Identification of candidate effector genes in the transcriptome of the rice root knot nematode *Meloidogyne graminicola*

ANNELIES HAEGEMAN, LANDER BAUTERS, TINA KYNDT, MOHAMMAD MASUDER RAHMAN† AND GODELIEVE GHEYSSEN*

Department of Molecular Biotechnology, Ghent University, B-9000, Ghent, Belgium

SUMMARY

Plant-parasitic nematodes secrete so-called effectors into their host plant which are able to suppress the plant's defence responses, alter plant signalling pathways and, in the case of root knot nematodes, induce the formation of giant cells. Putative effectors have been successfully identified by genomics, transcriptomics and proteomics approaches. In this study, we investigated the transcriptome of the rice root knot nematode *Meloidogyne graminicola* by 454 sequencing of second-stage juveniles as well as mRNA-seq of rice infected tissue. Over 350 000 reads derived from *M. graminicola* preparasitic juveniles were assembled, annotated and checked for homologues in different databases. From infected rice tissue, 1.4% of all reads generated were identified as being derived from the nematode. Using multiple strategies, several putative effector genes were identified, both pioneer genes and genes corresponding to already known effectors. To check whether these genes could be involved in the interaction with the plant, *in situ* hybridization was performed on a selection of genes to localize their expression in the nematode. Most were expressed in the gland cells or amphids of the nematode, confirming possible secretion of the proteins and hence a role in infection. Other putative effectors showed a different expression pattern, potentially linked with the excretory/secretory system. This transcriptome study is a good starting point to functionally investigate novel effectors derived from *M. graminicola*. This will lead to better insights into the interaction between these nematodes and the model plant rice. Moreover, the transcriptome can be used to identify possible target genes for RNA interference (RNAi)-based control strategies. Four genes proved to be interesting targets by showing up to 40% higher mortality relative to the control treatment when soaked in gene-specific small interfering RNAs (siRNAs).

INTRODUCTION

The complex interaction of plant-parasitic nematodes and their host plant(s) has been a subject of study for several decades. In

1998, the first 'parasitism gene', necessary for successful plant infection, was discovered, namely a nematode gene encoding β -1,4-endoglucanase capable of degrading cellulose in the plant cell wall (Smant *et al.*, 1998). Since then, other nematode plant cell wall-degrading enzymes have been found, as well as numerous other genes whose protein products play a role in the parasitic interaction with the host plant (reviewed by Haegeman *et al.*, 2012). These genes are nowadays referred to as 'effectors', in line with other plant pathology disciplines. The identification of putative effectors is a challenging task, and several techniques have been used for this purpose. As effectors are presumed to be secreted from the nematode into the plant tissue, an effective strategy to discover effectors is to identify nematode secretions directly by proteomics approaches. However, a more common approach is to investigate the transcriptome of infective and parasitic nematode stages, or specific gland cell tissue. Bioinformatics tools can subsequently predict putatively secreted proteins, which can be considered as good effector candidates.

One of the main staple foods is rice (*Oryza sativa* L.), with a worldwide production of over 672 million tonnes in 2010 (Food and Agriculture Organization, <http://faostat.fao.org/>). The rice root knot nematode *Meloidogyne graminicola* is the most damaging root knot nematode of rice, and is mainly distributed in South and South-East Asia, although it has also been reported in the USA, Brazil and Colombia (Bridge *et al.*, 2005). Recently, because of socio-economic and environmental changes, water shortages in South-East Asia have increased. This is gradually causing a shift from traditional paddy production systems, consuming large amounts of water, to water-saving rice production systems, such as direct wet seeding, intermittent irrigation, cultivation on raised beds and cultivation of aerobic rice varieties. Because of this change in culture system, the development of large populations of *M. graminicola* is favoured (De Waele and Elsen, 2007; Soriano *et al.*, 2000). Infective second-stage juveniles (J2s) invade rice roots just behind the root tip, where they form specialized feeding cells, called giant cells, causing galls predominantly at the root tips. A characteristic symptom of *M. graminicola* infection is the hook-like structure of the galls at the infected root tips. Its life cycle is completed in 19 days at a temperature of 22–29 °C. Through successive moults, females develop within the root and the egg masses remain inside the root. Hatched juveniles can either stay in the maternal gall or migrate within the same root.

*Correspondence: Email: godelieve.ghaysen@ugent.be

†Present address: Department of Biotechnology and Genetic Engineering (BGE), Mawlana Bhashani Science and Technology University (MBSTU), Santosh, Tangail-1902, Bangladesh.

The fact that a full life cycle can be completed without leaving the rice root is thought to be an adaptation to flooded conditions. If J2s do leave the root tissue, they cannot reinvade in flooded conditions, although they remain viable in the soil for at least 27 days, where they can enter new roots when the soil is drained. Above-ground symptoms include stunting and chlorosis of young plants prior to flooding. Submerged plants with serious infection are unable to grow above the water level (Bridge and Page, 1982).

Meloidogyne graminicola is a parasite of monocotyledons, with rice as the main host plant. As the genome of rice is available (Goff *et al.*, 2002), it is a good model to study the interaction between the nematode and its natural host. In addition, the genome sequences of two related root knot nematodes, namely *Meloidogyne incognita* (Abad *et al.*, 2008) and *Meloidogyne hapla* (Opperman *et al.*, 2008), were published a few years ago. These genomes can serve as tools to study genes important for successful infection of the plant. In this light, we decided to perform a transcriptome analysis on J2s of *M. graminicola* in order to identify putative novel effector genes. Furthermore, we analysed mRNA-seq data derived from nematode-infected rice tissue in order to identify nematode transcripts specifically expressed when the nematode resides inside the plant. These two approaches led to the identification of new candidate effector genes.

RESULTS

454 sequencing of *M. graminicola* J2s

Assembly and BLAST searches

The 454 sequencing run resulted in over 350 000 reads with an average length of 378 bp. After cleaning, vector trimming, rice sequence removal and assembly, these reads were condensed to 10 257 contigs longer than 100 bp (Table 1). The average length of these assembled sequences increased to 658 bp and ranged between 100 and 3056 bp. On average, each contig consisted of 21 reads.

To assign putative functions to the contigs, a BLAST search against the Swiss-prot and trEMBL databases was performed. This revealed that 7380 sequences (72%) showed significant similarity to sequences in these databases. As most data available for nematodes have not yet been annotated and are only available as expressed sequence tag (EST) data, we also performed similarity searches against all nematode ESTs. All nematode ESTs were divided into three smaller datasets containing sequences from

Fig. 1 Classification of contigs with significant similarity to nematode expressed sequence tags (ESTs) according to the nematode's lifestyle (APN, animal-parasitic nematodes; PPN, plant-parasitic nematodes; FLN, free-living nematodes).

free-living nematodes only, animal-parasitic nematodes only and plant-parasitic nematodes only. When the assembled sequences were BLASTed against all nematode ESTs, 6112 contigs (60%) showed hits. The distribution of these hits over the different nematode datasets according to the nematode's lifestyle is shown in Fig. 1. In total, 1332 contigs showed similarity exclusively to parasitic nematode ESTs.

