Article

Comparative genomics reveals novel target genes towards specific control of plant-parasitic nematodes

Priscila Grynberg[1,2], Roberto Coiti Togawa[1,2], Leticia Dias de Freitas[3], Jose Dijair Antonino[1,2,4], Corinne Rancurel[5], Marcos Mota do Carmo Costa[1], Maria Fatima Grossi-de-Sa[1,2,6], Robert N. G. Miller[2,3], Ana Cristina Miranda Brasileiro[1,2] Patricia Messenberg Guimaraes[1,2], Etienne G. J. Danchin[5*]

[1] Embrapa Recursos Genéticos e Biotecnologia, 70770-917, Brasília-DF, Brazil
[2] National Institute of Science and Technology – INCT PlantStress Biotech – EMBRAPA-Brazil
[3] Instituto de Ciências Biológicas, Universidade de Brasília, Campus Universitário Darcy Ribeiro, 70910-900, Brasília-DF, Brazil
[4] Departamento de Agronomia-Entomologia, Universidade Federal Rural de Pernambuco, 52171-900, Recife-PE, Brazil
[5] INRAE, Université Côte d'Azur, CNRS, ISA, Sophia-Antipolis, France
[6] Catholic University of Brasília, Genomic Sciences and Biotechnology PPG, 71966-700, Brasilia-DF, Brazil
[*] Correspondence: etienne.danchin@inrae.fr (E.G.J.D.)

Abstract: Plant-parasitic nematodes cause expressive annual yield losses to worldwide agricultural production. Most cultivated plants have no known resistance against nematodes and the few bearing a resistance gene can be overcome by certain species. The chemical methods that have been deployed to control nematodes were largely banned from use due to their poor specificity and high toxicity. Hence, there is an urgent need for the development of cleaner and more specific control methods. Recent advances in nematode genomics, including in phytoparasitic species, provide an unprecedented opportunity to identify genes and functions specific to these pests. Using phylogenomics, we compared 61 nematode genomes, including 16 for plant-parasitic species and identified more than 24,000 protein families specific to these parasites. In the genome of *Meloidogyne incognita*, one of the most devastating plant parasites, we found ca. 10,000 proteins with orthologs restricted only to phytoparasitic species and no further homology in protein databases. Among these phytoparasites-specific proteins, ca. 1,000 shared the same properties as known secreted effectors involved in essential parasitic functions. Of those, 68 were novel and showed strong expression during the endophytic phase of the nematode life cycle, based on both RNA-seq and RT-qPCR analyses. Besides effector candidates, transcription-related and neuro-perception functions were enriched in phytoparasites-specific proteins, revealing interesting targets for nematode control methods. This phylogenomics analysis, constitutes an unprecedented resource for the further understanding of the genetic basis of nematode adaptation to phytoparasitism and for the development of more efficient control methods.

Keywords: comparative genomics; plant-parasitic nematodes; phylogenomics; parasite-specific genes; pest control; de novo gene birth; horizontal gene transfers

1. Introduction

Plant parasitism has emerged at least four times independently in the phylum Nematoda and encompasses approximately 15% of the whole nematode biodiversity [1]. Collectively, phytoparasitic nematodes cause considerable damage to worldwide agriculture, although only a few species, mainly in the clade Tylenchida, are responsible for the majority of the losses. Root-knot nematodes (RKN) of the genus *Meloidogyne* are the most devastating plant-parasitic nematodes, with susceptible hosts covering more than 4,000 plant species, including the major economically important crops [2,3]. Among RKN, the tropical polyphagous apomictic species *M. incognita*, *M.*

javanica and *M. arenaria*, together with the facultative sexual species *M. hapla* from temperate regions, are considered the most damaging. Annual global crop production losses due to root infection by these sedentary endoparasites exceed 5%, equating to more than US$ 80 billion in lost revenue [4,5]. Most severe yield losses occur in tropical regions, with the RKN wide host range restricting options for disease control through crop rotation. Within RKN species, this wide host range is subdivided in 'host races' with distinct ranges of host compatibilities. Recent population genomics analyses showed that this division in 'host races' is not phylogenetically supported and most probably resulted from recurrent and independent adaptations [6]. Thus, the term "host race" is probably not well founded and the durability of crop rotation strategies is questioned. The employment of resistant cultivars to reduce the losses can also be limited, due to the occurrence of 'resistance-breaking' virulent RKN isolates and species mixtures [7]. Furthermore, due to their low specificity and potential risk for the environment and human health, chemical nematicides have mostly been banned from use across the world [8]. Hence, RKN control options are limited and advances in the development of novel sustainable and more specific control strategies are urgently required.

In RKN species, the infective second-stage juveniles (J2) penetrate the roots and migrate intracellularly to form specialized feeding cells which support the parasite throughout the sedentary life cycle. These different steps occur following the injection of secreted effector proteins produced in the nematode esophageal glands [3]. In the absence of host genes conferring resistance, nematode effectors will induce host cell differentiation into giant cells, which undergo nuclear enlargement, organelle proliferation, metabolic activation, cell-cycle alterations and cell-wall changes [9]. This continued delivery of effector proteins modulates suppression of plant defenses and the full development of the giant cell, mediating the feeding process and the resultant metabolic drain on the host [10].

Effectors of RKN and other plant-parasitic nematodes (PPN) target important host molecular components to facilitate parasitism, being involved in the degradation of the plant cell wall, uptake and processing of plant nutrients, manipulation of the plant defense system or in the establishment of feeding structures [11,12]. The evolutionary origin of nematode effectors has been only poorly investigated so far. Nonetheless, previous studies showed that certain effector proteins, including those involved in the degradation of the plant cell wall (e.g. cellulases, pectinases), are conserved in many phytoparasitic nematodes and seem to have been acquired via horizontal gene transfer (HGT) from bacteria and fungi [13,14]. Still, the majority of known and candidate effectors represent so-called 'orphan' proteins, lacking any recognizable homology in species other than the PPN [11]. Even when considering different species within the clade Tylenchida alone, besides the genes acquired in a common ancestor via HGT, there is little conservation between effector sets. For instance, a comparison of the effector sets of RKN and the cyst nematode *Globodera pallida*, showed only little overlap [15]. Despite lineage-specific genes and those acquired via HGT, some effectors can also emerge from evolutionary conserved genes and acquire modifications in transcriptional regulation or protein localization signals to eventually be co-opted for new parasitic-related functions [16]. Several effectors form expanded multigene families, probably as a result of positive selective pressure imposed by plant hosts [12,13].

Given their role in parasitism, one compelling method towards combating PPN consists in silencing the expression of effector genes [17–19]. Expanding the effector protein repertoire and dissecting functions is therefore important for the development of novel nematode control strategies.

Over the years, the availability of several RKN genomes provided new resources for the identification of complete sets of candidate effectors in these species [20–29]. Although nematode effectors are very diverse, many share some common features with other secreted animal proteins, such as the presence of N-terminal signal peptides for secretion, or the absence of transmembrane domains. These two features can be used as signatures for bioinformatics genome mining of putative secreted proteins (PSP) but the inclusion of other characteristics (e.g. gene expression patterns) is also required to identify candidate effectors among PSP. Furthermore, identification of signal

peptides for secretion is dependent on the quality of gene prediction (particularly at the 5' end) in the genomes under consideration, and certain known effector proteins do not bear signal peptides at all. For these reasons, complementary approaches have compared known effectors to negative sets of presumably non-effector proteins, in order to identify either an ensemble of characteristics of effectors via machine learning, for example in cyst nematodes [30], or degenerate protein motifs specific to RKN effectors [31].

Predictions of candidate effectors in the first available RKN genomes (*M. incognita* and *M. hapla*) were initially conducted using the presence of a signal peptide for secretion, of an RKN-specific motif and the absence of transmembrane domain as search filters [32]. Cross-referencing of this list with sets of proteins specific to RKN that were identified through a comparative genomics approach led to the identification of 15 novel candidate effectors without homologs in other species. RNA interference (RNAi) experiments subsequently showed a significant reduction of nematode infection symptoms in tomato host plants for 12 of these novel candidate effectors. Being specific to phytoparasites, while displaying the same characteristics to known effectors, and yielding reduced nematode infection symptoms following silencing, these genes constitute promising novel targets for the development of more specific and environmentally friendly control methods. Searching for genes specific to phytoparasites constituted an original strategy in comparison to the other studies, which rather started from *Caenorhabditis elegans* genes known to yield severe phenotypes when inactivated and searched for orthologs in plant parasitic nematodes [33]. Although this latter strategy maximizes the chance of severely affecting the PPN, their conservation across distantly related nematodes increases the risk of unintentional effects on the local nematofauna and possibly even beyond.

Although original, the comparative genomics study published in 2013 was limited to the only two genomes for plant-parasitic species available at the time (two RKN). Furthermore, the *M. incognita* genome employed in this original study [20] was later shown to be partially incomplete, with an improved version published afterwards [23]. Meanwhile, the availability of genome sequences for additional RKN species [24,26–29], as well as for other plant-parasitic nematode species, including cyst nematodes [15,34,35], stem and bulb nematodes [36,37], or the more distantly related pine wilt disease nematode [38], greatly expanded the diversity of genomes for plant-parasitic nematode species. Similar genome sequencing efforts were also accomplished in the other nematode clades, including in animal parasites and free-living species, offering a novel opportunity for nematode comparative genomics. Recently, a massive comparative genomics analysis of nematodes, centered on human and other animal parasites revealed the nematode-wide patterns of gene conservation, birth and loss, and allowed establishment of a prioritized list of new potential drug targets to combat these parasitic worms [39].

In the present study, we compared the genomes of 60 nematode species, carefully selected based on the completeness and quality of their predicted protein sets, to identify proteins exclusive to PPN. Our study encompassed 16 PPN species, and represents an unprecedented insight into phytoparasite genomics, as previous comparative studies used only two [32] or three [39] PPN genomes. We selected the genome of *M. incognita*, arguably one of the most damaging PPN, to further identify proteins having the same characteristics to known effectors, and thus possibly involved in plant parasitism. We employed a bioinformatics functional annotation and enrichment analysis to assess whether certain functional categories were significantly enriched in the set of PPN-specific genes and in candidate effectors. Cross-referencing the list of candidate *M. incognita* effectors with PPN-specific proteins, we aimed to identify novel RKN proteins amenable for the development of specific and sustainable control methods. Because *in-planta* RNAi-based strategies constitute a promising approach, we then further analyzed life-stage specific RNA-seq data in *M. incognita* to identify candidate effector genes expressed in the endophytic phases of the life cycle, and used qRT-PCR to confirm the expression patterns and validate them as promising targets. This comparative analysis constitutes a valuable resource for identification of new drug targets and to better understand how PPN manipulate their hosts.

2. Materials and Methods

2.1. Orthogroups inference

A total of 2,228,060 proteins were downloaded from 79 nematode genomes and two tardigrade outgroup species. We assessed the proteomes completeness using BUSCOv3, according to the Eukaryotic odb9 dataset [40], setting up a minimal threshold of 70% complete BUSCO genes for inclusion in further analysis. For each species, when multiple proteome versions were available, only those with the highest BUSCO completeness scores were retained. This resulted in a selection of 63 protein sets covering 1,607,105 proteins, from nematodes with free-living, plant, vertebrate and insect parasite lifestyles, as well as two outgroup free-living tardigrade species. We employed OrthoFinder (version 2.3.3) with default inflation parameters [41,42] to define orthogroups between the 63 different species, based on an all-against-all comparison of the proteomes with Diamond [43] run in sensitive mode. The list of analyzed species, with BUSCO completeness scores, sequences per species, nematode lifestyles, as well as source links for data download and references, is available at https://doi.org/10.15454/IIAQOW.

