Plant-parasitic nematodes

From genomics to functional analysis of parasitism genes

Joachim Jacob

UNIVERSITEIT GENT

FACULTY OF **BIOSCIENCE ENGINEERING**

Plant-parasitic nematodes:

from genomics to functional analysis

of parasitism genes

ir. Joachim Jacob

Promotor: Prof. dr. Godelieve Gheysen

Lab. Toegepaste Moleculaire Genetica
Dept. of Molecular Biotechnology
Faculty of Bioscience Engineering
Ghent University

Thesis submitted in fulfilment of the requirements
for the degree of Doctor (PhD) in Applied Biological Sciences

Dutch translation of the title:

Plantenparasitaire nematoden: van genoomanalyse tot functionele analyse van parasitismegenen

Cover: View of *Radopholus similis* nematodes through a binocular stereomicroscope. The overlaid words are generated by 'Wordle' [http://www.wordle.net/] based on the complete text of this thesis. Wordle bases the size of the words on the number of times they appear in the text. Every chapter starts with such 'word clouds', giving an overview of the key words.

Jacob, J (2009) Plant-parasitic nematodes: from genomics to functional analysis of parasitism genes. PhD thesis, Ghent University, Ghent, Belgium.

Jury Members

Promotor

Prof. dr. Godelieve Gheysen,
Dept. of Molecular Biotechnology,
Fac. Bioscience Engineering, Ghent University,
Gent, Belgium.

Promotion commission

Prof. dr. ir. Monica Höfte
Dept. of Crop Protection,
Fac. Bioscience Engineering, Ghent University,
Gent, Belgium.

Prof. dr. John Jones
Dept. of Plant Pathology,
Scottish Crop Research Institute,
Invergowrie, Dundee, Scotland, United Kingdom.

Dr. ir. Bartel Vanholme
Dept. of Plant Systems Biology,
VIB/Ghent University,
Gent, Belgium.

Dr. ir. Thomas Van Leeuwen
Dept. of Crop Protection,
Fac. Bioscience Engineering, Ghent University,
Gent, Belgium.

Prof. dr. Yves Van de Peer
Dept. of Bioinformatics & Evolutionary Genomics,
VIB/Ghent University,
Gent, Belgium.

Content

Abbreviations

AA	amino acid	NADH	nicotinamide adenine dinucleotide
AI	ambiguous intermediate model	NCR	non-coding region
ANOVA	analysis of variance	NEXT	nematode EST exploration tool
APN	animal-parasitic nematode	NFS	nematode feeding site
BLAST	basic local alignment search tool	MPSS	massive parallel signature sequencing
bp	base pairs		
CD model	codon disappearance model	nt	nucleotides
cDNA	complementary DNA	ORF	open reading frame
CM	chorismate mutase	PAL	phenylalanine ammonium lyase
COX	cytochrome oxidase	PCA	principal component analysis
Ct	threshold cycle	PC	principal component
CTAB	Cetyl trimethylammonium bromide	PCG	protein encoding genes
		PCR	polymerase chain reaction
CT-RT-PCR	circularization treatment-RT-PCR	PPN	plant-parasitic nematode
Cys	cysteine	Q-PCR	quantitative PCR
DAPI	4',6-diamidino-2-phenylindole	RI	reference index
DEPC	diethylpyrocarbonate	RIrER	reference index relative expression ratio
DIG	digoxigenin		
DMT	5-methoxy-N,N-dimethyltryptamine	RNA	ribonucleic acid
		RNAi	RNA interference
DNA	deoxyribonucleic acid	ROS	reactive oxygen species
DPBA	diphenyl-boric acid	rRNA	ribosomal RNA
dpi	days post infection	RSCU	relative synonymous codon usage
dsRNA	double stranded RNA	RT	room temperature
DUF	domain of unknown function	RT-PCR	reverse transcriptase - PCR
EF-Tu	elongation factor thermo-unstable	SA	salicylic acid
		SAGE	serial analysis of gene expression
EPM	ESTs per million	SDS	sodium dodecyl sulphate
EST	expressed sequence tag	SL	spliced leader
ET	ethylene	SOI	sample of interest
FLN	free-living nematode	SSC	saline-sodium citrate
FLP	FMRFamide-like peptide	TAE	tris-acetate-EDTA buffer
GA	gene arrangement	TAL	tyrosine ammonium lyase
GFP	green fluorescent protein	TBE	tris-borate-EDTA buffer
GO	gene ontology	tRNA	transfer RNA
GOI	gene of interest	TRP	transthyretin-related protein
HMM	hidden Markov model	TT	transparent testa
IAA	indole acetic acid	TTL	transthyretin-like
J	juvenile	TTR	transthyretin
JA	jasmonic acid	Tyr	tyrosine
kb	kilobase	UTR	untranslated region
kDa	kilodalton	WT	wild type
LMP	low melting point	X-gluc	5-bromo-4-chloro-3-indolyl-beta-D- glucuronic acid
mRNA	messenger RNA		
mt	mitochondrion		

Scope

This doctoral thesis contributes to the ongoing efforts to characterize important plant-parasitic nematodes at the molecular level and to unravel the parasitic interaction with their host plants.

The thesis focuses mainly on the nematode Radopholus similis*, an important parasite of banana. Since hardly any molecular data was available for this species, the generation and analysis of 'expressed sequence tags' (ESTs) was a starting point to explore various aspects of the transcriptome (chapter 2). Through extensive bioinformatic analyses, diverse interesting parasitism genes candidates were identified and some of them were further examined. An analysis of transthyretin-like genes in* R. similis *is described in chapter 3. This large nematode-specific gene family is often linked to nematode parasitism, and is in this chapter further characterized using various approaches.*

The analysis of the EST data provided also a glimpse of the unusual mitochondrial genome of R. similis*, and we took the chance to completely sequence and annotate this genome (chapter 4). The genetic code of this genome is slightly changed due to a codon reassignment, and based on comparisons with other nematode mitochondrial genomes, an evolutionary model for this reassignment was constructed (chapter 5).*

Although the majority of nematode sequences available in the databases today consist of ESTs, much of the information contained in ESTs is not readily accessible. Building on our experience gained from previous EST analyses, we developed a web-based tool that extracts relevant information contained in EST data, extending its usefulness for the scientific community (chapter 6). In addition, analysis of reference genes in nematode EST data revealed insight in the quantitative nature of EST data.

In contrast to previous chapters, the final chapter reports experiments performed on the plant side in the plant-nematode interaction, the incentive being the intriguing parasitism gene chorismate mutase, identified in plant-parasitic nematode Heterodera schachtii *(chapter 7). This gene plays most likely a role in the early stages of the plant-nematode interaction, and various experiments were performed to gain insight into those early events.*

To conclude, chapter 8 reflects about future perspectives and opportunities in plant-nematode research and control.

Chapter 1

General introduction

"Try to learn something about everything
and everything about something."
T.H. Huxley

General introduction   1

1. Plant-parasitic nematodes

Although they form the largest phylum of multicellular animals on earth, nematodes remain obscure to the majority of the public because most of them are harmless, free-living, microbivorous microscopically small worm-like animals. Occupying every possible ecological niche, including arctic environments, marine habitats and dry desert soils, these roundworms stay below the detection range of most except for those who specifically search for them. But in the course of evolution — presumably in an attempt to exchange a harsh unpredictable environment for more stable conditions — some of these small innocent worms evolved to adapt a parasitic way of life (Holterman et al. 2008). Virtually every multicellular organism existing on earth can be infected by one or more parasitic nematode species. As such, this group of nematodes ran into conflict with human interests, due to the dramatic effects they have on the yield of numerous crops, on the viability of farm animals and on the quality of life for millions of humans. Of the 20,000 described nematode species, some 4,000 species are parasites of plants (Weischer et al. 2000). Recent estimates of total crop losses caused by plant-parasitic nematodes (PPNs) are estimated at $80 billion annually, caused by the direct effects of parasitism as well as the indirect effects of spreading (viral) diseases (Agrios 2005). The PPNs evolved most likely from fungal feeders into plant-parasites with specialized feeding apparatus and life cycle. This evolutionary event has occurred independently at least three times in the phylum of nematodes (Blaxter et al. 1998; Holterman et al. 2008). Most of the PPNs of major economic importance belong to the Tylenchida, an order containing nematodes with various plant-parasitic strategies. The two nematodes considered in this thesis belong to this order.

The way by which PPNs infects the host root system forms the basis of the artificial classification of the root parasites into three major groups. Note that a few nematode species attack other plant parts as well, e.g. *Ditylenchus* spp. in stems. Some PPN species are called

Figure 1: Visualization of pharyngeal gland secretions (arrows) of the plant-parasitic nematode *Radopholus similis*. The stylet in the nematode's head is clearly visible. Nematodes were incubated in 400 µg/ml DMT (to stimulate secretion) and 0.005% coomassie G (to visualise secretion).

'ectoparasitic', adopting an assumed primitive mode of parasitism, since they do not enter the plant roots and remain at the outside to withdraw nutrients from the plant. In contrast, other so called 'endoparasitic' species actively enter the root system. Some of them migrate continuously through the roots, sucking out the cytoplasmic content of the root cells they encounter and causing ultimately large necrotic lesions ('migratory species'). Yet other endoparasitic species set up an elaborate long-lasting feeding structure consisting of modified root cells, through which the nematode obtains essential nutrients to complete its life cycle, while exhausting the plant's resources ('sedentary species'). All PPNs are equipped with sophisticated tools to fulfil their task. The most eye-catching is the stylet, a retractable hollow spear-like structure in the nematode's anterior end, controlled by protractor muscles and connected with the digestive system. In endoparasitic species, fierce stylet thrusts assist in migration through the plant root and in puncturing the rigid plant cell wall. Besides this structural adaptation to parasitism, PPNs make use of specialized biochemical tools to establish a successful parasitic interaction with the host plant. The secretions originating from three large unicellular pharyngeal glands are of utmost importance for the parasitism process (Figure 1). In Tylenchida, the glands consist of two subventral glands, which open into the pump chamber, and

one dorsal gland with an orifice at the base of the stylet (Figure 2). Identifying and characterizing the proteins contained in the secretions of these glands can shed light on how the nematode succeeds in parasitizing its host and — in the case of sedentary parasitic species — how the nematode elicits the most complex response in plant tissue by any parasite or pathogen known today, in order to establish a nematode feeding site. The genes coding for these secreted proteins are often termed 'parasitism genes' (Davis et al. 2000). Other tissues such as the epidermis, the rectal glands, the excretory/secretory system and the amphids (sensory organs in the head) also express genes that contribute to the so-called 'parasitome' (Jones et al. 2000; Robertson et al. 2000). During the last 10 years, considerable research efforts have been aimed at elucidating these parasitism genes (mostly of sedentary species) and different classes of parasitism genes are now being recognized (Vanholme et al. 2004). One class comprises genes with roles in host localisation, as proposed for SXP/RAL-2 proteins, which are expressed by diverse tissues as the amphids, epidermis and the subventral gland cells (Jones et al. 2000; Tytgat et al. 2005) (Figure 2). Another class of parasitism genes encodes for enzymes that modify and weaken the plant cell wall in order to facilitate migration through the plant root, such as β-1,4-endoglucanase, pectate lyase, xylanase, polygalacturonase and expansin. In sedentary nematodes, they are mostly expressed in the subventral pharyngeal gland cells, which show greatest activity during the migration phase (De Boer et al. 2002a; Popeijus et al. 2000a; Qin et al. 2004; Smant et al. 1998). Some of these proteins have also been identified in some migratory PPNs (Kikuchi et al. 2004; Uehara et al. 2001). Another class of parasitism genes aids in the feeding site initiation and maintenance, and is hence only identified in sedentary PPN. These genes originate from both pharyngeal gland types, and their products are assumed to be secreted directly into the cytoplasm of a selected plant root cell. Putative examples of these are genes encoding for a secreted chorismate mutase, a secreted ubiquitin-extension protein and a secreted Clavata3/ESR homologue (Tytgat et al. 2004; Vanholme et al. 2008; Wang et al. 2005). It is believed that the majority of the changes observed in the nematode feeding site — being a unique plant tissue based on gene expression profiling (Szakasits et al. 2008) — is due to the activity of secreted parasitism genes within the cytoplasm of the plant root cell. Illustrative for this, some ectoparasitic nematode genera (such as *Xiphinema*) establish a similar nematode feeding site, making solely contact with the host root cells through their exceptionally long stylet, which act as a duct for the secreted parasitism gene products (Wyss et al. 1988). Finally, forming the last class, some parasitism proteins assist in countering the plant defence responses, which is the only method a plant has

