

Bartel Vanholme · Makedonka Mitreva ·
Wim Van Criekinge · Marc Logghe · David Bird ·
James P. McCarter · Godelieve Gheysen

Detection of putative secreted proteins in the plant-parasitic nematode *Heterodera schachtii*

Received: 6 June 2005 / Accepted: 20 September 2005 / Published online: 28 December 2005
© Springer-Verlag 2005

Abstract The beet cyst nematode *Heterodera schachtii* is an important pathogen worldwide, but its molecular characterization has been limited to studying individual genes of interest. We undertook a high-throughput genomic approach and drastically increased the number of available sequences for this parasite. A total of 2,662 expressed sequence tags were grouped into 1,212 clusters representing a nonredundant catalog of *H. schachtii* genes. Implementing a bioinformatic workflow, we identified 50 sequences coding for candidate secreted proteins. All of these contain a putative signal peptide required for entry into the secretory pathway and lack any transmembrane domain. Included are previously postulated cell-wall-degrading enzymes and other parasitism-related genes. Moreover, we provide the first report of an arabinogalactan endo-1,4- β -galactosidase enzyme (EC 3.2.1.89) in animals. As sequence data increase at a rapid rate, developing

high-throughput genomic screening is a necessity. The *in silico* approach described here is an effective way to identify putative secreted proteins and prioritize candidates for further studies.

Abbreviations aa: amino acid · BLAST: basic local alignment search tool · CBD: cellulose-binding domain · CBP: cellulose-binding protein · cDNA: complementary DNA · EST: expressed sequence tag · HMM: hidden Markov model · J2: second stage juvenile · NN: neural network · nt: nucleotides · ORF: open reading frame · SL: spliced leader · SP: signal peptide · SPIT: secreted proteins identified from tags · SPW: standard pore water · TM: transmembrane

Introduction

Plant-parasitic nematodes are major parasites of most crops and cause considerable economic losses in agriculture (Müller 1999). Traditional management strategies, like crop rotation and host resistance, are often inefficient when compared to nematicidal agrochemicals, which in turn are neither environmentally safe nor nematode-specific. Hence, there is a strong demand for the development of novel control strategies. Characterizing parasites molecularly and unraveling the mechanisms these animals use to infect plants can aid in the development of new control strategies. Molecular studies have long been hampered by the species' obligate parasitic life cycle. However, research is now beginning to benefit from new methods for gene identification. Among the most powerful and cost-effective approaches for rapid gene discovery is the single-pass sequencing of randomly sampled cDNA clones or expressed sequence tags (ESTs). Different high-throughput EST projects have now brought the total number of publicly available sequences from parasitic nematodes to more than 300,000 (McCarter et al. 2003; Wylie et al. 2004). Organizing and analyzing these data by using standard similarity-based searches can greatly contribute both to basic understanding of nematode biology and applied

B. Vanholme · W. Van Criekinge · G. Gheysen (✉)
Molecular Biotechnology Department,
Faculty of Bioscience Engineering,
Ghent University,
Coupure links 653,
9000 Ghent, Belgium
e-mail: godelieve.gheysen@UGent.be
Tel.: +32-9-2645888
Fax: +32-9-2646219

M. Mitreva · J. P. McCarter
Department of Genetics,
Washington University School of Medicine,
4444 Forest Park Blvd.,
St Louis, MO 63108, USA

M. Logghe
DEVGEN NV,
Technologiepark 9,
9052 Zwijnaarde-Gent, Belgium

D. Bird
Center for the Biology of Nematode Parasitism,
North Carolina State University,
Box 7253, Raleigh, NC 27695, USA

research toward new means of nematode control (McCarter et al. 2003; Parkinson et al. 2004).

The beet cyst nematode *H. schachtii* is a sedentary endoparasitic nematode, and its life cycle consists of two phases: a soil stage and a plant stage. During the soil stage, the preparasitic juveniles (J2s) hatch from eggs and locate the roots of a suitable host. Following entry into the root, a single plant cell is selected and modified into a complex feeding site partly by cell-wall degradation and by fusion of neighboring protoplasts. At this stage, the nematode loses its ability to migrate through the plant and exploits the created feeding structure to withdraw nutrients during the remainder of its life cycle. It is unknown how the nematode manipulates the plant, but there is evidence that secretory proteins, produced by the nematode's subventral or dorsal pharyngeal glands and presumably injected into the plant cell, have an important role in this process (Vanholme et al. 2004). The identification of these proteins will aid in unraveling the parasitic strategy of this animal. Proteomic studies led to the purification and separation of *H. schachtii* secreted proteins by two-dimensional (2D) gel electrophoresis (De Meutter et al. 2001), but their ultimate identification was limited by a lack of sequence information. As recently as 2003, only four sequences from *H. schachtii* were available in the public database [two endoglucanases (GenBank AJ299386 and GenBank AJ299387), a vacuolar adenosinetriphosphatase (ATPase) subunit (GenBank AJ249961), and a hypothetical secreted protein (GenBank AJ297919)]. We undertook an EST approach and increased the number of available sequencing data 300-fold. The sequences originate from preparasitic J2s and consequently represent genes expressed at the onset of parasitism.

Although the data set was created to support the identification of proteins based on mass spectrometry tags obtained from purified secretions, we used the data set to identify putative secreted proteins directly by *in silico* analysis. This was facilitated by the presence of an amino terminal signal peptide (SP), which precedes the mature protein and is the trigger for export of many secretory proteins. SPs are on average 24 amino acids (aa) long but differ greatly in size (11–80 aa). Although their primary sequence is extremely variable, three short regions with distinct physicochemical properties can be distinguished: an amino terminal positively charged region, a central hydrophobic core, and a more polar carboxy terminal region preceding the signal peptidase cleavage site (Ladunga 2000). Using bioinformatic tools, it is possible to predict the presence of such a motif at the N-terminal end of premature proteins, and different EST projects have used such predictions (Harcus et al. 2004; Torto et al. 2003). We developed a computer program to facilitate screening for putative SP in open reading frames (ORFs) deduced from ESTs. As a result, we identified 50 putative secreted proteins, which could serve as a starting point for further experimental analyses.