We used the OrthoFinder option -m to generate multiple sequence alignments for each orthogroup using MAFFT [44], with phylogenetic trees generated with FastTree [45]. A species tree was inferred from gene trees using STAG, with the tree rooted using STRIDE [42]. The rooted species tree is available at https://doi.org/10.15454/MBMOQC. Global statistics of the OrthoFinder analysis and full results are available at https://doi.org/10.15454/ZGUP7N and https://doi.org/10.15454/ZAYJBC, respectively.

2.2. Identification of M. incognita putative secreted proteins and candidate effectors

Effector proteins are secreted by the nematode into the plant. In order to identify putative secreted proteins (PSP) in the *M. incognita* protein set, we searched for signal peptides as evidence for secretion using SignalP v4 [46]. This program improves the differentiation between transmembrane regions present in the N-terminal part of a protein and the presence of a signal peptide. To further circumvent this potential problem, predicted signal peptides were cropped, with the remaining part of the proteins subjected to TMHMM [47] for prediction of transmembrane regions. All proteins bearing a predicted signal peptide and no transmembrane region were considered as putative secreted proteins (PSP). Accession numbers of the proteins satisfying these two criteria are available at https://doi.org/10.15454/JCYZDI.

Pearson's Chi-squared test with Yates' continuity correction was applied to test whether the proportion of PSP was significantly different in PPN-specific predicted proteins vs. the whole predicted proteome. To further refine this set and identify candidate effectors, we searched for the presence of motifs enriched in known effectors in *M. incognita* [31]. We employed the MERCI software [31] to scan the whole set of *M. incognita* predicted proteins for the presence of at least one of the four known enriched motifs in the 100 first amino acids (list available at https://doi.org/10.15454/LMY6LV).

2.3. Taxonomic distribution of M. incognita homologs and detection of horizontal gene transfers

All *M. incognita* predicted proteins were compared against the NCBI's nr library using Diamond [43], with an E-value threshold of 0.01. In order to explore the taxonomic distribution of nr hits matching *M. incognita* proteins, we employed the taxonomic assignment algorithm of Diamond, based on inference of the last common ancestor (LCA) of hits returning a Diamond score not diverging by more than 10% from the score of the best hit.

We used Alienness [48] to detect putative horizontal gene transfers (HGT). Alienness parsed Diamond homology search results and retrieved taxonomic information associated with hits. The best metazoan and best non-metazoan hits were then identified to compute an Alien Index (AI), with self-hits to Tylenchomorpha (NCBI:txid33283) ignored. Proteins returning an AI>0 have higher similarity to non-metazoan than metazoan hits in nr and can include candidate HGT of non-metazoan origin. Proteins having an AI>0 and $\geqslant$70% identity with a non-metazoan hit were considered putative contaminants and eliminated from the list of candidate HGT. As AI$\geqslant$14

represents a good balance between precision and recall of phylogenetically confirmed HGT [48], we considered *M. incognita* proteins returning an AI⩾14 as likely to result from acquisition via HGT.

We then employed Pearson's Chi-squared test with Yates' continuity correction to test whether the proportion of putative HGT (AI>0 and AI⩾14) were significantly different in PPN-specific predicted proteins vs. the whole predicted proteome. Finally, we identified enriched GO terms associated with candidate HGTs, using the same methodology explained above. Results of the whole Alienness analysis are available at https://doi.org/10.15454/W6SWZH.

2.4. Functional annotation and Gene Ontology terms enrichment

All the predicted proteins from the *M. incognita* genome [23] were analyzed by InterProScan (v5.29-68.0) [49] for identification of conserved protein domains, with the option -iprlookup -goterms employed to assign Gene Ontology (GO) terms from the identified InterPro domains. Raw results of InterProScan annotation are available at https://doi.org/10.15454/9BFFKG.

We used the hypergeometric test within the FUNC program [50] to determine whether certain GO terms assigned via InterProScan annotation were significantly enriched in several subsets of *M. incognita* proteins. We made two independent analyses: (i) PPN-specific proteins - (a) PPN-specific proteins, (b) PPN-specific PSP and (c) PPN-specific candidate effectors and (ii) Proteins originating from HGT – (a) Possible HGT (AI>0) and (b) Likely HGT (AI⩾14).

We considered as significantly enriched those terms returning a false discovery rate (FDR) lower than 0.05 and a minimum of four genes in node. When overlaps between proteins associated with several enriched GO were identified, these proteins were joined together into representative groups. Each group was given a more general term that represents and / or explains the biological processes and / or molecular functions in which they participate. Poorly informative general terms (higher in the hierarchy, usually inferred by FUNC) were further excluded from the analysis when the majority of the corresponding proteins in a node were spread out across more specific GO terms. GO terms fold-enrichment (FE) values were calculated as the ratio between the observed and the expected genes frequency. The R packages ggplot2 [51] and pheatmap were used to plot a bubble chart and heatmaps of enriched GO terms. Rows were colored based on the GO terms and the above mentioned subsets of *M. incognita*.

We also used OrthoFinder results to assess whether *M. incognita* proteins corresponding to enriched GO terms were further conserved between RKN and other PPN (e.g. cyst, burrowing, and migratory plant parasitic nematodes). Because some genomes are polyploids, the raw numbers of proteins in each orthogroup were not taken into account for this analysis. Results were plotted into a heatmap using R's pheatmap package.

2.5. In-silico expression patterns of protein-coding genes throughout M. incognita life stages

In order to evaluate expression patterns of protein-coding genes, we analyzed data from a previously published life-stage-specific RNA-seq analysis of *M. incognita* during infection in tomato [23]. This analysis encompassed four different life stages sequenced in triplicate: (i) eggs, (ii) pre-parasitic second stage juveniles (J2), (iii) a mix of late parasitic J2, third stage (J3) and fourth stage (J4) juveniles, and (iv) adult females. Stages (i) and (ii) correspond to exophytic phases while (iii) and (iv) are endophytic phases. The original cleaned RNA-seq reads [23] were retrieved and re-mapped to the *M. incognita* annotated genome assembly, using a more stringent end-to-end option (i.e. no soft clipping) in 2-passes as explained in [52]. Expression data, as FPKM values, are available at https://doi.org/10.15454/YM2DHE.

2.6. Candidate effector gene expression validation

2.6.1. Nematode population and plant inoculation

A population of *M. incognita* race 3, previously identified using esterase and SCAR markers [53], was maintained on tomato (*Solanum lycopersicum* 'Santa Cruz') in greenhouse conditions. *M. incognita* eggs were extracted from 60 day-old infected roots according to ref [54], hatched *in vitro* and pre-parasitic J2 concentrations adjusted to 500/mL.

Bioassays were performed in triplicates, with twenty-day-old tobacco plants (*Nicotiana tabacum*) inoculated with 1,000 pre-parasitic J2, maintained at 25 °C, 12/12 h light/dark regime, and watered at 3-day intervals. Plants were harvested at 5, 10, 15 and 22 days after nematode inoculation (DAI). Roots were dissected under a binocular microscope to identify and collect RKN galls. Galls were immediately frozen in liquid nitrogen and maintained at -80 °C for RNA extraction.

2.6.2. Total RNA extraction and cDNA synthesis

Total RNA from eggs and pre-parasitic J2 were extracted using a DirectZol extraction kit (Zymo Research, Irvine, CA, USA), while for RNA extraction from *N. tabacum* root gall tissues at 5, 10, 15 and 22 DAI, we used the Quick-RNA™ Plant Miniprep (Zymo Research, USA), according to each manufacturer's standard protocol. After RNA extraction, samples were DNase I treated and cDNA synthesized using SuperScript II (Invitrogen, Carlsbad, CA, USA) and 1 μg of total RNA from each sample.

2.6.3. RT-qPCR analysis of expression patterns

The relative expression patterns of seven genes encoding candidate effectors, throughout the nematode infective life cycle, were determined by reverse transcription quantitative PCR (RT-qPCR) assays. Each reaction was performed in a final volume of 10 μL, containing 5 μL of iTaq™ Universal SYBR® Green qPCR mix (Bio-Rad, Hercules, CA, USA), 0.2 μM of each primer and 2 μL of cDNA (diluted 1:20). PCR amplifications were conducted on the StepOne Real-Time PCR Systems (Applied Biosystems, Foster City, CA, USA) using the following program: 52 °C for 2 min, 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. Primers were designed using the Primer3Plus software [55] using the default qPCR parameters (Table 1). Primer specificity was assessed by analyzing the dissociation curve of amplified products using SDS software, version 2.2.2 (Applied Biosystems, USA) and primer efficiencies were calculated using the online real-time PCR Miner tool [56]. The relative expression (RQ) of mRNA levels of target genes was calculated in samples using two reference genes (β-actin and 18S ribosomal subunit) by the SATqPCR web tool [57]. Experiments were conducted in biological and technical triplicate and the RQ was statistically tested by the analysis of variance (ANOVA) and Tukey's test ($P<0.05$).

Table 1. Primers used for RT-qPCR analysis.

Candidate effector	Primer sequence (5′ - 3′) Forward/Reverse	Amplicon size (bp)	Primer efficiency
Minc3s00010g00668	CTTGGTACCCTTACCGACCA/ AAGTCGCACTTGTTCTGGTG	110	0.91
Minc3s01635g25253	TGCAAAGATGGGTGTACAAT/ TTTGCACATCGCTTTCTCAC	113	0.94
Minc3s00010g00666	ACCGATTCGTTCAGTTCCAG/ CCTCATTATCCATCGGTGCT	89	0.91
Minc3s01217g21793	CAGGACGTTCGGTTCCAATA/ TGCTGTGGCATGACGTTTAG	122	0.89
Minc3s03880g35054	CGAAATGGGCGTAGAAAATG/ GTCGGCCATGTGGTACTTCT	83	0.88

Minc3s00020g01309	CCCAAAGCAATGCAACATAA/ ACGAATGTGCCGAAGAGAAT	97	0.87
Minc3s01206g21700	TCCAAATTGCGTGGTAGACA/ GCTTGTGAAATGCGTCAGAA	104	0.88
Minc3s00292g09561	AAGAGGAGTGTGGGGTTGTG/ GCTTGGAAGAATTGGACGAC	94	0.89

3. Results

3.1. The comparative genomics framework covers a diversity of nematode clades and lifestyles

A total of 63 proteomes predicted from 61 nematode and two outgroup tardigrade genomes passed the BUSCO eukaryota completeness inclusion threshold of 70% (annotated species list available at https://doi.org/10.15454/IIAQOW).

The 61 nematode genomes covered species in four of the five nematode clades (I, III-V) defined by Mark Blaxter and colleagues [58,59] and seven of the twelve nematode clades (2, 6, 8-12) defined by Hans Helder and colleagues [1,60]. Hereafter, B-clades and H-clades will refer to the clades defined by Mark Blaxter *et al.* and by Hans Heleder *et al.*, respectively. To date, no genome sequences are available for nematodes either in Enoplia (H-clade 1, B-clade II) or in H-clades 3-5 and 7.

In terms of lifestyle, the selected nematode genomes encompassed nine free-living species, distributed in H-clades 6, 9, 10b and 11, three insect parasites, in H-clades 2 and 10a, 32 vertebrate parasites, in H-clades 2, 8, 9 and 10b, as well as 16 plant parasites, in H-clades 10d and 12 (Figure 1). Although plant parasitism has emerged at least four times independently in the phylum Nematoda, in H-clades 1, 2, 10 and 12 [1,61], genomes for PPN are currently only available in H-clades 10d and 12. While H-clade 10d is only represented here by the pine wilt disease agent *Bursaphelenchus xylophilus*, PPN in H-clade 12 now benefit from a much larger sampling, with genomes characterized for sedentary endoparasitic root-knot nematodes (*Meloidogyne* spp) and cyst nematodes (*Heterodera* spp and *Globodera* spp), semi-endoparasitic reniform nematodes (*Rotylenchulus* spp) and migratory endoparasitic nematodes (*Radopholus* spp and *Ditylenchus* spp).