The genus *Meloidogyne* is relatively well studied with numerous ESTs and two genomes available for *M. hapla* and *M. incognita*. Phylogenetic analyses have revealed that the three species, *M. hapla*, *M. incognita* and *M. graminicola*, are in separate subgroups within the genus *Meloidogyne* (De Ley *et al.*, 2002). Therefore, it was interesting to compare how many sequences of *M. graminicola* had putative orthologues in other *Meloidogyne* species. A comparison with the data available for other *Meloidogyne* species resulted in 4570 sequences (44.6%) with significant similarity to *Meloidogyne* ESTs, 5714 (55.7%) to the *M. hapla* genome and 5275 (51.4%) to the *M. incognita* genome.

Orthologues between the *Meloidogyne* species were determined by a reciprocal BLAST strategy using the *M. graminicola* contigs and the predicted proteins from the two *Meloidogyne* genomes. The *M. graminicola* sequences had 2291 orthologues in *M. incognita* and 2536 in *M. hapla*. In comparison, this strategy resulted in 6580 orthologues between *M. incognita* and *M. hapla*. When all three species were compared, 1321 sequences were retained as core orthologues (Fig. 2). The corresponding genes could be considered as the most conserved in the genus *Meloidogyne*. In general, more orthologues were shared between *M. graminicola* and *M. hapla* than between *M. graminicola* and *M. incognita*.

	Sequences	Total bases	Average length (bp)
Total number of reads	351 311	132 957 899	378
High-quality reads >70 bp	320 094	125 412 518	392
High-quality reads >70 bp after pDNR-lib and rice removal	258 493	98 211 402	380
Contigs >100 bp after assembly	10 257	6 751 139	658

Table 1 Overview of sequencing run, quality and assembly.

Fig. 2 Venn diagram showing the number of orthologous sequences identified between the *Meloidogyne graminicola* (Mg) contigs and the predicted proteins from the *M. incognita* (Mi) and *M. hapla* (Mh) genomes.

Annotation and protein prediction

To assign putative functions to the *M. graminicola* sequences, an annotation was performed by linking the sequences to the protein identifier showing greatest similarity. In this way, 5919 sequences (57.7%) were successfully annotated. Ninety-two per cent of the annotated sequences had a match for over 30% of the length of the query sequence; 4733 of the protein identifiers assigned to the annotated sequences were unique, and were used to retrieve general gene ontology (GO) terms. Most annotated putative proteins occurred in the cytoplasm, had a function in nucleotide binding and were involved in transport (Fig. 3). To search for processes or functions that were more abundant in *M. graminicola* than in the free-living nematode *Caenorhabditis elegans*, the over- or under-representation of GO terms was determined in relation to all *C. elegans* GO terms. Of all *M. graminicola* sequences, 4838 (47.2%) had a similar *C. elegans* protein, corresponding to 2771 different genes. From the 'cellular component' category, it could be derived that transcripts occurring in the cytoplasm were more abundant in the *M. graminicola* sequences, whereas transcripts linked to the membrane occurred much less frequently. In the 'molecular function' category, the most overrepresented terms were catalytic, phosphatase, hydrolase and ATPase activity, as well as nucleotide binding, whereas receptor and signal transducer activity were the most under-represented terms. With regard to the 'biological process' category, the most pronounced differences were the over-representation of terms related to larval growth and development, whereas sensory perception and the regulation of transcription and metabolic processes were under-represented.

After protein prediction, 9112 putative proteins longer than 50 amino acids were retained (88% of the total amount of sequences). Of these putative proteins, 545 started with a methionine, had a signal peptide and did not have a transmembrane domain.

Effector identification and prediction

All known and annotated putatively secreted proteins from plant-parasitic nematodes were compared with the *M. graminicola* contigs to identify possible homologues. Putative homologues

were identified for the most common plant cell wall-modifying enzymes (β -1,4-endoglucanase, pectate lyase, xylanase, expansin and cellulose-binding protein), as well as for secreted proteins involved in the detoxification of reactive oxygen species (ROS) (peroxiredoxin, glutathione peroxidase, glutathione-S-transferase) (Table S2, see Supporting Information). Other putative homologues of effectors included transcripts encoding fatty acid- and retinol-binding protein, annexin, calreticulin, chitinase, transthyretin-like protein, mitogen-activated protein (MAP-1), venom allergen-like protein, galectin, C-type lectin and 14-3-3 protein. The *M. graminicola* contigs were also compared with the putative effector proteins identified in *M. incognita* by Huang *et al.* (2003). Surprisingly, only four of 36 genes seemed to have homologues in *M. graminicola*, namely 4D03, 19F07, 35E04 (all expressed in the dorsal gland) and 30G11 (an acid phosphatase expressed in the subventral glands) (Huang *et al.*, 2003). These putative effectors could be interesting to study further as they seem to be conserved in the genus *Meloidogyne*.

The usual approach for the identification of putative novel effectors is to check the proteins for the presence of a signal peptide and the absence of a transmembrane domain. However, in the case of transcriptome data, this approach is not always effective as numerous transcripts lack the 5' end and hence the region potentially encoding a signal peptide. Therefore, we combined the approach of looking for proteins with a signal peptide with an additional BLAST search that compared the *M. graminicola* sequences with putatively secreted proteins from the genomes of *M. hapla*, *M. incognita* and *Bursaphelenchus xylophilus*.

545 protein sequences contain a signal peptide and do not have a transmembrane domain. 1328 sequences show similarity to putatively secreted proteins from the genomes of *M. hapla*, *M. incognita* and *B. xylophilus*. When both resulting datasets were combined, 1684 different sequences could be considered as putative effectors. To further narrow down these putative effector sequences, a BLAST search was performed against all nematode ESTs, and all sequences that showed similarity to ESTs from free-living nematodes were removed. This resulted in a final set of 499 different putative effectors (Fig. 4).

Most of the identified putative effector sequences did not have an annotation based on searches against Swiss-prot and trEMBL, and hence we have no clue about their possible function. Some sequences had an annotation, and most were similar to the previously identified plant cell wall-modifying enzymes or other proteins identified in nematode secretions, such as FMRF amide-like peptide. When compared with nematode ESTs, 52 of the retained putative effectors showed similarity exclusively to ESTs from *Meloidogyne* species, whereas 72 also showed similarity to migratory nematode ESTs and 47 also resembled sequences from animal-parasitic nematodes. Only four sequences showed similarity exclusively to both cyst and root knot nematode ESTs; 324 sequences did not show any similarity to either nematode EST sequences or

Fig. 3 Percentage of occurrence of the top five gene ontology (GO) terms in the categories 'cellular component', 'molecular function' and 'biological process'.

Fig. 4 Approach to predict putative effector genes in the *Meloidogyne graminicola* sequences dataset. EST, expressed sequence tag; FLN, free-living nematodes; ORF, open reading frame.

Swiss-prot/trEMBL, but were retained because of their similarity to putative effectors from other *Meloidogyne* genomes and/or because of the presence of a putative signal peptide in the predicted protein.