Materials and methods

Collection and sterilization of nematodes

Heterodera schachtii was propagated on *Brassica oleracea* cv. Ramosa, grown in white sand supplemented with 4 g l^{-1} slow-release fertilizer (Osmocote, Scotts Europe BV, the Netherlands). Seeds were germinated in Arabaskets (Betatech, Ghent, Belgium), and 3 weeks after germination, the plants were transferred to a 2-l clay pot containing sand with cysts of a previous infection. The plants were grown in a growth chamber with a 16-h day/8-h night regime at 22°C for 10–20 weeks. The cysts were collected using sieves (1 mm^2 to retain organic material and 0.01 mm^2 to retain the cysts) and incubated in 3 mM ZnCl_2 to stimulate hatching of the J2s. The J2s were purified from debris by a 5-min centrifugation at $1,000\times g$ on a 35% sucrose step gradient followed by two washing steps in synthetic pore water (SPW; Smant et al. 1997), supplemented with 0.005% sodium dodecyl sulfate (SDS), and sterilized with $50 \mu\text{g ml}^{-1}$ chlorhexidine and 0.5 mg ml^{-1} cetrimonium bromide (30 min) followed by two washes with an excess of sterile SPW. After sterilization, motile nematodes were selected by letting them move through three layers of Mira cloth (Calbiochem, La Jolla, CA, USA)

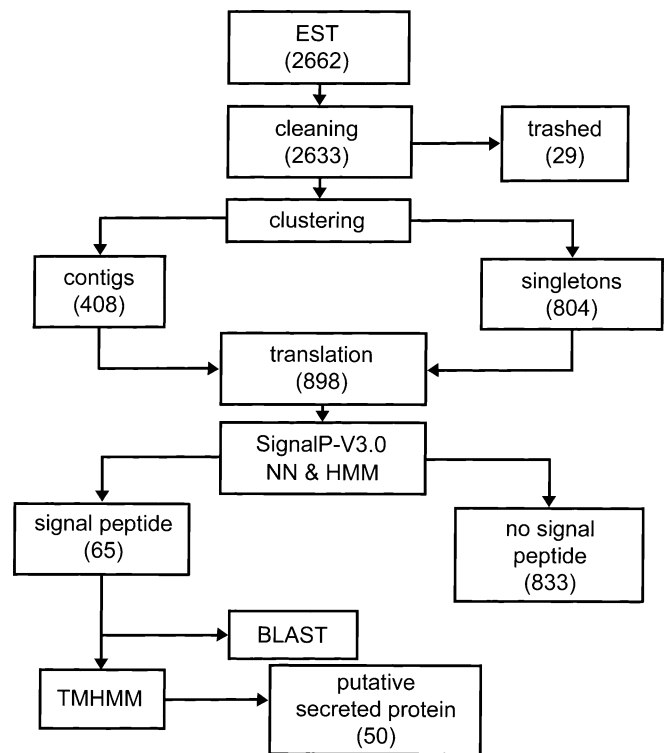


Fig. 1 Semiautomated processing pipeline for the identification of secreted proteins based on EST clusters and singletons. The number of ESTs retained after each step is listed. Details are described in the “Materials and methods” section

and placed in a 100- μ m sieve in a Petri dish with root exudate from cabbage. The cabbage root exudate was prepared by giving cabbage plants an excess of water and collecting the flow-through, filter-sterilized, and stored at 4°C till further use. Live nematodes were collected, washed in SPW, and concentrated by centrifugation (10 min at 1,000 \times g).

RNA extraction

Approximately 10⁶ freshly hatched J2s were ground in liquid nitrogen. Total RNA was isolated using LiCl extraction (Sambrook et al. 1989) and precipitated in ethanol. The RNA pellet was washed in 75% ethanol and resuspended in dimethyl pyrocarbonate-treated water. The quality of the RNA was checked by gel and spectrophotometric analysis.

Library construction, EST sequencing, and submission

cDNA was made by combining Dynabead (DynaBiotec, Oslo, Norway) oligo-dT priming with a modified protocol from the SMART cDNA library construction system (BD Biosciences, Palo Alto, CA, USA) (Mitreva et al. 2004a). Sequencing and processing prior to submission was performed as described by Mitreva et al. (2004b) and included cleaning the sequences using SeqClean (TIGR Gene Indices Software Tools; <http://www.tigr.org/tdb/tgi/software>) to remove contaminating vector sequences, poly A tails, or short sequences. All ESTs have been submitted to the Expressed Sequenced Tags database (dbEST; <http://www.ncbi.nlm.nih.gov/dbEST>). Information for clone request and sequence trace files is available at <http://www.nematode.net>.

Assembly and *in silico* analysis

Processed ESTs were assembled into multimember clusters of overlapping sequences using the TIGR gene indices clustering tool (TIGR Gene Indices Software Tools; Pertea et al. 2003) delivering CAP3 assemblies (Huang and Madan 1999) and singleton ESTs (without similarity to other ESTs). Additional analyses were done using customized Perl-based scripts and programs. One of them, called SPIT (Secreted Proteins Identified from Tags;

Table 1 Properties of the sequences obtained from the *H. schachtii* library

Number of ESTs	2,662
Total number of nt	1,285,177
Mean length	482 nt
Minimum length	54 nt
Maximum length	688 nt
Number of contigs	408
Number of singletons	804

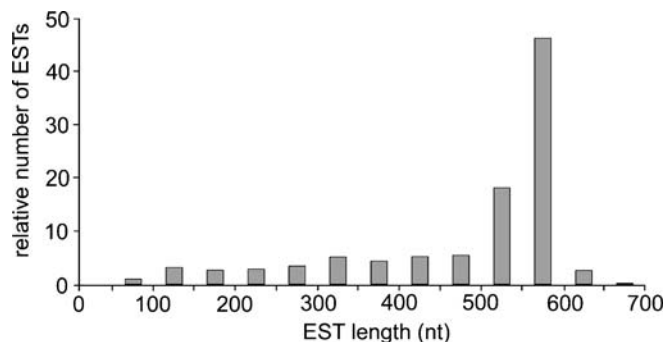


Fig. 2 Size distribution of the ESTs obtained from the *H. schachtii* cDNA library

Fig. 1), was developed to predict putative signal sequences for secretion on clustered EST data in conjunction with SignalP (<http://www.cbs.dtu/services/Signal>) (Bendtsen et al. 2004; Nielsen and Krogh 1998). SPIT translates the consensus sequence in six different reading frames and looks for the longest ORF preferentially starting with a methionine. The retained sequences (passing the threshold based on their relative length) are fed to SignalP. This program combines two signal prediction methods, SignalP-NN (based on neural networks) and SignalP-HMM (based on hidden Markov models). A signal sequence was considered present when it was predicted both by the artificial NN and HMM. Finally, the membrane topology prediction program TMHMM (Krogh et al. 2001) was used to exclude proteins with putative transmembrane (TM) regions. Identification of sequence similarity was performed using basic local alignment search tool (BLAST) analyses (matrix=BLOSUM62) against the nonredundant and dbEST databases (Altschul et al. 1990).