3.2. The phylogenomics analysis is consistent with the Nematode phylogeny

The species tree produced by OrthoFinder (Figure 1) was almost identical to the most recent phylogenomics analysis based on concatenated evolutionarily conserved protein coding genes [62] and consistent with the nematode classifications described by Mark Blaxter *et al.* [58,59] and Hans Helder *et al.* [1,60]. As expected, the two tardigrades were positioned as a monophyletic outgroup to the nematodes (Figure 1). Monophyly of all the nematode B-clades covered by our analysis (I, III, IV, V) was respected with strong STRIDE support values (all branch support values =1). Similarly, all the H-clades covered were monophyletic and supported by the same strong values, with the exception of H-clade 10, which was split into 4 subclades, exactly as in [62]. In addition to the clades monophyly, their relative branching orders were also consistent with the reference phylogeny and classification.

Considering that the Tylenchida group (H-clade 12) contains the most damaging PPN, we analyzed this clade in detail. In this group, the Anguinidae (*D. destructor* and *D. dipsaci*) hold the earliest branching position, in accordance with previous phylogenetic analyses [1,60,62]. A clear separation was also observed between the Meloidogynidae (RKN) and a group composed of the Heteroderidae (cyst nematodes) and their closest available relatives, namely the Rotylenchulidae (*R. reniformis*) and the Radopholinae (*R. similis*). Within the Meloidogynidae, a correct separation was observed between the three different RKN clades (I, II, III), as originally defined by Tandingan De Ley [63], and the branching order of these clades was respected.

Overall, the OrthoFinder-based tree topology received very high branch-level support values throughout, with most being higher than 0.9 (Figure 1). Only the relative positions of RKN within clade I displayed lower values, although the monophyly of this clade was itself supported by a value of 1.0 (see Discussion section for an explanation).
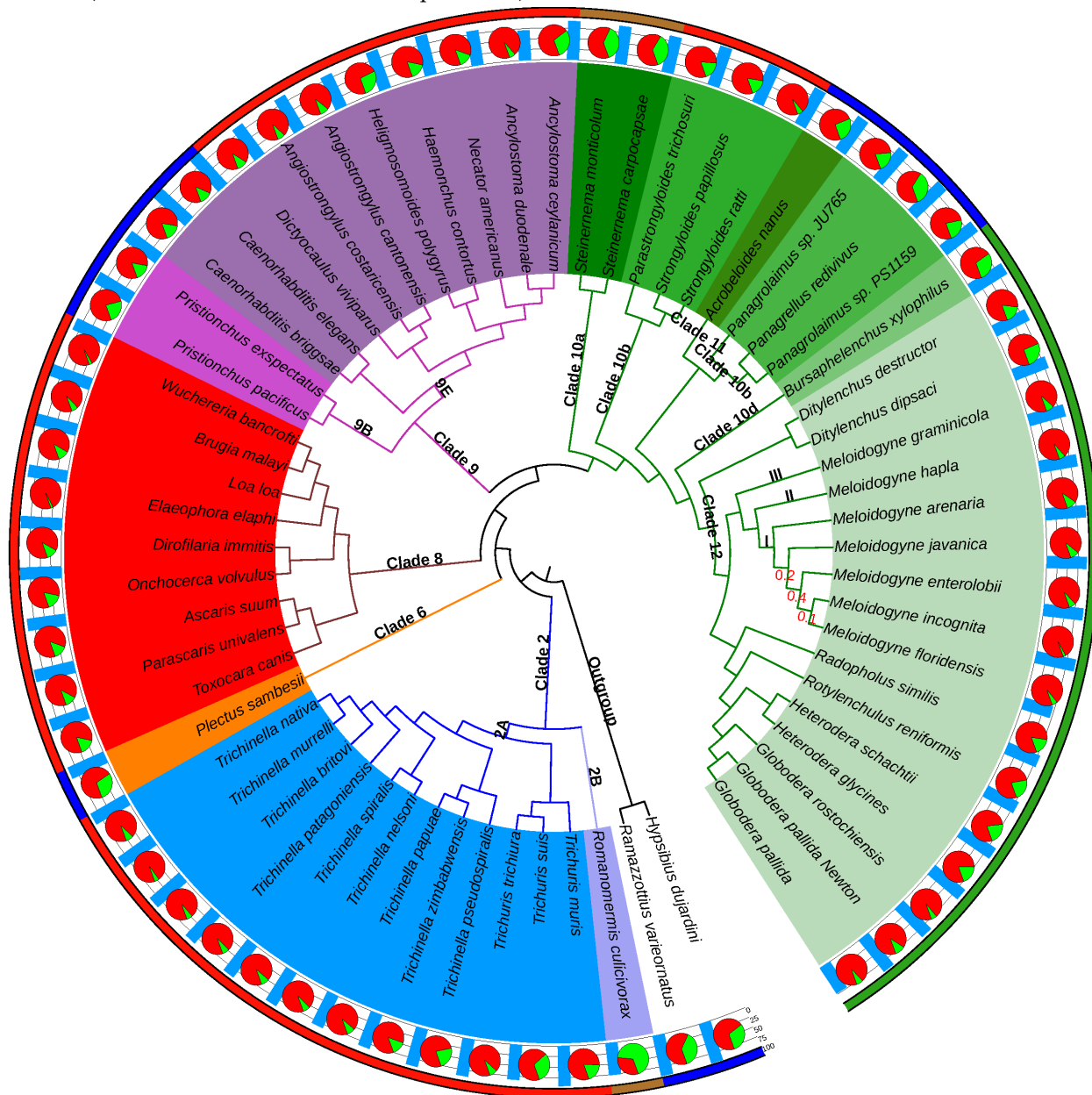

Figure 1. Annotated phylogenomic tree of the compared nematode genomes. Topology of the phylogenetic tree obtained from the OrthoFinder comparative analysis of 61 nematode genomes and two outgroup tardigrade genomes. Latin species names are given at the leaves. STRIDE support values <0.9 are provided at the corresponding branches in red (all the other values are ≥0.9). Nematode clade numbers, according to the Helder classification (H-clades), are provided at the corresponding branches. Roman numbers within clade 12, correspond to the RKN clades, according to De Ley [63]. For each species, the pie-charts represent the relative proportion of species-specific proteins (green) and proteins conserved in at least another species (red). The light blue bar histogram represents the percentage of complete eukaryotic BUSCO proteins in each proteome. The color-coded outermost circle is according to the species lifestyle (blue: free-living, brown: insect parasite, green: plant parasite, red: vertebrate parasite). This figure was generated with iTOL[64].

The overall high support of this phylogenomics-based tree topology, which is consistent with reference classifications, indicates that the phylogenetic signal present in our comparative genomics analysis allowed an accurate reconstruction of the nematode tree of life, therefore enabling a valid subsequent evolutionarily analysis.

3.3. Orthogroups encompass the vast majority of nematode proteins and reveal genes specific to plant parasites

More than 82.8% of the proteins from the 61 nematode and 2 tardigrade genomes (1,330,958/1,607,105 in total) were classified into 71,576 orthogroups by OrthoFinder (full results available at https://doi.org/10.15454/ZAYJBC). The median and average proportions of proteins assigned to orthogroups per species were 86.2 and 84.4%, respectively (full statistics available at https://doi.org/10.15454/ZGUP7N), therefore encompassing most of the diversity of proteins present across the different species selected for this analysis.

Overall, 289,165 proteins (<18% of the total) were species-specific and decomposed into two categories: (i) the great majority (276,147 proteins) could not be assigned to any orthogroup and represented species-specific singletons, while (ii) 13,018 proteins were assigned to single-species orthogroups and represent species-specific multicopy proteins. However, the proportion of species-specific proteins is not homogeneously distributed across the phylogenomics tree; with species being the sole representatives for their clades showing higher proportions of species-specific genes than species from well represented clades. For instance, with 68% of its proteins being species-specific, the most extreme case is *Romanomermis culicivorax*, the sole representative of H-clade 2B (Figure 1). The other species from poorly represented branches also tend to show higher proportions of species-specific genes (e.g. *Plectus sambesi*, the only representative of H-clade 6, has 29% of species-specific genes, and the two *Steinernema* species that are the sole representatives of H-clade 10a both have >35% of species-specific genes). In contrast, species from well represented clades showed a very low proportion of species-specific genes. For instance, with seven species from the RKN genus present, *M. incognita* has only 2.5% of species-specific genes. Similarly, *Elaeophora elaphi*, from H-clade 8, which is represented by 9 species, showed only 2.1% of species-specific genes.

As PPN species are the main focal point of our study, we explored these data more closely. We identified 24,397 orthogroups (covering 155,876 proteins) that were exclusive to PPN. Only five of these orthogroups (covering 527 proteins) were universal to all the PPN covered in this analysis, including *Bursaphelenchus xylophilus*, the only PPN not belonging to H-clade 12 (Tylenchida). Excluding *B. xylophilus*, we found 27 PPN-specific orthogroups (covering 1,750 proteins) that were universal to all the Tylenchida species genomes covered here. Being specific and widely conserved in PPN, these proteins are likely to cover core functions, possibly including those linked to plant parasitism.

In the RKN *M. incognita*, we identified 12,683 proteins that were specific to PPN species, with more than 91% (11,620) distributed in 6,526 PPN-specific orthogroups and only 1,063 being *M. incognita*-specific singleton proteins (https://doi.org/10.15454/IIAQOW).

3.4. A large set of M. incognita *proteins is specific to PPN*

According to the OrthoFinder results, approximately 12,700 *M. incognita* proteins had no ortholog in species other than PPN. However, this only holds true for the set of nematode and outgroup tardigrade species included in this comparative analysis. To further assess whether these proteins actually are orphans, without recognizable homology in other species beyond the nematode and tardigrades, we compared the whole set of *M. incognita* predicted proteins against the NCBI's nr library.

In the whole *M. incognita* predicted proteome, more than 72% (31,811/43,718) of the entries returned at least one hit at an E-value threshold of 0.01. Using the last common ancestor algorithm included in Diamond, a total of 330 different taxonomic identifiers (taxids) could be assigned from the hits. As expected, 68% of the hits were assigned nematode species taxids while metazoa as a whole represented 90% of taxid assignments. It should be noted that only 1,403 hits (<5%) were assigned to the Meloidogyne genus or more broadly to the Tylenchina clade, reflecting the poor

representation of plant-parasitic nematodes in the NCBI's nr databank. Complementary approaches such as the nematode-centered orthology search we carried out are thus necessary. Overall, eukaryota represented 94% of the hits taxonomic annotations, while unassigned cellular organisms and bacteria represented only 4% and 0.7%, respectively, and might correspond to horizontal gene transfers or contaminants in the genome of *M. incognita*. Conversely, 11,907 proteins (>27% of the predicted proteins) did not return any hit against nr and might represent orphan proteins. The taxonomic distribution of *M. incognita* hits against nr and the corresponding Krona graphs available at https://doi.org/10.15454/FROF42.

By comparing the 12,683 *M. incognita* PPN-specific proteins identified by OrthoFinder and the Diamond homology search, we found that only 22% (2,770) returned hits against the NCBI's nr databank. In total, only 222 taxids could be assigned and, in contrast to the whole proteome analysis, the phylum Nematoda was now minority and represented only 41% of the taxonomic assignments (of which 27% were Tylenchida). Overall, more than 78% (9,913/12,683) of the PPN-specific proteins identified by OrthoFinder returned no significant hit at all against the NCBI's nr database. These proteins lack homology in other nematodes and more broadly in other species. The accession numbers of *M. incognita* proteins specific to PPN and returning no significant hit against the NCBI's nr are available at https://doi.org/10.15454/48436V.

