Published in final edited form as:
Mol Biochem Parasitol. 2009 September ; 167(1): 32–40. doi:10.1016/j.molbiopara.2009.04.004.

Expressed sequence tags of the peanut pod nematode *Ditylenchus africanus*: the first transcriptome analysis of an Anguinid nematode

Annelies Haegeman[1], **Joachim Jacob**[1], **Bartel Vanholme**[1,3], **Tina Kyndt**[1], **Makedonka Mitreva**[2], and **Godelieve Gheysen**[1]

1 *Faculty of Bioscience Engineering (FBE), Department of Molecular Biotechnology, Ghent University, Coupure links 653, B-9000 Ghent, Belgium*

2 *The Genome Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO, USA*

Abstract

In this study, 4847 expressed sequenced tags (ESTs) from mixed stages of the migratory plant-parasitic nematode *Ditylenchus africanus* (peanut pod nematode) were investigated. It is the first molecular survey of a nematode which belongs to the family of the Anguinidae (order Rhabditida, superfamily Sphaerularioidea). The sequences were clustered into 2596 unigenes, of which 43% did not show any homology to known protein, nucleotide, nematode EST or plant-parasitic nematode genome sequences. Gene ontology mapping revealed that most putative proteins are involved in developmental and reproductive processes. In addition unigenes involved in oxidative stress as well as in anhydrobiosis, such as LEA (late embryogenesis abundant protein) and trehalose-6-phosphate synthase were identified. Other tags showed homology to genes previously described as being involved in parasitism (expansin, SEC-2, calreticulin, 14-3-3b and various allergen proteins). *In situ* hybridization revealed that the expression of a putative expansin and a venom allergen protein was restricted to the gland cell area of the nematode, being in agreement with their presumed role in parasitism. Furthermore, 7 putative novel candidate parasitism genes were identified based on the prediction of a signal peptide in the corresponding protein sequence and homologous ESTs exclusively in parasitic nematodes. These genes are interesting for further research and functional characterization. Finally, 34 unigenes were retained as good target candidates for future RNAi experiments, because of their nematode specific nature and observed lethal phenotypes of *Caenorhabditis elegans* homologs.

Keywords

transcripts; parasitism genes; gene ontology; anhydrobiosis; RNAi

Corresponding author: Godelieve Gheysen (E-mail: godelieve.gheysen@ugent.be),. Faculty of Bioscience Engineering (FBE), Department of Molecular Biotechnology, Ghent University, Coupure links 653, B-9000 Ghent, Belgium, tel 003292645888; fax 003292646219.

[3]present address: Department of Plant Systems Biology, Flanders Interuniversitary, Institute of Biotechnology, Technologiepark 927, B-9052 Zwijnaarde, Belgium

Note: EST sequence data reported in this paper are available in the GenBank™, EMBL and DDBJ databases under the accession numbers from **FE920352** to **FE925198**.

Introduction

Expressed sequence tag (EST) analysis is a relatively cheap and rapid method to obtain a first molecular impression of a species. The technique consists of a random selection of clones from a cDNA library and sequencing of their inserts. Although it is mainly used for gene discovery [1], it can also be used for other goals, e.g. estimation of gene expression level [2], detection of single nucleotide polymorphisms [3], or improving genome annotation [4]. In nematology, the technique is widely used and to date over one million ESTs from over 60 species are available (dbEST, NCBI GenBank). Some of these ESTs are derived from cDNA libraries generated from specific life stages or from specific tissues of the nematode. In the case of plant-parasitic nematodes, the transcriptional activity of the pharyngeal glands is of particular interest since gland proteins are injected in the plant tissue during the nematode-host interaction. EST analyses have led to the discovery of many of these parasitism genes that code for hydrolytic enzymes, such as pectate lyase [5] and xylanase [6]. They are involved in maceration of the plant cell walls during migration of the nematode in the plant tissue.

The peanut pod nematode *Ditylenchus africanus* (Wendt, Swart, Vrain and Webster, 1995) was first found in hulls and seeds of groundnut (*Arachis hypogea* L.) in South Africa [7]. It was initially identified as *Ditylenchus destructor*, the potato rot nematode, but experiments showed that it caused no damage to different potato varieties [8]. Therefore the South African population was eventually considered a new race, but later molecular data revealed that this organism rather had to be considered as a new species [9]. Within the nematode order of the Rhabditida, it belongs to the family of the Anguinidae and superfamily Sphaerularioidea [10]. *D. africanus* is a migratory endoparasite with peanut as the main host, causing black discoloration of the seeds and pods. Besides being a parasite of plants, it can also feed and reproduce on the hyphae of common plant pathogenic fungi such as *Aspergillus parasiticus*, *Fusarium oxysporum*, *Botrytis cinerea* and *Rhizoctonia solani*. It has a short life cycle of only 6–7 days at its optimal temperature of 28°C and can survive long periods of drought by anhydrobiosis [8]. Of several economically important genera in the Anguinidae family, *Ditylenchus* spp. have the widest impact on agriculture [11]. *Ditylenchus dipsaci*, a species complex with an extremely wide host range, is one of the most devastating plant-parasitic nematodes, especially in temperate regions (www.eppo.org). *D. destructor* is an important pest of potato tubers in Europe and North America, while *D. africanus* is a huge problem in the cultivation of groundnuts in South Africa. It was shown that 73% of the seeds collected from different growers was infected [7], and that this mainly has a qualitative effect on the groundnuts leading to a significantly lower income for the farmers [12].

D. africanus is an interesting nematode to subject to EST analysis for several reasons. First, no molecular knowledge is available for this species or any other member of the Anguinidae. Second, it has a different taxonomic classification in contrast to other EST studies examining plant-parasitic nematodes, which focused mainly on species from the superfamily Tylenchoidea (grouping cyst and root-knot nematodes as well as migratory nematodes such as *Radopholus* and *Pratylenchus*). Third, it is a plant-parasitic nematode facultatively feeding on fungi. Only one nematode with a similar feeding habit has been investigated using a comparable approach (*Bursaphelenchus xylophilus*; superfamily Aphelenchoidea) [13]. Finally, it adds to the EST dataset of migratory nematodes, therefore empowering comparative studies. The focus for small-scale EST projects is gradually shifting from sedentary nematodes towards migratory nematodes. In the last few years, EST data have become available for several migratory nematodes such as *Pratylenchus penetrans* [14], *Radopholus similis* [15], *B. xylophilus* [13] and *Xiphinema index* [16].

Materials and methods

1 Nematode culture, cDNA library construction and EST generation

D. africanus was cultured at 25°C on carrot discs in small petridishes (Ø 35 mm) under sterile conditions. Carrot discs were infected with approximately 100 individual nematodes. Six weeks after inoculation, nematodes were collected by rinsing the carrot discs with sterile water. RNA was extracted from approximately 10,000 individuals with TRIzol (Invitrogen, Carlsbad, CA, USA) as described by Jacob *et al.* [15]. A cDNA library was constructed using the SMART cDNA Library Construction Kit (Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions, starting from 1 μg RNA. The resulting *D. africanus* mixed stage library contained over 8×10^5 primary transformants. Random colonies were sequenced using the M13 forward or reverse primer at the Genome Center (Washington University, St.-Louis, MO, USA). Resulting sequences were submitted to the EST division of GenBank (dbEST).

2 Cleaning and clustering of the EST sequences

The sequences were cleaned using Seqclean (www.tigr.org) with a locally downloaded vector database and default parameter settings. The cleaned dataset was clustered using the TIGR Gene Indices Clustering Tool (TGICL) [17] and sequences were assembled by CAP3 [18] using default settings. The fragmentation (i.e. the percentage of unigenes which are redundant) was estimated with ESTstat [19]. The obtained unigenes served as a basis for following analyses.

3 Homology searches

Basic Local Alignment Search Tool (blast) analyses [20] were done both locally and by using netblast. Blastx and blastn searches were conducted with all unigenes against the NCBI protein and nucleotide database. Additionally, blastn and tblastx searches were done against the genomes of *Meloidogyne incognita* (http://meloidogyne.toulouse.inra.fr/) [21] and *Meloidogyne hapla* (http://www.hapla.org/) [22]. Since most data available for nematodes are in the EST database, a tblastx search against all nematode ESTs was done. In-house perl scripts parsed the resulting hits for species names and unigenes were subsequently classified into different categories (nematodes, invertebrates, plants, animals, fungi, prokaryotes) according to the species names derived from the blast hits. In parallel, a blastx search was conducted against proteins of the model organism *Caenorhabditis elegans*. Resulting top hits were searched for RNAi phenotypes using Wormbase [23]. For all *C. elegans* homologs with an RNAi lethal phenotype, Gene Ontology (GO) terms were retrieved and visualized with WEGO [24]. Blastn against all mitochondrial nematode genes revealed putative mitochondrial unigenes.