to react to infectio. It is believed that many secreted factors belong to this class, acting on many different points of the plant reaction. Examples of these proteins are peroxiredoxins, which are secreted by the epidermis and protect the nematode from reactive oxygen species produced by the plant (Jones et al. 2004; Robertson et al. 2000). The glycocalyx forms a sheet of a complex composition of glycoproteins on the cuticle and alters its composition depending on the life stage. This structure also helps to counter plant defences and to avoid detection by the host plants (Gravato-Nobre et al. 1999). Many identified putative parasitism genes are pioneer sequences (lacking known homologous sequences), consequently with as yet unidentified roles. Excitingly, some of these pioneers are targeted to the nucleus when expressed in plant cells (Elling et al. 2007a; Tytgat et al. 2004). For one of them the potential to interact with plant transcription factors is already proven (Huang et al. 2006b). With the identification of additional parasitism genes and their corresponding functions, this division in four classes should perhaps be extended as recently proposed (Bellafiore et al. 2008). Based on analysis of sensitive mass-spectrometry of secretions originating of the endoparasitic sedentary nematode *Meloidogyne incognita*, the total number of proteins contributing to the parasitism process is estimated to reach 500 (Bellafiore et al. 2008). It is clear that PPNs are very sophisticated parasites, equipped with exceptional tools to successfully parasitize plants. Although the following quote is originally based on results of *Meloidogyne* (which is sometimes seen as the most successful PPN due to its capability to infect over 3,000 plant species), it certainly applies to other PPNs as well: "Taken as a whole, these data reveal a nematode that attacks by stealth and deception, subverting plant host defences, and which carries a highly sophisticated array of weapons." (Bird et al. 2009).
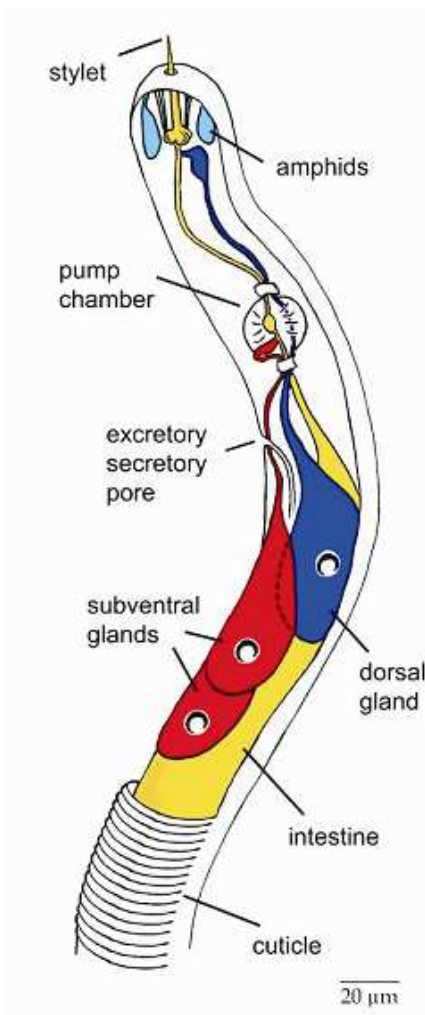

Figure 2: Schematic representation of the anterior region of a plant-parasitic nematode, with indication of body structures and cells important for parasitism (adapted from Vanholme et al. (2004)).

2. The burrowing nematode *Radopholus similis*

As for many important pests and plagues, the globalized world is one of the primary reasons for the widespread occurrence of the migratory plant-parasitic nematode *Radopholus similis* (Price 2006). Probably endogenous to Australasia, where it was first found in 1891 on banana plants in Fiji, *Radopholus* is thought to have entered Carribean and African (mainly Uganda) banana plantations through the import of infested planting material (Trinh et al. 2004). Today, it is an important pest in tropical and subtropical regions, able to infect over 360 plant species including banana, citrus, black pepper and sugar cane, and is viewed as the most costly pest affecting commercial banana plantations (mostly cultivar Cavendish) (Holdeman 1986; Sarah et al. 1996). Although classified as belonging to Tylenchida (class Chromadorea), the precise phylogenetic position of *Radopholus* within this order is still a matter of debate. The latest reports suggest an isolated position as a sister taxon of Hoplolaimidae (Bert et al. 2008). After hatching from the egg as a second stage juvenile (J2) (the developed J1 moults to J2 in the egg), *R. similis* immediately starts searching for a suitable plant host (Figure 3). After localisation and penetration of this host, the nematode will complete its life cycle by moulting to J3 and J4 juveniles, and finally to an adult female or male nematode, depending on the conditions (Trinh et al. 2004). Nematodes are found within the roots, as well as in the surrounding soil, dwelling the area for new plant roots to infect. The life cycle is completed in 18-20 days at 24-26°C. Females produce eggs at an average rate of 3 eggs a day. Strong sexual dimorphism occurs, in which the male is considered non-parasitic. Although amphimixis occurs, males form a minority in the population, the fraction which increases with adverse conditions. Females can also reproduce without the presence of males using a hermaphroditic mode of reproduction. A period of sperm generation (sperm cells are stored in the spermatheca) precedes the oogenesis, which continues for the remainder of the life cycle (Kaplan et al. 2000). All mobile stages of *Radopholus*, except for males, can parasitize a plant host, although juveniles often make use of cavities preformed by adult nematodes to enter an infected root. After entering the host's roots within 1 cm of the root tip, the nematode initially moves intercellularly, aided by thrusting movements of its stylet, causing contiguous cells to be separated. After reaching a depth of several cells, the nematode starts to perforate cell walls in order to start feeding. As observed in other migratory species, the single-cell-feeding can be divided into three different phases. First, the stylet is inserted slowly into the plant cell with increasingly deeper

thrusts. When the cytoplasm of a plant root cell is reached, salivation starts for a few minutes, after which the plant cell content is ingested, probably facilitated through prior action of injected pharyngeal gland secretions (Ferraz et al. 2002; Kurppa et al. 1985; Zunke 1990). The nematode moves in the root tissue by penetration of plant cells, fed upon or not, by boring a series of holes in the wall and forcing its way through. This behaviour causes large necrotic lesions in the cortex, in which secondary infections rapidly take place, mainly by *Fusarium oxysporum* and *Rhizoctonia solani*. The result of the nematode infection is the formation of black necrotic lesions and stunting and wilting of the host plant (known as the 'black head' disease in bananas). In severe cases, the infection often leads to toppling of the complete plant due to the weakened stem base (Figure 4). These effects can cause massive losses in crop production ranging from 5% to 75% (Ducharme 1968; Fogain et al. 1997; O'Bannon 1977; Price 2006; Sarah et al. 1996).

LIFE CYCLE STAGES of BURROWING NEMATODE (about 22 days) egg to adult

MATURE FEMALE WITH EGGS
MATURE MALE
NEWLY LAID UNDIVIDED EGG
2-CELL STAGE
3-CELL STAGE
4-CELL STAGE
8-CELL STAGE
16-CELL STAGE
EARLY TADPOLE STAGE

THIRD STAGE LARVA
SECOND STAGE LARVA (hatching from egg)
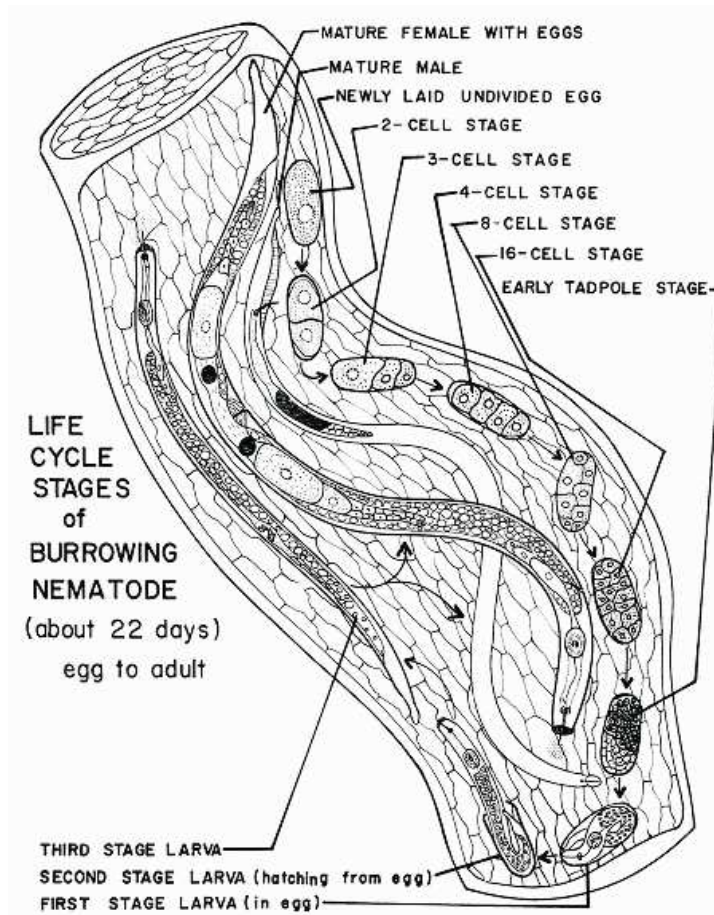FIRST STAGE LARVA (in egg)

Figure 3: Drawing of the life cycle of *Radopholus similis* (taken from Esser (1962)).

Once established in the field, it is very hard to nearly impossible to totally eradicate *R. similis*. Therefore, the foremost important measure against *Radopholus* infection is prevention of infection through the use of nematode-free or intensively cleaned planting material. The only efficient chemical nematode control is the use of e.g. methylbromide, organophosphates or carbamates. These coumpounds are extremely hazardous for the environment and banned for this reason (United Nations Environment Programme 1995). Since only very few resistant varieties are available, new sources of resistance are being sought (Elsen et al. 2004; Kalorizou et al. 2006; Quénéhervé et al. 2009; Stoffelen et al. 2000; Wuyts et al. 2007). However, since banana is propagated vegetatively and resistance can not easily be achieved through crossing, resistance through genetic engineering is the method of choice. Recently, transgenic banana expressing cystatin was proven to confer some level of resistance to *R. similis* infection, but it will take years before the first transgenic bananas reach the market (Atkinson et al. 2004)
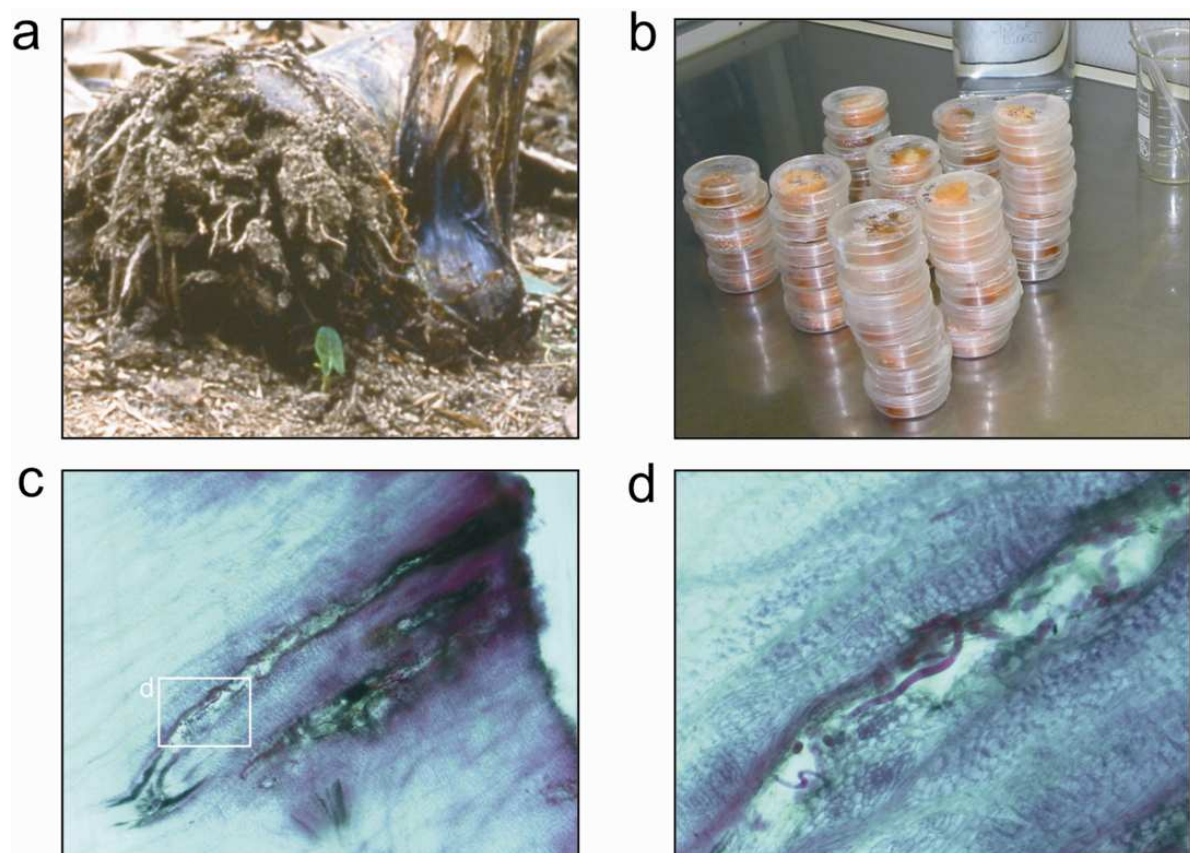

Figure 4: **a**, example of blackhead disease in banana (upright stem), next to a toppled tree (photo taken from http://www.ctahr.hawaii.edu/nelsons/banana/). **b**, culture of *Radopholus similis* on carrot discs. **c**, slices of carrot discs infected with *R. similis*, four weeks after inoculation, showing long stretched lesions. **d**, detail of a lesion (inset of c) in which nematodes of different life stages are visible. Slices of approximately 1 mm were made, stained in 0.2% acid fuchsin and destained with chloral hydrate.