3.5. Putative secreted proteins and candidate effectors are enriched in PPN-specific M. incognita proteins

We identified 2,811 putative secreted proteins (PSP) in the genome of *M. incognita*, based on the presence of a predicted N-terminal signal peptide for secretion and an absence of predicted transmembrane region. These PSP are enriched in the PPN-specific *M. incognita* proteome, with almost half of them (1,331/2,811) falling in this category, while PPN-specific proteins represent less than one third of the *M. incognita* proteins (Pearson's Chi-squared test with Yates' continuity correction = 489.63, df = 1, p-value < 2.2e^{-16}). The complete list of *M. incognita* PSP and those specific to PPN are available at https://doi.org/10.15454/JCYZDI.

Effector proteins are secreted by nematodes in plant tissues and support parasitism through degradation of the plant cell wall, manipulation of plant defense and cell cycle or through other mechanisms [12]. Although PSP will certainly include candidate effectors, protein motifs suggesting secretion alone is insufficient for effector identification. Indeed, many secreted proteins might be unrelated to plant parasitism or secreted within the nematode between different cells or organs. A previous analysis using degenerate amino acid classifications identified four protein motifs enriched specifically in the 100 first amino acids of known *M. incognita* effectors [31]. In the *M. incognita* genome, we identified 12,625 proteins harboring at least one of these effector-enriched motifs in their first 100 amino acids (list available at https://doi.org/10.15454/LMY6LV).

Cross-referencing this list with that of the PSP yielded a total set of 2,146 candidate effectors. Nearly half (1,039) of the candidate effectors were specific to PPN, according to OrthoFinder data, which represents a significant enrichment (Pearson's Chi-squared test with Yates' continuity correction: X-squared = 412.3, df = 1, p-value < 2.2e-16). The list of candidate effectors is available at https://doi.org/10.15454/CSTXU2.

3.6. Enriched functions in PPN-specific proteins, PSP and candidate effectors reveal new promising targets for RKN control

InterProScan annotation allowed the identification of at least one known protein domain, motif or signal in 39,353 *M. incognita* proteins (90% of the predicted proteins in this species). A total of 28,288 of these proteins had at least one detected InterPro domain, which allowed assignment of at least one GO term to 20,610 *M. incognita* proteins (covering 1,939 different GO terms in total). The full InterPro annotation results are available at https://doi.org/10.15454/9BFFKG.

We searched for significantly enriched GO terms in (i) the 12,683 *M. incognita* proteins specific to PPN identified by OrthoFinder, (ii) in the 1,331 PPN-specific PSP and (iii) in the 1,039 PPN-specific candidate effectors (Figure 2). Only 9.6% (1,222) of the PPN-specific proteins in *M. incognita* had at least one GO term assigned, which were distributed as follows: 1,139 terms in the

molecular function (MF), 654 in the biological process (BP) and 205 in the cellular component (CC) ontologies. For PPN-specific PSP and candidate effectors, only 7.7% (103) and 7.9% (83) had at least one GO term assigned. This illustrates the high proportion of potentially pioneer proteins in these PPN-specific datasets.
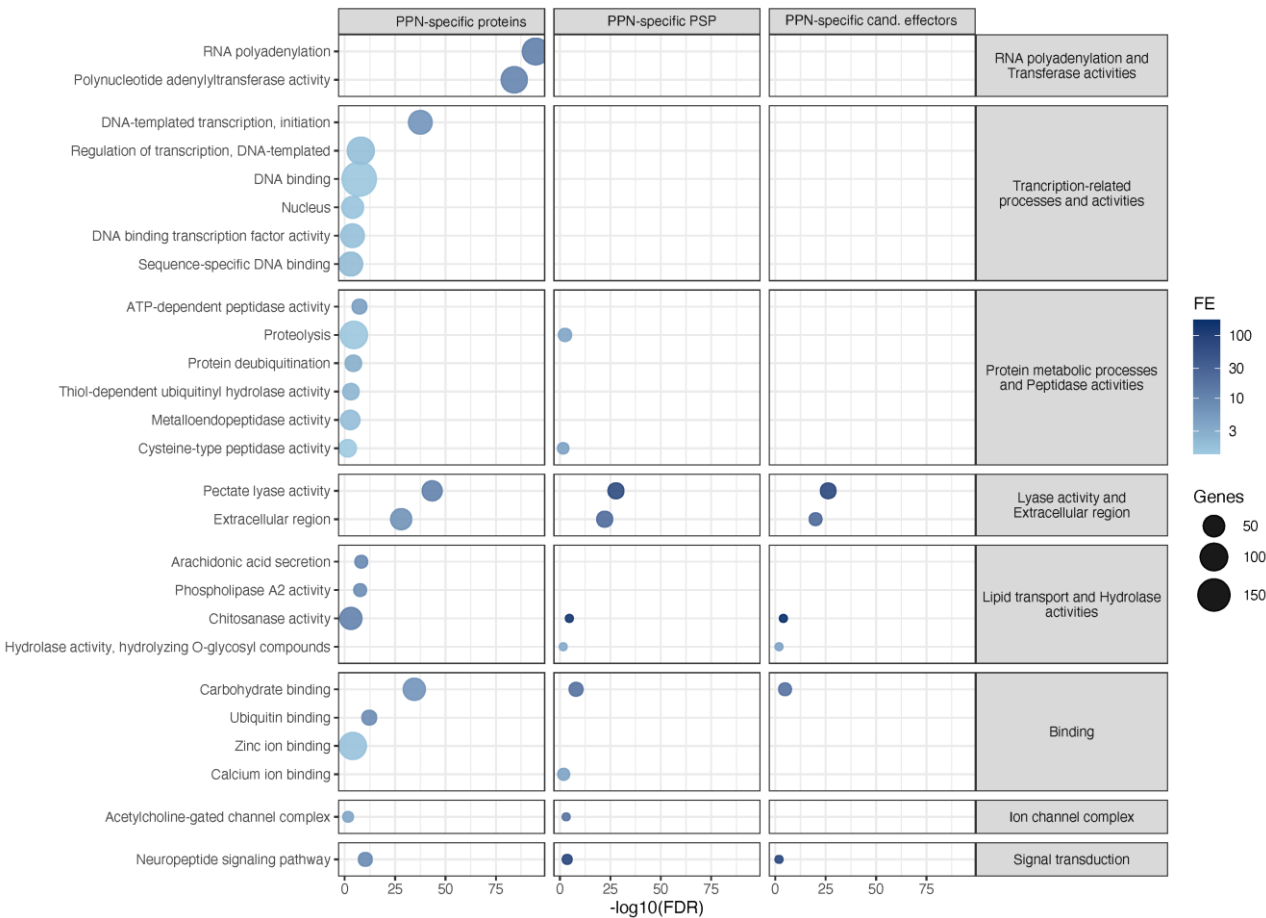

Figure 2. Enriched GO terms in plant-parasitic specific proteins, including putative secreted proteins and candidate effectors. Only GO terms with FDR<0.05 and at least four genes in the node are represented. Correlated terms are grouped together and represented by more general terms (right-side grey rectangles). Bubble sizes are according to the number of genes annotated with a specific GO term. Bubble positions are according to -log10(FDR) values. The darkness level of bubble colors is according to GO terms fold-enrichment (FE) values (see methods).

In the PPN-specific *M. incognita* proteins, FUNC identified 37, 83 and 45 overrepresented MF, BP and CC terms, respectively (raw and manually classified FUNC results available online at https://doi.org/10.15454/OTKID3). These terms were further grouped into 16, 7 and 3 over-represented MF, BP and CC representative terms. Because several protein sets overlapped different GO terms, with some of them belonging to the same GO sub-graphs, we further classified the enriched terms in the following eight main categories (Figure 2): (1) RNA polyadenylation and Transferase activities; (2) Transcription-related processes and activities; (3) Protein metabolic processes and Peptidase activities; (4) Lyase activity and Extracellular region; (5) Lipid transport and Hydrolase activity; (6) Binding; (7) Ion channel complex; and (8) Signal transduction. We then assessed whether the same categories were also enriched in the PSP and candidate effector subsets of PPN-specific *M. incognita* proteins (Figure 2). We also explored the RNA-seq based transcriptional profiles of genes belonging to each category through the construction and analysis of heatmaps (all data are available at https://doi.org/10.15454/OTKID3).

Terms related to post-transcriptional modification were strongly enriched in PPN-specific proteins. Surprisingly, 92 out of 99 *M. incognita* proteins annotated under RNA polyadenylation process (GO:0043631) and 88 out 95 proteins annotated under Polynucleotide adenylyltransferase

activity (GO:0004652) were PPN specific. These terms were the two most significantly overrepresented terms. Furthermore, a whole set of GO terms related to transcriptional processes was also enriched in the PPN-specific proteins, such as DNA-templated transcription, initiation (GO:0006352); Regulation of transcription; DNA-templated (GO:0006355); and terms related to DNA binding such as DNA binding function (GO:0003677); Sequence-specific DNA binding (GO:0043565), DNA binding transcription factor activity (GO:0003700); and Zinc ion binding (GO:0008270). The last one was plotted together with other Binding terms (Figure 2). Although all these terms related to transcriptional and post-transcriptional processes were highly enriched in the set of PPN-specific *M. incognita* proteins, none were found to be enriched in the PSP or candidate effectors subsets. This indicates these transcription-related proteins are not secreted and probably participate in functions within the nematode cells.

Several putative peptidases, which were assigned to the biological processes Proteolysis (GO:0006508) and Protein deubiquitination (GO:0016579), both critical post-translational mechanisms, were enriched in PPN-specific proteins. Additionally, three distinct classes of proteases were enriched: Metalloendopeptidase activity (GO:0004222), cysteine peptidases (Thiol-dependent ubiquitinyl hydrolase activity (GO:0036459) and Cysteine-type peptidase activity (GO:0008234) and threonine peptidase, represented by ATP-dependent peptidase activity (GO:0004176). While being overrepresented in PPN-specific proteins, most of these peptidase-related GO terms were not overrepresented in PSP or candidate effectors. The only two exceptions were the terms 'Proteolysis' and 'Cysteine-type peptidase activity', which were overrepresented in PSP but not in the candidate effectors.

Overall, only few GO terms overrepresented in PPN-specific proteins remained overrepresented in PPN-specific candidate effectors. These mainly included terms related to secreted carbohydrate-active enzymes known to be involved in the degradation of the plant cell wall [14], such as Hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553), Pectate lyase activity (GO:0030570) and Extracellular region (GO:0005576), as well as Carbohydrate binding (GO:0030246). In addition to terms related to plant cell wall digestion, two other terms were also found to be enriched in the PPN-specific candidate effectors, namely Chitosanase activity (GO:0016977) and Neuropeptide signaling pathway (GO:0007218).

3.7. Horizontal gene transfers are more abundant in PPN-specific proteins and contribute to some of their enriched functions

Alienness analysis [48] enabled the identification of 1,870 *M. incognita* proteins returning an Alien Index (AI) above 0; indicative of a better hit to non-metazoan than metazoan proteins in nr (cf. methods). The whole Alienness results are available at https://doi.org/10.15454/W6SWZH. Of these, 114 returned more than 70% identity to a non-metazoan hit, representing possible contaminations, according to the 70% rule [65]. As such, they were eliminated from further analyses, leaving a total of 1,756 proteins possibly acquired via HGT in *M. incognita*.

Nevertheless, an AI>0 alone only reflects a better hit to non-metazoan than metazoan proteins and is insufficient to assume acquisition via HGT. Higher AI values represent higher amplitudes of difference between the best non-metazoan and metazoan hits. An AI>14 represents a good balance between HGT prediction accuracy and recall in PPN [48]. Here, Alienness identified 404 *M. incognita* proteins returning an AI>14, thus likely to have been acquired via HGT from a non-metazoan source.

To estimate the accuracy of our HGT analysis, we compared the proteins identified here with an AI>14 to the phylogenetically supported HGT cases described in *M. incognita* and other RKN [48] (Table 2).

