Incorporates Additional Species-Specific Features. *Nucleic Acids Res.* **2012**, *40*, e161, doi:10.1093/nar/gks708.
29. Koutsovoulos, G.D.; Poullet, M.; Elashry, A.; Kozlowski, D.K.L.; Sallet, E.; Da Rocha, M.; Perfus-Barbeoch, L.; Martin-Jimenez, C.; Frey, J.E.; Ahrens, C.H.; et al. Genome Assembly and Annotation of Meloidogyne Enterolobii , an Emerging Parthenogenetic Root-Knot Nematode. *Sci. Data* **2020**, *7*, 324, doi:10.1038/s41597-020-00666-0.
30. Opperman, C.H.; Bird, D.M.; Williamson, V.M.; Rokhsar, D.S.; Burke, M.; Cohn, J.; Cromer, J.; Diener, S.; Gajan, J.; Graham, S.; et al. Sequence and Genetic Map of Meloidogyne Hapla: A Compact Nematode Genome for Plant Parasitism. *Proc Natl Acad Sci U A* **2008**, *105*, 14802–7.
31. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* **2019**, *20*, 238, doi:10.1186/s13059-019-1832-y.
32. Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y.C.; Laslo, P.; Cheng, J.X.; Murre, C.; Singh, H.; Glass, C.K. Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **2010**, *38*, 576–589, doi:10.1016/j.molcel.2010.05.004.
33. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: Scanning for Occurrences of a given Motif. *Bioinformatics* **2011**, *27*, 1017–1018, doi:10.1093/bioinformatics/btr064.
34. Vens, C.; Rosso, M.N.; Danchin, E.G. Identifying Discriminative Classification-Based Motifs in Biological Sequences. *Bioinformatics* **2011**, *27*, 1231–8, doi:10.1093/bioinformatics/btr110.
35. Maddison, W.P.; Maddison, D.R. *Mesquite: A Modular System for Evolutionary Analysis. Version 3.01.*; 2014;
36. Álvarez-Ortega, S.; Brito, J.A.; Subbotin, S.A. Multigene Phylogeny of Root-Knot Nematodes and Molecular Characterization of Meloidogyne Nataliei Golden, Rose & Bird, 1981 (Nematoda: Tylenchida). *Sci. Rep.* **2019**, *9*, 1–11, doi:10.1038/s41598-019-48195-0.
37. Danchin, E.G.; Rosso, M.N.; Vieira, P.; de Almeida-Engler, J.; Coutinho, P.M.; Henrissat, B.; Abad, P. Multiple Lateral Gene Transfers and Duplications Have Promoted Plant Parasitism Ability in Nematodes. *Proc Natl Acad Sci U A* **2010**, *107*, 17651–6.
38. Haegeman, A.; Jones, J.T.; Danchin, E.G. Horizontal Gene Transfer in Nematodes: A Catalyst for Plant Parasitism? *Mol Plant Microbe Interact* **2011**, *24*, 879–87, doi:10.1094/MPMI-03-11-0055.
39. Huang, G.; Gao, B.; Maier, T.; Allen, R.; Davis, E.L.; Baum, T.J.; Hussey, R.S. A Profile of Putative Parasitism Genes Expressed in the Esophageal Gland Cells of the Root-Knot Nematode Meloidogyne Incognita. *Mol Plant Microbe Interact* **2003**, *16*, 376–81.

40. Rutter, W.B.; Hewezi, T.; Abubucker, S.; Maier, T.R.; Huang, G.; Mitreva, M.; Hussey, R.S.; Baum, T.J. Mining Novel Effector Proteins from the Esophageal Gland Cells of Meloidogyne Incognita. *Mol. Plant. Microbe Interact.* **2014**, *27*, 965–974, doi:10.1094/MPMI-03-14-0076-R.
41. De Ley, I.T.; De Ley, P.; Vierstraete, A.; Karssen, G.; Moens, M.; Vanfleteren, J. Phylogenetic Analyses of Meloidogyne Small Subunit RDNA. *J Nematol* **2002**, *34*, 319–27.
42. Campbell, B.E.; Hofmann, A.; McCluskey, A.; Gasser, R.B. Serine/Threonine Phosphatases in Socioeconomically Important Parasitic Nematodes--Prospects as Novel Drug Targets? *Biotechnol. Adv.* **2011**, *29*, 28–39, doi:10.1016/j.biotechadv.2010.08.008.