3. The beet cyst nematode *Heterodera schachtii*

The beet cyst nematode *Heterodera schachtii* is a sedentary PPN. Its life cycle differs strongly from that of migratory nematodes. First, the freshly hatched infective J2 locates a suitable plant host, penetrates the epidermal cell layer and migrates intracellularly through the root tissue, assisted by vigorous stylet thrusts (Figure 5). When it arrives in the vicinity of the pericycle of the plant root, its behaviour changes: a single root cell is gently punctured and a cocktail of proteins is secreted, which initiates the transformation of the root cell into a nematode feeding site (NFS). The nematode remains sedentary from this phase on and stays associated with the NFS for the rest of its life. The feeding site is formed by dissolution of cell walls and subsequent incorporation of the neighbouring cells into an expanding syncytium. At regular intervals, the nematode feeds on this syncytium by carefully penetrating the cell and removing part of its contents. The syncytium typically shows high metabolic activity, and has a very high nutrient and energy demand, leading to stunting and wilting of the host plant and a subsequent severe reduction in yield. Importantly, initial observed symptoms are mostly difficult to assign to a nematode infection. During the sedentary phase, the nematode develops over J3 and J4 stages into adulthood. Females and males can be found, contributing to an amphimictic way of reproduction. At the end of their life, the female nematodes swell and become brown due to tanning of their cuticle, forming cysts packed with eggs. In this state, the protected eggs can survive for years in the field, being a potential threat for any host plant germinating in its vicinity.

4. Unraveling the molecular basis for nematode plant-parasitism

Identifying the substances in the secretions that contribute to the parasitism process of PPN nematodes is by no means a trivial task. Most of the PPNs – and especially the endoparasitic species – are inconveniently small worms (roughly 0.5 mm long and 20 μm diameter), a fact which hampers obtaining sufficient material for molecular experiments. In addition, the parasitic stages are buried deep into the host root system, requiring mostly laborious manual
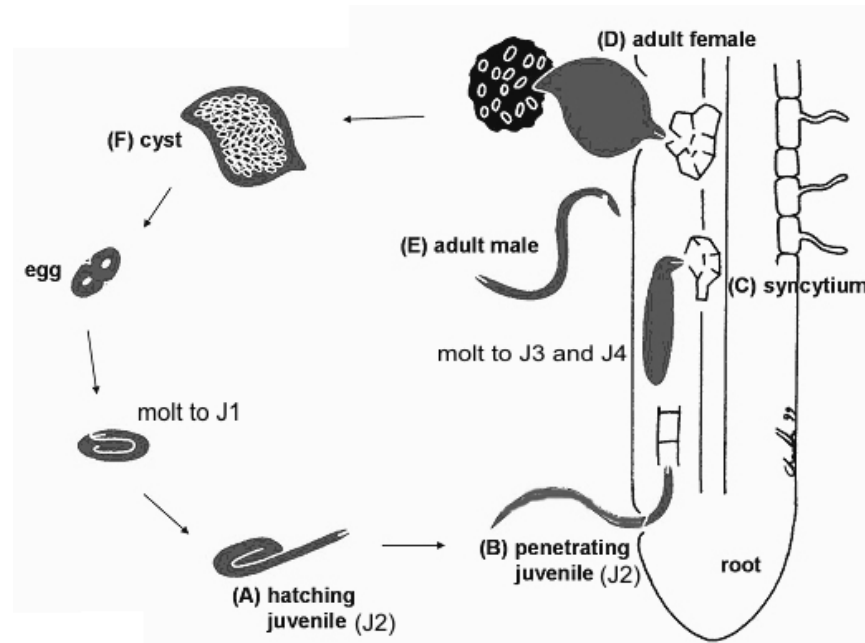
Figure 5: Life cycle of *Heterodera schachtii* (based on drawing of Dirk Charlton, http://plantsci.missouri.edu/mitchumlab/whatare.htm).

dissection of the desired stages. To deal with these major obstacles, various techniques have been applied to achieve identification of the factors controlling parasitism. Two major approaches to identify parasitism proteins and their corresponding genes can be distinguished, which have both proven their value: a 'protein-based' approach for which secretions (as the products of parasitism genes) of the nematode serve as a basis, and a 'sequence-based' approach which starts by analyzing gene expression to assign parasitism genes.

The protein-based approach was the first used. Early enzymatic tests on PPN secretions revealed activities such as cellulase, pectinase and proteinases (Giebel 1973). Antibodies raised against secretory granule proteins from the pharyngeal glands and screening for immunolocalisation to the glands succeeded for the first time in identifying a parasitism gene responsible for the cellulase activity in the secretions (De Boer et al. 1996a; De Boer et al. 1996b; Smant et al. 1998). However, this technique is very laborious and inefficient. When PPNs were found to induce secretions upon application of various chemical compounds (such as the serotonin analogue 5-methoxy-N,N-dimethyltryptamine (DMT) and resorcinol, McClure et al.

(1987)), relatively large amounts of induced secretions of millions of PPNs could be obtained. Two-dimensional polyacrylamide gel electrophoresis of such induced secretions and sequencing of the most abundant protein spots revealed proteins originating from the pharyngeal glands (Jaubert et al. 2002). Due to the advance in mass-spectroscopy techniques, a recent study succeeded in directly identifying proteins from induced secretions of the root-knot nematode *Meloidogyne incognita*. This study reached an unprecedented depth, identifying dozens of novel putative parasitism genes with expression in the pharyngeal gland cells (Bellafiore et al. 2008). Additionally, mass-spectroscopy can reveal also the presence of non-protein substances, such as cytokinins, which were detected in induced secretions of root-knot and cyst nematodes using this approach (De Meutter et al. 2003).

In contrast to the protein-based research, the sequence-based method is the most widely used approach for the identification of parasitism genes in PPNs. The first studies analyzed differential gene expression between cDNA pools constructed of parasitic second stage juveniles and other stages (by 'RNA fingerprinting' and 'cDNA-AFLP'). This has led to identification of some novel parasitism genes (Ding et al. 2000; Qin et al. 2000). To identify genes which contributed to the earlier observed enzymatic activities of the secretions, clones of a whole nematode cDNA library were expressed and functionally screened for expected enzymatic activities. A cellulase enzyme was identified this way (Yan et al. 2001). Furthermore, screening cDNA libraries – more specifically the expressed clones – with antibodies raised against PPN secretions proved another successful approach in the identification of parasitism genes (Fioretti et al. 2001; Prior et al. 2001). An exceptional approach for which cDNA was constructed of total host roots infected with nematodes and subsequently subtracted with non-inoculated host root and nematode egg cDNA, yielded also a few novel parasitism genes expressed in the gland cells (Tucker et al. 2005). It was recognized that cDNA libraries enriched in genes specifically expressed in the gland cells would be of more value. In a first attempt, a cDNA library from the gland cell region was screened for differential hybridisation with probes originating from the tail region, and resulted in the identification of the nematode parasitism gene chorismate mutase (Lambert et al. 1999). With advancing technology, specific cDNA libraries were constructed of microaspirated gland cell contents. Gao et al. (2001b) pioneered herein by constructing a gland-cell specific cDNA library combined with suppression subtractive hybridisation using intestinal cDNA. Further optimization of normalisation of gland cell-specific libraries was achieved through the use of secretion signal

selection (Wang et al. 2001), differential hybridizing the cDNA library on a macro-array with intestinal cDNA (Gao et al. 2003; Huang et al. 2003), and the use of solid-phase subtractive hybridisation using intestinal cDNA (Huang et al. 2004). Obtaining 'expressed sequence tags' (ESTs) from these and other nematode cDNA libraries contribute to an ever growing pool of transcriptomic data available in databases. Thorough comparisons of these EST data and directing bioinformatic searches based on characteristics of parasitism genes and expected functions achieved further identification of parasitism genes. (Jones et al. 2004; Jones et al. 2003; Popeijus et al. 2000b; Roze et al. 2008; Vanholme et al. 2006).

Combining the different techniques and data sources is the most promising direction for future identification of additional parasitome members. The availability of *Heterodera glycines* (soybean cyst nematode) micro-arrays (De Boer et al. 2002b; Elling et al. 2007b) allows measuring transcript abundance of different life stages, and combined with an extensive EST analysis novel putative parasitism genes were identified (Elling et al. 2009). In virtually all cases, however, combined prediction of signal peptide for secretion with confirmed expression in the gland cells remains the litmus test for a gene to be assigned a parasitism gene candidate. Additionally, its presence in (induced) secretions can be confirmed (Vanholme et al. 2005). Soon it became apparent that cataloguing sequences of parasitism genes alone would not improve our understanding of the parasitism process, since many parasitism genes lack homologous sequences (the so called 'pioneers'). However, some parasitism genes of PPNs bear homology to genes previously thought to be restricted to other organisms, such as (plant-pathogenic) bacteria and fungi, and animal-parasitic nematodes, and even plants (Bird et al. 2009; Olsen et al. 2003). This reflects the unique and diverse molecular tools PPN use for successful parasitism. Once a parasitism gene is discovered in one PPN species, homologues are searched in other PPNs, based on hybridisation-based screening of cDNA libraries, PCR amplification with degenerate primers on cDNA pools, and searching homologous genes in EST datasets (De Boer et al. 2002a; Gao et al. 2001a; Goellner et al. 2000; Mitreva et al. 2004; Rosso et al. 1999).

It must be noted that all of the abovementioned studies have been performed on sedentary nematodes, as they cause the greatest economical losses and have a scientifically very interesting parasitic behaviour. Little parasitism gene research has focussed on PPN species with other types of parasitic behaviour (Furlanetto et al. 2005; Kikuchi et al. 2005; Mitreva et al. 2004; Uehara et al. 2001). However, the accumulated knowledge and insights from

parasitism genes identified in sedentary PPN provide an ideal basis for the identification of parasitism genes in migratory PPN (Opperman et al. 1998). Additionally, exploring the occurrence of parasitism genes in other PPN species can significantly aid in the second step in parasitism gene research, elucidation of parasitism gene function. To understand the role of a nematode parasitism gene, combination of research in nematode and plant biology is indispensable, since PPNs live in such intimate contact with their plant host. Interestingly, this intimate contact is partly reflected in the occurrence of plant-specific domains in some parasitism genes (Bellafiore et al. 2008). The many novel sequences among the parasitism genes makes the elucidation of the function a most challenging task, which – currently – needs to be dealt with gene by gene (Bakhetia 2005; Bakhetia et al. 2007; Bakhetia et al. 2008; Elling et al. 2007a; Huang et al. 2006b; Rehman et al. 2009; Wang et al. 2005).

In conclusion, research in the molecular aspects of the parasitic interaction between PPN and their hosts has progressed enormously during the last decade. The general processes by which PPNs parasitize their hosts are becoming clear, but numerous issues still remain unsolved. The majority of the research results are obtained from only three genera of sedentary plant-parasitic nematodes, namely the root-knot nematodes *Meloidogyne*, and the cyst nematodes *Heterodera* and *Globodera*. The present thesis reports on the results of chorismate mutase, a parasitism gene specific for sedentary PPN. But the main focus lies in extending the research field to the migratory nematodes by analysis of *Radopholus similis* and reporting different aspects including – but not limited to – the parasitism process, using a sequence-based approach.

Chapter 2

Exploring the transcriptome of the
burrowing nematode
Radopholus similis

2

Parts of this chapter are published in:

Jacob J, Mitreva M, Vanholme B, Gheysen G (2008) Exploring the transcriptome of the burrowing nematode *Radopholus similis*. Mol Genet Genomics 280 : 1-17

Haegeman A, Vanholme B, Jacob J, Vandekerckhove T, Claeys M, Borgonie G, Gheysen G (2009) An endosymbiotic bacterium in a plant-parasitic nematode: Member of a new *Wolbachia* supergroup. Int J Parasitol 39:1045-1054.