To annotate gene ontology (GO) terms, BLAST2GO [25] was used on all unigenes using default parameters. Blastx value cut-off was chosen at E<1e-4. GOSlim view was used and GO graphs were generated with a node scoring filter of 25 for "molecular function", 50 for "biological process" and 12 for "cellular component".

4 Translation into putative proteins

All unigenes were translated with OrfPredictor [26]. The blastx output was used to select the correct reading frame for translation. The minimum amino acid number for predicted protein sequences was set to 40. Nucleotide sequences were trimmed to their coding parts and the overall GC content as well as the average GC content of the first (GC1), second (GC2) and third nucleotide (GC3) of the codons was calculated. Unigenes without predicted open reading frame were considered to be non-coding, and for these sequences only the GC content was calculated. Signal peptides for secretion in the predicted proteins were predicted by SignalP

3.0 [27]. A signal peptide was only assigned to a sequence when both the Hidden Markov Model (HMM) and the neural network predicted its presence. Moreover, a transmembrane domain search (http://protfun.net/services/TMHMM/) on the mature putative proteins revealed whether the protein was retained in the membrane.

5 Dot blot analysis

Forward (F) and reverse (R) primers were developed to amplify fragments of CL1, CL7, CL270, CL371, CL406, **FE921742** and **FE922861** (Table 1) by polymerase chain reaction (PCR). The reaction mixture contained 300 ng purified plasmids of the cDNA library as template, 0.5 μM of each primer, 4 mM dNTPs, 1.5 mM $MgCl_2$, 20 mM Tris-HCl (pH 8.3), 50 mM KCl and 1U of *Taq* DNA polymerase (Invitrogen, Carlsbad, CA, USA). PCR conditions were as follows: 35 cycles of 30s at 94°C, 30s at 54°C and 40s at 72°C. PCR products were loaded on a 0.5x TAE 1.5% agarose gel and the resulting fragments were ligated into pGEM-T (Promega, Madison, WI, USA) and the ligation mixture was used for transformation into *Escherichia coli* DH5α cells. Transformed cells were selected on LB agar plates supplemented with 100 μg/ml carbenicillin. Plasmids of positive colonies were extracted using the Nucleobond AX kit (Macherey-Nagel, Düren, Germany) and the inserts were sequenced at AGOWA (Berlin, Germany). Probes were generated by a PCR under conditions as described above with purified plasmid with the correct insert as template. Resulting PCR products were radioactively labeled with [α-^{32}P]-dCTP with the DecaLabel DNA Labeling Kit (Fermentas, St. Leon-Rot, Germany) according to the manufacturer's instructions. As a positive control, 100 ng of unlabeled PCR product was spotted on a Hybond N+ membrane (GE Healthcare, Uppsala, Sweden), together with 1.5 μg carrot RNA as a negative control and 1.5 μg *D. africanus* mixed population RNA, both extracted with TRIzol (Invitrogen) according to the manufacturer's instructions. The membrane was baked at 80°C for 2h and prehybridized for 1h at 65°C in hybridization solution (5x SSC, 5x Denhardt's solution, 0.5% SDS, 100 μg/ml salmon sperm DNA). After prehybridization, the hybridization buffer was replaced and the corresponding probe was added for hybridization overnight at 65°C. Washing was done at hybridization temperature for 5 min in 2x SSC with 0.1% SDS and 15 min in 2x SSC. The blots were exposed to a Fujifilm Imaging Plate for 6h and scanned with a FLA-5100 imaging system (Fujifilm, Düsseldorf, Germany).

6 *In situ* hybridization

A whole mount *in situ* hybridization was carried out according to Vanholme *et al.* [28] with minor modifications. Nematodes were fixed in 4% paraformaldehyde for 4h at room temperature, 16h at 4°C, and again 4h at room temperature. Permeabilization was performed by incubation in 0.5 mg/ml proteinase K (Merck, Whitehouse Station, NJ, USA) for 1h 30 min at room temperature. Hybridization was done overnight at 47°C. The templates for the production of DNA probes against unigenes with homology to expansin-like proteins (CL496) and venom allergen proteins (**FE920532**) were generated by a first PCR under standard conditions as described above on the plasmid pGEM-T containing the cDNA clone of the corresponding unigene. Subsequent linear PCRs using the first PCR products as templates with digoxigenin(DIG)-labeled oligonucleotides (Roche, Mannheim, Germany) and a single primer, generated single strand DNA probes (F-primers: sense probes; R-primers: antisense probes) (Table 1).

Results

1 Cleaning and cluster analysis

A total of 4847 ESTs were obtained from a cDNA library prepared from mixed stages of the plant-parasitic nematode *D. africanus.* Cleaning the EST sequences resulted in the removal of 602 ESTs, due to vector contamination, poly(A) tails and sequences shorter than 100

nucleotides. The average length of the 4245 retained ESTs was 286 bp. Clustering to reduce data redundancy resulted in 2456 ESTs forming 807 contigs (merged overlapping sequences) classified into 778 clusters (groups of contigs with minor sequence variation) and 1789 singletons, resulting in a total of 2596 unigenes. After clustering, the average length of a unigene increased to 335 bp and the average GC content is 38.9%. A graphical representation of the cluster size distribution is shown in Figure 1. The largest single cluster represents 84 ESTs, which is 2.0% of the total number of ESTs. Most of the clusters (88.0%) consist of 4 or less ESTs. The fragmentation or underclustering of the dataset was estimated at 3.5%, which means that the dataset represents approximately 2505 genes.

2 Transcript abundance and homology searches

The 20 most abundant unigenes in the EST dataset represent 9.6% of the total number of ESTs (Table 2). Eight of these unigenes, including cluster 1, showed no significant hit in any of the blast searches conducted (E<1e-3). In Table 3 unigenes are shown which have homologs involved in anhydrobiosis, oxidative stress and parasitism. Classifying the unigenes based on species names of blastx homologous sequences revealed 298 animal specific sequences of which 160 unigenes were nematode specific. These nematode specific unigenes were classified according to their homologs in the blastx search and in the nematode EST database (Figure 2). Most of these sequences have homologs in both free-living, animal-parasitic and plant-parasitic nematodes, suggesting a role in general nematode development and metabolism. Only one unigene (CL496) has homologous protein sequences exclusively from plant-parasitic nematodes, i.e. an expansin (top blastx hit: expansin-like protein BAG16537 from *B. xylophilus*, E=1e-20). Unigenes with homologs to animal-parasitic nematodes exclusively (4) are all homologous to proteins with unknown function. Twelve unigenes are found in both animal- and plant-parasitic nematodes exclusively, and include several allergen homologous sequences (Table 3) as well as sequences with unknown function. Remarkably, some unigenes had unexpected blastx homologs originating from prokaryotic sequences exclusively (15). Similarly, two unigenes matched fungal hypothetical proteins only: one (**FE921742**) with hits from *Coprinopsis cinerea* (E=3e-04) and the other (**FE922861**) with hits from fungal plant pathogens (*Gibberella* sp., *Sclerotinia* sp. and *Botryotinia* sp.; E=4e-14). Searching the genomes of *M. incognita* and *M. hapla* revealed a tblastx homolog for respectively 30.6% and 35.4% of the unigenes (E<1e-3). Thirty unigenes of putative mitochondrial origin were retrieved corresponding to 40 ESTs or 0.94% of the total dataset. The GC content of the putative mitochondrial unigenes is 34.5%, which is lower than the overall GC content of 38.9% (Figure 3).

For 1113 unigenes (42.9%), no homologous sequences could be found in either protein, nucleotide or nematode EST databases or in the genomes of *M. incognita* and *M. hapla* (E<1e-3). These sequences had a shorter average sequence length (260 bp) compared to unigenes with homologs (391 bp) (Figure 3), which is the most obvious reason why no homology is found. These unigenes of unknown origin could correspond either to non-coding sequences (e.g. UTRs), regulatory and structural RNA or to novel protein coding genes.

3 Gene Ontology

To have a general overview of the functions of the unigenes, Gene Ontology (GO) terms were assigned to the unigenes with BLAST2GO. Of the 931 unigenes for which BLAST2GO could identify at least one blast hit, 801 were annotated with at least one GO term. The number of GO terms per unigene varied from 1 to 53. In total, 2648 different GO terms were retrieved: 1658 of biological process, 653 of molecular function and 337 of cellular component (Figure 4). In the molecular function category the majority of the putative proteins are involved in protein binding (43.7%) whereas other GO terms occur between 5.8% (hydrolase activity) and 10.9% (structural constituent of ribosome). For the biological process category, 23.0% of the

terms are involved in embryonic development and 17.9% in larval development. The most prevalent cellular component terms are protein complex (25.1%), integral to membrane (16.7%) and mitochondrion (13.6%).