43. Malagón, D.; Benítez, R.; Kasny, M.; Adroher, F.J. Peptidases in Parasitic Nematodes. A review. In *Parasites: Ecology, Diseases and Management*; 2013; pp. 61–102 ISBN 978-1-62257-692-0.
44. Hussey, R.S.; Davis, E.L.; Baum, T.J. Secrets in Secretions: Genes That Control Nematode Parasitism of Plants. *Braz. J. Plant Physiol.* **2002**, *14*, 183–194, doi:10.1590/S1677-04202002000300002.
45. Vieira, P.; Shao, J.; Vijayapalani, P.; Maier, T.R.; Pellegrin, C.; Eves-van den Akker, S.; Baum, T.J.; Nemchinov, L.G. A New Esophageal Gland Transcriptome Reveals Signatures of Large Scale de Novo Effector Birth in the Root Lesion Nematode Pratylenchus Penetrans. *BMC Genomics* **2020**, *21*, 738, doi:10.1186/s12864-020-07146-0.
46. McCarter, J.P.; Mitreva, M.D.; Martin, J.; Dante, M.; Wylie, T.; Rao, U.; Pape, D.; Bowers, Y.; Theising, B.; Murphy, C.V.; et al. Analysis and Functional Classification of Transcripts from the Nematode Meloidogyne Incognita. *Genome Biol.* **2003**, *4*, R26, doi:10.1186/gb-2003-4-4-r26.
47. Shukla, N.; Yadav, R.; Kaur, P.; Rasmussen, S.; Goel, S.; Agarwal, M.; Jagannath, A.; Gupta, R.; Kumar, A. Transcriptome Analysis of Root-Knot Nematode (Meloidogyne Incognita)-Infected Tomato (Solanum Lycopersicum) Roots Reveals Complex Gene Expression Profiles and Metabolic Networks of Both Host and Nematode during Susceptible and Resistance Responses. *Mol. Plant Pathol.* **2018**, *19*, 615–633, doi:10.1111/mpp.12547.
48. Petitot, A.-S.; Dereeper, A.; Da Silva, C.; Guy, J.; Fernandez, D. Analyses of the Root-Knot Nematode (Meloidogyne Graminicola) Transcriptome during Host Infection Highlight Specific Gene Expression Profiling in Resistant Rice Plants. *Pathogens* **2020**, *9*, 644, doi:10.3390/pathogens9080644.
49. Cotton, J.A.; Lilley, C.J.; Jones, L.M.; Kikuchi, T.; Reid, A.J.; Thorpe, P.; Tsai, I.J.; Beasley, H.; Blok, V.; Cock, P.J.A.; et al. The Genome and Life-Stage Specific Transcriptomes of Globodera Pallida Elucidate Key Aspects of Plant Parasitism by a Cyst Nematode. *Genome Biol.* **2014**, *15*, R43, doi:10.1186/gb-2014-15-3-r43.
50. Huang, X.; Xu, C.-L.; Yang, S.-H.; Li, J.-Y.; Wang, H.-L.; Zhang, Z.-X.; Chen, C.; Xie, H. Life-Stage Specific Transcriptomes of a Migratory Endoparasitic Plant Nematode, Radopholus Similis Elucidate a Different Parasitic and Life Strategy of Plant Parasitic Nematodes. *Sci. Rep.* **2019**, *9*, 6277, doi:10.1038/s41598-019-42724-7.
51. Dash, M.; Dutta, T.K.; Phani, V.; Papolu, P.K.; Shivakumara, T.N.; Rao, U. RNAi-Mediated Disruption of Neuropeptide Genes, Nlp-3 and Nlp-12, Cause Multiple Behavioral Defects in Meloidogyne Incognita. *Biochem. Biophys. Res. Commun.* **2017**, *490*, 933–940, doi:10.1016/j.bbrc.2017.06.143.
52. Shivakumara, T.N.; Dutta, T.K.; Mandal, A.; Rao, U. Estimation of Lipid Reserves in Different Life Stages of Meloidogyne Incognita Using Image Analysis of Nile Red-Stained Nematodes. *Nematology* **2019**, *21*, 267–274, doi:10.1163/15685411-00003212.