Exploring the transcriptome of the burrowing nematode *Radopholus similis*

2

1. Abstract

Unravelling the transcriptome of parasites can provide insight into the parasitism process and lead to more efficient control measures. For the first high throughput molecular characterization of the plant-parasite *Radopholus similis*, 5,853 expressed sequence tags from a mixed stage population were generated. For subsequent analysis, 1,154 tags from the EST division of GenBank were added, resulting in a total of 7,007 ESTs representing approximately 3,200 genes. The mean GC content of the nucleotides at the third codon position was calculated to be as high as 64.8%, the highest for nematodes reported to date. Remarkably, no evidence was found for the presence of spliced leader sequences commonly occurring in nematodes, despite the use of various approaches. BLAST searches resulted in about 70% of the unigenes having homology to (DNA and protein) sequences from the GenBank database. Approximately one quarter of the unigenes seems to be derived from housekeeping genes, whereas one third did not match to any known sequence. Nevertheless, roughly 40% of the latter are predicted to be coding, pointing to putative novel protein encoding genes. Functional annotation of the sequences by GO annotation and KEGG biochemical pathway mapping revealed an abundance of genes involved in reproduction and development. Most interestingly, tags derived from an endosymbiotic *Wolbachia* were found. This is the first time this endosymbiont has been discovered in a plant-parasitic nematode, and subsequent microscopic observations established its presence in the ovaria. The parasitic life style of this nematode is reflected in the presence of tags derived from genes with a putative role in parasitism, such as genes coding for cell wall degrading enzymes, and proteins involved in detoxification of reactive oxygen species and host recognition. In addition, several unigenes had homology to parasitism genes with unknown function of other parasitic nematode species. One of the identified novel parasitism genes is expressed in the pharyngeal gland cells.

2. Introduction

The generation of expressed sequence tags (EST) is a very cost-effective method to generate large amounts of transcriptomic data. In short, the technique consists of construction of a cDNA library, and single-pass sequencing of randomly picked clones. However, the sequences typically are of relatively low quality as a number of biases are introduced during the complex process (further explained in chapter 6). A plethora of bioinformatic tools for standard EST analysis currently exists (Nagaraj et al. 2007). Accompanied by a thorough understanding of the generation of EST sequences, these tools can deliver valuable information. Information contained in EST sequences is useful for obtaining an impression of the molecular composition of species (McCarter et al. 2003), for identifying tissue, developmental stage or organism specific genes (Chen et al. 2006; Dubreuil et al. 2007), for estimating the level of gene expression (digital northern) (Liu et al. 2006a; Munoz et al. 2004), for annotating genes for genome analysis (Blumenthal et al. 2002; Yan et al. 2005) and for facilitating proteome analysis (Bellafiore et al. 2008; Liu et al. 2006b). To support the on-going research on *Radopholus similis*, we generated and analyzed EST sequences from mixed stages in order to gain a first insight into the transcriptome of *R. similis*.

3. Results

a. Dataset characteristics

A total of 5,853 new EST sequences were obtained, having a slightly higher average sequence length compared to the ESTs of *R. similis* already deposited in the dbEST division of GenBank (deposited by Irina Ronko and Dr. Makedonka Mitreva, see table 1). Analytical processing of both sets combined (removal of vector sequences, poly(A) tails and sequences < 100 bp) resulted in 6,800 ESTs, and subsequent clustering (merging overlapping sequences together into 'contigs') established a 13% increase in sequence length. This final set of unigenes contains 1,008 contigs, grouped into 989 clusters (enclosing sequences with minor sequence variations), and 2,659 'singletons' (non-redundant EST sequences). With growing cluster size (i.e. the number of ESTs contained in a cluster), the number of clusters decreases logarithmically

(Figure 6). A certain degree of 'fragmentation' - also called underclustering (the extent to which ESTs belonging to the same gene are not clustered) - could be expected in our final dataset and was estimated by ESTstat to be as high as 15.8% (Wang et al. 2006). Another method as described by Mitreva and associates (2004a) resulted in a comparable estimation of 12,9%. Due to this fragmentation error, our dataset represents at most 3,194 genes, which is approximately 16% of the total gene number, if the genome is assumed similar to *C. elegans* (C. elegans Genome Sequencing Consortium 1998).

Table 1: Dataset characteristics on DNA and protein level

Datasets	No. of sequences	Relative No.*	Length ± Standard Deviation**	Remarks
Starting ESTs	7,007	1.92	394±157 nt	Dataset for processing
Internal ESTs	5,853	1.60	396±154 nt	New sequences
External ESTs	1,154	0.32	386±170 nt	Retrieved from dbEST
Processed ESTs	6,800	1.85	395±157 nt	Removal vector, polyA etc.
Unigenes	3,667	1.00	449±195 nt	Result of clustering
Contigs	1,008	0.28	565±234 nt	989 clusters, from 4141 processed ESTs (61% of ESTs)
Singletons	2,659	0.73	405±158 nt	Not clustered processed ESTs (39% of ESTs)
Translations	2,755	0.75	118±59 aa	Obtained protein translations
Prot4EST	510	0.14	129±57 aa	BLASTx homology, but no FrameD translation
FrameD	2,245	0.61	111±62 aa	FrameD translation prediction
5' truncated	1,040	0.28	106±51 aa	Containing 5' truncated coding part and 3' UTR
3' truncated	396	0.11	131±62 aa	Containing 5' UTR and 3' truncated coding part
Full length	431	0.12	111±55 aa	Containing full length coding sequence
Internal part	378	0.10	154±58 aa	Containing internal piece of coding sequence

* Relative number to the 'unigene' dataset (set as '1') on which the analysis was done.
** nt: nucleotide; aa: amino acid

b. BLASTx analysis

A BLASTx search against the GenBank non-redundant protein sequences resulted in hits for 2,130 unigenes (58.1%; unless stated otherwise, all percentages are relative to the total unigene set). An overview of the extensive analysis is shown in Figure 7. Of these, 1,710 (46.6%) had an E-value lower than 1E-05 (more significant match), whereas the remaining 420 sequences (11.5%) were only retained with higher E-values (between 1E-05 and 1E-01). Ribosomal proteins were of the most abundant top-hits (n=176 or 4.8% of the unigenes). 535 unigenes (14.6%) matched sequences originating from both eukaryotes and prokaryotes, and 622 unigenes (17.0%) matched solely to sequences from all major eukaryotic lineages. The wide occurrence of these unigenes suggests a role in basal cell metabolism. Surprisingly, 14 unigenes gave a plant-specific hit. Since *Radopholus similis* was cultured on carrot disks (*Daucus carota*), the
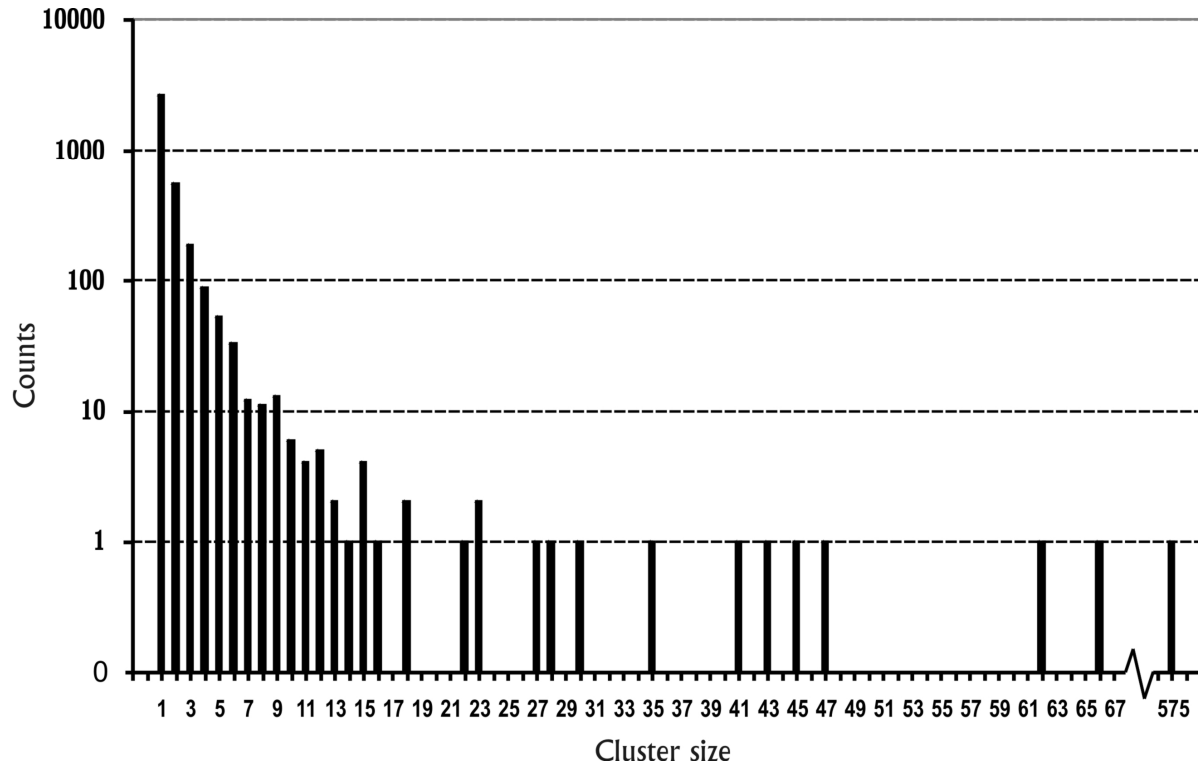
Figure 6. Graphical representation of the cluster size distribution.

presence of contaminating plant tissue could explain these sequences, although remarkably the top-hit sequences originated from different plant species (with an E-value range between 1E-10 and 0.1). Of the remaining unigenes with a BLASTx-hit, 428 exclusively matched animal sequences (11.8%), of which 328 (8.9%) were nematode-specific. Seven of the nematode-specific unigenes were found to match exclusively sequences of plant-parasitic nematodes and 8 matched both plant- and animal-parasitic nematode sequences. Special attention was paid to the largest clusters, as they correspond most likely to highly expressed genes in *R. similis*. The BLASTx results of the largest clusters are reported in Table 2: commonly known highly expressed genes are found (such as actin, sec-2), but some pioneer or unknown sequences are present as well. A considerable subset of our unigenes (n=1,537; 41.9%) gave no BLASTx-hits with an E-value cut-off set as high as 1E-01. One striking feature of these unigenes is their shorter average sequence length (354 bp ± 166) compared to the unigenes with matches (518 bp ± 185) (p-value two sample T-test < 0.001) (Figure 7). For many of these unigenes the short sequence length is the cause for lacking BLASTx homology, as their E-values will not reach the preset threshold.