4 Translation into putative proteins

Unigenes were translated into putative protein sequences on the basis of the blastx output or based on the longest ORF in case of no blastx homology. For 1769 unigenes a putative ORF of minimum 40 AA was found, while 827 were predicted to be non-coding by OrfPredictor. Overall GC contents of the coding sequences is 41.4%, while GC1, GC2 and GC3 are 50.8; 38.5 and 48.5%, respectively. The overall GC content of putative non-coding sequences is 33.6% (Figure 3). A signal peptide for secretion was found in 101 of the 1769 putative proteins (5.7%). Of these, 90 had no transmembrane domain (5.1%). These are possibly being secreted by the nematode and hence could play a role in parasitism. Only 13 showed similarity to known, mostly hypothetical, proteins (blastp search; E<1e-3). The majority (72) of these 90 putative secreted proteins did not show any similarity to other known nematode ESTs (tblastn search; E<1e-3). Only 7 candidate parasitism genes were retained which had homology restricted to parasitic nematode ESTs (Table 4).

5 RNAi phenotypes

RNAi phenotypes were assigned to the unigenes by homology with *C. elegans* proteins. 965 unigenes had a *C. elegans* homolog (blastx search; E<1e-3). After removing redundant protein hits, 815 different *C. elegans* proteins were retained. For 473 or 58.0% of the total number of unigenes with a *C. elegans* homolog, an RNAi phenotype was retrieved from the Wormmart database. The majority (333) of these proteins have a lethal phenotype (WBPhenotype0000050, WBPhenotype0000054, or WBPhenotype0000062). GO terms for these *C. elegans* proteins were retrieved and the terms of the second level of the “biological process” category are shown in Figure 5. All of them have the GO terms “developmental process” and “multicellular organismal process”, suggesting that the corresponding genes are involved in the general development of the organism. Other GO terms that occur in the majority of these genes are “reproduction”, “metabolic process”, “cellular process”, “growth” and “biological regulation”. 366 unigenes (14.1% of the total number of unigenes) correspond to *C. elegans* genes with lethal phenotype. A subset of these unigenes (34) are promising future candidates for parasitic control due to their nematode specific nature.

6 Dot blot analysis

To find out whether the largest clusters without any homology were really present in the mRNA pool, a dot blot analysis was carried out. Correct fragments were obtained by PCR on the cDNA library for both CL1 and CL7. The dot blot analysis showed a clear signal for both fragments against RNA extracted from *D. africanus* nematodes, while no signal was obtained against carrot RNA (Figure 6). This shows that these clusters are not artifacts or contaminants from the cDNA library since they are also present in independently isolated nematode RNA.

Some sequences showed similarity to prokaryotic or fungal sequences exclusively in a blastx search, for example CL270, CL371, CL406, **FE921742** and **FE922861**. To check if these sequences are derived from contamination, a similar dot blot analysis was done. For three of the five selected sequences (CL406, **FE921742** and **FE922861**), the corresponding fragment could not be amplified from the cDNA library. This most likely means that these sequences are not present in the cDNA library, and are probably the result of contamination. For CL270 and CL406, the correct fragment could be cloned, however in a dot blot, no signal was obtained neither for nematode nor carrot RNA (Figure 6). This suggests that CL270 and CL406 are from contaminating origin, although it cannot be ruled out that the detection limit of the dot blot is too low to detect the corresponding mRNA.

7 *In situ* hybridization

An *in situ* hybridization with probes against CL496 (putative expansin-like protein) and **FE920532** (putative venom allergen protein) was carried out. Both probes showed staining in the gland cell area of the nematode, while the corresponding sense probes showed no staining (Figure 7). Whether the expression was located in the subventral or dorsal gland cells could not be determined. These results suggest that both proteins are possibly being secreted by the nematode and hence could play a role in parasitism.

Discussion

An increasing amount of sequence information derived from plant-parasitic nematodes is becoming available through EST projects, especially from sedentary nematodes from the superfamily Tylenchoidea within the order of the Rhabditida. In contrast, *D. africanus* is a migratory endoparasite belonging to the family of the Anguinidae, superfamily Sphaerularioidea. This family comprises important seed and stem nematodes such as *Anguina tritici* and *D. dipsaci*. This study is the first molecular survey of a nematode in this family, *D. africanus*, by means of a small scale EST analysis. Only 57% of the clustered unigenes showed homology to known proteins, nematode ESTs and the genomes of the plant-parasitic nematodes *Meloidogyne incognita* and *M. hapla*. This means that at least 43% of all unigenes did not have any homology to either all known proteins, nucleotide sequences, nematode ESTs or the *Meloidogyne* genomes. The amount of unigenes without homology is unusually high, as compared to for example *B. xylophilus* (27% unigenes without homology) [13], a nematode which is phylogenetically further removed from the heavily sampled cyst and root-knot nematodes than *D. africanus*. The most obvious reason for this high amount of orphan sequences is the relatively short average length of the inserts of the cDNA library. The shorter the insert, the less likely it is to find a significant homolog. Moreover, the length distribution of unigenes with homology is similar to unigenes which are predicted to be coding. Other sequences without homology are shorter in length and could be either non-coding regions, regulatory or structural RNAs or pioneer sequences. The sequences predicted to be non-coding have a lower average GC content (34%) compared to coding sequences (41%). In *C. elegans*, the GC content of coding sequences is around 45%, while it is only 34% in intergenic regions [29]. These similar numbers suggest that a large portion of the *D. africanus* non-coding sequences are 5′ or 3′ UTRs. On the other hand, since we are dealing with the first sequences from a representative of the Anguinidae family, it is to be expected that there are a number of species or family specific unigenes without any homology in currently available sequence databases. Interestingly, a considerable number of these unknown unigenes seems to be highly expressed. Eight of the 20 most abundant clusters including the most abundant cluster (CL1) have no homologs at all. A dot blot for two of these unknown, highly expressed unigenes confirmed that these are truly present in the nematode mRNA pool. A high fraction of the largest clusters lacking homology has also been observed in other nematode EST datasets, and these ESTs are thought to be of mitochondrial origin [15,30]. Other highly expressed genes represented by a large number of ESTs are for example major sperm protein (important in the motility of sperm), ferritin (iron-storage protein), tropomyosin (an actin-filament regulator) and *sec-2* or *far* (fatty acid binding protein, previously hypothesized to be involved in defense; [31]). The latter two have been reported earlier as highly abundant transcripts in ESTs [1,13, 15,32]. Gene ontology mapping revealed that most unigenes are involved in the biological processes of reproduction and development. This is not surprising, since *D. africanus* has a short life cycle of only 6–7 days and has a high reproductive rate.

Interestingly, nematode specific unigenes predominantly had EST homologs from animal-parasitic species (95%), while 79% also had homologs in plant-parasitic nematodes. This apparent closer relationship to animal-parasitic nematodes can be due to the fact that there is

more sequence information available for animal-parasitic compared to plant-parasitic nematodes (24,575 vs. 1227 proteins and 294,843 vs. 143,883 ESTs). Only one unigene (CL496) had homologs within plant-parasitic nematodes exclusively, (an expansin identified in *B. xylophilus*). In addition, CL8 shows the highest similarity to an EST from *Meloidogyne chitwoodi* with homology to expansins. Although plant-parasitic nematodes are known to secrete an arsenal of different cell wall modifying enzymes [33], unigenes similar to expansin genes were the only ones found in the EST dataset. The expression of CL496 was demonstrated to be in the gland cell region of the nematode, pointing towards the secretion of the protein. The relatively low occurrence of cell wall modifying enzymes in EST datasets was previously also observed for the migratory nematodes *R. similis* [15] and *P. penetrans* [14]. This observation suggests that cell wall modifying enzymes only form a very small part of the transcriptome. Some unigenes showed similarity to putative parasitism genes such as *sec-2* [31], calreticulin [34] and 14-3-3b protein [35]. However all of these enzymes should be more thoroughly investigated in order to find out their true role in parasitism. For example, calreticulin and 14-3-3b protein have been linked to the formation of a feeding site by sedentary nematodes [34]. The fact that these secreted proteins are also present in migratory nematodes could suggest a more general role in parasitism. The unigenes were searched for new putative parasitism genes which should code for proteins equipped with a signal peptide. In addition, the best candidates should only have homologs in parasitic nematode ESTs. Although this approach rules out the finding of species or family specific parasitism genes, since it is dependent on available nematode EST data, this condition was included to reduce the number of false positives. Seven promising novel candidate parasitism genes were found (Table 4). These genes are interesting for future functional studies and could reveal new insights about plant parasitism.