Of the 15 previously known families of HGT in RKN, 13 could be retrieved with an AI>14, including all the plant cell wall-degrading enzyme families previously described in RKN [13] and in the PPN-specific section. Two other known HGT families, namely candidate cyanate lyases and expansin-like proteins, returned positive AI values, although below the AI>14 threshold value. In total, as most form multigene families, the known HGT cases alone represented 124 (>30%) of the 404

proteins with an AI>14. The rest of the proteins share the same AI properties to the known HGT and might represent new cases not yet validated by phylogenetic analyses.

Table 2. Phylogenetically confirmed HGT cases in RKN retrieved by Alienness.

Confirmed HGT	Representative InterPro domain	Highest AI	# with AI>14	Process[1]
GH5_2 Cellulases	Glycoside hydrolase, family 5 (IPR001547)	32.57	21	PCW degradation
PL3 Pectate Lyase	Pectate lyase PlyH/PlyE-like (IPR004898)	121.94	39	PCW degradation
GH30 xylanase	Glycoside hydrolase family 30 (IPR001139)	191.63	19	PCW degradation
GH28 Polygalacturonase	Glycoside hydrolase, family 28 (IPR000743)	302.29	15	PCW degradation
GH43 Candidate Arabinanase	Glycoside hydrolase, family 43 (IPR006710)	174.53	3	PCW degradation
Expansin-like proteins	RlpA-like protein, double-psi beta-barrel domain (IPR009009)	8.54	0	PCW degradation
Candidate Isochorismatase	Isochorismatase-like (IPR000868)	51.56	3	Def. Manipulation
Chorismate mutase	Chorismate mutase II, prokaryotic-type (IPR002701)	30.85	4	Def. Manipulation
GH32 invertase	Glycoside hydrolase, family 32 (IPR001362)	306.79	6	Nutrient processing
VB5 PanC	Pantoate-beta-alanine ligase (IPR003721)	137.24	2	Nutrient processing
Candidate GSI Glutamine Synthase	Candidate GSI Glutamine Synthase (IPR008146)	248.7	5	Nutrient processing
NodL-like	Hexapeptide transferase, conserved site (IPR018357)	136.85	2	Feed. site induction
Candidate L-threonine Aldolase	Aromatic amino acid beta-eliminating lyase/threonine aldolase (IPR001597)	271.61	3	Unknown
Candidate Phosphoribosyltransferase	Phosphoribosyltransferase domain (IPR000836)	154.59	2	Unknown
Candidate Cyanate Lyase	Cyanate lyase, C-terminal (IPR003712)	3.49	0	Detoxification

[1] PCW: plant cell wall; Def. manipulation: manipulation of plant defense; Feed. site induction: induction of the nematode feeding site (from plant giant cells).

To explore the functional impact of genes possibly acquired via HGT, we performed a GO terms enrichment on the set of proteins returning AI scores >0 and >14 (methods, https://doi.org/10.15454/OTKID3 and Figure 3). As expected, terms related to the degradation of the plant cell wall, such as Pectate Lyase activity, Polygalacturonase activity, Carbohydrate metabolic process or Hydrolase activity, hydrolyzing O-glycosyl compounds, were enriched both in the possible (AI>0) and more stringent candidate (AI>14) HGT sets. Because these cell wall-degrading enzymes are secreted *in planta* by the nematode, the term Extracellular region was also significantly enriched. Besides these enzymes, other enriched terms in the possible and candidate HGT sets corresponded to previously reported HGT cases, such as Pantothenate biosynthetic process, which corresponds to PanC, a protein probably involved in the biosynthesis of vitamin B5 [66], or

Glutamate ammonia ligase activity, corresponding to GSI [67,68], the exact function of which is still unclear in RKN.
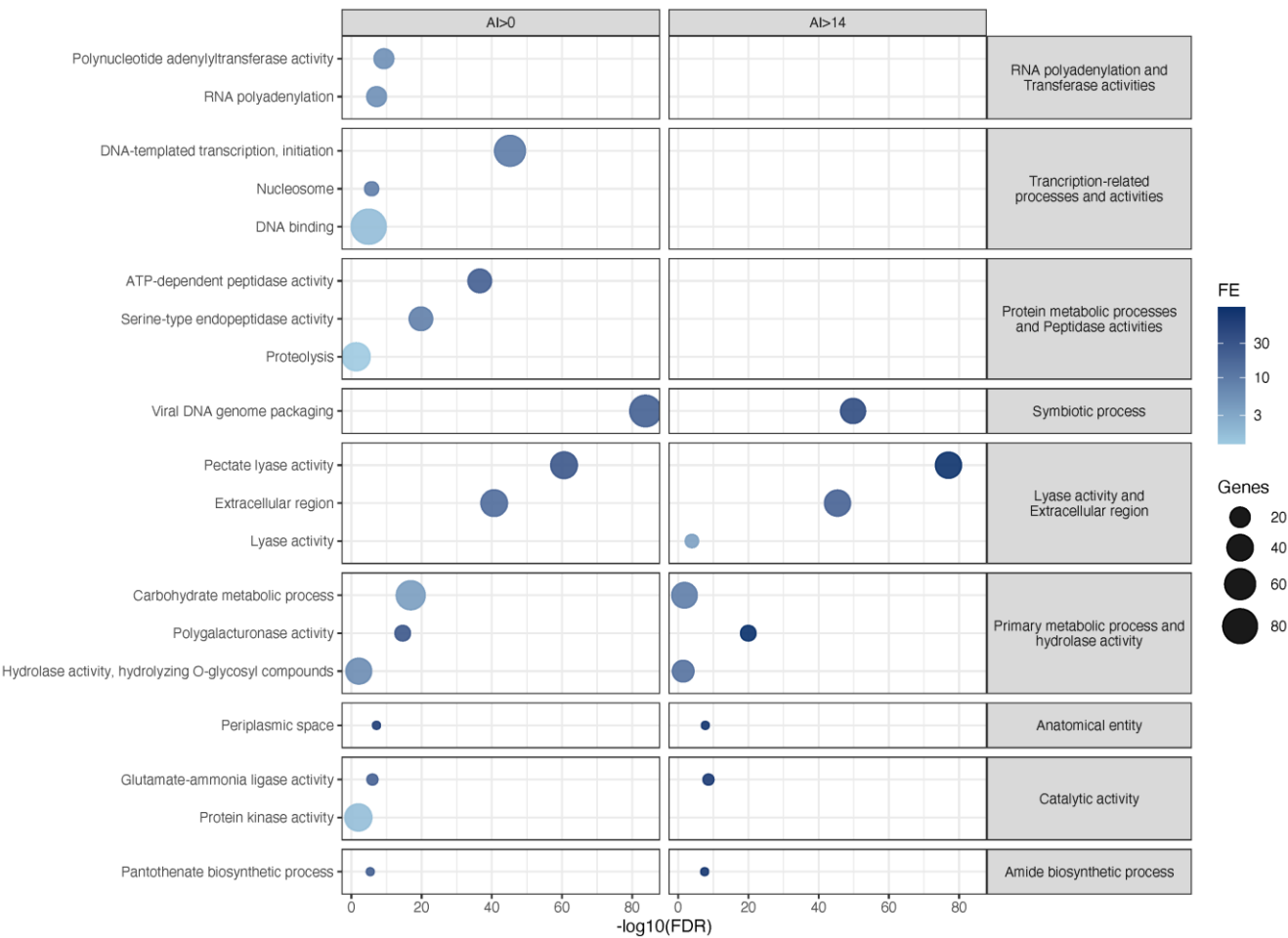

Figure 3. Enriched GO terms in proteins possibly acquired via horizontal transfers. Left side: possible HGT – AI>0; right side: likely HGT – AI>14. Only GO terms with FDR<0.05 and at least four genes in the node are represented. Correlated terms are grouped together and represented by more general terms (right-side grey rectangles). Bubble sizes are according to the number of genes annotated with a specific GO term. Bubble positions are according to -log10(FDR) values. The darkness of bubble colors is according to GO terms fold-enrichment (FE) values (see methods).

Other terms, not corresponding to previously known HGT, were enriched in the set of possible HGT (AI>0). Interestingly, several of these were also enriched in the PPN-specific set presented in the previous section. These included, those related to transcriptional or post-transcriptional processes (DNA binding, DNA-templated transcription, initiation, RNA polyadenalyation), Protein kinase activity, as well as terms related to protein degradation (Proteolysis, Serine-type endopeptidase and ATP-dependent peptidase activities). However, none of these terms were further enriched in the set of candidate HGT (AI>14), as most of the corresponding proteins returned positive AI values but <14. Whether they actually represent new HGT cases of non-metazoan origin would need to be further phylogenetically assessed in future analyses. Actually, only three novel terms (Viral DNA genome packaging, Lyase activity and Periplasmic space) were enriched in the candidate HGT set. The term 'Viral DNA genome packaging' corresponded to a whole set of proteins bearing an InterPro domain named 'Adenovirus packaging protein 1', with no evident role in relation to plant parasitism. The enriched terms Lyase activity and Periplasmic space were associated with proteins bearing an Alginate lyase domain. Alginate is a polysaccharide usually

found in brown algae and whether the RKN possess enzymes that can degrade alginate would be interesting to assess in the future.

Exploring the possible donors of candidate HGT cases showed that 83% (334/404) were of possible bacterial origin, 5% from plants, 4% from fungi, 5% from other non-metazoan eukaryotes, 1.5% from Stramenopiles, while Archaea and Viruses each represented <0.5% of the possible donors. By restricting the analysis to taxonomic information supported by at least three consecutive blast hits, Bacteria represented 94% of the candidate donors, Plants 2%, Fungi 1%; other non-metazoan Eukaryotes 3%, while the other categories (Stramenopiles, Viruses and Archaea) now disappeared, being supported by a maximum of two consecutive hits.

To estimate the contribution of HGT to the set of PPN-specific genes in *M. incognita*, we cross-referenced the list of PPN-specific proteins with those returning an AI>0. Only 756 or 14% of the 12,683 PPN-specific proteins in *M. incognita* returned an AI>0, of which 148 had an AI>14 and thus likely to have been acquired via HGT. Hence, only 1% of the PPN-specific proteins have likely been acquired via HGT and possibly up to 14%. Nevertheless, we note that the proportion of possible or candidate HGT in the set of PPN-specific proteins represents a significant enrichment compared to the rest of *M. incognita* proteins (p-values of the X^2 tests being $2.2E^{-16}$ and $8.48E^{-4}$ for AI>0 and AI>14, respectively). Moreover, proteins possibly acquired via HGT contribute to some of the enriched functions in PPN-specific proteins, including in candidate effectors.

3.8. New candidate effectors amenable to nematode control via in planta *inactivation*

One effective strategy towards controlling infection by phytoparasitic nematodes is engineering plants to produce double stranded RNAs that, in turn, will specifically silence nematode genes essential for parasitism via RNAi-based technologies [69]. A prerequisite is to target nematode genes that are expressed while the parasite is infecting plant tissues, such as genes encoding effectors [17–19].

Hence, from the 1,039 PPN-specific candidate effector proteins, we selected those encoded by genes with substantial RNA-seq support for expression in endophytic phases of the nematode life cycle. We required the genes to have expression values in these phases above the median (>0.78 and 0.73 for J3-J4 and adult females, respectively). We further required that the average expression in endophytic phases (J3-J4 and adult females) was higher than the average expression during exophytic phases (eggs and pre-parasitic J2).

Aiming at targeting genes important for parasitism with a potential broad effect on RKN, we only retained proteins conserved in at least two more RKN species than *M. incognita*.

In addition, to facilitate downstream biological validations and assays, we eliminated candidates that originated from multigene families. Indeed, designing specific probes for targeting one single gene in a multigene family can be challenging and monitoring silencing effects would be more difficult, with possible compensation of the loss of activity / function. Given that the *M. incognita* genome is triploid, many genes are present in three homoelogous copies [23]. Consequently, all the proteins present in more than three copies were considered as possibly coming from a multigene family and were therefore eliminated.