Results

The sequencing of 2,662 tags from a root-extract stimulated second-stage juvenile *H. schachtii* cDNA library resulted in the submission of 1,285,177 nucleotides to dbEST between June and September 2003. General

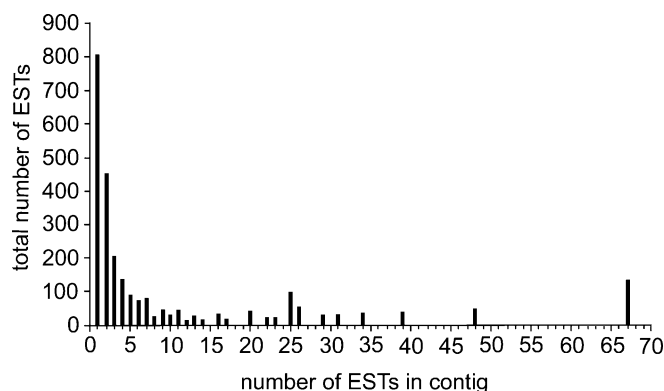


Fig. 3 Histogram showing the distribution of 2,633 ESTs by cluster size. For example, there were four clusters of 25 ESTs each, thus representing 100 ESTs

Table 2 Summary of the 20 most abundantly represented transcripts in the EST data set obtained from the cDNA library

Number	Name	ESTs	Homology	<i>E</i> value
1	CL1contig2	67	Putative SEC-2 protein (<i>Globodera pallida</i>)	1e-48
2	CL2contig1	67	Hypothetical protein C50D2.8 (<i>C. elegans</i>)	4e-19
3	CL3contig1	48	putative gland protein G34B08 (<i>H. glycines</i>)	3e-31
4	CL5contig1	39	Calponin homolog (<i>Onchocerca volvulus</i>)	2e-41
5	CL6contig1	34	–	–
6	CL4contig3	31	Putative phosphatidylinositol 4-phosphate 5-kinase (<i>Oryza sativa</i>)	2e-04
7	CL7contig1	29	Hypothetical protein C48A7.2 (<i>C. elegans</i>)	2e-49
8	CL8contig1	26	Secreted glutathione peroxidase (<i>G. rostochiensis</i>)	1e-98
9	CL9contig1	26	Cyclin M2 (<i>C. elegans</i>)	1e-12
10	CL10contig1	25	–	–
11	CL11contig1	25	Hypothetical protein C35D10.5 (<i>C. elegans</i>)	4e-07
12	CL12contig1	25	Hypothetical protein Y34B4A.7 (<i>C. elegans</i>)	4e-06
13	CL13contig1	25	–	–
14	CL15contig1	23	Putative calponin (<i>M. incognita</i>)	0.0
15	CL14contig1	22	Hypothetical protein Y49E10.20 (<i>C. elegans</i>)	1e-22
16	CL17contig1	20	–	–
17	CL18contig1	20	Pectate lyase (<i>Streptomyces avermitilis</i>)	1e-39
18	CL19contig1	17	JNK interacting protein (<i>C. elegans</i>)	6e-44
19	CL16contig2	16	28S ribosomal RNA gene (<i>H. glycines</i>)	0.0
20	CL20contig1	16	60S ribosomal protein L3 (<i>D. melanogaster</i>)	4e-74

Homology searches were performed using BLASTX

properties of the sequenced tags are presented in Table 1. Figure 2 shows the distribution of the relative number of ESTs in relation to their length and illustrates that the majority of the ESTs are longer than 500 nt. A consequence of the random sampling of cDNA clones for sequencing is a general redundancy of the data set because certain genes will be sampled multiple times. Such tags usually correlate with highly abundant transcripts in an organism under study at the point of sampling (Audic and Claverie 1997). For example, 67 ESTs had significant homology to a gene coding for the fatty acid-binding protein FAR-1 (also called SEC-2; GenBank Y09293), reflecting the abundant expression of this gene in J2s (Prior et al. 2001; Vanholme et al. 2002). To compress the size and increase the quality of the data set, an *in silico* nonredundant database was derived by clustering the entries. From a total of 2,628 high-quality sequences, 1,824 were assembled in 408 multimember clusters, varying in size from 2 ESTs (454 cases) to 67 ESTs (2 cases) (Fig. 3). The redundancy of the library, as estimated by the number of reads that were assembled into clusters, was approximately 68%, and highly redundant sequences (occurring >10 times) made up approximately 27% of all successful reads (Table 2). Clustering increased the overall length from 482 nt for an individual tag to 589 nt for a contig. Since it is not possible to determine whether each cluster represents a single gene or a mixture of transcripts from closely related genes, a cluster was defined as a group of related sequences. Adding the 804 singletons, the data set represents an estimated 1,212 genes of *H. schachtii*, which is approximately 6.4% of the total gene repertoire based on the assumption that *H. schachtii* contains a similar number of genes as *Caenorhabditis*

elegans (~19,000; The *C.elegans* Sequencing Consortium 1998).

Despite the fact that a considerable number of nematode transcripts are equipped with a highly conserved 22-nucleotide *trans*-spliced leader (SL) sequence [use of SL1 by transcripts is estimated at 70% in *C. elegans* (Blumenthal and Steward 1997) and 60% in *Globodera rostochiensis* (Dautova et al. 2001)], only one out of the 2,662 ESTs contains such sequence (gi|32325592; GenBank CD750667). In a similar study performed on ESTs of *Meloidogyne incognita*, an SL was found in 33 clusters obtained from 5,700 ESTs (McCarter et al. 2003). And in a less stringent analysis, a 5'-truncated form of the SL was found in 37 of 1,234 ESTs obtained from *Nippostrongylus brasiliensis* (Harcus et al. 2004). Although higher than our observation, this is still far below the expected 60–70%. The presence of truncated cDNAs or lower-quality sequence at read ends

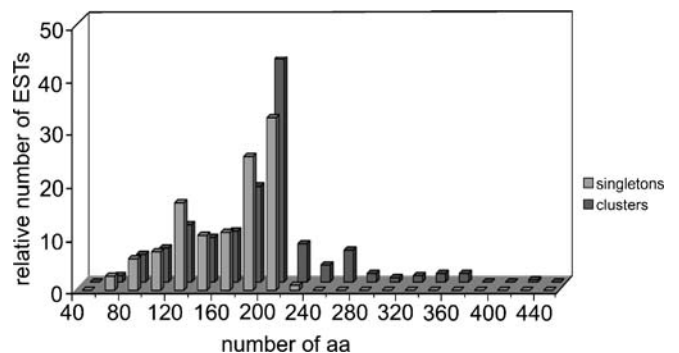


Fig. 4 Histogram showing the distribution of clusters and singletons by the size of their ORF (in amino acids)