Target genes for RNA interference (RNAi)

One of the proposed strategies to control plant-parasitic nematodes is *in planta* RNAi (Gheysen and Vanholme, 2007). To identify genes that are suited as RNAi targets, we searched for *C. elegans* homologues and selected those with lethal RNAi phenotypes assuming that the orthologue of *M. graminicola* could also have a lethal RNAi phenotype. To minimize off-target effects in humans, other plants and insects in the environment, we ensured, by performing different BLAST searches, that any similarity between the gene and plant, human or insect sequences was absent. Of all the *M. graminicola* sequences, 4838 showed similarity to a *C. elegans* protein, corresponding to 2771 different genes. RNAi phenotypes

could be retrieved for 1592 genes, 1008 of which had a lethal phenotype. These potential RNAi target genes were compared with all plant proteins, and this resulted in 93 genes without any similarity to plant proteins. The latter genes were screened against all human proteins, resulting in 29 genes without similarity to human proteins. These were subsequently screened for proteins with similarity to insect proteins, resulting in 22 sequences without similarity. In conclusion, 22 *M. graminicola* sequences were interesting RNAi targets as they showed similarity to a *C. elegans* gene with a lethal RNAi phenotype, and did not show similarity to plant, human or insect proteins. Four of these RNAi targets had over 20 reads in the *M. graminicola* dataset; these are shown in Table 2. The function of most of these genes is unknown. Three of these genes seem to be nematode specific, whereas *gob-1* also resembles bacterial sequences.

The lethality of small interfering RNAs (siRNAs) targeting each of the four retained genes was evaluated by a soaking experiment. The results of one of these experiments are shown in Fig. 5. *Meloidogyne graminicola* treated with siRNAs against green fluorescence protein (GFP) showed a significantly higher mortality than nematodes soaked in water, despite the relatively low concentration of siRNAs used (25 ng/μL). Dalzell *et al.* (2009) also observed a toxic effect of non-nematode-derived double-stranded RNA (dsRNA) on *M. incognita*, although this effect was less pronounced than that in this study. Nevertheless, the samples treated with siRNAs against the four target genes all showed a higher mortality relative to the GFP siRNA-treated sample. All samples were significantly different from the GFP siRNA treatment, except for nematodes treated with siRNAs against contig5704. The same experiment was performed three times, resulting in variable mortality levels, but, in all experiments, the four samples treated with siRNAs to nematode genes showed a smaller number of living nematodes than the GFP siRNA-treated sample.

Table 2 *Meloidogyne graminicola* contigs with a number of reads higher than 20 that are suitable RNA interference (RNAi) targets because they have a lethal RNAi phenotype in *Caenorhabditis elegans*, but do not show similarity to plant, human and insect sequences. The *M. graminicola* (Mg) contig, its number of reads, the *C. elegans* homologue, its RNAi phenotype and the description of the *C. elegans* protein are shown.

Mg contig	No. reads	<i>C. elegans</i> homologue	RNAi phenotype	Description <i>C. elegans</i> protein
contig00149	193	pqn-83	Larval lethal	Prion-like-(Q/N-rich)-domain-bearing protein family member
contig05704	79	C44C1.1	Larval lethal	Hypothetical protein
contig06825	42	F36A2.7	Embryonic lethal	Hypothetical protein
contig06992	21	gob-1	Lethal	Gut obstructed or defective family member

Fig. 5 Percentage of living nematodes after 20 h of soaking with gene-specific small interfering RNAs (siRNAs) (25 ng/μL). Treatments that are significantly different ($P < 0.05$) from the green fluorescent protein (GFP) treatment are indicated by stars.

Identification of nematode transcripts from mRNA-seq data of infected rice tissue

Recently, Kyndt *et al.* (2012) have studied the rice transcriptome up on *M. graminicola* infection by performing mRNA-seq on rice gall tissue. Approximately one-half of the sequence reads produced (16 283 494) did not map on the rice genome and hence probably contained nematode-derived transcripts. In this study, all unmapped sequence reads from the samples infected by *M. graminicola* were assembled into 10 199 contigs. These were checked for similarities to rice protein sequences in Swiss-prot and trEMBL, as these databases may contain other rice sequence information from the annotated genome; 6435 contigs did not match rice sequences and were subsequently screened for nematode-derived contigs by checking for similarity to: (i) the predicted proteins from the genomes of *C. elegans*, *M. hapla* and *M. incognita*; (ii) all nematode ESTs; and (iii) the contigs derived from the *M. graminicola* J2 454 sequencing data. This resulted in a total of 3865 putative nematode-derived contigs (Fig. 6). These sequences consisted of 438 027 reads, corresponding to 2.7% of the unmapped sequences, or 1.4% of the total amount of sequences generated (30 769 323). The most abundant nematode gene in the data was collagen, which consisted of several contigs (5240 reads in total), corresponding to 1.2% of the nematode transcripts. From the hits with *C. elegans*, associated GO terms could be derived

Fig. 6 Approach to identify nematode-derived sequences from the mRNA-seq data from rice infected tissue. EST, expressed sequence tag; J2, second-stage juvenile; N, No; Y, Yes.

and counted. This revealed that most identified nematode sequences were involved in the biological processes 'embryonic development ending in birth or egg hatching', 'nematode larval development', 'locomotion' and 'growth'.

Of the 3865 nematode-derived contigs, 2511 showed similarity to 962 different contigs from the J2 transcriptome data. Of the 499 putative effectors identified from the J2 transcriptome data, 23 were present in the RNA sequence data derived from galls, confirming their expression when the nematode is inside the plant tissue. Most of these transcripts (18) were more abundant in the galls sampled at 7 days after infection relative to 3 days after infection, whereas five transcripts showed the reverse pattern; 1490 contigs showed similarity exclusively to parasitic nematode sequences, and hence these are potential nematode effectors. As a result of the relatively short nature of the sequences, the prediction of signal peptides from putative protein sequences was considered to be unreliable. Therefore, we believe that the nematode-derived sequences identified in the gall tissue should be considered as additional data to the 454 transcriptome data of J2s, e.g. to prioritize effector candidates or to check at what time point they are more abundantly expressed in the plant tissue.

A total of 2570 sequences were not derived from rice or from the nematode (Fig. 6). Contaminating sequences from other organisms were expected to be present as the infections were not performed under sterile circumstances. They were more closely examined by BLASTing to Swiss-prot and trEMBL databases to check from which organism they could have originated. Only 612 sequences had a hit, and 93% of these pointed to a bacterial origin, corresponding to 16 198 reads, or 0.05% of the total amount of generated sequences. Approximately one-half of these bacterial sequences (52%) were derived from *Burkholderia* species, suggesting that the investigated rice tissues experienced a minor infection from this plant-pathogenic bacterium, although the symptoms of this disease (leaf sheath browning and grain rot) were not observed. The remaining sequences without any similarities could either be artefacts from the sequencing or assembly, or could have been too short to detect any significant similarity in the databases (mean length of 103 bp).