c.   Homologues in nematode ESTs

To find homologues in the transcriptional data of other nematodes (ESTs of all nematodes excluding *R. similis*), a tBLASTx-search was performed (E-value cut-off of 1E-05). The E-value cut-off for the tBLASTx-search was set lower than the E-value cut-off for the BLASTx-search (i.e. 1E-05 compared to 1E-01), since for the majority of the unigenes a consistently lower top-hit E-value with tBLASTx (i.e. more significant) was found compared to BLASTx (Figure 8). This tBLASTx search revealed 2,305 hits to nematode EST sequences, of which 560 (15.3%) with homologous EST sequences exclusively in plant-parasitic nematode (PPN), 106 unigenes (2.9%) exclusively in animal-parasitic nematodes (APN), and 147 (4.0%) unigenes exclusively in both APN and PPN. As seen for the sequences with and without BLASTx-hit, a comparable difference in sequence length was observed between unigenes with and without tBLASTx-hit: unigenes having homologues in the nematode ESTs are generally longer (408 bp ± 169), compared to those without homologous counterparts (344bp ± 163) (p-value two sample T-test < 0.005). The persistence of this difference in sequence length, points to an important influence of sequence length in finding homologues based on BLAST searches, arguing for a thorough quality check of the used cDNA library. Furthermore, the tBLASTx-search revealed a large portion of the unigenes (n=408 or 11.1%) without BLASTx homology to known proteins, but with homology to EST sequences of other nematodes (Figure 7). Notably the majority of the PPN specific EST sequences lack a BLASTx-hit (367 of 560 unigenes, or 65.5%). Despite these efforts to identify unigenes on basis of homology, unigenes without hits (either BLASTx or tBLASTx; the so called 'orphans') constitute still a large portion (n=1,128 or 30.8%, see Figure 7).
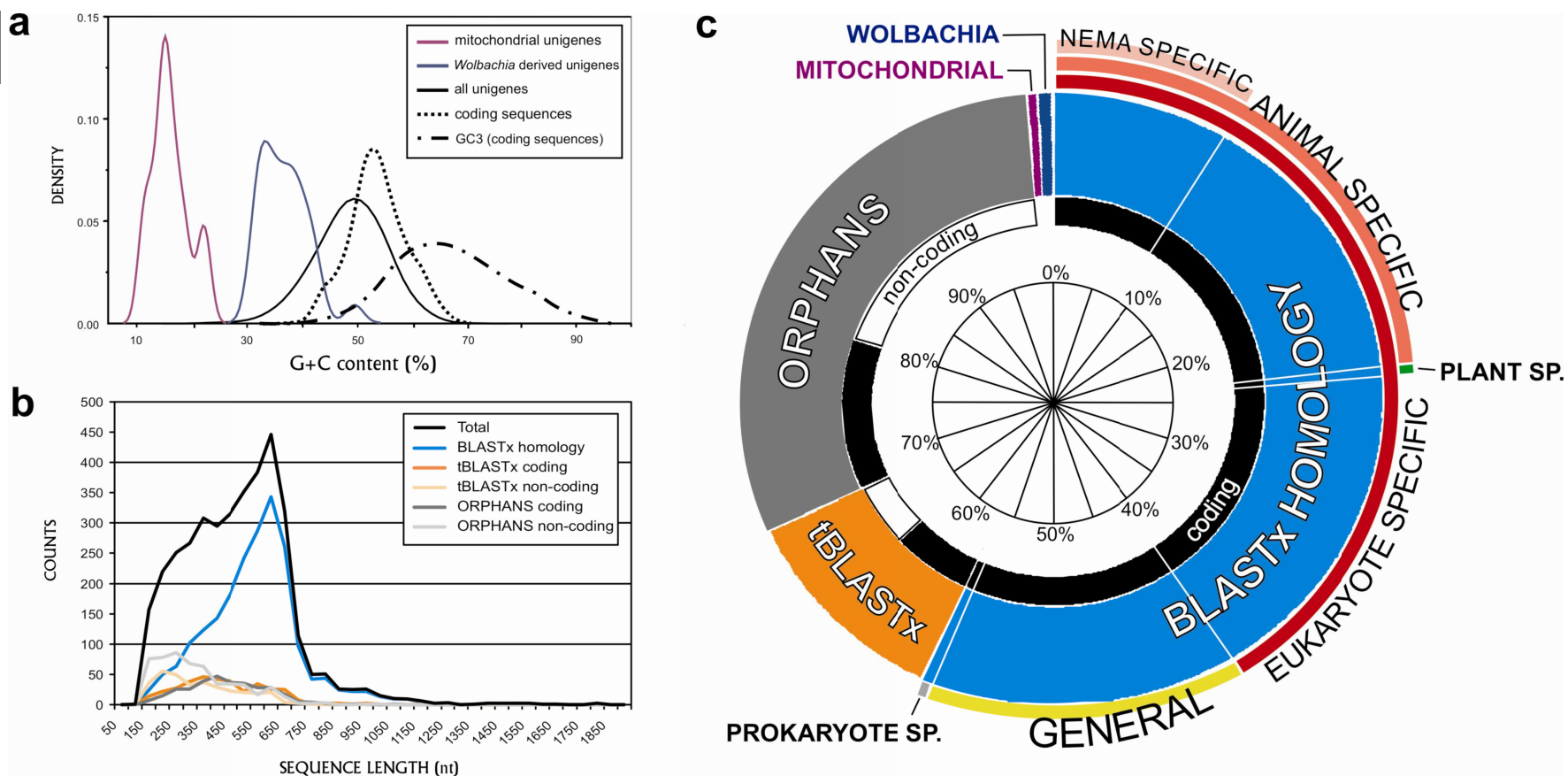
Figure 7: Overview of the EST analysis. **a,** density lines of the GC content of the three different subsets (mitochondrial, *Wolbachia* and nuclear derived) unigenes; **b,** Length distributions of different subsets of unigenes, represented in c; **c,** Schematic classification of unigenes with BLASTx homology (E-value cut-off of 1e-1; *blue*), with indication of the portion of unigenes with nematode-specific hits ("nema specific"), animal-specific hits, plant-specific hit ("plant sp."), eukaryote specific hits, prokaryote specific hits and unigenes with hits to both pro- and eukaryote sequences ("general"). The *orange part* marked with 'tBLASTx' indicates unigenes without BLASTx homology, but with tBLASTx homology to nematode ESTs. The *gray* part marked with "orphans" indicates the subset of unigenes without BLASTx or tBLASTx homology. The *inner black* and *white* circle corresponds to coding (*black*) and non-coding (*white*) prediction of the unigenes based on FrameD and BLASTx homology. The small portion of mitochondrial and *Wolbachia* derived sequences are indicated in the circle as *violet* and *dark blue*.

Table 2: BLASTx report of the 15 largest clusters

Unigene ID	Size*	GC %	Nuc**	BLASTx top-hit	SPECIES	E-value
Cluster 1	575	14.77	NO	No hit	-	-
Cluster 2	66	52.04	YES a	Hypothetical protein CBG12084	*Caenorhabditis briggsae*	1E-53
Cluster 3	62	40.19	YES b	unnamed protein product	*Homo sapiens*	6E-34
Cluster 4	47	47.69	YES a	Inhibitor of Cell Death family member (*icd-1*)	*Caenorhabditis elegans*	1E-52
Cluster 5	45	45.22	YES a	Translationally-controlled tumor protein homolog (*tctp*)	*Caenorhabditis briggsae*	2E-72
Cluster 6	43	53.95	YES c	No hit	-	-
Cluster 7	41	50.98	NO	No hit	-	-
Cluster 8	33	57.92	YES a	actin	*Caenorhabditis elegans*	1E-180
Cluster 9	30	50.00	NO	hypothetical protein	*Macaca fascicularis*	5E-14
Cluster 10	28	57.74	YES a	No hit	-	-
Cluster 11	27	45.83	YES a	type-1 cytochrome c	*Ascaris suum*	2E-50
Cluster 12	23	53.98	YES a	SEC-2 protein	*Globodera pallida*	6E-60
Cluster 13	22	55.08	YES a	P22U	*Dirofilaria immitis*	5E-40
Cluster 14	22	52.59	YES d	No hit	-	-
Cluster 15	18	44.93	YES a	F25H2.5	*Caenorhabditis elegans*	4E-60

* Size: the number of ESTs clustered to form the cluster
** Nuclear-coding protein, prediction by FrameD; a: full length; b: internal part of protein sequence; c: only N-terminal part of the protein sequence present; d: only C-terminal part of the protein sequence present;

d. Annotation of the unigenes

Annotation of "gene ontology" (GO) terms helps to categorize unigenes based on their putative function. We used the user-friendly BLAST2GO program to explore the 3,667 *R. similis* unigenes (Conesa et al. 2005). This annotation method is based on sequence homology determined by BLAST searches. Consequently, for 1,920 unigenes (52.4%) BLAST2GO could not find a homologous sequence and no mapping could be retrieved for an additional 259 sequences (7.0%) with BLAST homology. After mapping a total of 5,501 GO terms, 812 sequences (22.1% of the unigenes; note that this corresponds to an estimated 3.5% of the total
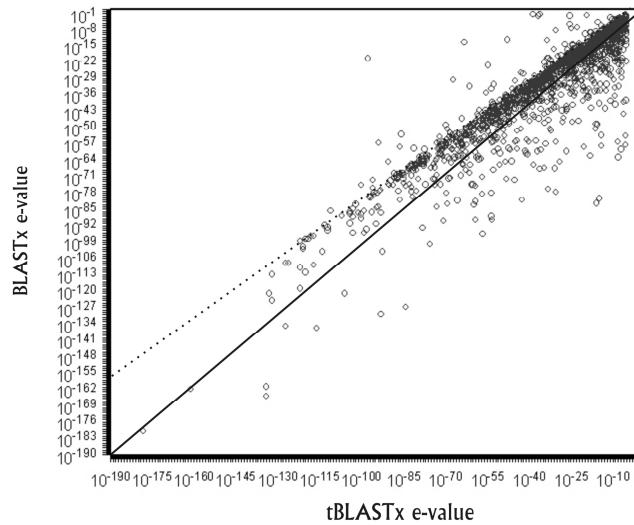
Figure 8: Comparison of the E-values obtained by BLASTx-search (Y-axis) and tBLASTx-search (X-axis). Every dot represents one unigene. Unigenes without hits in either the two searches are not represented.

number of genes in *R. similis*) were successfully annotated, with a bias towards sequences with longer lengths (Figure 9). Analyzing the 'biological process' main category and its child terms by annotation scores (assigned by BLAST2GO), revealed that the 'embryonic development', 'growth' and 'reproduction' GO terms are the most represented, followed by terms involved in basal cell metabolism (Figure 9). This could be caused by the high reproductive rate of *R. similis* and the high percentage of females (and developing eggs) in the population. In the main GO category 'molecular function', the 'protein binding' term covers most of the pie (~37% of the terms), followed by 'structural molecule activity', 'RNA binding', 'hydrolase activity' and 'nucleotide binding'. Many ribosomal proteins encoding unigenes are assigned to the 'protein binding' term, but also highly expressed genes coding for structural molecules (such as actin) and regulatory molecules (such as transcription factors) are categorized under this term. These are all abundant in the dataset, providing an explanation for the bias to 'protein binding'. Regarding the main GO category 'cellular component', the term 'ribosome' is most represented, constituting together with the term 'cytosol' almost the half of the total terms in this category. Taking the parasitic life style of *R. similis* into account, interesting sequences are likely to be found under the term 'extracellular region', since the nematode secretes a cocktail

of potentially interesting proteins into the plant to modulate the parasitism process. However, since the GO terms are annotated using homology and scarcely any parasitism gene of plant-parasitic nematodes to date have GO terms, most parasitism genes are likely not to be annotated. Also the presence of proteins secreted by the gut, epidermis and the nervous system are grouped under 'extracellular region'. In addition to the GO annotation, mapping of the unigenes to KEGG biochemical pathways was performed (Kanehisa et al. 2004). This revealed a significant enrichment in pathways involved in protein folding and associated processing (p = 5.44E-08) and ubiquitin mediated proteolysis (p = 2.62E-07), compared to the *C. elegans* reference set. The reason could lie in the bias of our (relatively small) EST dataset towards highly expressed genes, compared to the more equilibrated *C. elegans* dataset. In summary, the GO annotation of the unigenes is a representation of the biology of *R. similis* and the characteristics of the cDNA library, and is unsuited for detection of genes associated with parasitism.

e. Assigning RNAi phenotypes

Exploring the RNAi phenotypes in the *R. similis* unigene dataset can lead to interesting insights into potential control strategies based on disrupting gene expression. Using BLASTx (cut-off value 1E-05), 1638 (44.7%) unigenes were found having a homologous *C. elegans* gene. Of those *C. elegans* genes, 659 (18.0%) have an observed RNAi phenotype. Comparing GO terms of this set to the complete unigene set, shows a marked increase in genes involved in 'biological regulation' (GO:0065007) (an enrichment from 1% in the complete GO annotation (see previous paragraph) to 10% in the RNAi GO annotation), at the cost of general biological processes, such as 'metabolic' (GO:0008152) and 'cellular process' terms (GO:0009987) (from 19 to 10 and 18 to 12% respectively). In the 'cellular component' GO category, an increase is seen in 'macromolecular complex' (GO:0032991) (from 7% to 13%) and 'organelle part' (GO:0044422) (from 3% to 8%), while the largest decreases are for 'organelle' (GO:0043226) (from 18% to 22%) and 'extracellular region' (GO: GO:0005576) (from 2% to 0.5%). In contrast, the GO term distribution of 'molecular function' does not show any remarkable difference. Three quarters of the RNAi phenotypes (n=506, 76.8%) report a lethal effect. Compared to the total gene set with RNAi phenotypes, the genes with lethal RNAi phenotypes are significantly enriched in the GO terms 'macromolecular complex' (GO:0032991), 'developmental process' (GO:0032502), 'growth' (GO:0040007) and