Several nematode parasitism genes have been proposed to originate through horizontal gene transfer (HGT) [36]. New HGT candidates should have homology with prokaryotes or fungi exclusively, although prokaryotic sequences could also point to contamination or the presence of an endosymbiont [15]. Several unigenes identified in our EST dataset showed similarity to prokaryotes exclusively, but most of these unigenes show homology to widely occurring non endosymbiotic species, suggesting some bacterial contamination. For other unigenes resembling prokaryotic and fungal sequences, it could not be demonstrated that these were present in independently isolated RNA of the nematode, hence these are probably of contaminating origin as well.

Our EST dataset contains several genes that are predicted to be involved in anhydrobiosis, a dehydrated state of the nematode. One of these genes is a LEA (late embryogenesis abundant protein) homolog, proteins found to be highly expressed in the anhydrobiotic nematodes *Aphelenchus avenae* and *Steinernema carpocapsae*, as well as in plants where they prevent protein aggregation during seed desiccation [37,38]. In addition, fatty acid desaturase, stomatin and trehalose-6-phosphate synthase were found, mainly involved in maintaining the integrity of the cell membrane [38]. Other unigenes commonly upregulated in anhydrobiotic stages are involved in oxidative stress, such as superoxide dismutase, glutathione-S-transferase and peroxiredoxin. These enzymes and also LEA proteins are represented by a quite large number of ESTs, which indicates that oxidative stress as well as drought stress are important factors affecting the nematode. In agreement with this finding, as much as 8% of the GO mappings in the “molecular function” category (node score>25), are associated with oxidoreductase activity.

Another group of unigenes showed remarkable similarities to polyprotein allergens from animal-parasitic nematodes and the venom allergen protein (VAP) of *Heterodera glycines*. Gao *et al.* [39] found that the latter is expressed in the gland cells, which was confirmed for a *vap* homolog in *D. africanus* by *in situ* hybridization. In *M. incognita* it was shown that a venom

allergen AG5-like protein is expressed during preparasitic and parasitic J2 stages exclusively [40]. Animal-parasitic allergen genes have been associated with the immune response of the host animal [41]. However, their presence in plant-parasitic nematodes suggests a more general role for these proteins in parasitism.

The RNAi gene knockout technique has shown promising results in sedentary plant-parasitic nematodes [42]. Transcript levels of a certain gene can be reduced when dsRNA homologous to that gene is fed to the nematode. Interesting RNAi targets for experiments with *D. africanus* were identified by looking for *C. elegans* homologs with lethal RNAi phenotypes. Retained unigenes were all involved in major biological processes (development, metabolic processes, reproduction) as revealed by the Gene Ontology terms of their *C. elegans* homologs. Most of these unigenes have homologs both in free-living and animal-parasitic nematodes, due to the limited genomic data available for plant-parasitic nematodes. Unigenes showing homology to nematodes exclusively (34) are interesting targets for control measures since these are very unlikely to have off-target effects on other organisms. Future experiments will have to prove the efficacy of these candidate RNAi targets and whether or not RNAi has similar effects in migratory nematodes as in sedentary nematodes, although the technique was shown to be effective in the migratory nematode *R. similis* [43].

In conclusion, these EST data from *D. africanus* are the first molecular data from a nematode from the Sphaerularioidea. Although the dataset suffered from the short insert length, it may still provide useful information about biological processes such as parasitism or anhydrobiosis. Furthermore, the data contribute to the ESTs from migratory nematodes, interesting for comparative analyses. Nevertheless, in the future, better quality sequences should be generated by either ESTs or new generation sequencing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Prof. A. McDonald and Dr. H. Fourie (ARC-GCI, Potchefstroom, South Africa) for kindly providing the initial *Ditylenchus africanus* culture. This research was supported by a grant from Ghent University (GOA 01G00805), a grant to A.H. from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) and a grant to J.J. from Ghent University (BOF). B.V. and T.K. are postdoctoral researchers at Ghent University (BOF). Work at Washington University School of Medicine was supported by NIH research grant AI 46593 to M.M.

Reference list

1. 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. Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*. Genome Biol 2003;4:R26. [PubMed: 12702207]
2. Munoz ET, Bogarad LD, Deem MW. Microarray and EST database estimates of mRNA expression levels differ: The protein length versus expression curve for *C. elegans*. BMC Genomics 2004;5:30. [PubMed: 15134588]
3. Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M. Mining SNPs from EST databases. Genome Res 1999;9:167–74. [PubMed: 10022981]
4. Bailey LC, Searls DB, Overton GC. Analysis of EST-driven gene annotation in human genomic sequence. Genome Res 1998;8:362–76. [PubMed: 9548972]
5. de Boer JM, McDermott JP, Davis EL, Hussey RS, Popeijus H, Smant G, Baum TJ. Cloning of a putative pectate lyase gene expressed in the subventral esophageal glands of *Heterodera glycines*. J Nematol 2002;34:9–11. [PubMed: 19265900]

6. Mitreva-Dautova M, Roze E, Overmars H, de Graaff L, Schots A, Helder J, Goverse A, Bakker J, Smant G. A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode *Meloidogyne incognita*. Mol Plant Microbe In 2006;19:521–9.
7. De Waele D, Jones BL, Bolton C, van den Berg E. *Ditylenchus destructor* in hulls and seeds of peanut. J Nematol 1989;21:10–5. [PubMed: 19287571]
8. De Waele D, Venter C, McDonald AH. The peanut pod nematode, *Ditylenchus africanus*. Nematology Circular 1997;218:1–6.
9. Wendt KR, Swart A, Vrain TC, Webster JM. *Ditylenchus africanus* sp-N from South-Africa - A morphological and molecular characterization. Fund Appl Nematol 1995;18:241–50.
10. De Ley, P.; Blaxter, M. Systematic position and phylogeny. In: Lee, DL., editor. The Biology of Nematodes. London: Taylor & Francis; 2002. p. 1-30.
11. Duncan, LW.; Moens, M. Migratory endoparasitic nematodes. In: Perry, RN.; Moens, M., editors. Plant Nematology. Oxfordshire: CABI publishing; 2006. p. 123-52.
12. Mc Donald AH, Loots GC, Fourie H, De Waele D. A microplot study on *Ditylenchus africanus* population densities and damage symptoms on groundnut in relation to commercial yields. Nematology 2005;7:647–53.
13. Kikuchi T, Aikawa T, Kosaka H, Pritchard L, Ogura N, Jones JT. Expressed sequence tag (EST) analysis of the pine wood nematode *Bursaphelenchus xylophilus* and *B. mucronatus*. Mol Biochem Parasit 2007;155:9–17.
14. Mitreva M, Elling AA, Dante M, Kloek AP, Kalyanaraman A, Aluru S, Clifton SW, Bird DM, Baum TJ, McCarter JP. A survey of SL1-spliced transcipts from the root-lesion nematode *Pratylenchus penetrans*. Mol Genet Genomics 2004;272:138–48. [PubMed: 15338281]
15. Jacob J, Mitreva M, Vanholme B, Gheysen G. Exploring the transcriptome of the burrowing nematode *Radopholus similis*. Mol Genet Genomics 2008;280:1–17. [PubMed: 18386064]
16. Furlanetto C, Cardle L, Brown DJF, Jones JT. Analysis of expressed sequence tags from the ectoparasitic nematode *Xiphinema index*. Nematology 2005;7:95–104.
17. Pertea G, Huang XQ, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 2003;19:651–2. [PubMed: 12651724]
18. Huang XQ, Madan A. CAP3: A DNA sequence assembly program. Genome Res 1999;9:868–77. [PubMed: 10508846]
19. Wang JPZ, Lindsay BG, Cui LY, Wall PK, Marion J, Zhang JX, dePamphilis CW. Gene capture prediction and overlap estimation in EST sequencing from one or multiple libraries. BMC Bioinformatics 2005;6:300. [PubMed: 16351717]
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol 1990;215:403–10. [PubMed: 2231712]
21. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchi EGJ, Deleury E, Perfus-Barbeoch L, Anthouard V, Artiguenave F, Blok VC, Caillaud MC, Coutinho PM, Dasilva C, De Luca F, Deau F, Esquibet M, Flutre T, Goldstone JV, Hamamouch N, Hewezi T, Jaillon O, Jubin C, Leonetti P, Magliano M, Maier TR, Markov GV, 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 TJ, Blaxter M, Bleve-Zacheo T, Davis EL, Ewbank JJ, Favery B, Grenier E, Henrissat B, Jones JT, Laudet V, Maule AG, Quesneville H, Rosso MN, Schiex T, Smant G, Weissenbach J, Wincker P. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nat Biotechnol 2008;26:909–15. [PubMed: 18660804]
22. Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, Cohn J, Cromer J, Diener S, Gajan J, Graham S, Houfek TD, Liu Q, Mitros T, Schaff J, Schaffer R, Scholl E, Sosinski BR, Thomas VP, Windham E. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. P Natl Acad Sci USA 2008;105:14802–7.
23. Rogers A, Antoshechkin I, Bieri T, Blasiar D, Bastiani C, Canaran P, Chan J, Chen WJ, Davis P, Fernandes J, Fiedler TJ, Han M, Harris TW, Kishore R, Lee R, Mckay S, Mulller HM, Nakamura C, Ozersky P, Petcherski A, Schindelman G, Schwarz EM, Spooner W, Tuli MA, Van Auken K, Wang D, Wang X, Williams G, Yook K, Durbin R, Stein LD, Spieth J, Sternberg PW. WormBase 2007. Nucleic Acids Res 2008;36:D612–D617. [PubMed: 17991679]

24. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 2006;34:W293–W297. [PubMed: 16845012]
25. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005;21:3674–6. [PubMed: 16081474]
26. Min XJ, Butler G, Storms R, Tsang A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res 2005;33:W677–W680. [PubMed: 15980561]
27. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95. [PubMed: 15223320]
28. Vanholme B, De Meutter J, Tytgat T, Gheysen GDC, Vanhoutte I, Gheysen GDR. An improved method for whole-mount *in situ* hybridization of *Heterodera schachtii* juveniles. Parasitol Res 2002;88:731–3. [PubMed: 12122430]
29. Shin H, Hirst M, Bainbridge MN, Magrini V, Mardis E, Moerman DG, Marra MA, Baillie DL, Jones SJM. Transcriptome analysis for *Caenorhabditis elegans* based on novel expressed sequence tags. BMC Biol 2008:6. [PubMed: 18226223]
30. Dubreuil G, Magliano M, Deleury E, Abad P, Rosso MN. Transcriptome analysis of root-knot nematode functions induced in the early stages of parasitism. New Phytol 2007;176:426–36. [PubMed: 17692078]
31. Prior A, Jones JT, Blok VC, Beauchamp J, McDermott L, Cooper A, Kennedy MW. 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 2001;356:387–94. [PubMed: 11368765]
32. Vanholme B, Mitreva M, Van Criekinge W, Logghe M, Bird D, McCarter JP, Gheysen G. Detection of putative secreted proteins in the plant-parasitic nematode *Heterodera schachtii*. Parasitol Res 2006;98:414–24. [PubMed: 16380840]
33. Vanholme B, De Meutter J, Tytgat T, Van Montagu M, Coomans A, Gheysen G. Secretions of plant-parasitic nematodes: a molecular update. Gene 2004;332:13–27. [PubMed: 15145050]
34. Jaubert S, Milac AL, Petrescu AJ, de Almelda-Engler J, Abad P, Rosso MN. In planta secretion of a calreticulin by migratory and sedentary stages of root-knot nematode. Mol Plant Microbe In 2005;18:1277–84.
35. Jaubert S, Laffaire JB, Ledger TN, Escoubas P, Amri EZ, Abad P, Rosso MN. Comparative analysis of two 14-3-3 homologues and their expression pattern in the root-knot nematode *Meloidogyne incognita*. Int J Parasitol 2004;34:873–80. [PubMed: 15157770]
36. Jones JT, Furlanetto C, Kikuchi T. Horizontal gene transfer from bacteria and fungi as a driving force in the evolution of plant parasitism in nematodes. Nematology 2005;7:641–6.
37. Browne J, Tunnacliffe A, Burnell A. Anhydrobiosis - Plant desiccation gene found in a nematode. Nature 2002;416:38. [PubMed: 11882885]
38. Tyson T, Reardon W, Browne JA, Burnell AM. Gene induction by desiccation stress in the entomopathogenic nematode *Steinernema carpocapsae* reveals parallels with drought tolerance mechanisms in plants. Int J Parasitol 2007;37:763–76. [PubMed: 17306805]
39. Gao B, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS. Molecular characterisation and expression of two venom allergen-like protein genes in *Heterodera glycines*. Int J Parasitol 2001;31:1617–25. [PubMed: 11730789]
40. Ding X, Shields J, Allen R, Hussey RS. Molecular cloning and characterisation of a venom allergen AG5-like cDNA from *Meloidogyne incognita*. Int J Parasitol 2000;30:77–81. [PubMed: 10675748]
41. Murray J, Gregory WF, Gomez-Escobar N, Atmadja AK, Maizels RM. Expression and immune recognition of *Brugia malayi* VAL-1, a homologue of vespid venom allergens and *Ancylostoma* secreted proteins. Mol Biochem Parasit 2001;118:89–96.
42. Lilley CJ, Bakhetia M, Charlton WL, Urwin PE. Recent progress in the development of RNA interference for plant parasitic nematodes. Mol Plant Pathol 2007;8:701–11.
43. Haegeman A, Vanholme B, Gheysen G. Characterization of a putative endoxylanase in the migratory plant-parasitic nematode *Radopholus similis*. Mol Plant Pathol. 200910.1111/J.1364–3703.2009.00539.X

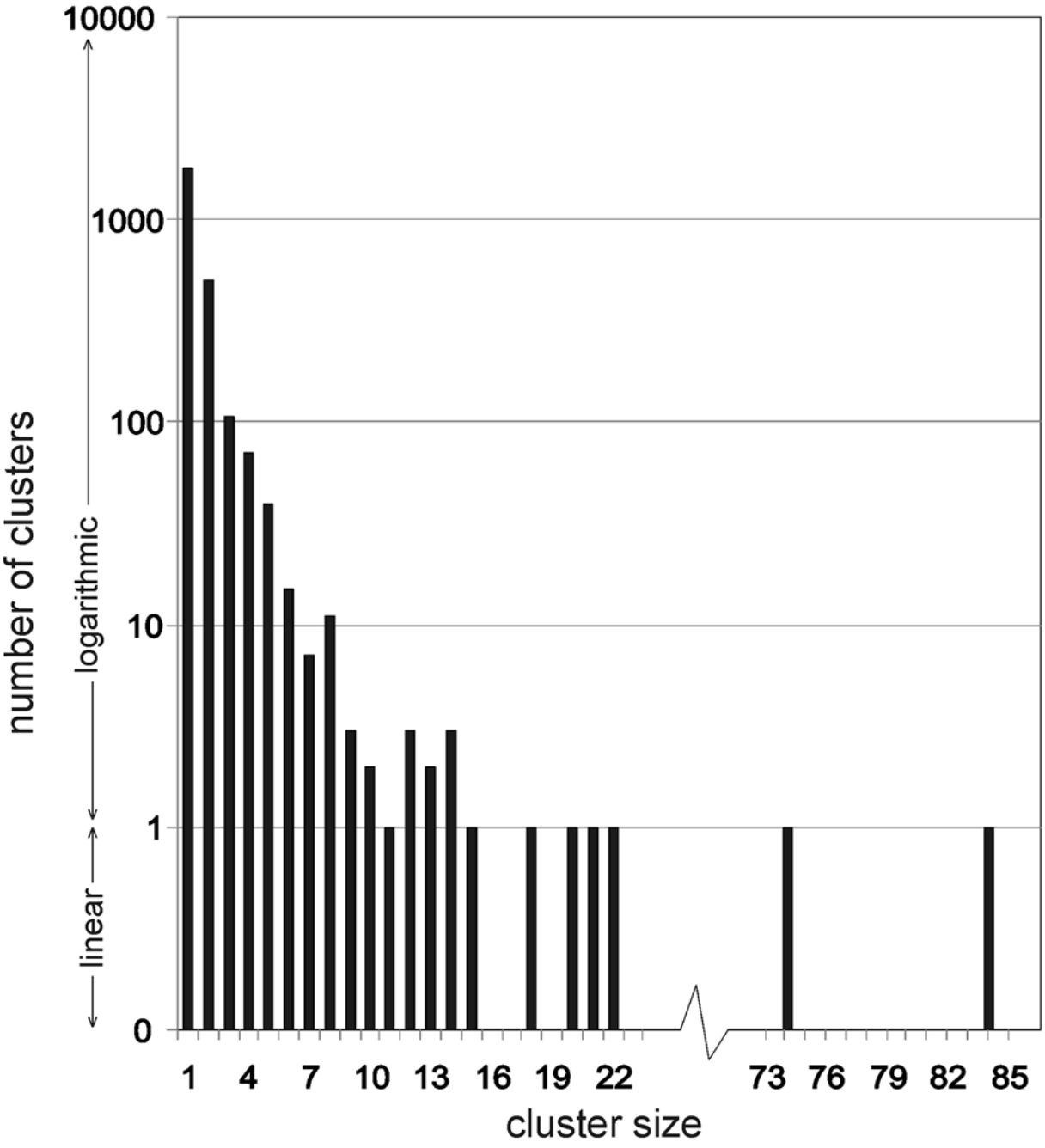

Figure 1.
Cluster size distribution. The chart shows the number of clusters with a particular cluster size (the number of ESTs present in this cluster) in the EST dataset of *Ditylenchus africanus*.