***In situ* hybridization**

An *in situ* hybridization was performed for 26 putative effector genes, selected from the list of 499 putative effectors identified from the J2 transcriptome data. Only genes that had a relatively high coverage (number of reads > 20) were selected, as the *in situ* hybridization protocol has a low sensitivity and success largely depends on the expression level of the gene. Five of the tested genes were also present in the sequences derived from infected rice. The negative control was checked carefully and did not show any signal (data not shown). Most of the genes tested did not show any clear signal in the *in situ* hybridization. Only when the signal was observed in over approximately 50% of the cut nematodes was it considered to be genuine. For six genes, including a pectate lyase, a positive signal was observed in the subventral glands of the nematode (Fig. 7). Three genes showed expression in the dorsal gland, one of which was a homologue to the *M. incognita* 19F07 protein that has been shown previously to be expressed in the dorsal gland (Huang *et al.*, 2003) (Fig. 7). Two other genes showed a similar expression pattern in an unidentified organ in proximity to the gland cells, stretching laterally towards the tail of the nematode (Fig. 7). Finally, one gene showed expression in the amphids (Fig. 7). Transcripts of this gene were also identified in the infected rice tissue (7 days after infection).

DISCUSSION

In this article, we have studied the transcriptome of the root knot nematode *M. graminicola*, and present an effector identification pipeline for this species. Using a 454 FLX Titanium Platform, over 350 000 reads were generated derived from preparasitic J2s. After assembly, 10 257 contigs longer than 100 bp remained with a relatively high average length of 658 bp. As a result of these relatively long assembled transcripts, 72% of the sequences

showed similarity to proteins in Swiss-prot and trEMBL, which is considerably higher than in other nematode 454 EST projects (Cantacessi *et al.*, 2010; Haegeman *et al.*, 2011; Nicol *et al.*, 2012). Unexpectedly, when comparing the *M. graminicola* sequences with all predicted protein sequences of *M. hapla* and *M. incognita*, the number of homologous sequences was relatively low, 56% and 51%, respectively. These low similarities could have several explanations. The transcriptome data also contain 5' and 3' untranslated regions, which are less conserved. Another reason could be that the three *Meloidogyne* species belong to different phylogenetic clades, hence showing considerable genetic differences (De Ley *et al.*, 2002). Despite the relatively low similarities to the predicted proteins of *M. incognita* and *M. hapla*, 58% of all sequences were successfully annotated based on sound descriptions of BLAST hits against Swiss-prot and trEMBL. These annotated sequences were functionally categorized into GOs, revealing that most putative proteins occur in the cytoplasm, have a function in the binding to nucleotides and metal ions, or are involved in transport, translation or oxidation–reduction processes. Checking over- and under-represented GO terms, as opposed to the GO terms of all *C. elegans* proteins, showed that the *M. graminicola* dataset had a high number of sequences involved in nucleotide binding, as well as numerous sequences with catalytic, phosphatase, hydrolase and ATPase activity. Terms in larval development and growth and development were also over-represented, but this was expected as the sequences were derived from J2s.

When studying a plant parasite, the main objective is to understand its interactions with the host plant and, more specifically in this case, to identify so-called effectors or proteins that are secreted into the plant tissue to establish a successful infection. To look for effectors, several approaches were combined in this study. First, putatively secreted proteins described in other plant-parasitic nematodes were downloaded, and putative homologues to these proteins were searched for in the *M. graminicola* data. This resulted in the identification of several known effectors, such as plant cell wall-modifying proteins, as well as several other putatively secreted proteins from different plant-parasitic nematodes. Surprisingly, only very few homologues to pioneer effectors identified in *M. incognita* (Huang *et al.*, 2003) were present in the *M. graminicola* transcriptome. This suggests that the transcriptome from *M. graminicola* is not complete, e.g. because of the low expression levels of some genes, or that *M. graminicola* uses a diverged set of effectors, possibly as a result of the fast evolution of these proteins. In a second approach, putative novel effectors were identified from the dataset by combining several search strategies. As the *M. incognita*, *M. hapla* and *B. xylophilus* genomes are available, all predicted proteins from these three species were searched for putative effector genes, and putative homologues to these genes were identified in the *M. graminicola* data. In addition, sequences with similarity to parasitic nematode ESTs exclusively, as well as sequences with a putative signal

<i>In situ</i> hybridization	result	contig	FL	#AA	SP	nema EST hits	reads	annotation by BLASTX	Mi, Mh, Bx effector homologue
	SvG	contig06513	Y	264	Y	PPN (M+S)	79	Pectate lyase	Minc03618 Mh_Contig400_8836..9978 BUX.s01259.23
	SvG	contig01965	Y	226	Y	PPN (Mel.)	165	C-type lectin	Minc03867 Mh_Contig357_7141 1..72190
	SvG	contig03718	Y	187	Y	PPN (Mel.)	37	C-type lectin	Minc01589 Mh_Contig188_16304..17367
	SvG?	contig01318	N	277+	?	PPN(Mel.)	51	C-type lectin	Minc0241 1 Mh_Contig296_6244..7980
	SvG	contig06883	Y	247	Y	PPN+APN	45	Ancylostoma secreted prot	Minc00970 Mh_Contig481_59954..61740 BUX.s01653.397
	SvG	contig01796	Y	224	Y	PPN (Mel.)	158	-	-
	DG	contig06410	Y	316	Y	PPN (Mel.)	60	Oesophageal gland cell prot 32 (19F07)	Minc17182
	DG	contig05581	N	344+	?	PPN (Mel.)	35	-	Minc03328
	DG	contig03015	Y	090	Y	no hits	149	-	-
	?	contig01832	N	128+	?	PPN (Mel.)	457	-	Minc13358 Mh_Contig1001_7693..9508
	?	contig00937	N	333+	?	PPN (M+S)	124	-	Minc13358 Mh_Contig1001_7693..9508
	A	contig09259	Y	148	Y	PPN (Mel.)	22	-	-

Fig. 7 *In situ* hybridization of *Meloidogyne graminicola* second-stage juvenile (J2) nematodes. The contig number is given, as well as the location of the signal (A, amphids; DG, dorsal glands; SvG, subventral glands), whether or not the full-length (FL) sequence is available (N, no; Y, yes), the number of amino acids (AA) of the predicted putative protein, whether or not there is a signal peptide (SP) present, the type of hits in a search against all nematode expressed sequence tags (ESTs) (APN, animal-parasitic nematodes; Mel., *Meloidogyne* species only; M + S, migratory and sedentary plant-parasitic nematodes; PPN, plant-parasitic nematodes), the number of reads, the annotation and the potential homologues in *M. incognita* (Mi), *M. hapla* (Mh) and *Bursaphelenchus xylophilus* (Bx) putative effectors.