Table 3 Putative secreted proteins retained by SPIT from contigs and singletons of the cDNA library

	ESTs	nt	aa	Start	SP	Homology	E value
CLUSTERS							
CL003contig1	48	773	227	M	23	Putative gland protein G34B08 (<i>H. glycines</i>)	3e-31
CL003contig2	12	707	195	M	23	Putative gland protein G34B08 (<i>H. glycines</i>)	1e-16
CL004contig3	31	1948	648	M	24	Putative phosphatidylinositol 4-phosphate 5-kinase (<i>Oryza sativa</i>)	2e-04
CL008contig1	26	766	240	M	19	Secreted glutathione peroxidase (<i>Globodera rostochiensis</i>)	1e-98
CL013contig1	25	607	189	M	21	–	–
CL018contig1	20	585	178	M	20	Pectate lyase (<i>Streptomyces</i> sp.)	1e-40
CL045contig1	2	351	108	M	23	–	–
CL045contig2	5	401	124	M	23	–	–
CL060contig1	6	434	111	M	19	–	–
CL107contig1	4	535	163	M	25	PFG377 protein (<i>Plasmodium falciparum 3D7</i>)	2e-05
CL112contig1	3	515	145	M	18	–	–
CL162contig1	3	778	253	M	23	β -1,4-endoglucanase 2 (<i>H. schachtii</i>)	4e-62
CL164contig1	3	586	184	M	22	ABC transporter, substrate-binding protein (<i>Erwinia carotovora</i>)	3e-38
CL174contig1	3	587	195	–	16	–	–
CL221contig1	2	310	87	M	27	β -1,4-endoglucanase 5 (<i>H. glycines</i>)	5e-46
CL222contig1	2	566	185	M	31	Putative cuticular collagen (<i>Globodera pallida</i>)	2e-42
CL255contig1	2	425	76	M	27	–	–
CL264contig1	2	548	162	M	33	–	–
CL279contig1	2	349	111	M	18	β -1,4-endoglucanase 1 (<i>H. schachtii</i>)	3e-06
CL282contig1	2	579	192	–	25	Hypothetical protein CBG18123 (<i>Caenorhabditis briggsae</i>)	4e-20
CL292contig1	2	545	164	M	20	–	–
CL320contig1	2	576	178	M	16	Cathepsin L-like cysteine proteinase (<i>H. glycines</i>)	7e-87
CL323contig1	2	570	184	M	24	Chitinase (<i>H. glycines</i>)	1e-87

Table 3 (continued)

	nt	aa	Start	SP	Homology	<i>E</i> value
SINGLETONS						
gi_33140501	556	139	M	21	Cellulose binding protein (<i>H. glycines</i>)	1e-09
gi_33140337	214	51	M	17	–	–
gi_33140262	586	176	M	19	–	–
gi_33140261	554	112	M	27	–	–
gi_33140204	453	86	M	24	Hypothetical protein CBG17404 (<i>C. briggsae</i>)	1e-7
gi_33140151	559	139	M	23	–	–
gi_33140147	577	164	M	24	Secreted venome allergen-like protein vap-1 (<i>H. glycines</i>)	4e-75
gi_33140137	564	125	M	21	Hypothetical protein C33A11.4 (<i>C. elegans</i>)	2e-45
gi_33140127	563	177	M	18	Secreted venome allergen-like protein vap-2 (<i>H. glycines</i>)	1e-107
gi_33139876	563	168	M	26	–	–
gi_33139816	304	69	M	22	Putative hypodermis secreted protein (<i>G. rostochiensis</i>)	4e-18
gi_33139547	540	164	M	22	Translocon associated protein (<i>C. elegans</i>)	4e-31
gi_33139334	562	160	M	22	–	–
gi_33139095	547	166	M	26	–	–
gi_33139057	534	99	–	29	–	–
gi_32325599	584	168	M	16	–	–
gi_32325558	564	177	M	17	–	–
gi_32325419	290	73	M	30	–	–
gi_32325259	586	195	–	27	UNC-44 ankyrins (<i>Fusobacterium nucleatum</i>)	5e-16
gi_32325225	593	173	M	23	–	–
gi_32325127	353	110	M	24	–	–
gi_32324951	557	176	M	20	Hypothetical protein CBG06065 (<i>C. briggsae</i>)	4e-12
gi_32324929	555	99	M	22	Hypothetical protein B0303.9 (<i>C. elegans</i>)	–
gi_32324875	466	142	M	21	Arabinogalactan endo-1, 4- β -galactosidase (<i>Xanthomonas</i> sp)	1e-36
gi_32324496	579	128	M	22	Aminomethyl transferase (<i>Anopheles gambia</i>)	2e-07
gi_32324295	590	196	–	22	Putative steroid dehydrogenase F11A5.12 (<i>C. elegans</i>)	3e-33
gi_32324088	341	97	–	17	–	–

The presence of a putative start codon (M) and the length of the putative SP is indicated. Homology searches were performed using BLASTX

can partly explain this discrepancy. Alternatively, the absence of SL sequences may be the result of the experimental approach undertaken. Most projects, aiming to clone putative parasitism genes (e.g., Dautova et al. 2001; Gao et al. 2003; Huang et al. 2003; this work), are based on the SMART technology, in which the 5'-end of the mRNA is extended with an oligonucleotide during first-strand synthesis. The altered 5'-cap of the SL sequence, consisting of a trimethylguanosine instead of a monomethylguanosine (Thomas et al. 1988), can be responsible for a negative selection against SL-containing mRNA sequences. Indeed, several reports mention that the altered cap structure can be responsible for altered biochemical properties (Fernandez et al. 2002; Lall et al. 2004). With the number of identified and characterized putative parasitism genes increasing, it is noticed that most lack a SL sequence, indicating a possible negative correlation between SL and parasitism genes (Gao et al. 2003). Nevertheless, controversy will remain concerning this relation as long as the number of parasitism sequences characterized is limited.

The algorithm SPIT (Fig. 1), developed for automated identification of putative secreted proteins, was applied to the complete *H. schachtii* EST data set. Out of 1,212 sequences, 898 were retained (314 contigs and 584 singletons) based on the criteria that they contained an ORF encoding a peptide of at least 50 aa. The distribution of their length (Fig. 4) illustrates the benefits of the clustering process. Contigs have generally longer ORFs compared to singletons; 19.1% (60/314) and 0.8% (5/584) of the ORFs obtained from contigs and singletons, respectively, code for proteins of more than 200 aa. The fidelity of *in silico* translation was checked by comparing the aa composition of the theoretically constructed proteins to the aa frequency of other organisms (see supplementary information). The calculated aa composition of *H. schachtii* was more similar to eukaryotes than to prokaryotes. In addition, the codon preference was more like that of the related species *Heterodera glycines* than that of *C. elegans* [Codon Usage Database (<http://www.kazusa.or.jp/codon>)]. This indicates that the nucleotide sequences (cluster consensus and singletons) were likely to be correctly translated. Moreover, 61.5% (552/898) of the ORFs could be aligned with the proteome of *C. elegans* (at $E < 10e^{-3}$).