53. Bellafiore, S.; Shen, Z.; Rosso, M.-N.; Abad, P.; Shih, P.; Briggs, S.P. Direct Identification of the Meloidogyne Incognita Secretome Reveals Proteins with Host Cell Reprogramming Potential. *PLOS Pathog.* **2008**, *4*, e1000192, doi:10.1371/journal.ppat.1000192.
54. Gahoi, S.; Gautam, B. Genome-Wide Analysis of Excretory/Secretory Proteins in Root-Knot Nematode, Meloidogyne Incognita Provides Potential Targets for Parasite Control. *Comput. Biol. Chem.* **2017**, *67*, 225–233, doi:10.1016/j.compbiolchem.2017.01.014.
55. Antonino de Souza Júnior, J.D.; Ramos Coelho, R.; Tristan Lourenço, I.; da Rocha Fragoso, R.;

Barbosa Viana, A.A.; Lima Pepino de Macedo, L.; Mattar da Silva, M.C.; Gomes Carneiro, R.M.; Engler, G.; de Almeida-Engler, J.; et al. Knocking-Down Meloidogyne Incognita Proteases by Plant-Delivered DsRNA Has Negative Pleiotropic Effect on Nematode Vigor. *PloS One* **2013**, *8*, e85364, doi:10.1371/journal.pone.0085364.
56. Shivakumara, T.N.; Papolu, P.K.; Dutta, T.K.; Kamaraju, D.; Chaudhary, S.; Rao, U. RNAi-Induced Silencing of an Effector Confers Transcriptional Oscillation in Another Group of Effectors in the Root-Knot Nematode, Meloidogyne Incognita. *Nematology* **2016**, *18*, 857–870.
57. Bell, C.A.; Lilley, C.J.; McCarthy, J.; Atkinson, H.J.; Urwin, P.E. Plant-Parasitic Nematodes Respond to Root Exudate Signals with Host-Specific Gene Expression Patterns. *PLoS Pathog.* **2019**, *15*, e1007503, doi:10.1371/journal.ppat.1007503.
58. Rehman, S.; Butterbach, P.; Popeijus, H.; Overmars, H.; Davis, E.L.; Jones, J.T.; Goverse, A.; Bakker, J.; Smant, G. Identification and Characterization of the Most Abundant Cellulases in Stylet Secretions from Globodera Rostochiensis. *Phytopathology* **2009**, *99*, 194–202.
59. Espada, M.; Silva, A.C.; Eves van den Akker, S.; Cock, P.J.A.; Mota, M.; Jones, J.T. Identification and Characterization of Parasitism Genes from the Pinewood Nematode Bursaphelenchus Xylophilus Reveals a Multilayered Detoxification Strategy. *Mol. Plant Pathol.* **2016**, *17*, 286–295, doi:10.1111/mpp.12280.
60. Gillet, F.-X.; Bournaud, C.; Antonino de Souza Júnior, J.D.; Grossi-de-Sa, M.F. Plant-Parasitic Nematodes: Towards Understanding Molecular Players in Stress Responses. *Ann. Bot.* **2017**, *119*, 775–789, doi:10.1093/aob/mcw260.
61. Vieira, P.; Kamo, K.; Eisenback, J.D. Characterization and Silencing of the Fatty Acid- and Retinol-Binding Pp-Far-1 Gene in Pratylenchus Penetrans. *Plant Pathol.* **2017**, *66*, 1214–1224, doi:https://doi.org/10.1111/ppa.12664.
62. Dubreuil, G.; Deleury, E.; Magliano, M.; Jaouannet, M.; Abad, P.; Rosso, M.-N. Peroxiredoxins from the Plant Parasitic Root-Knot Nematode, Meloidogyne Incognita, Are Required for Successful Development within the Host. *Int. J. Parasitol.* **2011**, *41*, 385–396, doi:10.1016/j.ijpara.2010.10.008.
63. Castagnone-Sereno, P.; Deleury, E.; Danchin, E.G.; Perfus-Barbeoch, L.; Abad, P. Data-Mining of the Meloidogyne Incognita Degradome and Comparative Analysis of Proteases in Nematodes. *Genomics* **2011**, *97*, 29–36, doi:10.1016/j.ygeno.2010.10.002.