The application of all the above criteria led to the selection of 89 candidate effector proteins in *M. incognita* that were specific to PPN species, according to OrthoFinder (list available at https://doi.org/10.15454/WBZZ5M.

Of these 89 *M. incognita* candidate effector proteins, only 21 returned hits against the NCBI's nr, of which 13 corresponded to previously published *M. incognita* effectors and one to a putative *M. javanica* effector (Diamond hit accession numbers in bold in https://doi.org/10.15454/WBZZ5M). Because we aimed at identifying novel candidate effectors, lacking homology in other species, the 21 *M. incognita* proteins returning hits were eliminated.

The 68 remaining proteins returned no hits against the NCBI's nr library and represent novel PPN-specific candidate effectors.

3.9. Expression validation of candidate effector genes

Many important effectors are developmentally regulated, with an increase in expression during endophytic phases of the RKN life cycle [70,71]. From the 68 novel candidate effectors identified in this study, and expressed in endophytic stages, according to *in silico* RNA-seq data, we randomly selected seven genes (Table 1) for independent validation of expression pattern in *M. incognita* during parasitism on *N. tabacum*. In addition, a previously identified candidate *M. incognita* effector, meeting the same selection criteria, Minc3s00292g09561 (= Minc02654) was included as a positive control in expression analysis. This gene was previously described as overexpressed in *M. incognita* parasitic stages [72].

Using RT-qPCR analysis, we monitored gene expression in eggs, pre-parasitic J2 (ppJ2), and endophytic *M. incognita* stages at 5, 10, 15, and 22 days after nematode inoculation (DAI). Expression modulation was compared with the in silico RNA-seq whole transcriptome data for nematode life-cycle stages described in the previous section (Egg, ppJ2, pJ2-J3-J4 and Female). Overall, all seven candidate genes displayed significant increased expression in at least one of the endophytic phases of the life cycle (5, 10, 15 or 22 DAI), in comparison to their expression in the egg and pre-parasitic J2 stages (Figure 4). These expression trends are consistent with the RNA-seq-derived expression patterns. Furthermore, RT-qPCR analysis of the positive control gene (Figure 4 H) also showed expression modulation patterns matching those previously described [72].

Similarly to this previously described effector, the effector candidates Minc3s03880g3504, Minc3s00020g1309 and Minc3s01206g21700 displayed a sharp increase in their expression during the juvenile endophytic stages pJ2-J3-J4, and continued to be expressed in the adult female stage albeit at lower level for Minc3s03880g3504 (Fig 4- E, F, G). Therefore, these genes constitute interesting targets for in-planta delivered RNAi strategy.

In addition to the parasitic J2 stage, *M. incognita* nematodes have relatively short and regular feeding cycles at adult female stages [73]. We found four effector candidates, Minc3s00016g00668, Minc3s01635g25253, Minc3s00010g00666 and Minc3s1217g21793, which returned highest expression in females (22 DAI) with low to no expression at earlier stages (Figure 4 A, B, C and D). These proteins are probably involved in the feeding site maintenance, and therefore also constitute potential targets for RNAi.
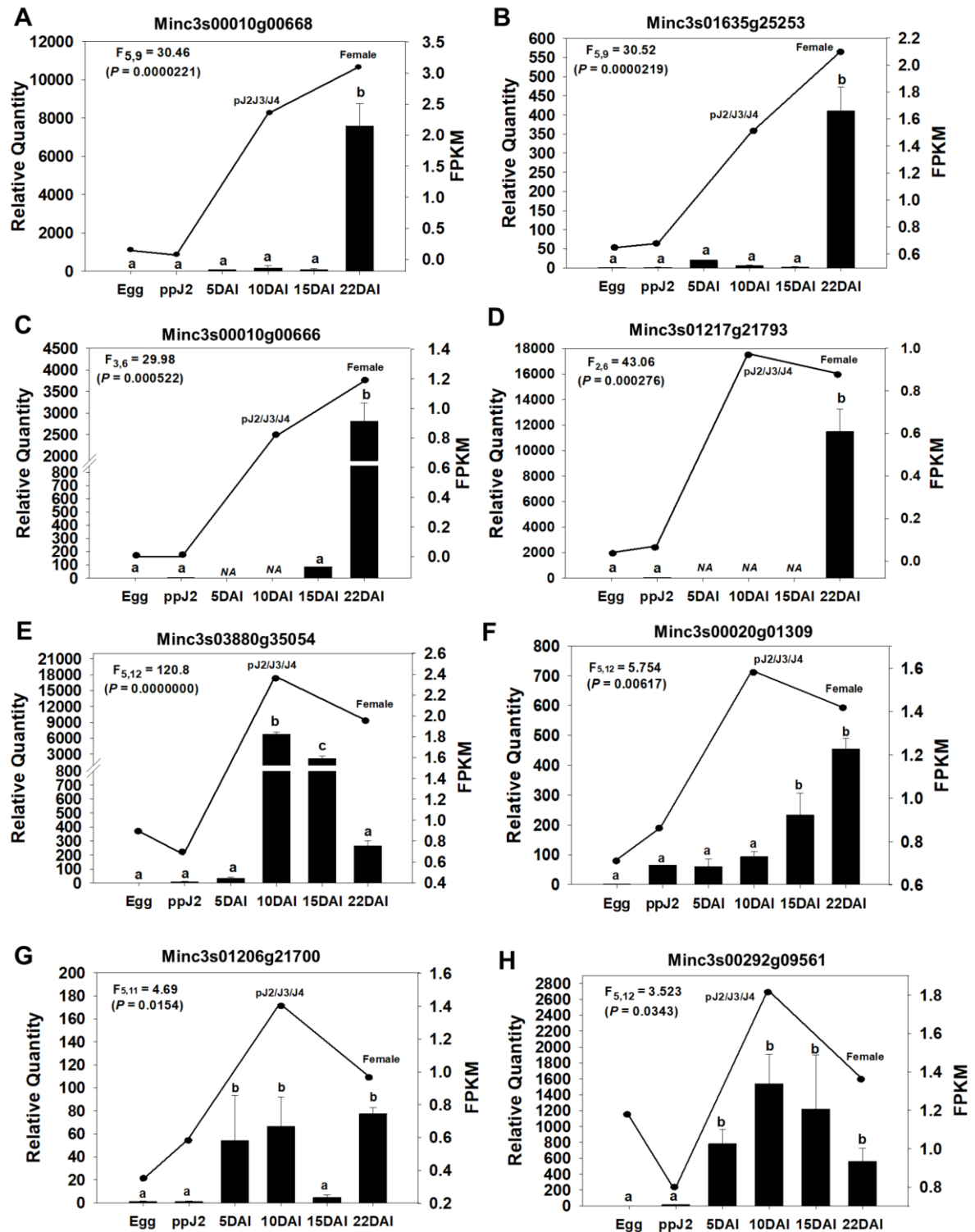
Figure 4. RNA-seq and RT-qPCR expression profiles of PPN-specific candidate effector genes during *M. incognita* life cycle. Relative mRNA levels of eight genes (A-H) at different stages of the nematode life cycle during its compatible interaction with *Nicotiana tabacum*. RT-qPCR data are shown in bars (mean ± standard error) with values in the left "y" axis and analyzed in six stages: egg, ppJ2 and endophytic phases at 5, 10, 15 and 22 days after inoculation (DAI). RNA-seq *in silico* data are shown as a line plot with values displayed at the right "y" axis (FPKM) and analysed in four stages: egg, ppJ2 and endophytic phases pJ2/J3/J4 and female. RT-qPCR expression data were analyzed using SATqPCR tool [57], with expression values statistically analyzed using ANOVA and Tukey's test (P<0.05). Different letters indicate statistically significant diferences between samples.

4. Discussion

4.1. PPN-specific proteins reveal novel candidate effectors and other promising functions for nematode control

Using an extensive comparative genomics analysis, we identified a whole set of novel candidate effectors in *M. incognita* that have no recognizable homologs in species other than PPN. Because they occur in several *Meloidogyne* or other PPN from H-clade 12, these genes probably play important roles in plant parasitism. Furthermore, as they are exclusive to PPN, the specific targeting of these genes would reduce the risk of unintentional off-target effects on other species. Finally, as they are expressed during endophytic phases of the nematode parasitic life cycle, these genes are amenable to *in planta* RNAi silencing and thus constitute a promising set of targets for the development of cleaner and more specific strategies to control nematode pests.

Although we focused on candidate effectors, representing the most obvious targets, functional analysis of the ensemble of PPN-specific proteins revealed other novel and interesting candidates. As expected, because the majority of these PPN-specific proteins lack recognizable homology, only a few possess a predicted conserved domain that allows functional annotation. Indeed, only 9.6% of the PPN-specific proteins in *M. incognita* had at least one GO term assigned, in comparison to more than 47% for the whole proteome. This illustrates the high proportion of potential pioneer proteins without evident functional annotation in the PPN-specific datasets.

Despite that, among the few PPN-specific proteins with functional annotation, some terms were significantly enriched, with functions deserving further attention in the future. For instance, a series of GO terms related to transcriptional and post-transcriptional activities were enriched in PPN-specific proteins, but not further enriched in the secreted or candidate effector subsets, indicating an internal function in the plant parasitic nematode. Interestingly, these terms were also enriched in the possible HGT protein set (AI>0), indicating higher similarity to non-metazoan than metazoan proteins. Whether they represent possible acquisition via horizontal gene transfer will need further phylogenetic investigation. Regardless of their origin, these proteins might be involved in the regulation of PPN-specific functions, including control of effector gene expression or development of organs directly involved in parasitism.

Another interesting term, 'Neuropeptide signaling pathway' was enriched in PPN-specific proteins and associated with four proteins bearing a signal peptide for secretion, no transmembrane region and a 'FMRFamide related peptide family' domain. FMRFamide-like peptides (FLPs) are widely conserved neuropeptides in nematodes that likely play a variety of roles from locomotion to feeding [74]. These *M. incognita* FLP proteins have no predicted homologs in nematodes other than plant parasites and might therefore be involved in phytoparasitic-specific functions, such as sensing and recognition of the host plant and attraction towards the roots.

Finally, terms related to metallo and cysteine peptidase activity were also enriched in the set of *M. incognita* PPN-specific proteins. In nematodes, peptidases of these families are involved in molting and ecdysis processes, as a new cuticle is synthesized five times during nematode development [75]. The metalloendopeptidase genes we identified in *M. incognita* are mainly expressed in eggs and in pre-parasitic J2 larvae while the cysteine-type peptidases are highly expressed in late parasitic forms and females. As in the root-knot nematodes and other PPN several molts take place within the plant tissue, these peptidases are probably also involved in PPN-specific molts or cuticle modifications. Preventing nematode molting in the soil or within plant tissues would lead to developmental arrest and failure of parasitism.

Overall, the present comparative genomics analysis not only revealed new and interesting candidate effectors but also identified other families of proteins that improve our understanding of the arsenal of genes employed by nematodes to manipulate their hosts or that are essential for development and survival. Now being publicly available, these data constitutes a valuable resource to fuel future research in the field of plant - nematode interactions.

Reinforcing their interest, the PPN-specific candidate effectors seem to show similar expression patterns on different *M. incognita* isolates and host plants. The selected candidate effectors in *M. incognita* that were more highly expressed in endophytic stages of the parasitic life cycle, according

to previously generated RNA-seq data [23] were also more highly expressed during endophytic phases in the RT-qPCR assays. The RNA-seq assay had been performed with an *M. incognita* strain originally from Mexico (stain Moreols) on a susceptible *S. lycopersicum* host (tomato, var. *esculentum* cv. St Pierre) whereas the RT-qPCR assays was conducted with an *M. incognita* strain from Brazil as well as a different compatible host plant (tobacco). This consistency in expression suggests the examined genes are broadly important for *M. incognita* infection on different plant hosts and strengthen their potential as targets for the development of efficient nematode control methods.