'multicellular organismal process' (GO:0032501) (Figure 5). Apparently, disrupting expression of genes belonging to these GO classifications has a greater chance to be more detrimental for the nematode. Likewise, different GO terms were identified for which disrupted gene expression of the member genes is expected to have a less profound influence on nematode survival, such as 'enzyme regulator' (GO:0030234), 'molecular transducer' (GO:0060089) and 'cell surface receptor linked signal transduction' (GO:0007166).
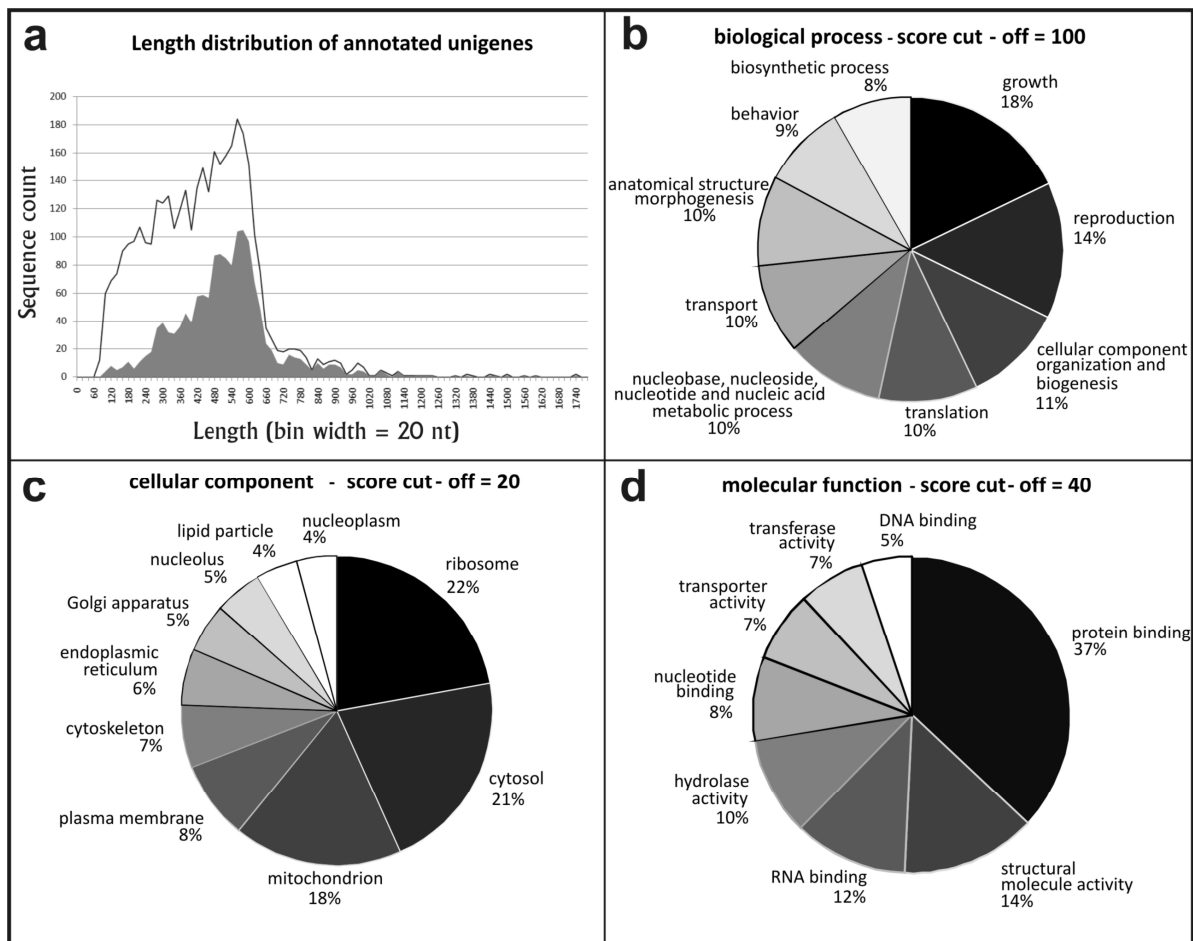

Figure 9: Summary of Gene Ontology annotation. **a**, the subset of the unigenes on which the GO annotation was based (filled area) is biased to longer sequences compared to the total unigene set (black line). **b**, **c** and **d**, annotation results of BLAST2GO, with indication of applied score-cut-off.

f. Translation of unigenes

It was reported before that coding DNA sequences of *R. similis* are GC-rich (GC content of 60%). This is mainly the result of the high mean GC content of the nucleotides at the third position of each codon (GC3%), which is approximately 63% (Cutter et al. 2006; Haegeman et al. 2007). This feature poses a potential problem for translation of the DNA sequences using standard sequence translation rules. This is due to the fact that every stop codon starts with a uracil (U) (or thymine (T) on a DNA level: TAG, TGA and TAA). A high GC3% means that most of the nucleotides at the last position of the codons are G or C. Hence the reverse complementary codons start mostly with G or C, reducing the likelihood to encounter a stop codon. It was noticed that translation based on the longest open reading frame will therefore result frequently in translating the wrong strand.

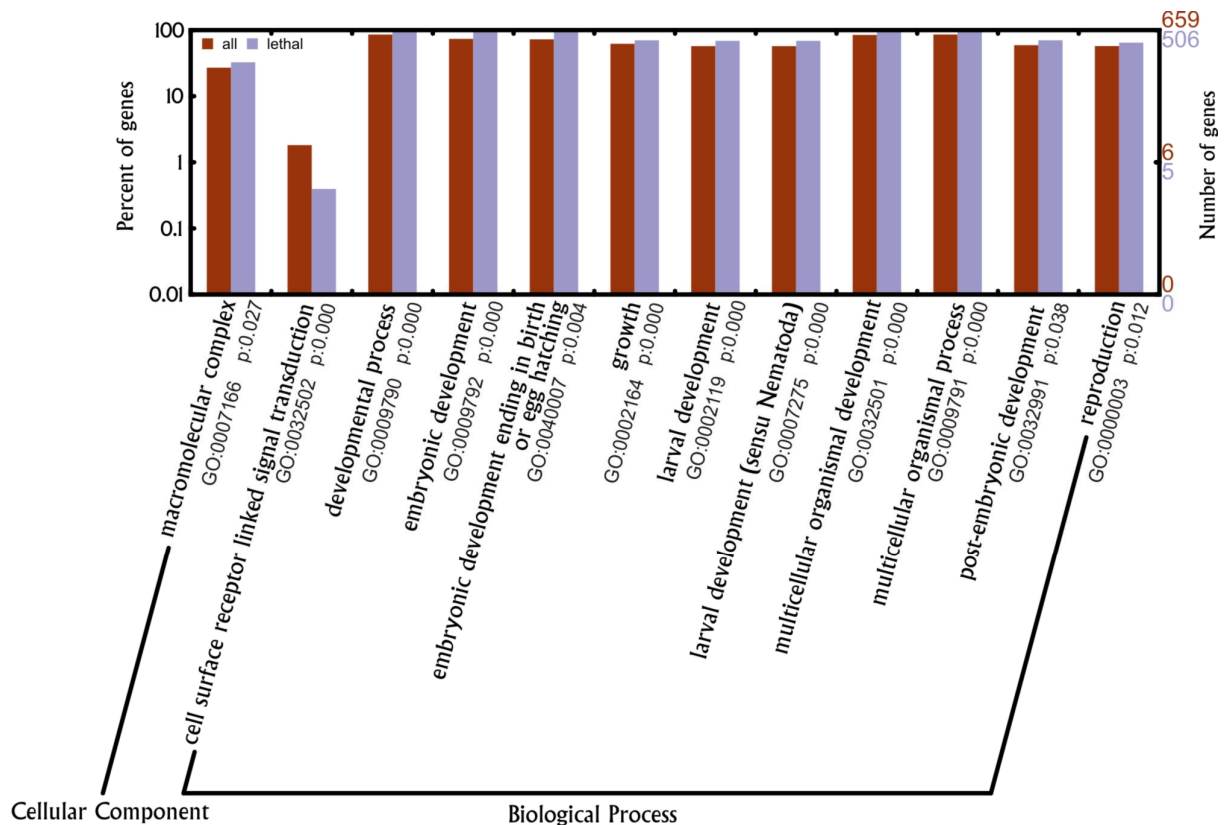Figure 10: GO terms of genes with predicted RNAi phenotypes. Red bars represent the total set with RNAi phenotypes. The gray bars represent the subset with predicted lethal RNAi phenotypes. Only the GO categories (up to level 5) which are significantly decreased or increased in the 'lethal' subset (Pearson Chi square test p-value < 0.05, with p-values indicated) are represented. Analysis and output by WEGO (Ye et al. 2006).

Translation prediction programs as FrameD (which uses a Markov model) have been designed to deal with a high GC3%: it predicts coding regions and corrects frame-shifts to obtain reliable translations, based on a set of training sequences (Schiex et al. 2003). We trained FrameD with 15 kb of coding nucleotides from manually collected full length ORFs based on the BLASTx-report and cloned *R. similis* genes submitted to GenBank. Testing the performance of FrameD, it classified all 226 unigenes containing 3' untranslated regions (selected by the presence of a poly(A) tail and polyadenylation signal) as non-coding, indicative of a low false positive error. On the other hand, the false negative error is rather high, since FrameD classified only 1317 (77.0%) of the 1710 unigene sequences with homologous protein sequences (BLASTx E-value cut-off of 1E-05) as coding. On the total unigene set, FrameD classified 2,245 (61.2%) unigenes as coding, and 1,422 (38.8%) unigenes as non-coding. Almost half of the FrameD translated unigenes were truncated at the 5' end, and one fifth had a full-length open reading frame (see Table 1). As expected from the false negative error rate, 38.8% of the 'non-coding' unigenes (552 sequences of 1422) actually have BLASTx-hits. By classifying these sequences also as coding, the total number of protein coding unigenes reaches 2,797 (76.3%). To obtain a translation for the unigenes with BLASTx-homology, but without a FrameD translation, the Prot4EST translation program was used (Wasmuth et al. 2004). Calculation of the GC content of the coding part of the unigenes resulted in an overall GC-percentage of 53.7%, a GC1% of 55.5%, a GC2% of 40.7% and a GC3% of 64.8%. Comparison of the *R. similis* translations with translations of ESTs of other plant-parasitic nematodes (obtained with Prot4EST) was performed through BLASTp searches (E-value 0.1). In addition, we compared with the Wormpep database, containing all predicted proteins of the *C. elegans* genome (C. elegans Genome Sequencing Consortium 1998). For each translated sequences set, approximately 20% matched to *R. similis* proteins, irrespective of the size of the translated sequence set. For example, 24.0% of Wormpep matched to *R. similis* translations. On the contrary, only 64.7% of the *R. similis* translations matched to Wormpep sequences. The bigger the translation set, the more *R. similis* translations are matched, as clearly visible in Figure 12. Still, 35.3% of the *R. similis* translations are not found in the Wormpep database. Apparently, these different small EST datasets are very heterogeneous in their composition and comparison between them will most likely not yield informative results.
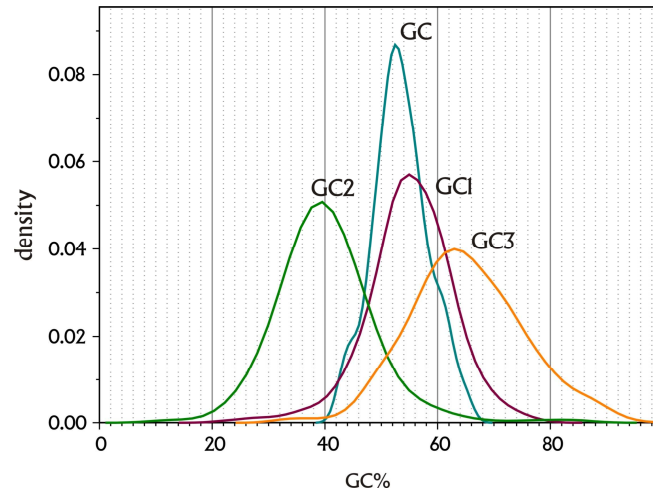
Figure 11: GC content of protein-coding unigenes of *R. similis*, with GC1, GC2 and GC3 the GC content of the nucleotides of the first, second and third position of the codons respectively. The Y axis represents the density line, a representation of a continuous distribution in which the area under graph equals one.
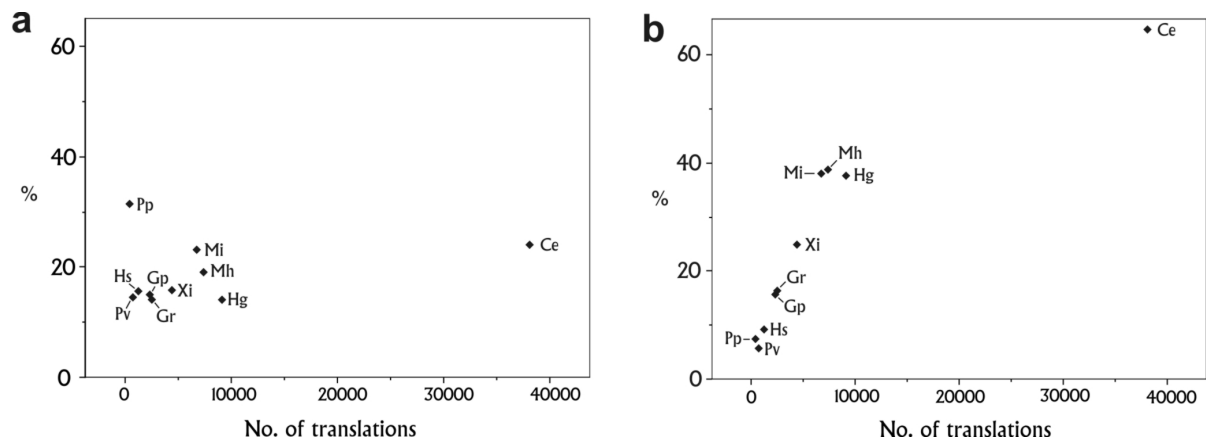

Figure 12: Comparison of the obtained translations of *R. similis* ESTs to other nematode EST translations. Each point represents a translated EST dataset. **a**, percentage of the translations (in Y-axis) matched to *R. similis* ESTs. Irrespective of the size of the translated EST (in X-axis), approximately 20% of the translations matches to *R. similis* translations. **b**, percentage of the *R. similis* translations (in Y-axis) matched to the translated EST sets (size in X-axis). Clearly, the larger the translated EST set used, the more *R. similis* translations are hit. Pp: *Pratylenchus penetrans*, 419 translations; Pv: *P. vulnus*, 726 translations; Hs: *Heterodera schachtii*, 1,247 translations; Gp: *Globodera pallida*, 2,324 translations; Gr: *G. rostochiensis*, 2,516 translations; Mi: *Meloidogyne incognita*, 6,734 translations; Mh: *M. hapla*, 7,380 translations; Hg: *Heterodera glycines*, 9,122 translations; Ce: *Caenorhabditis elegans*, 38,096 translations, from Wormpep (May 2006).