peptide, were retained, which ultimately resulted in a list of 499 putative effector proteins. Fifty-two putative effectors showed similarity to *Meloidogyne* ESTs only, suggesting that there are a number of effectors that originated relatively recently in the genus *Meloidogyne*, which reflects the fast evolution of these proteins. Other genes did not show any similarity to any sequence in the databases. These 'pioneer' genes with unknown functions and genes that show similarity to parasitic nematode sequences exclusively are of interest for further study. In a first step towards the elucidation of the function of these genes, confirmation is needed to determine whether the corresponding proteins are being secreted from the nematode. *In situ* hybridization is a way to check where certain genes are expressed. When the expression is located in the gland cells of the nematode, it is probable that the proteins are secreted into the plant tissue. We performed an *in situ*

hybridization on a subset of genes from our putative effector list. Several genes, including a pectate lyase, were clearly expressed in the subventral gland cells, indicating that the corresponding proteins are probably secreted into the plant tissue. Three of the putative proteins expressed in the subventral glands consist of a signal peptide linked to a C-type lectin domain. A C-type lectin from *M. chitwoodi* is also expressed in the subventral glands of preparasitic juveniles (Roze *et al.*, 2008) and a C-type lectin of *Heterodera glycines* is expressed in the hypodermis of adult females (de Boer *et al.*, 2002). C-type lectins occur in large numbers in all nematodes, and have a role in protection against possible pathogens (Zelensky and Gready, 2005). In animal-parasitic nematodes, C-type lectins have been shown to be secreted into the host by infective stages, possibly to mediate immunological interactions with the host (Harcus *et al.*, 2009). The

three C-type lectin sequences of *M. graminicola* show similarity exclusively to *Meloidogyne* ESTs, suggesting a possible role in parasitism, for example by interfering with plant defence. One of these transcripts was also identified in the mRNA-seq data from infected rice tissue, with higher expression levels in galls 3 days after infection relative to galls 7 days after infection. This is consistent with its expression in the subventral glands, which are known to be more active during the earlier stages of infection (Davis *et al.*, 2000). Three other putative effector genes showed staining in the dorsal gland cell, whereas two others showed staining laterally along the gland cells, stretched towards the posterior part of the nematode. It is not clear to which tissue the latter staining pattern corresponds. If it concerns the excretory/secretory system, the putative proteins could be secreted from the nematode as well, and could be involved in parasitism. It has been shown that the excretory/secretory system can indeed be a source of surface components in plant-parasitic nematodes (Bird *et al.*, 1988). Finally, one pioneer gene was expressed in the amphids, a localization that has also been linked to effectors (Vieira *et al.*, 2011).

The most obvious way to study the transcriptome of a plant parasite is to study the parasite in its natural environment, namely in the host plant. New techniques developed in the last decade have helped this form of multi-species transcriptomics, and these studies are becoming increasingly important in understanding plant–parasite relationships (Schenk *et al.*, 2012). Multi-species transcriptomics have so far mainly been performed by microarrays, e.g. Moy *et al.* (2004) developed a microarray that contained both soybean and *Phytophthora sojae* transcripts. Recently, high-throughput RNA sequencing has opened up new possibilities for the study of host–parasite interactions. Kyndt *et al.* (2012) generated mRNA-seq data from rice gall tissue induced by *M. graminicola* in order to study the responses of the host plant. Although it was not the initial purpose of the study, it was interesting to determine whether nematode transcripts could be identified from such data. After careful selection, it was deduced that 1.4% of all generated reads were designated to be derived from the nematode. The major drawback of the technique was the short nature of the sequences (76 bp). Without a reference genome available, it is very difficult to assign annotations to the transcripts and impossible to predict putative effectors from these data. Therefore, in this particular study, 454 sequencing would probably have been better suited to identify sequences from a nonmodel organism inside the host plant. A similar analysis was performed in coffee plants infected by a rust and, in this case, 20% of the reads generated by 454 sequencing were derived from the pathogen and the contigs were relatively long, which allowed the prediction of putative effectors (Fernandez *et al.*, 2012). In our case, the generated mRNA-seq data are useful in combination with the transcriptome generated from the preparasitic J2s. For example, it could be deduced that 23 of the identified 499 effectors were also found in the mRNA-seq data on galls. This proves that these putative

effector genes are expressed by the nematode whilst the nematode resides in the plant tissue, potentially by later stages than J2.

Although the identification of putative effectors was the main objective of this study, the transcriptome can also be used for other purposes. One of these is to look for genes that are interesting targets for parasite control. If we can identify proteins that are likely to be essential for the nematode's survival, these could be targets for knockdown strategies, such as *in planta* RNAi. Several candidate genes were identified on the basis of the following criteria: (i) the homologous *C. elegans* gene has a lethal RNAi phenotype; (ii) the gene is relatively highly expressed; and (iii) the gene does not show similarity to sequences from plants, insects or humans, to minimize off-target effects of RNAi. Four *M. graminicola* genes were predicted to be promising candidates for RNAi control strategies, although functional information for most of these genes is lacking. Soaking experiments targeting these four genes showed that gene-specific siRNAs indeed had a lethal effect on the nematodes, with up to 40% less survival compared with nematodes soaked in GFP siRNA. Despite the low concentration of siRNAs used, the siRNAs against GFP appeared to be toxic when compared with the water-treated nematodes. This confirms the observations of Dalzell *et al.* (2009), who showed similar off-target effects when treating *M. incognita* with low concentrations of non-nematode-derived dsRNA. When the soaking experiment was independently repeated, the lethality of siRNAs varied slightly, although the treated nematodes all showed a significant decrease in viability relative to nematodes treated with GFP siRNA in at least one of the repeated experiments. Variable and toxic effects of dsRNA and siRNA have been observed previously, and are largely dependent on the nematode species and the targeted gene (reviewed by Lilley *et al.*, 2012). Nevertheless, the soaking experiments clearly show that the four identified genes are indeed good targets for RNAi-based control strategies.

In conclusion, the 454 sequencing technique is well suited for transcriptomic studies on nonmodel organisms with little sequence information available. The data obtained will be useful as a starting point to study specific genes or gene families possibly involved in parasitism or suitable as RNAi targets. mRNA-seq on plant tissue infected by nematodes also resulted in a substantial amount of nematode-derived sequences. Although the reads were too short to assemble contigs sufficiently large for protein prediction, the contigs were still useful in confirming the presence of certain nematode transcripts inside the plant.

EXPERIMENTAL PROCEDURES

454 sequencing and data analysis on *M. graminicola* J2s

Nematode collection, cDNA preparation and sequencing

J2 stages of *M. graminicola* were collected from an *in vivo* culture on rice plants by a modified Baermann method. RNA was isolated using TRI

reagent (Sigma, St Louis, MO, USA), and first-strand cDNA synthesis and amplification were performed with the Mint Universal cDNA synthesis kit (Evrogen, Moscow, Russia). The amplified cDNA was normalized using the TRIMMER kit (Evrogen) and subsequently transformed into the vector pDNR-lib (Clontech, Mountain View, CA, USA) to obtain a cDNA library. The cDNA library was sequenced on one-half of a picotitre plate on a 454 FLX Titanium Platform (Roche, Branford, CT, USA) by the company LGC Genomics (Berlin, Germany) according to the manufacturer's instructions. The raw sequence data were submitted to the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) and has accession number SRA049061.