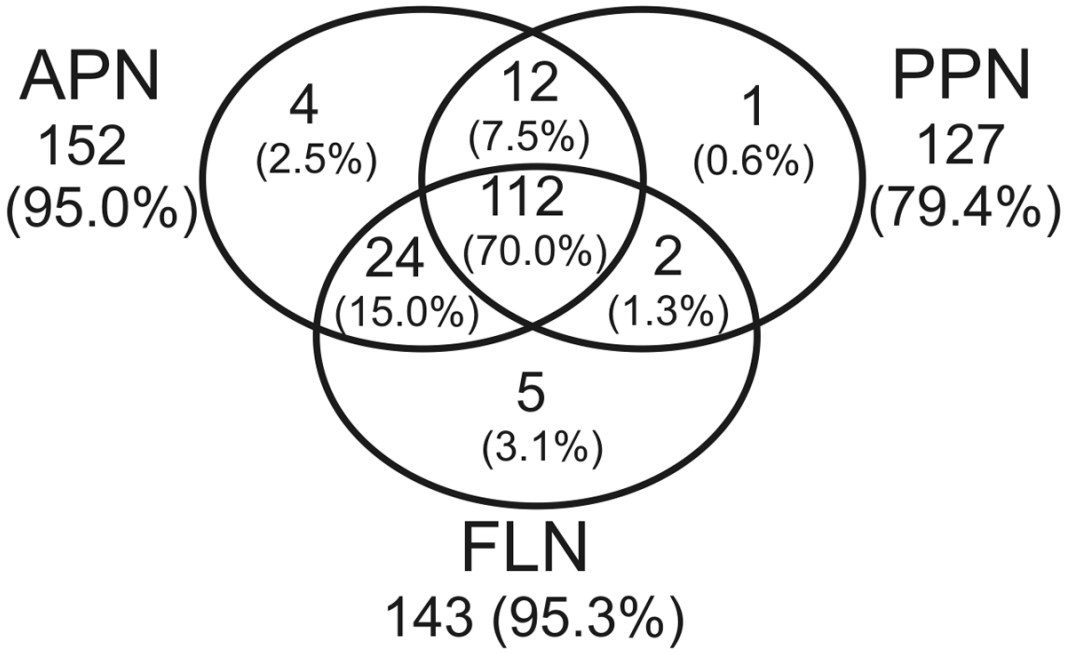

Figure 2.
Classification of the 160 nematode specific unigenes with blastx homology (E<1e-3) into animal-parasitic (APN), plant-parasitic (PPN) and free-living (FLN) nematodes based on nematode EST homology. The number of unigenes in each class is given, including the percentage of the total nematode specific genes.

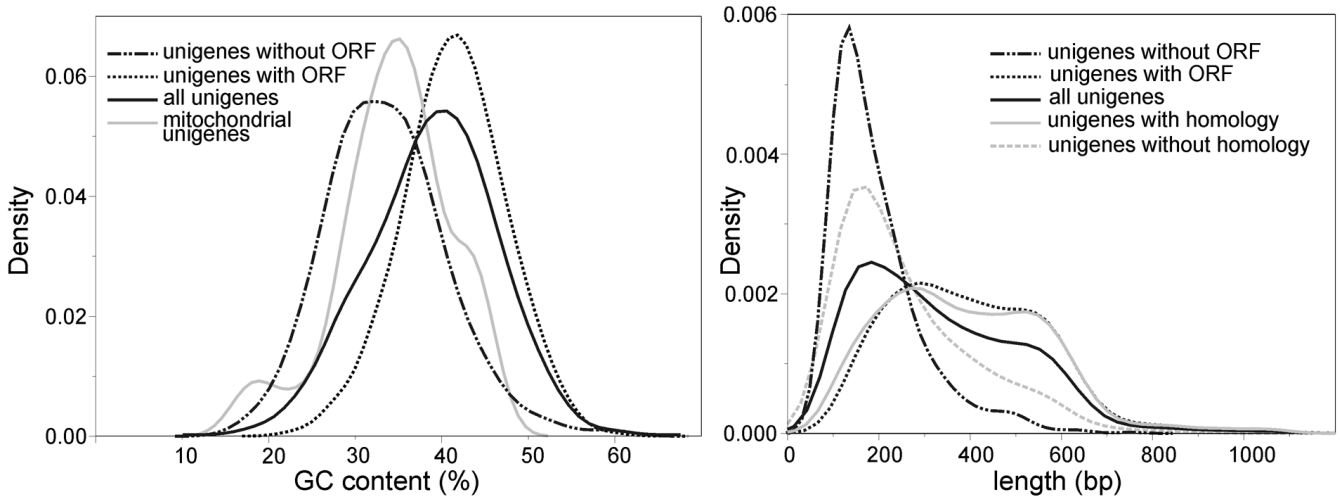

Figure 3.
Density plots of the GC content and length of different subsets of unigenes.

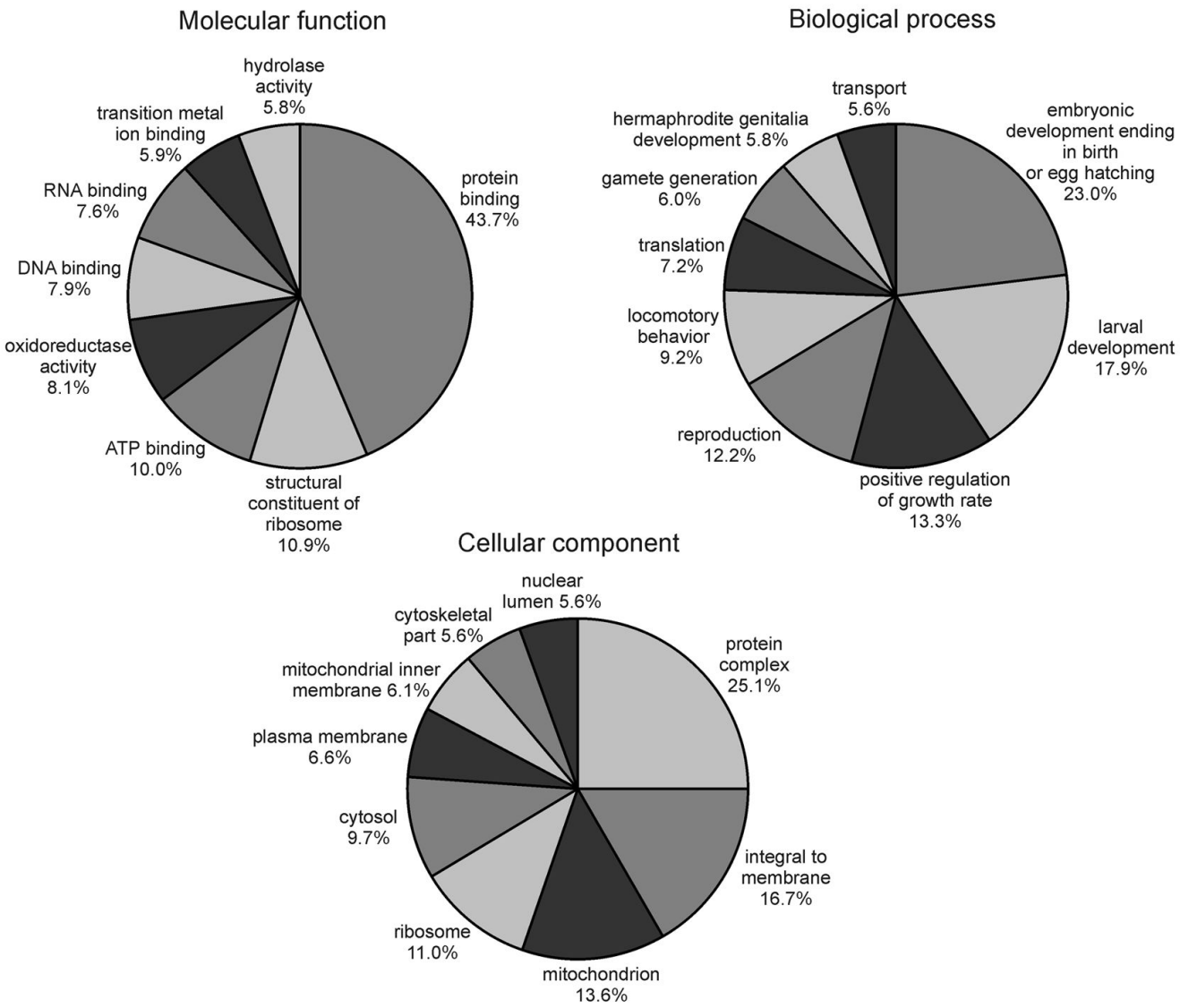

Figure 4.