Searching secretion signal peptides in the translated sequences revealed that 216 of the 2,755 translations (7.8%) were predicted to contain a signal peptide, of which 156 (5.6% of the translations) lacked a transmembrane region. Based on these predictions, 4.3% of our unigene dataset could encode for secreted proteins. Remarkably, the translation prediction shows that 447 orphan sequences are predicted to be coding (39.6% of the orphans). These interesting unigenes are in all probability novel protein encoding genes, without any known homologous sequence in the database to date. But on the other hand and equally remarkable, a large fraction of the unigene sequences (at most 682 sequences or 18.6%) is predicted to be not protein-coding, forcing us conclude that they are derived from non-coding RNA.

g. *Trans*-spliced leaders

One of the major eccentricities in the molecular biology of nematodes is the widespread occurrence of operons and *trans*-spliced leader sequences, the latter which are frequently found at the start of transcripts (Guiliano et al. 2006). To investigate the occurrence of the currently known *trans*-spliced leader sequences in our unigenes, searches were performed based on common features of a reference set of spliced leader (SL) sequences reported in Guiliano and Blaxter (2006). Surprisingly, in the first approach using BLASTn-searches, none of the *R. similis* unigenes in our dataset matched to any SL sequence, whereas in a control dataset of *Meloidogyne incognita* ESTs (in which *trans*-splicing is known to occur), 293 sequences were matched to a total of 9 different SL sequences of our reference set. After this negative result, a second approach was applied using a pattern which was able to match 76% of the reference SL sequences. Remarkably, this pattern was unable to match any *R. similis* unigene. On the contrary, this pattern found motifs in 12.2% of the control data set of *M. incognita* ESTs (376 out of 3,098 ESTs). To exclude the possibility that by chance none of our unigenes belongs to a nematode gene family which is frequently *trans*-spliced, we searched very strong homologues of the *trans*-spliced *M. incognita* ESTs (detected by the pattern search) in *R. similis* unigenes and in *C. elegans* ESTs. Five *R. similis* unigenes (CL13, CL18, CL25, CL112 and CL929) were found with very strong homology to *trans*-spliced *M. incognita* ESTs (ranging from 60% to 93% identity at the protein level). For each of the five cases, the unigene and the *trans*-spliced *M. incognita* EST were aligned (on a DNA level), together with the strongest *C. elegans* homologue. In all five cases, this alignment revealed the occurrence of similar *trans*-spliced leader sequences on both the *M. incognita* EST and *C. elegans* gene, but
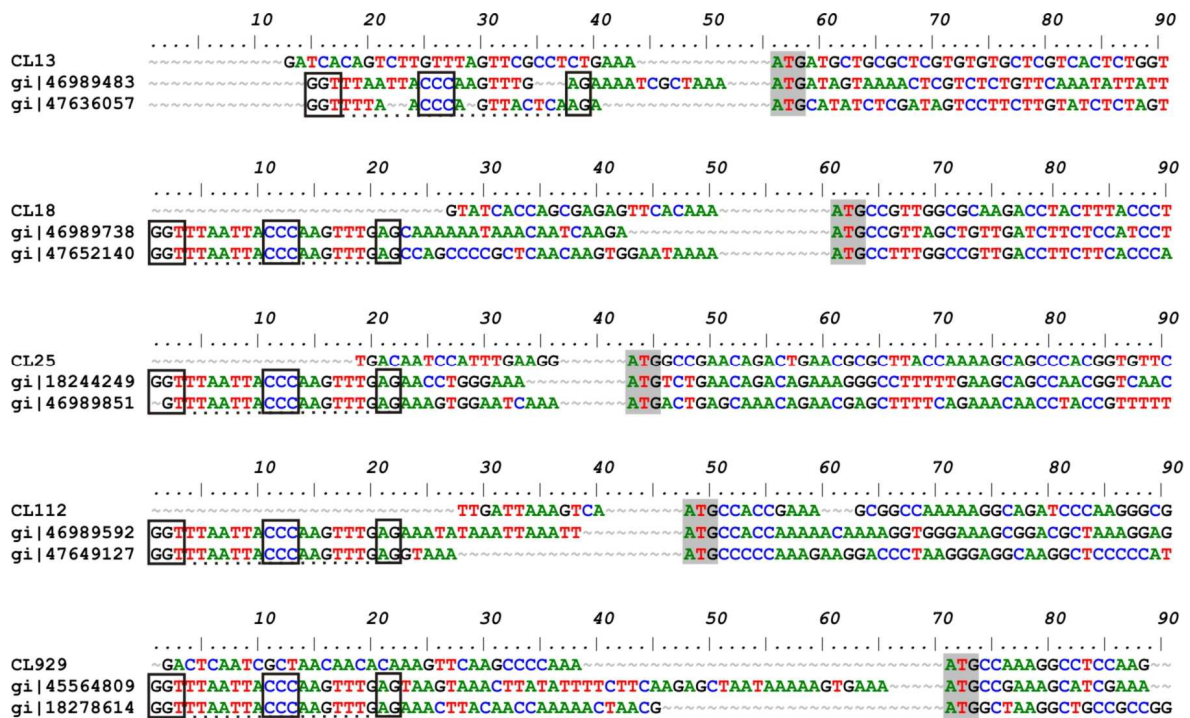
Figure 13: Alignment of the 5' end of five *R. similis* unigenes (first sequence) with the *M. incognita* EST with strongest homology on protein level (second sequence) and *C. elegans* EST with strongest homology on protein level (third sequence). Conserved sequence features of the *trans*-spliced leader sequences are indicated with black boxes, the start ATG codon is shaded in gray.

absence of such a sequence on the 5' UTR sequence of the *R. similis* unigene (Figure 13). Closer inspection of the 5' UTR sequences of the analyzed unigenes could not reveal any sequence similarity, neither with the *M. incognita* or *C. elegans* homologues, nor with each other.

h. Characteristics of the 3' UTR

Since a portion of EST sequences contain a polyadenylation tail, information can be obtained about the 3' untranslated region (UTR) sequence. EST data of *Caenorhabditis elegans* has been used to yield insight into 3' end formation (Hajarnavis et al. 2004). In this model species, the T nucleotide content steadily increases in the 3'UTR until the polyadenylation signal is encountered. The polyadenylation tail follows this signal after 10 to 18 nt, with an average of 14 nt. In *C. elegans*, 57% of the polyadenylation signals are AATAAA motifs (Hajarnavis et al.

2004). In *R. similis* ESTs, 978 ESTs (14.3 % of the cleaned EST set) have polyadenylation tails, of which we could extract 218 (22.3%) sequences with an AATAAA motif. Similar to *C. elegans*, a steadily increase of T nucleotides towards the polyadenylation signal was observed (Figure 14). The average spacer length was 12 nt, ranging from 8 to 18 nt (Figure 15).
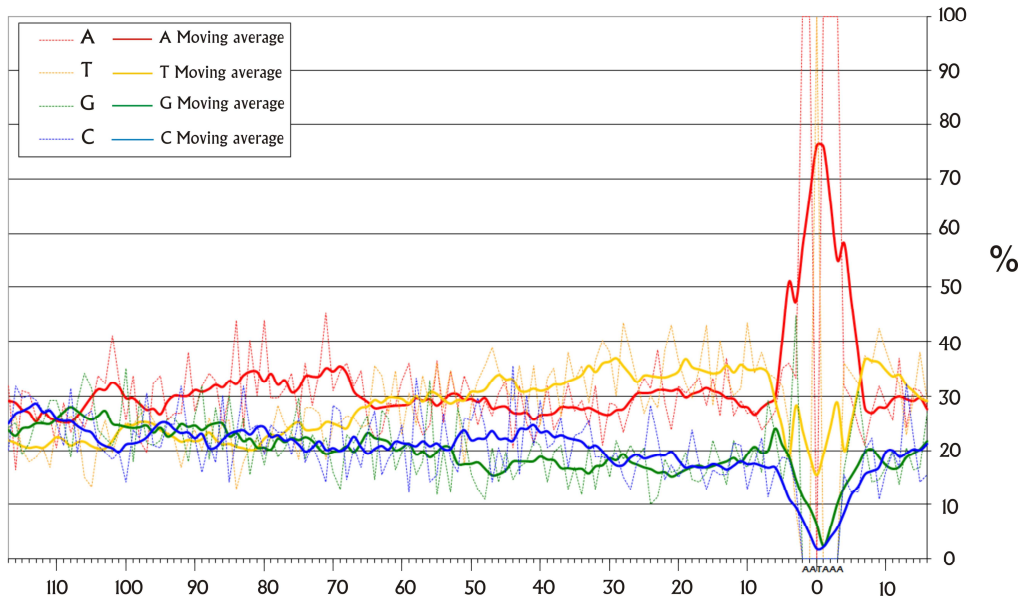
Figure 14: Graphical representation of the alignment of 281 ESTs containing an AATAAA polyadenylation signal. All sequences are aligned on this signal sequence.
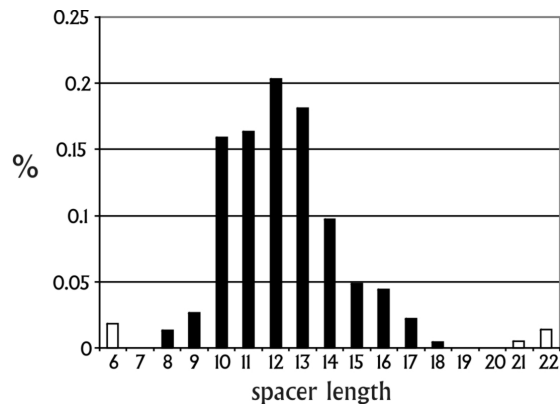
Figure 15: Length distribution of the spacer sequences identified on ESTs containing AATAAA polyadenylation signals. The spacer is the sequence between the AATAAA signal and the poly(A) tail.

i. Occurrence of mitochondrial ESTs

Depending on the cDNA library construction method, a remarkably high fraction of ESTs can be of mitochondrial origin and can even be used as a guideline for the sequencing of the mitochondrial genome (Gissi et al. 2003; Jex et al. 2008). We searched our unigene dataset for ESTs most likely derived from the mitochondrial genome, and found five unigenes with significant similarity to various nematode mitochondrial genes (Table 3).

The AT richness (83.6%) of these sequences was very high compared to the mean AT content of the total unigene set (51.2%; Figure 7). One of the major characteristics of mitochondrial genomes is their low GC content, which is usually between 20% and 30% (He et al. 2005). The different GC content of nuclear unigenes compared to mitochondrial unigenes can therefore be used to predict the origin of the EST. The GC content density line of the total unigene dataset follows a normal distribution, pointing to a similar source of the unigenes. However a bias from the normal distribution is observed at one end of the curve, representing 21 unigenes with a lower GC content (from 10% to 26% GC, Figure 16), most likely all origi-

Table 3: Summary of the unigenes with lowest GC content, putatively derived from the mitochondrial genome.