An analysis based on "Signal P" revealed the presence of an SP for secretion in 31 contigs and 34 singletons. Fifteen of the sequences were eliminated based on the presence of a putative TM region as predicted by TMHMM (<http://www.cbs.dtu.dk/services/TMHMM>), indicating that they are integral membrane proteins rather than secreted. Hence, only 5.6% (50/898) of the predicted proteins are putative-secreted (Table 3), lower than the 10% observed in some proteomes (Ladunga 2000). Different reasons could have decreased this number. First, the accurate prediction of secretory proteins from partial sequences relies on the presence of the 5'-end of the transcript and the localization of the correct translation initiation site. For example, SPIT was unable to retain a single clone out of the 67 ESTs

(CL1contig2) encoding for the secretory protein Hs-FAR-1. Alignment subsequently showed that these clones were all truncated in their 5'-domain. Another example is EST gi|32324822 (GenBank CD750279), which is similar to Mi-CRT1 (GenBank AF402771, $E = 1e^{-41}$), a gene of *M. incognita* coding for a calreticulin. Mi-CRT1 is expressed in the subventral glands of the nematode (Jaubert et al. 2002), and the corresponding protein is injected in the feeding site during the parasitic interaction (Jaubert et al. 2005). Also, this EST is lacking the 5'-end. Second, to decrease the number of false positives, we used rather stringent selection criteria and retained only those sequences that scored above the cutoff value for both the NN and HMM model. Finally, some proteins lack a typical hydrophobic SP but contain internal signal sequences, which allow them to be secreted by the normal secretory pathway or other type of secretion independent from the classical secretory pathway (Bendtsen et al. 2004; Nickel 2003). To circumvent these drawbacks, Mitreva et al. (2004c) predicted secretory and TM candidates from *Trichinella* ESTs based on homology to other putative secreted proteins rather than finding the SP directly in the EST sequences.

Discussion

By constructing a cDNA library from preparasitic nematodes, we ensured that the obtained transcripts encode genes expressed during the initial stage of parasitism such as plant penetration and migration. In addition, to enrich the library for putative parasitism genes, the nematodes were stimulated with root exudates to simulate the plant environment. This treatment results in the accumulation of secretory granules and transcriptional activity in the pharyngeal glands of the cyst nematodes (Blair et al. 1999; Perry et al. 1989; Smant et al. 1997). It would be interesting to extend the EST approach toward early parasitic stages when the nematode induces the feeding site. However, the use of these stages is hampered by their inaccessibility, as they are enclosed inside the plant root. Although it is feasible to dissect the nematodes manually from the root, it is difficult to do this when large quantities are needed to construct a representative cDNA library (de Boer et al. 2002a).

Organization, interpretation, and characterization of the generated sequences requires the creation and implementation of a bioinformatic approach, which can be applied uniformly across the data set. For this purpose, a computer program (SPIT) was written in Perl to predict secreted proteins from the data set. A similar approach (PexFinder) was undertaken previously to identify secreted proteins from an EST library of the plant pathogen *Phytophthora infestans* (Torto et al. 2003). However, our *in silico* pipeline was made uniform by combining different steps in a single script and expanding the algorithm to facilitate additional analyses. One of the advantages of the program is the ability to retain sequences missing the start codon. For example,

contig CL376Contig1 [a homolog of a gene coding for a putative gland protein of *H. glycines* (GenBank AF500031)] was retained, although it was lacking a start codon. Due to the presence of a putative TM domain in this protein, the sequence was not included in Table 3. It is obvious that both approaches, analyzing partial sequences and gene sets or undertaking different approaches to screen the same type of data, may inflate or deflate the number of proteins containing an SP for secretion contained in a species' genome. For example, an alternative tag-filtering approach performed on ESTs obtained from the nematode *N. brasiliensis* revealed an SP in 17.6% (176/997) of the retained ORFs (Harcus et al. 2004). However, the sequences were not further analyzed for putative TM regions, and one would expect that a considerable number of the retained sequences will be trapped in the cellular membrane and cannot be considered as secreted.

Out of the 50 putative secreted proteins, 23 (46%) were novel (Table 3). The fact that many putative parasitism genes are novel sequences (Gao et al. 2003) makes them intriguing candidates for further characterization. Our semiautomated pipeline was validated by the identification of ESTs coding for parasitism factors sharing homology with previously identified secreted proteins from other plant-parasitic nematodes. Therefore, we are confident that our findings hold equally true for the "pioneer" sequences. The function of the putative homologues is indicative of a role in plant parasitism, so we will briefly discuss some of the interesting candidates.

Cell-wall-modifying proteins

The polypeptide deduced from the EST gi|33140501 (GenBank CF101434) shows similarities to a cellulose-binding protein (CBP) of *H. glycines* (Hg-CBP-1; GenPept AAR01198; $E=1e-9$) and, accordingly, was named Hs-CBP. Normally, these sequences are domains of larger proteins such as endo- or exoglucanases. In such cases, the cellulose-binding domain (CBD) is linked with a catalytic domain and regulates the interaction of the enzyme with the substrate cellulose. Remarkably, the transcript does not encode any other domain. Although it is possible that our information is deduced from a truncated cDNA, this is unlikely. The presence of several in- and out-of-frame stop codons at the 3'-end of the sequence indicates that the EST possibly contains the complete ORF coding for a protein of 139 aa. In addition, similar proteins were detected in other plant-parasitic nematodes [e.g., *H. glycines* (GenPept AAN32887) and *M. incognita* (GenPept AAC05133)], and the sequence is more similar to these sequences than to the CBD of nematode endoglucanases. Previous analyses (Ding et al. 1998; Gao et al. 2004) indicate that the CBPs are secreted by plant-parasitic nematodes and may have a role in parasitism. The specific function of this protein is not yet known, but recombinant CBD was found to modulate the elongation of different plant cells in vitro (Shpigel et al. 1998). If indeed involved in cell expansion, Hs-CBP has a

similar function as expansin, which was recently detected in secretions of cyst nematodes (Gr-EXPB1; GenPept CAC83611) (Qin et al. 2004). This functional similarity is accompanied by sequence similarity (32% identity between Gr-EXPB1 and Hs-CBP at the protein level).