Cleaning and assembly

The resulting SFF file was converted to a fasta and a qual file using the SFF converter (v1.0.1) from the Galaxy Project (Goecks *et al.*, 2010), including removal of the low-quality ends of the sequences. The sequences were subsequently processed with the NGS QC Toolkit version 2.2.3 (Patel and Jain, 2012). Reads with a minimum length of 70 bp and a minimum quality score of 20 over more than 70% of the read length were retained. The following adapter sequences were removed: 454 sequencing primer A (GCCTCCCTCGGCCATCAG), 454 sequencing primer B (GCCTTGCAGCCCGCTCAG), SMART 5' adapter (AAGCAGTGGTATCAACGCAGAGTACGCGGG) and SMART 3' adapter (AAGCAGTGGTATCAACGCAGAGTACT). Vector contamination was determined by performing a BLASTN search (bit score > 30) against the pDNR-LIB sequence, and the resulting vector pieces were subsequently clipped from all reads using custom Perl scripts. Contaminating rice sequences were found by a BLASTN search (bit score > 50) against all rice coding sequences ('all.cds', version 7.0) downloaded from the rice genome annotation website (<http://rice.plantbiology.msu.edu/>). Reads that showed over 98% identity to a rice sequence for over 75% of the length of the read were removed. Small reads (<70 bp) were also removed. The assembly using the preprocessed data was performed using the GS *de novo* assembler (Roche, version 2.6) with default settings.

BLAST searches

Assembled sequences longer than 100 bp were retained and BLASTed (BLASTX) locally against the Swiss-prot/trEMBL protein database (March 2012). In all BLAST searches mentioned above, a bit score of 50 was used as a cut-off to identify significant hits.

All publicly available nematode ESTs were downloaded and classified according to the lifestyle of the nematode, resulting in three EST databases of free-living nematodes (FLN), animal-parasitic nematodes (APN) and plant-parasitic nematodes (PPN). The data included all sequences available from NCBI's dbEST, as well as contigs from recent 454 sequencing projects on *Pratylenchus coffeae* (Haegeman *et al.*, 2011), *Pratylenchus thornei* (Nicol *et al.*, 2012) and *Hirschmanniella oryzae* (in-house data) (data summarized in Table S1, see Supporting Information). A TBLASTX of the *M. graminicola* sequences was performed against each of these EST databases to identify parasite-specific sequences. The *M. graminicola* nucleotide sequences were also searched with BLASTN and BLASTX against the genome contig data and the predicted proteins, respectively, both downloaded from the genome project websites for *M. incognita* and *M. hapla* (http://www.inra.fr/meloidogyne_incognita, <http://www.hapla.org>) (Abad *et al.*, 2008; Opperman *et al.*, 2008).

To find orthologues from *M. graminicola* to *M. incognita* and *M. hapla*, a reciprocal BLAST strategy was used. All combinations of BLAST searches (BLASTX and TBLASTN) were performed and sequences were considered to be orthologous when the best BLAST match selected the original sequence as top hit in the reciprocal BLAST. For *M. incognita* and *M. hapla*, the predicted protein sequences were used, whereas, for *M. graminicola*, the search was performed with the assembled nucleotide sequences longer than 100 bp.

Annotation and protein prediction

Annotation was performed based on the top hit from the BLASTX search result against the Swiss-prot/trEMBL database. Top hits from Swiss-prot sequences with the terms 'unknown', 'putative', 'uncharacterized', 'hypothetical', 'similar', 'predicted' and 'probable' in their description were not considered. For the sequences with successful annotation, the protein identifiers of the BLASTX top hits were used to retrieve GO terms using QuickGO (<http://www.ebi.ac.uk/QuickGO/GAnnotation>). To check for enriched GO terms as opposed to all *C. elegans* proteins, the following approach was used. The most homologous *C. elegans* proteins were determined for all *M. graminicola* sequences in a local BLASTX search (bit score > 50) against all *C. elegans* proteins downloaded from WormBase (<http://www.wormbase.org>, release WS227). The *C. elegans* gene identifiers were loaded into Cytoscape 2.8.2 (Shannon *et al.*, 2003) and GO terms were retrieved. Using the Cytoscape plugin BiNGO 2.44 (Maere *et al.*, 2005), enriched or depleted GO terms, as opposed to the full *C. elegans* GO dataset, were calculated using standard settings ($P < 0.05$).

Putative proteins were predicted using OrfPredictor (Min *et al.*, 2005). The presence of signal peptides, in combination with the absence of transmembrane domains, was predicted using SignalP 4.0 (Petersen *et al.*, 2011).

Effector prediction and selection

Sequences from known annotated secreted proteins from different plant-parasitic nematodes were downloaded from GenBank (for accession numbers, see Table S2), as well as the pioneer effector sequences identified in *M. incognita* by Huang *et al.* (2003). These were BLASTed (TBLASTN) against the *M. graminicola* contigs to find potential homologues. In all BLAST searches mentioned below, a bit score cut-off of 50 was used.

To search for putative novel effector genes, we decided to combine two approaches in order to obtain a diverse set of putative effectors. All predicted proteins were checked for the presence of a signal peptide and the absence of a transmembrane domain, as described above. In addition, all sequences were checked for putative homologues by a BLASTX search against all putatively secreted proteins derived from plant-parasitic nematode genomes. The latter dataset was constructed by combining the predicted proteins from *M. incognita*, *M. hapla* and *B. xylophilus* that had a predicted signal peptide and lacked a transmembrane domain. The obtained putative effectors were subjected to additional TBLASTX searches against different EST databases: namely, a database containing ESTs exclusively from *Meloidogyne* species, one containing all plant-parasitic nematode ESTs, except for *Meloidogyne* species, one containing ESTs exclusively from migratory plant-parasitic nematodes (genera *Bursaphelenchus*, *Xiphinema*, *Radopholus*, *Pratylenchus*, *Hirschmanniella*, *Aphelenchus*, *Ditylenchus* and *Anguina*) and one with ESTs exclusively from sedentary nematodes (genera *Heterodera*, *Globodera*, *Meloidogyne*). By

comparing the number of hits in each TBLASTX search, the sequences could be categorized as being potentially specific to the genus *Meloidogyne* or to sedentary nematodes.

Target genes for RNAi

Several subsequent BLAST searches were performed to identify suitable *M. graminicola* target genes for RNAi. In all BLAST searches, a bit score cut-off of 50 was used, unless stated otherwise. In a first local BLASTX search against all *C. elegans* proteins downloaded from WormBase (<http://www.wormbase.org>, release WS227), the most homologous *C. elegans* genes were identified for the *M. graminicola* sequences. Next, the WormBase identifiers of presumably homologous *C. elegans* genes were used to retrieve all RNAi phenotypes from the WormMart section of WormBase (release WS220). Genes coupled to lethal phenotypes were subsequently retrieved from the dataset. The sequences of these *C. elegans* genes with a lethal RNAi phenotype were downloaded from WormBase and BLASTed (bit score > 40) against the nonredundant protein database including plant sequences only (Embryophyta, NCBI txid3193). Genes without hits against the plant sequences were subsequently BLASTed (bit score > 40) against the nonredundant human protein database (NCBI txid9606) and against the nonredundant insect protein database (NCBI txid6960). Only genes with lethal phenotypes and without hits against plant, human and insect sequences were retained as possible targets for RNAi.