Representation of gene ontology (GO) mappings for all unigenes as calculated by BLAST2GO. Different pie charts are given for the terms molecular function, biological process and cellular component with a node score cut-off of 25, 50 and 12 respectively.

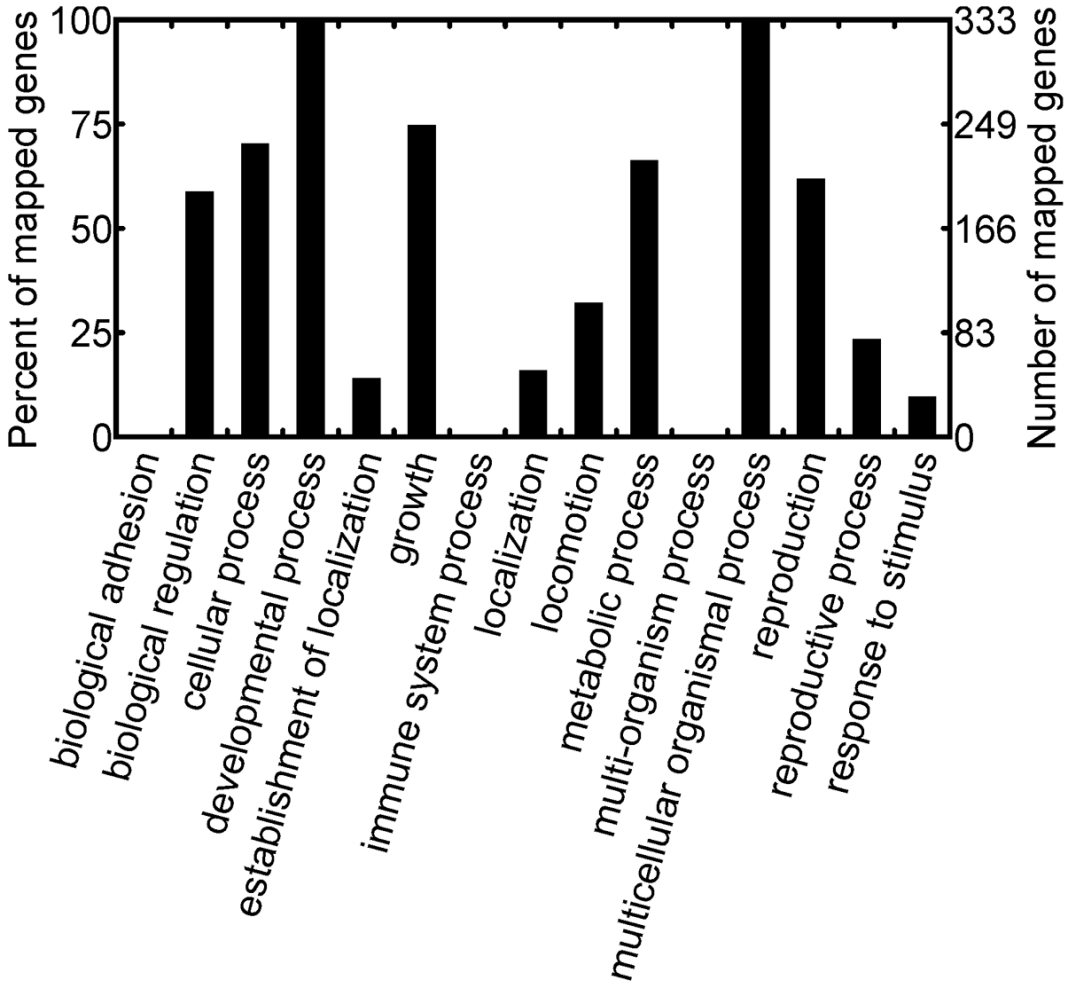

Figure 5.
Graphical representation of the occurrence of GO terms of biological process (level 2) among the *C. elegans* homologs of *Ditylenchus africanus* unigenes with a lethal RNAi phenotype

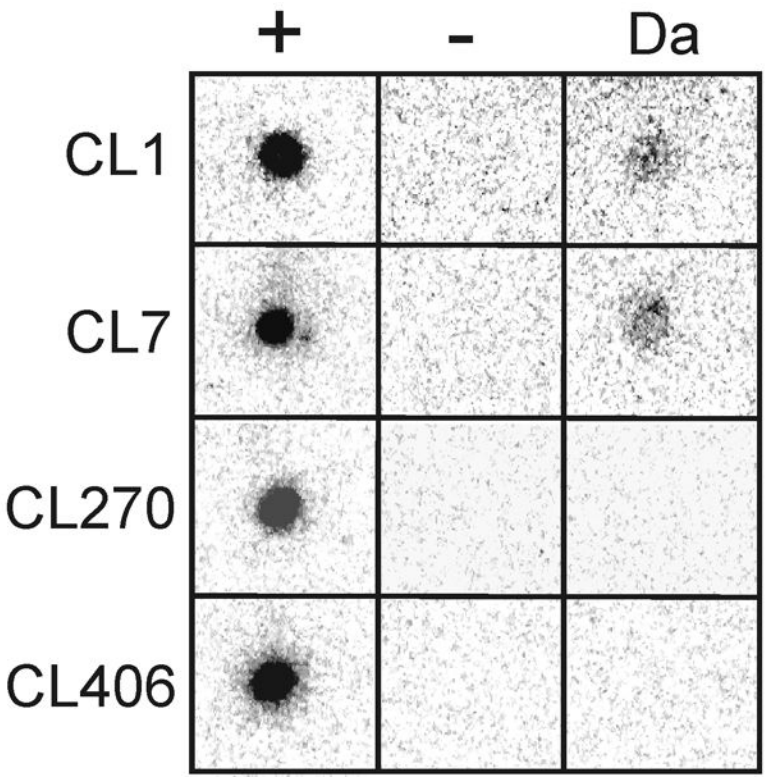

Figure 6.
Dot blot analysis with probes against the selected unigenes CL1, CL7, CL270 and CL406. (+) PCR fragment of the corresponding probe; (−) carrot RNA; (Da) RNA isolated from mixed stages of *Ditylenchus africanus*

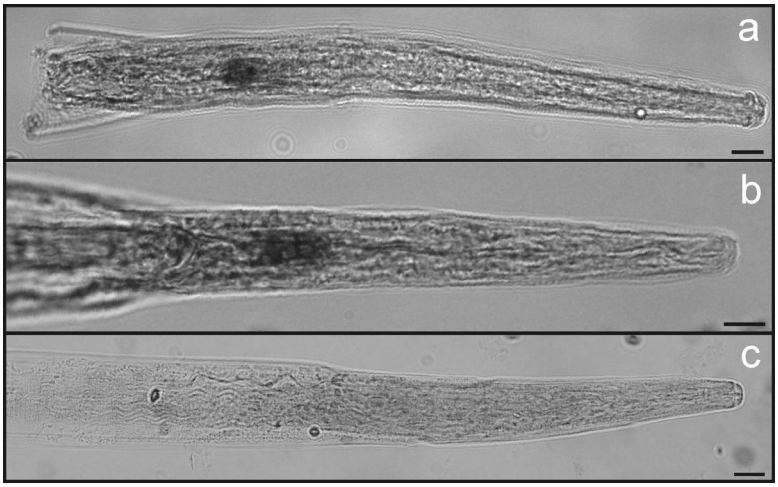

Figure 7.
In situ hybridization on *Ditylenchus africanus*. (a) CL496 (expansin-like protein) antisense probe; (b) **FE920532** (venom allergen protein) antisense probe; (c) CL496 sense probe (negative control). Scale bars: 15 μm.