Three contigs (CL126contig1, CL221contig1, and CL279contig1) showed homology to endoglucanases. The sequence of CL221contig1 is similar to Hg-ENG5 (GenBank AY336935; $E=5e-46$), an endoglucanase gene of *H. glycines*. Hg-ENG5 differs significantly in its catalytic domain compared to the other cloned endoglucanases of *H. schachtii* [Hs-ENG1 (GenPept CAC12958) and Hs-ENG2 (GenPept CAC12959)] (De Meutter et al. 1999). In contrast to these endoglucanases, Hg-ENG5 possesses activity toward xylan (Gao et al. 2004), which is a major component of hemicellulose of the plant cell wall. CL279contig1 showed only weak similarity to previously cloned endoglucanases (Hs-ENG1; $E=3e-6$), and the homology was restricted to the CBD of Hs-ENG1. As expected, based on the similarity with the CBD, homology searches revealed some degree of similarity with Gr-EXPB1 (GenPept CAC83611; $E=9e-3$). Although the protein might function as a CBP, the protein is more similar to a CBD than to the previously described Hs-CBP (EST gi|33140501; GenBank CF101434). The isolation of a full-length clone would indicate if this contig is coding for a fragment of an endoglucanase or is functioning as an independent CBP. If the EST indeed originates from an endoglucanase-encoding transcript, the enzyme will consist of a CBD extended with the catalytic domain. Such topology is not yet described for nematode endoglucanases, in which CBDs are normally present as C-terminal extensions of the catalytic domains.

Another contig (CL018contig1), comprised of 20 ESTs, showed homology with pectate lyase. Remarkably, the deduced polypeptide sequence is more related to pectate lyases of fungi [*Streptomyces* sp ($E=1e-40$) and *Magnaporthe grisea* ($E=3e-28$)] or bacteria [*Bacillus* sp ($E=3e-28$)] than to pectate lyases of related cyst nematodes (Gr-PEL1; GenPept AAF80746; $E=2e-16$ /Hg-PEL2; GenPept AAM74954; $E=1e-14$ /Hg-PEC; GenPept AAK08974; $E=8e-14$) (de Boer et al. 2002b; Popeijus et al. 2000), indicating that the sequence codes for a pectate lyase distinct from the others described in cyst nematodes. The similarity to bacterial genes have been found for other genes coding for cell-wall-degrading enzymes in nematodes and have led to the speculation of horizontal gene transfer between bacteria and nematodes (Smant et al. 1998).

Finally, one of the retained ESTs (gi|32324875; GenBank CD750306) showed similarity to arabinogalactan endo-1,4- β -galactosidases (*Xanthomonas campestris*; GenPept AAM42894; 62% BLASTX identity). Those enzymes hydrolyze the 1,4- β -galactosidic linkage in arabinogalactans of the plant cell wall and have not yet been identified in animals. Although additional experiments are needed to confirm the endogenous origin of this gene, a database search revealed an EST of *H. glycines* (gi|15770814; GenBank BI749012) coding for a similar enzyme (74%

BLASTN identity). This finding decreases the possibility of contamination because the chance that a similar contaminating organism was copurified with both *H. schachtii* and *H. glycines* samples is unlikely. Moreover, detecting a novel cell-wall-degrading enzyme supports the finding that plant-parasitic nematodes secrete a complex cocktail of cell-wall-degrading enzymes to tackle the different components of the highly resistant cell wall to “soften” the plant tissue, which facilitates their migration. The large number of ESTs coding for cell-wall-modifying proteins mirrors the known high expression level of these genes in other Tylenchids. Currently, these are the best studied parasitism factors from plant-parasitic nematodes.

Other parasitism-related genes

Two ESTs forming CL323contig1 encode an ORF of 184 aa which showed 95% identity to the N-terminal part of *H. glycines* chitinase (Hs-CHI-1; GenPept AAN14978) (Gao et al. 2002). Chitin, the substrate of this enzyme, is an unbranched polysaccharide polymer consisting of β -1,4-linked *N*-acetyl-D-glucosamine and forms an integral part of the nematode eggshell. A chitinase-derived EST was also found in a cDNA library from preparasitic *M. incognita* (GenBank BE239118) (Dautova et al. 2001). These tags could be residual from hatching, during which large amounts of chitinase are produced. However, Schwekendiek et al. (1999) showed the expression of chitinases during different parasitic stages of *H. glycines*, indicating that chitinase may be expressed after hatching. Moreover, Gao et al. (2002) analyzed a chitinase of *H. glycines* in more detail, and they showed that the transcript of this gene accumulated specifically in the subventral gland cells of parasitic stages, but this could not be detected in eggs or freshly hatched J2s. These observations indicate that this enzyme could have other functions besides the cleavage of chitin during hatching. Although a role in moulting is considered as a possibility, the expression in the pharyngeal glands points toward a role in the parasitic interaction.

Two ESTs, gi|33140127 (GenBank CF101060) and gi|33140147 (GenBank CF101080), showed similarity to genes coding for venom allergen-like proteins. Members of this class are similar to the allergen 5 (AG-5) family of extracellular proteins isolated from hymenopteran venom (Fang et al. 1988). Although the function of these proteins is unknown, it is suggested that they play a central role in the transition of the nematode from free living toward parasitic stages (Hawdon et al. 1996). A sequence analysis of the deduced polypeptides confirmed that the transcripts are derived from different genes sharing only 45% identity. gi|33140147 was more related to Hg-VAP-1 (*H. glycines*; GenBank AF374388; 84% identical), whereas gi|33140127 showed a higher similarity with Hg-VAP-2 (*H. glycines*; GenBank AY033601; 97% identical). In *H. glycines*, homologous genes are expressed in the subventral glands of preparasitic and parasitic stages (Gao et al. 2001). Their developmental expression in conjunction with the presence

of a putative signal for secretion suggests that these proteins may have a role in plant–host interaction.

Expressed sequence tag gi|33139816 (GenBank CF100749) showed homology to Gr-SXP-1 (GenBank AJ271910; 82% identical), a gene expressed in the epidermis of *G. rostochiensis* and coding for a secreted protein (Jones et al. 2000). Gr-SXP-1 belongs to the family of SXP/RAL-2 proteins with unknown function but which are nematode-specific, which makes them suitable as targets for control. SXP proteins are also candidate antigens for the vaccination of animal-parasitic nematodes (Wang et al. 1997). In plant-parasitic nematodes, only two other members of this family were described: Gr-AMS-1 (GenPept CAB66341) (Jones et al. 2000) expressed in the amphids of *G. rostochiensis*, and Mi-SXP-1 (GenPept AAR35032) (Tytgat et al. 2005) expressed in the subventral glands of *M. incognita*.