Four genes were predicted to be interesting targets for RNAi (see Results section). To check whether the silencing of these genes could truly have a lethal effect on the nematode, a soaking experiment with gene-specific siRNAs was set up. The design and construction of the siRNAs were performed according to the Silencer siRNA construction kit (Ambion, Foster City, CA, USA). Sense (S) and antisense (AS) primers that were used as templates were as follows: contig0149_siRNA_AS, AATAACAGAAACCGAGGTATGCCTGTCTC; contig0149_siRNA_S, AACATACCTCGGTTTCTGTACCTGTCTC; contig5704_siRNA_AS, AACCTTTTGTGCGAGTTCCTGTCTC; contig5704_siRNA_S, AAGAACTCCGACTAAAAAAGGCTGTCTC; contig6825_siRNA_AS, AAGATAAATTGTAGAGGCTCTGTCTC; contig6825_siRNA_S, AACAGCCTCTAACAATTATCCCTGTCTC; contig6992_siRNA_AS, AAGAGTTCTCCATTGACCCCTGTCTC; contig6992_siRNA_S, AAGGTCAAATGGAGAACCTCCCTGTCTC; GFP_siRNA_AS, AAACATTCTCGCCACAAGCTCCTGTCTC; GFP_siRNA_S, AAAGCTGTGGCCGAGAATGTCTGTCTC. Nematodes were extracted from infected rice tissue by a classic Baermann funnel method. After 2 days of extraction, infective J2s were collected and soaked for 20 h with gene-specific siRNAs at a final concentration of 25 ng/μL. As control treatments, nematodes were soaked in water or in 25 ng/μL siRNAs targeting GFP. After soaking, the viability of the worms was evaluated by counting the number of living and dead nematodes in each sample. Statistical comparisons between samples were made using the Mann–Whitney *U*-test in SPSS software version 20 (Chicago, IL, USA) ($P < 0.05$). This experiment was repeated twice.

Identification of nematode transcripts in mRNA-seq data from infected rice tissue

In a previous study, rice transcriptional changes on nematode infection with *M. graminicola* (3 and 7 days after infection) were identified using

mRNA-seq (Illumina sequencing) (Kyndt *et al.*, 2012). Over 16 million sequence reads from the rice samples infected with *M. graminicola* did not map on the rice genome (build MSU6.0). In this study, these reads were further characterized by assembling them using Velvet v 1.1.07 (Zerbino and Birney, 2008) with variable *k*-mer lengths ($k = 43, 47, 53, 57$). The program default parameters were used with the addition of 'read_trkg yes'. The contig files from each assembly were merged using the multi-K method described in Surget-Groba and Montoya-Burgos (2010). As an extra filter, the resulting contigs were BLASTed (BLASTX) against all rice proteins submitted to Swiss-prot and trEMBL. Contigs without hits against the rice sequences were further analysed by BLASTing them against the following different datasets: all coding protein sequences from the genomes of *C. elegans*, *M. incognita* and *M. hapla* (BLASTX), all nematode ESTs (TBLASTX) and all *M. graminicola* contigs as derived from the J2 data as described above (BLASTX). From these BLAST results, nematode-derived transcripts were distinguished from rice transcripts and transcripts with unknown origin. The latter sequences were BLASTed to Swiss-prot and trEMBL to try to identify the function and/or species of origin. Bit score cut-offs in all BLAST searches mentioned above were chosen at 50. The infected rice tissue transcriptome data has accession number GSE35843. (Gene Expression Omnibus, NCBI)

In situ hybridization

Several contigs that were identified as putative effectors were amplified from *M. graminicola* J2 cDNA under standard polymerase chain reaction (PCR) conditions using the primers listed in Table S3 (see Supporting Information). PCR products were ligated into the pGEM-T vector (Promega, Madison, WI, USA), and transformed into *Escherichia coli* DH5α cells. Positive colonies identified by colony PCR were grown, and plasmids were extracted using the Nucleobond AX kit (Macherey-Nagel, Düren, Germany). Digoxigenin (DIG)-labelled probes were made by amplifying the fragments from the corresponding plasmids using the primers above, followed by a subsequent PCR using the F-primer only (negative control) or the R-primer only, in the presence of DIG-labelled deoxynucleoside triphosphates (dNTPs) (Roche). An *in situ* hybridization was performed on J2s of *M. graminicola* according to the protocol of de Boer *et al.* (1998) using a hybridization temperature of 47 °C instead of 55 °C.

ACKNOWLEDGEMENTS

Annelies Haegeman and Tina Kyndt are postdoctoral fellows funded by the Research Foundation Flanders (FWO-Vlaanderen). This work was partly carried out using the Stevin Supercomputer Infrastructure at Ghent University, funded by Ghent University, the Hercules Foundation and the Flemish Government – Department EWI. Financial support for this project was received from Ghent University (BOF GOA 01GA0805 and 01B00308).

REFERENCES

- Abad, P., Gouzy, J., Aury, J.M., Castagnone-Sereno, P., Danchin, E.G.J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V.C., Caillaud, M.C., Coutinho, P.M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J.V., Hamamouch, N., Hwezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T.R., Markov, G.V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Segurens, B., Steinbach, D., Tytgat, T., Ugarte,