4.2. Evolutionary origin of PPN-specific genes

Comparative analysis of 61 nematode genomes revealed 12,683 *M. incognita* proteins with no orthologs in nematodes other than PPN. One can thus wonder about the origin of these numerous proteins, which appear to be absent from the other nematodes. These proteins are unlikely to result from overpredictions by gene calling software, as the vast majority are supported by substantial expression data and are conserved in multiple PPN genomes. Further homology search against the NCBI's nr and Alienness analysis showed HGT to have likely contributed to 1% (AI>14) of the 12,683 PPN-specific genes in *M. incognita,* and up to 14% if we consider possible HGT (AI>0). Some of the functions enriched in the PPN-specific genes, including those related to degradation of the plant cell wall or transcription, are clearly due to genes possibly acquired by HGT. Candidate HGT also revealed new proteins not previously described in PPN, such as 'adenovirus packaging protein 1' and 'alginate lyase'. The origin and possible function of these proteins in Meloidogyne and other PPN deserve further investigation.

Although HGT are significantly enriched in the PPN-specific dataset, this only explains the origin for a minority of the genes (maximum 14%). This is consistent with the observation that 88% of the PPN-specific *M. incognita* proteins return no significant hit at all against the NCBI's nr library. Thus, these proteins more likely represent lineage-specific orphan genes, with two main hypotheses to explain their evolutionary origin [76]. The first postulates that duplications followed by rearrangements and extremely rapid sequence divergence can lead to gene copies lacking recognizable homology to other sequences. A second hypothesis, of *de novo* gene birth, had long been overlooked, but is now gaining greater attention. Under this hypothesis, protein-coding genes can emerge *de novo* from non-genic regions by two main mechanisms: one called RNA first and the other called ORF first [77]. RNA first postulates that an ORF can emerge by accumulation of mutations in an originally transcribed non-coding region. ORF first, in contrast, postulates that a coding but untranscribed region can acquire cis-mutations (e.g. insertion of a transposable element) and become transcriptionally active. As these PPN-specific proteins represent a substantial part of the proteome and include known and new candidate effectors, their contribution both to the genome and biology of the RKN is important. These findings are consistent with previous reports indicating that most of the effectors are 'orphan' proteins, lacking homology outside the PPN [11]. Further investigation into the evolutionary origin of these PPN-specific genes and of their actual impact on nematode biology is recommended. Available genomes from close outgroups of the Tylenchida PPN will be a prerequisite to search for similarities to the PPN-specific genes in these species. Hopefully, such data can be expected in the near future, enabling investigations of their ancientness and whether they emerged from pre-existing non-genic regions.

4.3. An unprecedented representation of PPN in a comprehensive and phylogenetically accurate comparative framework

Although comparative phylogenomics analyses of nematodes have previously been published, these studies only encompassed two [32] or three [39] genomes for PPN species, limiting insight in genome evolution associated with plant parasitism. In the present study, we conducted a phylogenomics analysis that covered a diversity of nematode clades and lifestyles and included 16 PPN species. This phylogenomics comparison offers an unprecedented representation of phytoparasitic nematodes and constitutes a new opportunity to study genomic singularities at the protein family level, associated to the evolution of plant parasitism in nematodes.

However, comparative analysis can only express its full potential if supported by an underlying phylogenetically correct framework. We thus compared the phylogeny of the species obtained by our phylogenomics analysis to the reference nematode classification systems.

Two main classification systems have been developed to describe the nematode diversity. The system described by Mark Blaxter and colleagues [58,59] defined five clades (I-V) that we have referred to as B-clades. This system is based on a limited number of species but employs concatenated multiple marker genes. In parallel, the classification system described by Hans Helder and colleagues defined 12 clades [1,60] and, although it comprises the largest diversity of nematodes to date (including PPN); it is based on only one single marker gene (SSU rRNA). We have referred to this classification system as H-clades (1-12). The backbones of these two classification systems have recently been reinforced and further confirmed by a phylogenomics analysis based on a concatenated multi-gene alignment comprising 108 nematode species [62], representing a good balance between the phylogenetic information in the concatenated markers and species diversity.

The phylogenetic species tree we obtained by OrthoFinder analysis was remarkably consistent with the recently published phylogenomics study of nematodes [62], and by extension with the two historical classification systems in B and H clades. This implies that the dominant phylogenetic signal obtained from the protein-level phylogenies of the ensemble of orthogroups, taken as a whole, recapitulates the reference nematode phylogeny. Hence, the comparative analysis of the protein sets, including at the level of clade-specific proteins and proteins shared by multiple species, is supported by a phylogenetically correct background, allowing evolutionarily relevant comparisons. Not only was the OthoFinder topology almost identical to that of [62]; but the support values for the different clades were all >0.9 (except in one branch), indicating a strong signal supporting the phylogeny. The only support values lower than 0.9 concerned the relative branching order of the RKN within clade I, *sensu* De Ley (*M. arenaria*, *M. enterolobii*, *M. floridensis*, *M. incognita* and *M. javanica*). This observation is not particularly surprising, as these RKN species underwent complex hybridization and polyploidization events, with a probable same common or very similar maternal origin and distinct paternal progenitors [23,24,29,78]. As a result, some genome regions are highly similar between these species, while copies of the region are highly divergent within a same species. These complex homology relationships complicate both one-to-many and many-to-many orthogroup analyses in these allopolyploid species. Nonetheless, because the gene copies are more similar to each other within RKN clade I than to homologs in the other RKN clades, most of the orthogroup trees will show a monophyletic group encompassing all these allopolyploid nematode species. This explains the high support for the monophyly of clade I in Figure 1. However, depending on the gene and homoeolog gene copy under consideration in a given orthogroup, the relative positions of the corresponding proteins between the four allopolyploid RKN species will vary and lead to lower support values. This lower support for relative positions of species among clade I RKN have no influence on our comparative analysis, as we mainly focused on *M. incognita* genes specific to PPN and did not draw conclusions on phylogenomic differences between the four clade I RKN.

5. Conclusions

Using comparative genomics analysis on a broad sample of nematode species, selected for their diversity of phylogenetic position and lifestyles, as well as the quality of their genomes, we have identified a whole set of proteins specific to plant-parasitic species. Focusing more specifically on the model and devastating root-knot nematode *M. incognita*, we found that ca. 10,000 genes (a quarter of the predicted proteins) are only found in other plant-parasitic nematodes. A small proportion of these proteins show similarity with non-metazoan species and might originate from HGT. However, the vast majority returned no significant similarity at all, and their origin remains so far elusive. Because most of these proteins are conserved in multiple PPN species and the corresponding genes are supported by expression data, they are unlikely to result from in-silico over-predictions. Although de novo gene birth is a process that has recently attracted greater attention, elucidation of the origin of these PPN-specific genes will need further and more focused investigation. Regardless of their origin, some functions enriched in these PPN-specific genes point towards new processes

such as neuro-signaling, molting, or transcription, that have not yet been investigated as targets for development of control strategies of phytoparasitic nematodes.

Targeting PPN-specific genes important for parasitism appears as a promising prospect in the quest towards controlling these devastating plant pests while minimizing negative environmental impact. Within this perspective, candidate effector proteins constitute obvious target genes, given their importance in parasitism. Here, we identified ca. 1,000 PPN-specific proteins in *M. incognita* sharing the same characteristics as known effectors. Targeting specific genes of an endoparasite can be conducted through plant transformation and RNA-mediated gene silencing. Towards this, we further filtered the dataset to retain a number of novel PPN-specific candidate effectors expressed while the nematode is within the plant. Employing RT-qPCR, we confirmed their higher expression during endophytic phases of parasitism on a second host species, suggesting a conserved and important role in plant parasitism.

Collectively these datasets of proteins constitute promising resources towards a better understanding of the genetic basis of nematode adaptation to phytoparasitism and development in the future of more efficient and specific nematode control strategies.

References

1.   Holterman, M.; Karegar, A.; Mooijman, P.; Megen, H. van; Elsen, S. van den; Vervoort, M.T.W.; Quist, C.W.; Karssen, G.; Decraemer, W.; Opperman, C.H.; et al. Disparate gain and loss of parasitic abilities among nematode lineages. *PLOS ONE* **2017**, *12*, e0185445, doi:10.1371/journal.pone.0185445.

2.   Decraemer, W.; Hunt, D.J. Structure and classification. In *Plant nematology*; CABI, 2013; pp. 3–39 ISBN 978-1-78064-151-5.

3.   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.

4.   Ramadan, K.M. Interaction of root-knot nematode (Meloidogyne javanica) and tomato as affected by hydrogen peroxide. *J. Plant Prot. Res.* **2008**, *48*, 181–187, doi:10.2478/v10045-008-0021-x.

5. Nicol, J.M.; Turner, S.J.; Coyne, D.L.; Nijs, L. den; Hockland, S.; Maafi, Z.T. Current Nematode Threats to World Agriculture. In *Genomics and Molecular Genetics of Plant-Nematode Interactions*; Jones, J., Gheysen, G., Fenoll, C., Eds.; Springer Netherlands: Dordrecht, 2011; pp. 21–43 ISBN 978-94-007-0433-6.

6. Koutsovoulos, G.D.; Marques, E.; Arguel, M.-J.; Duret, L.; Machado, A.C.Z.; Carneiro, R.M.D.G.; Kozlowski, D.K.; Bailly-Bechet, M.; Castagnone-Sereno, P.; Albuquerque, E.V.S.; et al. Population genomics supports clonal reproduction and multiple independent gains and losses of parasitic abilities in the most devastating nematode pest. *Evol. Appl.* **2020**, *13*, 442–457, doi:10.1111/eva.12881.

7. Sikora, R.A.; Claudius-Cole, B.; Sikora, E.J. Nematode parasites of food legumes. In *Plant Parasitic Nematodes in Subtropical and Tropical Agriculture*; CABI North American Office: Boston, MA, 2018; pp. 290–345.

8. Chitwood, D.J. Nematicides. In *Encyclopedia of Agrochemicals*; American Cancer Society, 2003 ISBN 978-0-471-26363-0.

9. Gheysen, G.; Mitchum, M.G. How nematodes manipulate plant development pathways for infection. *Curr. Opin. Plant Biol.* **2011**, *14*, 415–421, doi:10.1016/j.pbi.2011.03.012.

10. Kyndt, T.; Vieira, P.; Gheysen, G.; de Almeida-Engler, J. Nematode feeding sites: unique organs in plant roots. *Planta* **2013**, *238*, 807–818, doi:10.1007/s00425-013-1923-z.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. Lilley, C.J.; Maqbool, A.; Wu, D.; Yusup, H.B.; Jones, L.M.; Birch, P.R.J.; Banfield, M.J.; Urwin, P.E.; Akker, S.E. den Effector gene birth in plant parasitic nematodes: Neofunctionalization of a housekeeping glutathione synthetase gene. *PLOS Genet.* **2018**, *14*, e1007310, doi:10.1371/journal.pgen.1007310.

17. Huang, G.; Allen, R.; Davis, E.L.; Baum, T.J.; Hussey, R.S. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proc Natl Acad Sci U A* **2006**, *103*, 14302–6.

18. Niu, J.; Liu, P.; Liu, Q.; Chen, C.; Guo, Q.; Yin, J.; Yang, G.; Jian, H. Msp40 effector of root-knot nematode manipulates plant immunity to facilitate parasitism. *Sci. Rep.* **2016**, *6*, 1–13, doi:10.1038/srep19443.

19. Shi, Q.; Mao, Z.; Zhang, X.; Ling, J.; Lin, R.; Zhang, X.; Liu, R.; Wang, Y.; Yang, Y.; Cheng, X.; et al. The Novel Secreted Meloidogyne incognita Effector MiISE6 Targets the Host Nucleus and Facilitates Parasitism in Arabidopsis. *Front. Plant Sci.* **2018**, *9*, doi:10.3389/fpls.2018.00252.