64. Subramanian, P.; Oh, B.-J.; Mani, V.; Lee, J.K.; Lee, C.-M.; Sim, J.-S.; Koo, J.C.; Hahn, B.-S. Differential Metabolic Profiles during the Developmental Stages of Plant-Parasitic Nematode Meloidogyne Incognita. *Int. J. Mol. Sci.* **2017**, *18*, doi:10.3390/ijms18071351.
65. 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.; et al. Genome Sequence of the Metazoan Plant-Parasitic Nematode Meloidogyne Incognita. *Nat. Biotechnol.* **2008**, *26*, 909–915, doi:10.1038/nbt.1482.
66. Li, X.; Yang, D.; Niu, J.; Zhao, J.; Jian, H. De Novo Analysis of the Transcriptome of Meloidogyne Enterolobii to Uncover Potential Target Genes for Biological Control. *Int. J. Mol. Sci.* **2016**, *17*, 1442, doi:10.3390/ijms17091442.
67. Chen, C.; Cui, L.; Chen, Y.; Zhang, H.; Liu, P.; Wu, P.; Qiu, D.; Zou, J.; Yang, D.; Yang, L.; et al. Transcriptional Responses of Wheat and the Cereal Cyst Nematode Heterodera Avenae during Their Early Contact Stage. *Sci. Rep.* **2017**, *7*, 14471, doi:10.1038/s41598-017-14047-y.
68. Sabeh, M.; Duceppe, M.-O.; St-Arnaud, M.; Mimee, B. Transcriptome-Wide Selection of a Reliable Set of Reference Genes for Gene Expression Studies in Potato Cyst Nematodes (Globodera Spp.). *PLOS ONE* **2018**, *13*, e0193840, doi:10.1371/journal.pone.0193840.
69. Yang, Z.; Wafula, E.K.; Honaas, L.A.; Zhang, H.; Das, M.; Fernandez-Aparicio, M.; Huang, K.; Bandaranayake, P.C.G.; Wu, B.; Der, J.P.; et al. Comparative Transcriptome Analyses Reveal Core Parasitism Genes and Suggest Gene Duplication and Repurposing as Sources of Structural Novelty. *Mol. Biol. Evol.* **2015**, *32*, 767–790, doi:10.1093/molbev/msu343.
70. Hunt, V.L.; Hino, A.; Yoshida, A.; Kikuchi, T. Comparative Transcriptomics Gives Insights into the

Evolution of Parasitism in Strongyloides Nematodes at the Genus, Subclade and Species Level. *Sci. Rep.* **2018**, *8*, 5192, doi:10.1038/s41598-018-23514-z.
71. Hunt, V.L.; Tsai, I.J.; Coghlan, A.; Reid, A.J.; Holroyd, N.; Foth, B.J.; Tracey, A.; Cotton, J.A.; Stanley, E.J.; Beasley, H.; et al. The Genomic Basis of Parasitism in the Strongyloides Clade of Nematodes. *Nat. Genet.* **2016**, *48*, 299–307, doi:10.1038/ng.3495.
72. Rai, K.M.; Balasubramanian, V.K.; Welker, C.M.; Pang, M.; Hii, M.M.; Mendu, V. Genome Wide Comprehensive Analysis and Web Resource Development on Cell Wall Degrading Enzymes from Phyto-Parasitic Nematodes. *BMC Plant Biol.* **2015**, *15*, 187, doi:10.1186/s12870-015-0576-4.
73. Tanaka, S.E.; Dayi, M.; Maeda, Y.; Tsai, I.J.; Tanaka, R.; Bligh, M.; Takeuchi-Kaneko, Y.; Fukuda, K.; Kanzaki, N.; Kikuchi, T. Stage-Specific Transcriptome of Bursaphelenchus Xylophilus Reveals Temporal Regulation of Effector Genes and Roles of the Dauer-like Stages in the Lifecycle. *Sci. Rep.* **2019**, *9*, 1–13, doi:10.1038/s41598-019-42570-7.
74. International Helminth Genomes Consortium Comparative Genomics of the Major Parasitic Worms. *Nat. Genet.* **2019**, *51*, 163–174, doi:10.1038/s41588-018-0262-1.