- E., van Ghelder, C., Veronico, P., Baum, T.J., Blaxter, M., Bleve-Zacheo, T., Davis, E.L., Ewbank, J.J., Favery, B., Grenier, E., Henrissat, B., Jones, J.T., Laudet, V., Maule, A.G., Quesneville, H., Rosso, M.N., Schiex, T., Smant, G., Weissenbach, J. and Wincker, P. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* **26**, 909–915.
- Bird, A.F., Bonig, I. and Bacic, A. (1988) A role for the excretory system in secretory nematodes. *J. Nematol.* **20**, 493–496.
- de Boer, J.M., Yan, Y., Smant, G., Davis, E.L. and Baum, T.J. (1998) In-situ hybridization to messenger RNA in *Heterodera glycines*. *J. Nematol.* **30**, 309–312.
- de Boer, J.M., McDermott, J.P., Wang, X.H., Maier, T., Qu, F., Hussey, R.S., Davis, E.L. and Baum, T.J. (2002) The use of DNA microarrays for the developmental expression analysis of cDNAs from the oesophageal gland cell region of *Heterodera glycines*. *Mol. Plant Pathol.* **3**, 261–270.
- Bridge, J. and Page, S.L.J. (1982) The rice root-knot nematode, *Meloidogyne graminicola*, on deep water rice (*Oryza sativa* subsp. *indica*). *Rev. Nematol.* **5**, 225–232.
- Bridge, J., Plowright, R.A. and Peng, D. (2005) Nematode parasites of rice. In: *Plant Parasitic Nematodes in Subtropical and Tropical Agriculture* (Luc, M., Sikora, R.A. and Bridge, J., eds), pp. 87–130. Wallingford, Oxfordshire: CABI Publishing.
- Cantacessi, C., Mitreva, M., Campbell, B.E., Hall, R.S., Young, N.D., Jex, A.R., Ranganathan, S. and Gasser, R.B. (2010) First transcriptomic analysis of the economically important parasitic nematode, *Trichostrongylus colubriformis*, using a next-generation sequencing approach. *Infect. Genet. Evol.* **10**, 1199–1207.
- Dalzell, J.J., McMaster, S., Johnston, M.J., Kerr, R., Fleming, C.C. and Maule, A.G. (2009) Non-nematode-derived double-stranded RNAs induce profound phenotypic changes in *Meloidogyne incognita* and *Globodera pallida* infective juveniles. *Int. J. Parasitol.* **39**, 1503–1516.
- Davis, E.L., Hussey, R.S., Baum, T.J., Bakker, J. and Schots, A. (2000) Nematode parasitism genes. *Annu. Rev. Phytopathol.* **38**, 365–396.
- De Ley, I.T., De Ley, P., Vierstraete, A., Karssen, G., Moens, M. and Vanfleteren, J. (2002) Phylogenetic analyses of *Meloidogyne* small subunit rDNA. *J. Nematol.* **34**, 319–327.
- De Waele, D. and Elsen, A. (2007) Challenges in tropical plant nematology. *Annu. Rev. Phytopathol.* **45**, 457–485.
- Fernandez, D., Tisserant, E., Talhinas, P., Azinheira, H., Vieira, A., Petitot, A.S., Loureiro, A., Poulain, J., Da, S.C., Silva, M.C. and Duplessis, S. (2012) 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant–rust interaction. *Mol. Plant Pathol.* **13**, 17–37.
- Gheysen, G. and Vanholme, B. (2007) RNAi from plants to nematodes. *Trends Biotechnol.* **25**, 89–92.
- Goecks, J., Nekrutenko, A. and Taylor, J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86. doi:10.1186/gb-2010-11-8-r86.
- Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R.L., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchinson, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J.P., Miguel, T., Paszkowski, U., Zhang, S.P., Colbert, M., Sun, W.L., Chen, L.L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y.S., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, K., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalima, T., Oliphant, A. and Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, **296**, 92–100.
- Haegeman, A., Joseph, S. and Gheysen, G. (2011) Analysis of the transcriptome of the root lesion nematode *Pratylenchus coffeae* generated by 454 sequencing technology. *Mol. Biochem. Parasitol.* **178**, 7–14.
- Haegeman, A., Mantelin, S., Jones, J.T. and Gheysen, G. (2012) Functional roles of effectors of plant-parasitic nematodes. *Gene*, **492**, 19–31.
- Harcus, Y., Nicoll, G., Murray, J., Filbey, K., Gomez-Escobar, N. and Maizels, R.M. (2009) C-type lectins from the nematode parasites *Heligmosomoides polygyrus* and *Nippostrongylus brasiliensis*. *Parasitol. Int.* **58**, 461–470.
- Huang, G.Z., Gao, B.L., Maier, T., Allen, R., Davis, E.L., Baum, T.J. and Hussey, R.S. (2003) A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode *Meloidogyne incognita*. *Mol. Plant–Microbe Interact.* **16**, 376–381.
- Kyndt, T., Denil, S., Haegeman, A., Trooskens, G., Bauters, L., Van Criekeing, W., De Meyer, T. and Gheysen, G. (2012) Transcriptional reprogramming by root knot and migratory nematode infection in rice. *New Phytol.* **196**, 887–900.
- Lilley, C.J., Davies, L.J. and Urwin, P.E. (2012) RNA interference in plant parasitic nematodes: a summary of the current status. *Parasitology*, **139**, 630–640.
- Maere, S., Heymans, K. and Kuiper, M. (2005) BINGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. *Bioinformatics*, **21**, 3448–3449.
- Min, X.J., Butler, G., Storms, R. and Tsang, A. (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* **33**, W677–W680.
- Moy, P., Qutob, D., Chapman, B.P., Atkinson, I. and Gijzen, M. (2004) Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. *Mol. Plant–Microbe Interact.* **17**, 1051–1062.
- Nicol, P., Gill, R., Fosu-Nyarko, J. and Jones, M.G.K. (2012) De novo analysis and functional classification of the transcriptome of the root lesion nematode, *Pratylenchus thornei*, after 454 GS FLX sequencing. *Int. J. Parasitol.* **42**, 225–237.
- Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T.D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B.R., Thomas, V.P. and Windham, E. (2008) Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. USA*, **105**, 14 802–14 807.
- Patel, R.K. and Jain, M. (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE*, **7**, e30619.
- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.
- Roze, E., Hanse, B., Mitreva, M., Vanholme, B., Bakker, J. and Smant, G. (2008) Mining the secretome of the root-knot nematode *Meloidogyne chitwoodi* for candidate parasitism genes. *Mol. Plant Pathol.* **9**, 1–10.
- Schenk, P.M., Carvalhais, L.C. and Kazan, K. (2012) Unraveling plant–microbe interactions: can multi-species transcriptomics help? *Trends Biotechnol.* **30**, 177–184.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.
- Smant, G., Stokkermans, J.P.W., Yan, Y.T., de Boer, J.M., Baum, T.J., Wang, X.H., Hussey, R.S., Gommers, F.J., Henrissat, B., Davis, E.L., Helder, J., Schots, A. and Bakker, J. (1998) Endogenous cellulases in animals: isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc. Natl. Acad. Sci. USA*, **95**, 4906–4911.
- Soriano, I.R.S., Prot, J.C. and Matias, D.M. (2000) Expression of tolerance for *Meloidogyne graminicola* in rice cultivars as affected by soil type and flooding. *J. Nematol.* **32**, 309–317.
- Surget-Groba, Y. and Montoya-Burgos, J.I. (2010) Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res.* **20**, 1432–1440.
- Vieira, P., Danchin, E.G.J., Neveu, C., Crozat, C., Jaubert, S., Hussey, R.S., Engler, G., Abad, P., de Almeida-Engler, J., Castagnone-Sereno, P. and Rosso, M. (2011) The plant apoplast is an important recipient compartment for nematode secreted proteins. *J. Exp. Bot.* **62**, 1241–1253.
- Zelenksy, A.N. and Gready, J.E. (2005) The C-type lectin-like domain superfamily. *FEBS J.* **272**, 6179–6217.
- Zerbino, D.R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1 Downloaded nematode expressed sequence tag (EST) data, including the species name, the number of ESTs, the lifestyle of the nematode, the National Center for Biotechnology Information (NCBI) txid and the database.

Table S2 Table summarizing putative homologues to all currently known putative effectors, including their (speculative) function, accession number, number of *Meloidogyne graminicola* contigs with hits, number of reads corresponding to these contigs, bit score in BLAST, the species in which the effector has been found (*Aa*, *Aphelenchus avenae*; *Bm*, *Bursaphelenchus mucronatus*;

Bx, *Bursaphelenchus xylophilus*; Da, *Ditylenchus africanus*; Gm, *Globodera Mexicana*; Gp, *Globodera pallid*; Gr, *Globodera rostochiensis*; Gt, *Globodera tabacum*; Hg, *Heterodera glycines*; Hs, *Heterodera schachtii*; Ma, *Meloidogyne arenaria*; Mc, *Meloidogyne chitwoodi*; Mh, *Meloidogyne hapla*; Mi, *Meloidogyne incog-*

nita; Mj, *Meloidogyne javanica*; Pc, *Pratylenchus coffeae*; Pp, *Pratylenchus penetrans*; Rr, *Rotylenchulus reniformis*; Rs, *Radopholus similis*; Xi, *Xiphinema index*) and corresponding references. **Table S3** Primer sequences used for cloning putative effector genes for the performance of *in situ* hybridization.