20. 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.

21.   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.

22.   Bird, D.M.; Williamson, V.M.; Abad, P.; McCarter, J.; Danchin, E.G.; Castagnone-Sereno, P.; Opperman, C.H. The genomes of root-knot nematodes. *Annu Rev Phytopathol* **2009**, *47*, 333–51.

23.   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.

24.   Szitenberg, A.; Salazar-Jaramillo, L.; Blok, V.C.; Laetsch, D.R.; Joseph, S.; Williamson, V.M.; Blaxter, M.L.; Lunt, D.H. Comparative Genomics of Apomictic Root-Knot Nematodes: Hybridization, Ploidy, and Dynamic Genome Change. *Genome Biol. Evol.* **2017**, *9*, 2844–2861, doi:10.1093/gbe/evx201.

25.   Sato, K.; Kadota, Y.; Gan, P.; Bino, T.; Uehara, T.; Yamaguchi, K.; Ichihashi, Y.; Maki, N.; Iwahori, H.; Suzuki, T.; et al. High-Quality Genome Sequence of the Root-Knot Nematode Meloidogyne arenaria Genotype A2-O. *Genome Announc.* **2018**, *6*, e00519-18, doi:10.1128/genomeA.00519-18.

26.   Somvanshi, V.S.; Tathode, M.; Shukla, R.N.; Rao, U. Nematode Genome Announcement: A Draft Genome for Rice Root-Knot Nematode, Meloidogyne graminicola. *J. Nematol.* **2018**, *50*, 111–116, doi:10.21307/jofnem-2018-018.

27.   Susič, N.; Koutsovoulos, G.D.; Riccio, C.; Danchin, E.G.J.; Blaxter, M.L.; Lunt, D.H.; Strajnar, P.; Širca, S.; Urek, G.; Stare, B.G. Genome sequence of the root-knot nematode Meloidogyne luci. *J. Nematol.* **2020**, *52*, 1–5, doi:10.21307/jofnem-2020-025.

28.   Phan, N.T.; Orjuela, J.; Danchin, E.G.J.; Klopp, C.; Perfus-Barbeoch, L.; Kozlowski, D.K.; Koutsovoulos, G.D.; Lopez-Roques, C.; Bouchez, O.; Zahm, M.; et al. Genome structure and content of the rice root-knot nematode (Meloidogyne graminicola). *Ecol. Evol.* **2020**, *n/a*, doi:10.1002/ece3.6680.

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.   Gardner, M.; Dhroso, A.; Johnson, N.; Davis, E.L.; Baum, T.J.; Korkin, D.; Mitchum, M.G. Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode Heterodera glycines. *Sci. Rep.* **2018**, *8*, 2505, doi:10.1038/s41598-018-20536-5.

31.   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.

32.   Danchin, E.G.J.; Arguel, M.-J.; Campan-Fournier, A.; Perfus-Barbeoch, L.; Magliano, M.; Rosso, M.-N.; Da Rocha, M.; Da Silva, C.; Nottet, N.; Labadie, K.; et al. Identification of Novel Target Genes for Safer and More Specific Control of Root-Knot Nematodes from a Pan-Genome Mining. *PLoS Pathog* **2013**, *9*, e1003745, doi:10.1371/journal.ppat.1003745.

33.   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.

34.   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.

35. Masonbrink, R.E.; Maier, T.R.; Muppirala, U.; Seetharam, A.S.; Lord, E.; Juvale, P.; 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. *bioRxiv* **2018**, 391276, doi:10.1101/391276.

36. Zheng, J.; Peng, D.; Chen, L.; Liu, H.; Chen, F.; Xu, M.; Ju, S.; Ruan, L.; Sun, M. The Ditylenchus destructor genome provides new insights into the evolution of plant parasitic nematodes. *Proc. R. Soc. B Biol. Sci.* **2016**, *283*, 20160942, doi:10.1098/rspb.2016.0942.

37. Mimee, B.; Lord, E.; Véronneau, P.-Y.; Masonbrink, R.; Yu, Q.; Akker, S.E. den The draft genome of Ditylenchus dipsaci. *J. Nematol.* **2019**, *51*, 1–3, doi:10.21307/jofnem-2019-027.

38. Kikuchi, T.; Cotton, J.A.; Dalzell, J.J.; Hasegawa, K.; Kanzaki, N.; McVeigh, P.; Takanashi, T.; Tsai, I.J.; Assefa, S.A.; Cock, P.J.A.; et al. Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen Bursaphelenchus xylophilus. *PLoS Pathog* **2011**, *7*, e1002219, doi:10.1371/journal.ppat.1002219.

39. International Helminth Genomes Consortium Comparative genomics of the major parasitic worms. *Nat. Genet.* **2019**, *51*, 163–174, doi:10.1038/s41588-018-0262-1.

40. Waterhouse, R.M.; Seppey, M.; Simão, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548, doi:10.1093/molbev/msx319.

41. Emms, D.M.; Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157, doi:10.1186/s13059-015-0721-2.

42. Emms, D.M.; Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238, doi:10.1186/s13059-019-1832-y.

43. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60, doi:10.1038/nmeth.3176.

44. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**, *30*, 772–80, doi:10.1093/molbev/mst010.

45. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **2010**, *5*, e9490, doi:10.1371/journal.pone.0009490.

46. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786, doi:10.1038/nmeth.1701.

47. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L.L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *J. Mol. Biol.* **2001**, *305*, 567–580, doi:10.1006/jmbi.2000.4315.

48. Rancurel, C.; Legrand, L.; Danchin, E.G.J. Alienness: Rapid Detection of Candidate Horizontal Gene Transfers across the Tree of Life. *Genes* **2017**, *8*, 248, doi:10.3390/genes8100248.

49. 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.

50. 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.

51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag: New York, NY, 2016; ISBN 978-3-319-24277-4.

52. Kozlowski, D.K.; Hassanaly-Goulamhoussen, R.; Rocha, M.D.; Koutsovoulos, G.D.; Bailly-Bechet, M.; Danchin, E.G. Transposable Elements are an evolutionary force shaping genomic plasticity in the parthenogenetic root-knot nematode Meloidogyne incognita. *bioRxiv* **2020**, 2020.04.30.069948, ver. 4 peer-reviewed and recommended by PCI Evolutionary Biology., doi:10.1101/2020.04.30.069948.

53. Randig, O.; Bongiovanni, M.; Carneiro, R.M.D.G.; Castagnone-Sereno, P. Genetic diversity of root-knot nematodes from Brazil and development of SCAR markers specific for the coffee-damaging species. *Genome* **2002**, *45*, 862–870, doi:10.1139/g02-054.

54. Hussey, R.; Barker, K. A comparison of methods of collecting inocula of Meloidogyne spp., including a new technique. *Plant Dis. Report.* **1973**, *57*, 1025–1028.

55. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J.A.M. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **2007**, *35*, W71–W74, doi:10.1093/nar/gkm306.

56. Zhao, S.; Fernald, R.D. Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. *J. Comput. Biol.* **2005**, *12*, 1047–1064, doi:10.1089/cmb.2005.12.1047.

57. Rancurel, C.; van Tran, T.; Elie, C.; Hilliou, F. SATQPCR: Website for statistical analysis of real-time quantitative PCR data. *Mol. Cell. Probes* **2019**, *46*, 101418, doi:10.1016/j.mcp.2019.07.001.

58. Blaxter, M.L.; Ley, P.D.; Garey, J.R.; Liu, L.X.; Scheldeman, P.; Vierstraete, A.; Vanfleteren, J.R.; Mackey, L.Y.; Dorris, M.; Frisse, L.M.; et al. A molecular evolutionary framework for the phylum Nematoda. *Nature* **1998**, *392*, 71–75, doi:10.1038/32160.

59. Blaxter, M.; Koutsovoulos, G.; Jones, M.; Kumar, S.; Elsworth, B. Phylogenomics of Nematoda. In *Next Generation Systematics*; Olson, P.D., Hughes, J., Cotton, J.A.E., Eds.; Systematics Association Special Volume Series; Cambridge University Press, 2016; pp. 62–83.

60. van Megen, H.; van den Elsen, S.; Holterman, M.; Karssen, G.; Mooyman, P.; Bongers, T.; Holovachov, O.; Bakker, J.; Helder, J. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* **2009**, *11*, 927–950.

61. Bird, D.M.; Jones, J.T.; Opperman, C.H.; Kikuchi, T.; Danchin, E.G.J. Signatures of adaptation to plant parasitism in nematode genomes. *Parasitology* **2015**, *142*, S71–S84, doi:10.1017/S0031182013002163.

62. Smythe, A.B.; Holovachov, O.; Kocot, K.M. Improved phylogenomic sampling of free-living nematodes enhances resolution of higher-level nematode phylogeny. *BMC Evol. Biol.* **2019**, *19*, 121, doi:10.1186/s12862-019-1444-x.

63. 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.

64. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259, doi:10.1093/nar/gkz239.

65. Ku, C.; Martin, W.F. A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70 % rule. *BMC Biol.* **2016**, *14*, 89, doi:10.1186/s12915-016-0315-9.

66. Craig, J.P.; Bekal, S.; Niblack, T.; Domier, L.; Lambert, K.N. Evidence for Horizontally Transferred Genes Involved in the Biosynthesis of Vitamin B-1, B-5, and B-7 in Heterodera glycines. *J. Nematol.* **2009**, *41*, 281–290.

67. Scholl, E.H.; Thorne, J.L.; McCarter, J.P.; Bird, D.M. Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biol.* **2003**, *4*, R39, doi:10.1186/gb-2003-4-6-r39.

68. Paganini, J.; Campan-Fournier, A.; Da Rocha, M.; Gouret, P.; Pontarotti, P.; Wajnberg, E.; Abad, P.; Danchin, E.G.J. Contribution of Lateral Gene Transfers to the Genome Composition and Parasitic Ability of Root-Knot Nematodes. *PLoS ONE* **2012**, *7*, e50875, doi:10.1371/journal.pone.0050875.

69. Dutta, T.K.; Banakar, P.; Rao, U. The status of RNAi-based transgenic research in plant nematology. *Front. Microbiol.* **2015**, *5*, doi:10.3389/fmicb.2014.00760.

70. 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.

71. 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.

72. 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.

73. Miyashita, N.; Yabu, T.; Kurihara, T.; Koga, H. The Feeding Behavior of Adult Root-knot Nematodes (Meloidogyne incognita) in Rose Balsam and Tomato. *J. Nematol.* **2014**, *46*, 296–301.

74. Johnston, M.J.G.; McVeigh, P.; McMaster, S.; Fleming, C.C.; Maule, A.G. FMRFamide-like peptides in root knot nematodes and their potential role in nematode physiology. *J. Helminthol.* **2010**, *84*, 253–265, doi:10.1017/S0022149X09990630.

75. Page, A.P.; Stepek, G.; Winter, A.D.; Pertab, D. Enzymology of the nematode cuticle: A potential drug target? *Int. J. Parasitol. Drugs Drug Resist.* **2014**, *4*, 133–141, doi:10.1016/j.ijpddr.2014.05.003.

76. Tautz, D.; Domazet-Lošo, T. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **2011**, *12*, 692–702, doi:10.1038/nrg3053.

77. McLysaght, A.; Guerzoni, D. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Philos. Trans. R. Soc. B Biol. Sci.* **2015**, *370*, 20140332, doi:10.1098/rstb.2014.0332.

78. Lunt, D.H.; Kumar, S.; Koutsovoulos, G.; Blaxter, M.L. The complex hybrid origins of the root knot nematodes revealed through comparative genomics. *PeerJ* **2014**, *2*, e356, doi:10.7717/peerj.356.