Table 1

Primers used for dot blot analysis and *in situ* hybridization

Primer	Primer sequence	Primer	Primer sequence
CL1-F	TTGTGAATTTCGGCTCACTG	CL1-R	AAGTTTTTCCTGCCGAGTTG
CL7-F	CAGCAGACTACTGGGCATCA	CL7-R	TTCCTTCGGCTGGAATAATG
CL14-F	TCGATATGATCTGCGAGCTG	CL14-R	GCGGATGATTATGACGAGGT
CL270-F	TCTTCAGGTGATGGTCGATG	CL270-R	TTGTTCCATATAAATAATCTTTGCATT
CL371-F	TGGGTTGTTGTGAAGGTCAA	CL371-R	TCCGTTTTCTTTTGACACTCTG
CL406-F	CACCAGGTCCAGCCATTAGT	CL406-R	AAGCTGGACCGGAAGAAGAT
CL496-F	GTTTCCCGTGGCTAACAAAT	CL496-R	CATCCAGAATTTTGTTCACTGC
FE920532F	CAACCAACCTGTCTACCAGAGG	FE920532R	TCAGCCTGATCCGATAGTGCAAG
FE921742F	AATTAACCGGGGTTGGAAAA	FE921742R	GAAAACCGGTTCGAAGGTG
FE922861F	CAAGACTGACTACGGCCACA	FE922861R	CCATAGTCCTCACCGTGCTT

Table 2

The 20 most prevalent unigenes in the *Ditylenchus africanus* EST dataset

CL	ESTs	Top hit species and descriptor	Accession	E-value	db	%id	%GC
1	84	no hit (longest ORF: 159 bp)					32.9
2	74	*Meloidogyne incognita* major sperm protein homolog	**BM880927**	6.00E-79	EST	91.9	51.5
3	22	*Drosophila melanogaster* IP15837p	**ABC86319**	1.00E-07	Prot	54.2	39.4
4	21	*Pratylenchus vulnus* ferritin homolog	**CV200280**	4.00E-33	EST	53.8	45.1
5	20	*Ascaris suum* translation elongation factor homolog	**CB014976**	1.00E-141	EST	89.8	51.0
6	18	*Heligmosomoides polygyrus* tropomyosin homolog	**EU131541**	1.00E-100	Nuc	92.4	51.4
7	15	no hit (longest ORF: 996 bp)					36.5
8	14	*Meloidogyne chitwoodi* putative expansin-like protein homolog	**CB831016**	6.00E-16	EST	46.4	43.1
9	14	*Caenorhabditis elegans* D1086.9	**NP_001023754**	3.00E-05	Prot	34.1	35.7
10	14	*Zeldia punctata* fatty acid and retinol binding protein homolog	**AW783768**	4.00E-72	EST	76.3	46.4
11	13	*Pratylenchus penetrans* LEA5 protein homolog	**BQ627245**	1.00E-21	EST	55.4	47.8
12	13	*Anisakis simplex* hypothetical protein homolog	**EH005299**	6.00E-41	EST	38.4	42.4
13	12	*Globodera rostochiensis* hypothetical protein homolog	**BM356077**	1.00E-35	EST	49.2	43.1
14	12	no hit (longest ORF: 204 bp)					43.8
15	12	no hit (longest ORF: 114 bp)					53.5
16	11	no hit (longest ORF: 171 bp)					40.7
17	10	no hit (longest ORF: 72 bp)					41.9
18	10	no hit (longest ORF: 213 bp)					35.0
19	9	*Ascaris suum* translationally controlled tumor protein homolog	**CB039336**	8.00E-85	EST	75.8	44.4
20	9	no hit (longest ORF: 204 bp)					39.2

From left to right: cluster number, number of ESTs the cluster contains, best blast hit description (either derived from blastx against the Protein division of GenBank with cutoff 1e-3, blastn against the Nucleotide division of GenBank with cut-off 1e-5 or tblastx against all nematode ESTs with cut-off 1e-3), accession number, blast E-value, database of the top blast hit (Protein, Nucleotide or EST), percentage protein identity and GC content

Table 3

Some selected unigenes based on homology results

Unigene	Top hit species and descriptor	Accession	E-value	db	%id
Anhydrobiosis					
CL23	*Aphelenchus avenae* LEA1	**Q95V77**	6.00E-23	Prot	48.4
CL11	*Pratylenchus penetrans* LEA5 homolog	**BQ627245**	1.00E-21	EST	55.4
CL79	*Pratylenchus vulnus* fatty acid desaturase (fat-6) homolog	**EL890688**	1.00E-83	EST	77.4
FE922042	*Haemonchus contortus* C-type lectin homolog	**CB333334**	6.00E-10	EST	35.3
FE921359	*Caenorhabditis elegans* C-type lectin (clec-53)	**NP_491247**	4.00E-04	Prot	28.2
CL475	*Globodera rostochiensis* stomatin (sto-5) homolog	**EE268121**	4.00E-51	EST	74.8
FE922269	*Aphelenchus avenae* trehalose 6-phosphate synthase homolog	**AJ811572**	1.00E-11	Nuc	75.6
Oxidative stress					
FE922062	*Wuchereria bancrofti* Cu/Zn-superoxide dismutase homolog	**CK726411**	7.00E-07	EST	54.8
FE925119	*Bursaphelenchus xylophilus* glutathione reductase homolog	**CJ987986**	3.00E-45	EST	75.9
CL450	*Globodera rostochiensis* glutathione S-transferase homolog	**BM344760**	1.00E-18	EST	36.8
FE924154	*Meloidogyne incognita* glutathione S-transferase homolog	**CK983784**	4.00E-23	EST	42.4
CL500	*Dictyocaulus viviparus* peroxiredoxin homolog	**EV850422**	5.00E-28	EST	87.0
CL130	*Heterodera glycines* peroxiredoxin homolog	**CA940959**	4.00E-98	EST	75.4
CL452	*Ancylostoma caninum* superoxide dismutase	**EX542471**	6.00E-13	EST	61.4
CL226	*Chlamydomonas reinhardtii* superoxide dismutase	**XM_001699025**	1.00E-21	Nuc	83.3
CL21	*Haemonchus contortus* superoxide dismutase	**CB018828**	8.00E-41	EST	74.7
Allergens					
CL752	*Toxoascaris leonina* ABA-1 allergen homolog	**ES880608**	6.00E-10	EST	54.0
FE923927	*Meloidogyne paranaensis* polyprotein allergen homolog	**CN477518**	5.00E-11	EST	42.9
CL43	*Globodera pallida* major allergen homolog	**BM415278**	3.00E-33	EST	52.6
FE923357	*Heterodera glycines* vap-1	**AAK60209**	5.00E-23	Prot	61.7
Putative parasitism genes					
CL10	*Zeldia punctata* SEC-2 protein homolog	**AW783768**	4.00E-72	EST	76.3
FE922606	*Pratylenchus vulnus* transthyretin-like protein homolog	**EL889277**	2.00E-20	EST	64.9
CL678	*Globodera rostochiensis* transthyretin-like protein homolog	**BM355233**	5.00E-10	EST	80.0

Unigene	Top hit species and descriptor	Accession	E-value	db	%id
CL536	*Meloidogyne chitwoodi* transthyretin-like protein homolog	**CD682816**	5.00E-52	EST	72.6
CL112	*Meloidogyne chitwoodi* transthyretin-like protein homolog	**CF801754**	6.00E-10	EST	48.3
CL665	*Haemonchus contortus* calreticulin homolog	**AI723603**	2.00E-69	EST	86.4
CL573	*Strongyloides ratti* 14-3-3b protein homolog	**FC819840**	1.00E-70	EST	57.5
CL301	*Globodera rostochiensis* 14-3-3b protein homolog	**EE267463**	1.00E-103	EST	95.8
CL8	*Meloidogyne chitwoodi* putative expansin-like protein homolog	**CB831016**	6.00E-16	EST	46.4
CL496	*Bursaphelenchus xylophilus* expansin-like protein homolog	**CJ981766**	4.00E-26	EST	57.5

From left to right: best blast hit description (either derived from blastx with cut-off 1e-3, blastn with cut-off 1e-5 or tblastx against all nematode ESTs with cut-off 1e-3), accession number, blast E-value, database of the top blast hit (Protein, Nucleotide or EST) and percentage protein identity.

Table 4

Novel candidate parasitism genes

Unigene	EST hits	Top hit species	Accession	E-value	%id	AA	SP	start M?
CL454	PN	*Haemonchus contortus*	CB012354	2.00E-09	35.8	173	17	yes
CL578	APN	*Brugia malayi*	BE758361	4.00E-04	36.2	84	22	yes
FE922893	APN	*Ascaris suum*	CB014479	6.00E-06	51.2	43	18	yes
FE925095	PN	*Ascaris suum*	CB014479	3.00E-06	48.8	72	18	no
FE925011	PN	*Meloidogyne incognita*	BQ519557	1.00E-07	62.3	71	24	yes
FE924589	PPN	*Heterodera glycines*	CB299361	6.00E-07	37.7	139	16	yes
FE922853	PPN	*Bursaphelenchus xylophilus*	CJ981856	2.00E-04	40.0	108	23	yes

Putative proteins of these unigenes have a predicted signal peptide and tblastn searches against nematode ESTs (E<1e-4) reveal hits with parasitic nematodes only (PPN: EST hits in plant-parasitic nematodes only, APN: EST hits in animal-parasitic nematodes only, PN: EST hits in both animal- and plant-parasitic nematodes). The top hit species, accession number, blast E-value, percentage protein identity, the number of amino acids, the length of the putative signal peptide (SP) and the presence of a start methionine and are given.