CL320contig1 showed similarity to genes coding for cathepsin L-like cysteine proteinase. This protein is secreted in the intestine of parasitic nematodes and is most likely involved in digestion (Lilley et al. 1996) and indirectly involved in parasitism (Jasmer et al. 2003). Strategies toward developing nematode-resistant plants based on the expression of cysteine proteinase inhibitors are being pursued (Lilley et al. 2004; Urwin et al. 1997).

Conclusion

The analysis presented demonstrates that single-pass sequencing of random clones, although generating incomplete sequences, is a powerful tool for rapid identification of genes that may mediate important aspects in the plant–nematode interaction. By screening the data set to select tags coding for secreted proteins, interesting candidate parasitism genes were selected, which warrant further characterization. Furthermore, these preliminary findings confirmed that a proper implementation of developed bioinformatic tools can help prioritize targets for potential nematode control.

Acknowledgments We thank Claire Murphy and Mike Dante for their role in library generation and technical support. *H. schachtii* EST sequencing was supported by a US National Science Foundation Plant Genome Award 0077503 to D.B. of North Carolina State University and Sandra Clifton of Washington University. Research at Ghent University was supported by the European Union (NONEMA EC, QLT-CT 1999-01501) and the Fund for Scientific Research-Flanders.

Appendix

Description of additional data files

All nucleotide sequence data have been submitted to the Genbank dbEST database. The sequences and programs (Perl script) can be freely downloaded from <http://www.MolecularBiotechnology.ugent.be/publications/Vanholme2005B>

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Audic S, Claverie JM (1997) The significance of digital gene expression profiles. *Genome Res* 7:986–995
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795
- Blair L, Perry RN, Oparka K, Jones J (1999) Activation of transcription during the hatching process of the potato cyst nematode *Globodera rostochiensis*. *Nematology* 1:103–111
- Blumenthal T, Steward K (1997) RNA processing and gene structure. In: Riddle TL, Blumenstiel B, Meyer BJ, Priess JR (eds) *C. elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Dautova M, Rosso MN, Abad P, Gommers FJ, Bakker J, Smant G (2001) Single pass cDNA sequencing—a powerful tool to analyse gene expression in pre-parasitic juveniles of the southern root-knot nematode *Meloidogyne incognita*. *Nematology* 3:129–139
- de Boer JM, McDermott JP, Wang XH, Maier T, Qu F, Hussey RS, Davis EL, Baum TJ (2002a) The use of DNA microarrays for the developmental expression analysis of cDNAs from the esophageal gland cell region of *Heterodera glycines*. *Mol Plant Pathol* 3:261–270
- de Boer JM, McDermott JP, Davis EL, Hussey RS, Smant G, Baum TJ (2002b) Cloning of a putative pectate lyase gene expressed in the subventral esophageal glands of *Heterodera glycines*. *J Nematol* 33:9–11
- De Meutter J, Tytgat T, Van der Schueren E, Smant G, Schots A, Coomans A, Van Montagu M, Gheysen G (1999) Cloning of two endoglucanase genes from *Heterodera schachtii*. *Meded Rijksuniv Gent Fak Landbouwkd Toegep Biol Wet* 63:619–623
- De Meutter J, Vanholme B, Bauw G, Tytgat T, Gheysen G (2001) Preparation and sequencing of secreted proteins from the pharyngeal glands of the plant parasitic nematode *Heterodera schachtii*. *Mol Plant Pathol* 2:297–301
- Ding X, Shields J, Allen R, Hussey RS (1998) A secretory cellulose-binding protein cDNA cloned from the root-knot nematode (*Meloidogyne incognita*). *Mol Plant Microbe Interact* 11:952–959
- Fang KS, Vitale M, Fehlner P, King TP (1988) cDNA cloning and primary structure of a white-face hornet venom allergen, antigen 5. *Proc Natl Acad Sci U S A* 85:895–899
- Fernandez C, Gregory WF, Loke P, Maizels RM (2002) Full-length-enriched cDNA libraries from *Echinococcus granulosus* contain separate populations of oligo-capped and trans-spliced transcripts and a high level of predicted signal peptide sequences. *Mol Biochem Parasitol* 122:171–180
- Gao B, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS (2001) Molecular characterisation and expression of two venom allergen-like protein genes in *Heterodera glycines*. *Int J Parasitol* 31:1617–1625
- Gao B, Allen R, Maier T, McDermott JP, Davis EL, Baum TJ, Hussey RS (2002) Characterisation and developmental expression of a chitinase gene in *Heterodera glycines*. *Int J Parasitol* 32:1293–1300
- Gao B, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS (2003) The parasitome of the phytonematode *Heterodera glycines*. *Mol Plant Microbe Interact* 16:720–726
- Gao B, Allen R, Davis EL, Baum TJ, Hussey RS (2004) Molecular characterisation and developmental expression of a cellulose-binding protein gene in the soybean cyst nematode *Heterodera glycines*. *Int J Parasitol* 34:1377–1383
- Harcus YM, Parkinson J, Fernandez C, Daub J, Selkirk ME, Blaxter ML, Maizels RM (2004) Signal sequence analysis of expressed sequence tags from the nematode *Nippostrongylus brasiliensis* and the evolution of secreted proteins in parasites. *Genome Biol* 5:R39
- Hawdon JM, Jones BF, Hoffman DR, Hotez PJ (1996) Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. *J Biol Chem* 271:6672–6678
- Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:868–877
- Huang G, Gao B, Maier T, Allen R, Davis EL, Baum TJ, Hussey RS (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
- Jasmer DP, Govere A, Smant G (2003) Parasitic nematode interactions with mammals and plants. *Annu Rev Phytopathol* 41:245–270
- Jaubert S, Ledger TN, Laffaire JB, Pottie C, Abad P, Rosso MN (2002) Direct identification of stylet secreted proteins from root-knot nematodes by a proteomic approach. *Mol Biochem Parasitol* 121:205–211
- Jaubert S, Milac AL, Petrescu AJ, de Almeida-Engler J, Abad P, Rosso MN (2005) In planta secretion of a calreticulin by migratory and sedentary stages of root-knot nematode. *Mol Plant Microbe Interact* 18:1277–1283
- Jones JT, Smant G, Block V (2000) SXP/RAL2 proteins of the potato cyst nematode *Globodera rostochiensis*: secreted proteins of the hypodermis and amphids. *Nematology* 2:887–893
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580
- Ladunga I (2000) Large-scale predictions of secretory proteins from mammalian genomic and EST sequences. *Curr Opin Biotechnol* 11:13–18
- Lall S, Friedman CC, Jankowska-Anyszka M, Stepinski J, Darzynkiewicz E, Davis RE (2004) Contribution of trans-splicing, 5'-leader length, cap-poly(A) synergism, and initiation factors to nematode translation in an *Ascaris suum* embryo cell-free system. *J Biol Chem* 279:45573–45585
- Lilley CJ, Urwin PE, McPherson MJ, Atkinson HJ (1996) Characterization of intestinally active proteinases of cyst-nematodes. *Parasitology* 113 (Pt 4):415–424
- Lilley CJ, Uwin PE, Atkinson HJ (2004) Preferential expression of a plant cystatin at nematode feeding site confers resistance to *Meloidogyne incognita* and *Globodera pallida*. *Plant Biotechnol J* 2:3–12
- McCarter JP, Mitreva MD, Martin J, Dante M, Wylie T, Rao U, Pape D, Bowers Y, Theising B, Murphy CV, Kloek AP, Chiapelli BJ, Clifton SW, Bird DM, Waterston RH (2003) Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*. *Genome Biol* 4:R26
- Mitreva M, Elling AA, Dante M, Kloek AP, Kalyanaraman A, Aluru S, Clifton SW, Bird DM, Baum TJ, McCarter JP (2004a) A survey of SL1-spliced transcripts from the root-lesion nematode *Pratylenchus penetrans*. *Mol Genet Genomics* 272:138–148
- Mitreva M, McCarter JP, Martin J, Dante M, Wylie T, Chiapelli B, Pape D, Clifton SW, Nutman TB, Waterston RH (2004b) Comparative genomics of gene expression in the parasitic and free-living nematodes *Strongyloides stercoralis* and *Caenorhabditis elegans*. *Genome Res* 14:209–220
- Mitreva M, Jasmer DP, Appleton J, Martin J, Dante M, Wylie T, Clifton SW, Waterston RH, McCarter JP (2004c) Gene discovery in the adenophorean nematode *Trichinella spiralis*: an analysis of transcription from three life cycle stages. *Mol Biochem Parasitol* 137:277–291
- Müller J (1999) The economic importance of *Heterodera schachtii* in Europe. *Helminth* 36:205–213
- Nickel W (2003) The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. *Eur J Biochem* 270:2109–2119