75. Davis, E.L.; Hussey, R.S.; Mitchum, M.G.; Baum, T.J. Parasitism Proteins in Nematode–Plant Interactions. *Curr. Opin. Plant Biol.* **2008**, *11*, 360–366, doi:10.1016/j.pbi.2008.04.003.
76. Jones, J.T.; Furlanetto, C.; Bakker, E.; Banks, B.; Blok, V.; Chen, Q.; Phillips, M.; Prior, A. Characterization of a Chorismate Mutase from the Potato Cyst Nematode Globodera Pallida. *Mol. Plant Pathol.* **2003**, *4*, 43–50.
77. Nguyen, C.-N.; Perfus-Barbeoch, L.; Quentin, M.; Zhao, J.; Magliano, M.; Marteu, N.; Rocha, M.D.; Nottet, N.; Abad, P.; Favery, B. A Root-Knot Nematode Small Glycine and Cysteine-Rich Secreted Effector, MiSGCR1, Is Involved in Plant Parasitism. *New Phytol.* **2018**, *217*, 687–699, doi:10.1111/nph.14837.
78. Cardoso, J.M.S.; Fonseca, L.; Abrantes, I. α-l-Fucosidases from Bursaphelenchus Xylophilus Secretome—Molecular Characterization and Their Possible Role in Breaking Down Plant Cell Walls. *Forests* **2020**, *11*, 265, doi:10.3390/f11030265.
79. Tang, J.; Ma, R.; Zhu, N.; Guo, K.; Guo, Y.; Bai, L.; Yu, H.; Hu, J.; Zhang, X. Bxy-Fuca Encoding α-L-Fucosidase Plays Crucial Roles in Development and Reproduction of the Pathogenic Pinewood Nematode, Bursaphelenchus Xylophilus. *Pest Manag. Sci.* **2020**, *76*, 205–214, doi:https://doi.org/10.1002/ps.5497.
80. Kranse, O.; Beasley, H.; Adams, S.; Silva, A.P. da; Bell, C.; Lilley, C.; Urwin, P.; Bird, D.; Miska, E.; Smant, G.; et al. Towards Genetic Modification of Plant-Parasitic Nematodes: Delivery of Macromolecules to Adults and Expression of Exogenous MRNA in Second Stage Juveniles. *bioRxiv* **2020**, 2020.07.15.193052, doi:10.1101/2020.07.15.193052.
81. Arroyo-Velez, N.; González-Fuente, M.; Peeters, N.; Lauber, E.; Noël, L.D. From Effectors to Effectomes: Are Functional Studies of Individual Effectors Enough to Decipher Plant Pathogen Infectious Strategies? *PLOS Pathog.* **2020**, *16*, e1009059, doi:10.1371/journal.ppat.1009059.
82. Iqbal, S.; Fosu-Nyarko, J.; Jones, M.G.K. Attempt to Silence Genes of the RNAi Pathways of the Root-Knot Nematode, Meloidogyne Incognita Results in Diverse Responses Including Increase and No Change in Expression of Some Genes. *Front. Plant Sci.* **2020**, *11*, doi:10.3389/fpls.2020.00328.
83. Lilley, C.J.; Davies, L.J.; Urwin, P.E. RNA Interference in Plant Parasitic Nematodes: A Summary of the Current Status. *Parasitology* **2012**, *139*, 630–640, doi:10.1017/S0031182011002071.
84. Jones, D.A.B.; John, E.; Rybak, K.; Phan, H.T.T.; Singh, K.B.; Lin, S.-Y.; Solomon, P.S.; Oliver, R.P.; Tan, K.-C. A Specific Fungal Transcription Factor Controls Effector Gene Expression and Orchestrates the Establishment of the Necrotrophic Pathogen Lifestyle on Wheat. *Sci. Rep.* **2019**, *9*, 15884, doi:10.1038/s41598-019-52444-7.
85. Eves-van den Akker, S.; Birch, P.R.J. Opening the Effector Protein Toolbox for Plant–Parasitic Cyst Nematode Interactions. *Mol. Plant* **2016**, *9*, 1451–1453, doi:10.1016/j.molp.2016.09.008.
86. Vieira, P.; Gleason, C. Plant-Parasitic Nematode Effectors — Insights into Their Diversity and New Tools for Their Identification. *Curr. Opin. Plant Biol.* **2019**, *50*, 37–43,

doi:10.1016/j.pbi.2019.02.007.