- Nielsen H, Krogh A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol* 6:122–130
- Parkinson J, Mitreva M, Whitton C, Thomson M, Daub J, Martin J, Schmid R, Hall N, Barrell B, Waterston RH, McCarter JP, Blaxter ML (2004) A transcriptomic analysis of the phylum Nematoda. *Nat Genet* 36:1259–1267
- Perry RN, Zunke U, Wyss U (1989) Observations on the response of the dorsal and subventral oesophageal glands of *Globodera rostochiensis* to hatching stimulation. *Revue Nématol* 12:91–96
- Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003) TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19:651–652
- Popeijus H, Overmars H, Jones J, Blok V, Goverse A, Helder J, Schots A, Bakker J, Smant G (2000) Degradation of plant cell walls by a nematode. *Nature* 406:36–37
- Prior A, Jones JT, Blok VC, Beauchamp J, McDermott L, Cooper A, Kennedy MW (2001) A surface-associated retinol- and fatty acid-binding protein (Gp-FAR-1) from the potato cyst nematode *Globodera pallida*: lipid binding activities, structural analysis and expression pattern. *Biochem J* 356:387–394
- Qin L, Kudla U, Roze EH, Goverse A, Popeijus H, Nieuwland J, Overmars H, Jones JT, Schots A, Smant G, Bakker J, Helder J (2004) Plant degradation: a nematode expansin acting on plants. *Nature* 427:30
- Sambrook J, Fritsch EF, Martin J (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Schwekendiek A, Maier TR, Womack CR, Byrne DH, de Boer JM, Davis EL, Baum TJ (1999) Initial characterisation of endochitinase genes of the plant-parasitic nematode *Heterodera glycines*. *J Nematol* 31:568
- Shpigel E, Roiz L, Goren R, Shoseyov O (1998) Bacterial cellulose-binding domain modulates in vitro elongation of different plant cells. *Plant Physiol* 117:1185–1194
- Smant G, Goverse A, Stokkermans JP, de Boer JM (1997) Potato root diffusate-induced secretion of soluble, basic proteins originating from the esophageal glands of potato cyst nematodes. *Phytopathology* 87:839–845
- Smant G, Stokkermans JP, Yan Y, de Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ, Henriessat B, Davis EL, Helder J, Schots A, 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 U S A* 95:4906–4911
- The *C. elegans* Sequencing Consortium (1998) The genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012–2018
- Thomas JD, Conrad RC, Blumenthal T (1988) The *C. elegans* trans-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. *Cell* 54:533–539
- Torto TA, Li S, Styer A, Huitema E, Testa A, Gow NA, van West P, Kamoun S (2003) EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res* 13:1675–1685
- Tytgat T, Vercauteren I, Vanholme B, De Meutter J, Vanhoutte I, Gheysen G, Borgonie G, Coomans A, Gheysen G (2005) An SXP/RAL2 protein produced by the subventral pharyngeal glands in the plant parasitic root-knot nematode *Meloidogyne incognita*. *Parasitol Res* 95: 50–54
- Urwin PE, Lilley CJ, McPherson MJ, Atkinson HJ (1997) Resistance to both cyst and root-knot nematodes conferred by transgenic *Arabidopsis* expressing a modified plant cystatin. *Plant J* 12:455–461
- Vanholme B, De Meutter J, Tytgat T, Gheysen GD, Vanhoutte I, Gheysen GD (2002) An improved method for whole-mount in situ hybridization of *Heterodera schachtii* juveniles. *Parasitol Res* 88:731–733
- Vanholme B, De Meutter J, Tytgat T, Van Montagu M, Coomans A, Gheysen G (2004). Secretions of plant-parasitic nematodes: a molecular update. *Gene* 332:13–27
- Wang SH, Zheng HJ, Dissanayake S, Cheng WF, Tao ZH, Lin SZ, Piessens WF (1997) Evaluation of recombinant chitinase and SXP1 antigens as antimicrofilarial vaccines. *Am J Trop Med Hyg* 56:474–481
- Wylie T, Martin JC, Dante M, Mitreva MD, Clifton SW, Chinwalla A, Waterston RH, Wilson RK, McCarter JP (2004) Nematode.net: a tool for navigating sequences from parasitic and free-living nematodes. *Nucleic Acids Res* 32:D423–D426