Unigene ID	Size*	Unigene Length (nt)	%GC	BLASTx homology
Cluster 1 Contig1	551	684	14.77	No
Cluster 1 Contig2	18	344	18.61	No
Cluster 1 Contig3	4	329	17.63	No
Cluster 1 Contig4	2	335	15.53	No
Cluster 21 Contig1	15	1518	15.95	Cytochrome oxidase subunit 2
Cluster 86 Contig1	6	1756	14.18	NADH dehydrogenase subunit 5
Cluster 337 Contig1	3	984	15.05	Cytochrome oxidase subunit 1
Cluster 421 Contig1	3	796	12.57	No
Cluster 429 Contig1	3	1020	22.75	Cytochrome oxidase subunit 1
aaa88a11.g1 (EY193009)	1	370	15.41	No
aaa89f05.g1 (EY193136)	1	134	14.93	No
aab01c04.g1 (EY193989)	1	538	19.15	Cytochrome oxidase subunit 3
aab09f05.g1 (EY194661)	1	292	11.99	No
aab12g04.g1 (EY194846)	1	138	14.5	No
aaa04a05.g1 (EY190346)	1	292	21.92	No
aaa05b09.g1 (EY190790)	1	285	13.34	No
aaa13g09.g1 (EY191363)	1	310	21.62	No
aaa15b02.g1 (EY191468)	1	314	16.88	No
aaa19e11.g1 (EY191779)	1	343	16.91	No
aaa57f06.g1 (EY192478)	1	179	11.18	No
aaa71h02.g1 (EY192644)	1	156	10.9	No

* Size: the number of ESTs clustered to form the cluster

nating from the mitochondrial genome as the 5 mitochondrial unigenes are also included (summarized in Table 3). These results provide information about another curiosity of the unigene dataset, the largest cluster in our data set, cluster 1. This cluster represents a disproportionately large amount (n=575, or 8.5%) of the total number of ESTs. Despite this, it shows no significant homology to any known sequence using various BLAST approaches against various databases, and is also predicted not to be protein coding by different prediction programs. Using cluster 1 specific primers, we succeeded in amplifying this unigene from a cDNA pool constructed from DNAse treated RNA, making genomic DNA contamination very unlikely (data not shown). In addition, genomic contamination is likely to appear as singletons and not as a cluster. Based on identified mitochondrial transcripts and the
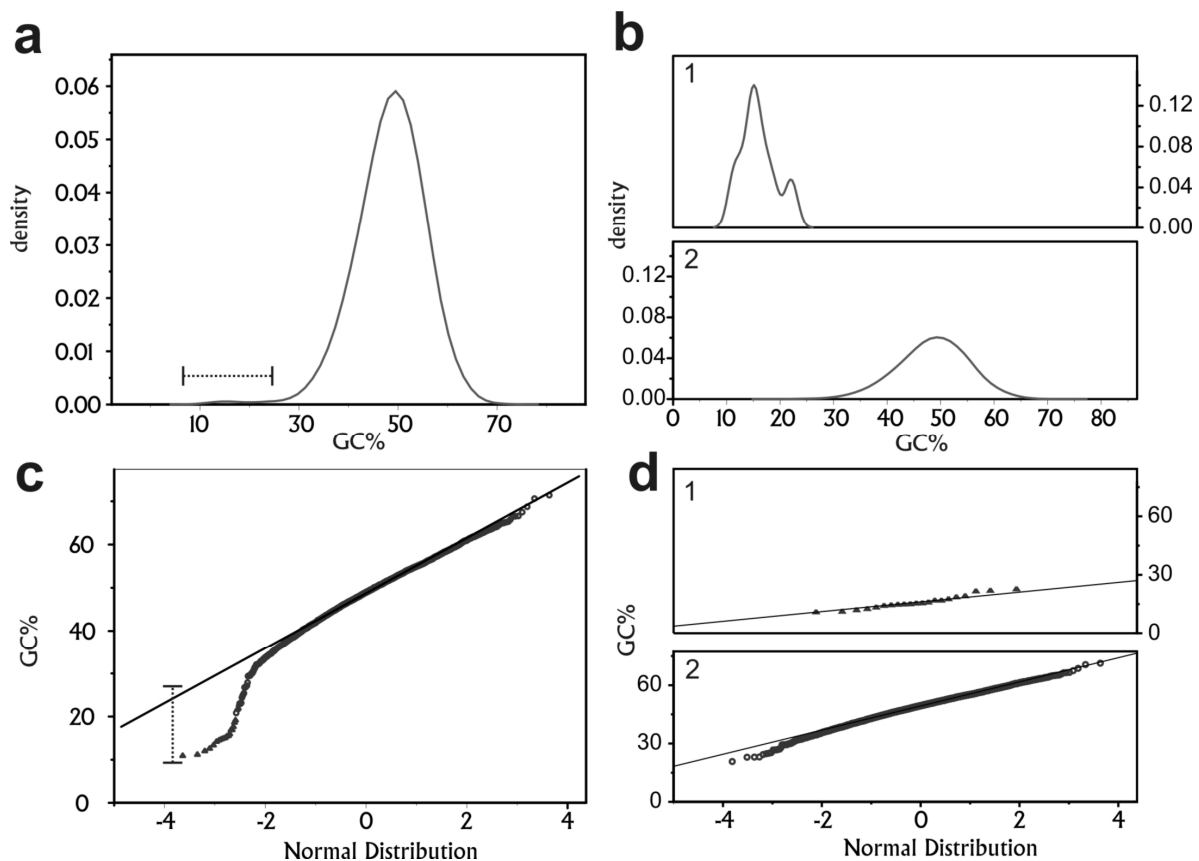
Figure 16: GC analysis of the unigenes. **a,** GC percentage density line of total unigene set. The subset with low GC content is indicated by a dashed line. **b,** GC percentage density line of the subset (**b1**) and the remaining unigenes (**b2**). **c,** quantile-quantile plot (QQ plot) with normal distribution line of total unigene set. The subset with low GC content is indicated by a dashed line. **d,** QQ plot of the subset (**d1**) and the remaining unigenes (**d2**).

2

low GC content, it is possible that cluster 1 is derived from the mitochondrial genome. In clustered EST datasets of other plant-parasitic nematodes (*Bursaphelenchus* and *Pratylenchus*), the largest clusters represent also a disproportionately large part of the EST dataset (Kikuchi et al. 2007; Mitreva et al. 2004). Remarkably, checking these sequences revealed homology to mitochondrial genes (*cytochrome oxidase subunit I, III* and *cytochrome b*). However, certainty about cluster 1 will only be achieved with the complete sequence of the *R. similis* mitochondrial genome.

j. Unigenes with similarity to *Wolbachia* sequences

A subset of 43 unigenes (2%) was selected based on the homology exclusively to prokaryotic sequences. Although the possibility exists that some of these unigenes are the result of contamination, 18 significant matches to genes of the endosymbiotic *Wolbachia* species were found. Further investigation of all the unigenes (having homologues not only in prokaryotic species) revealed another 12 unigenes with BLASTx top-hits to *Wolbachia* (see Table 4). The mean GC content for these 29 sequences is 36.9% ± 4.2, similar to previously reported GC percentages of *Wolbachia* sequences (Foster et al. 2005). These results suggest an endosymbiotic presence of *Wolbachia* within *R. similis*. As a first approach to confirm its presence, fluorescence microscopy on DAPI (4',6-diamidino-2-phenylindole) stained *R. similis* nematodes was performed, as has been done for other endosymbiotic species (Vandekerckhove et al. 2002). Small signals were observed in the ovaria, indicative of the presence of an endosymbiont (Figure 17). Since *Wolbachia* has been shown in other nematodes to reside in the ovaria, the observed signals could be due to this endosymbiont. Further confirmation was obtained by transmission electron microscopy, revealing numerous endosymbiotic bacteria with various shapes located in the ovaria. In many cases, the three membranes could be distinguished, two of bacterial origin and one of host origin.

k. Unigenes putatively involved in parasitism

Based on homology to genes of other parasitic species, some unigenes were identified which could be involved in the parasitism process (Table 5). These included putative parasitism genes previously characterized in other plant-parasitic nematodes. Two unigenes encoding plant cell wall degrading enzymes were found: an endoglucanase and a xylanase. Both enzymes are secreted to soften the plant tissue during the migration of the nematode through the host

roots. Those enzymes have been identified in numerous plant-parasitic nematodes (Haegeman et al., 2007; Ledger et al., 2006; Smant et al., 1998) and their role in parasitism has been extensively studied. Other unigenes showed homology to parasitism genes coding for enzymes
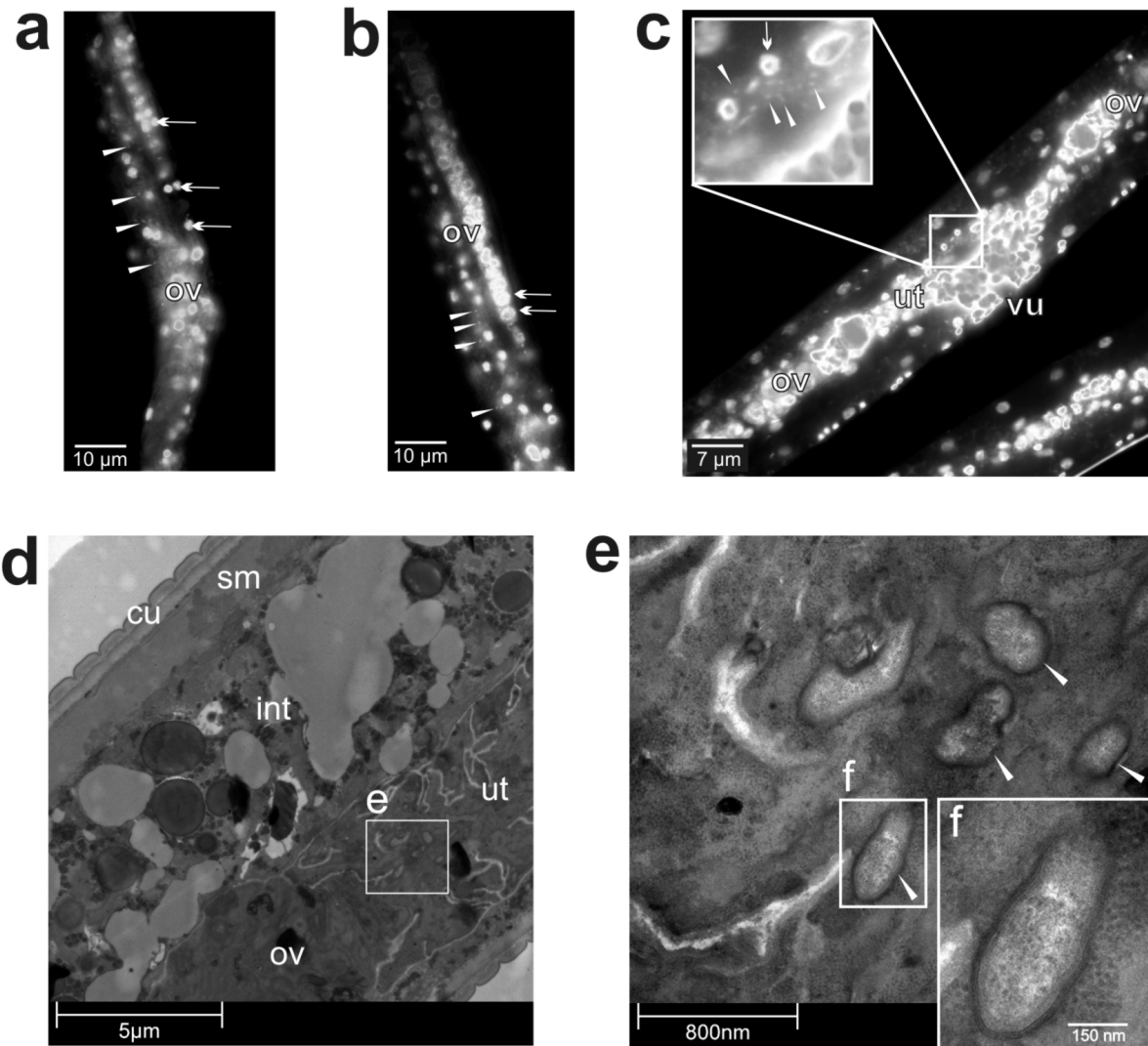
Figure 17: Localisation of *Wolbachia* endosymbiont in *Radopholus similis* females. **a**, **b** and **c**, DAPI staining of *Radopholus similis* ovaries. Examples of host nuclei are indicated by arrows. *Wolbachia* endosymbionts are indicated by arrowheads. **d**, transmission electron microscopy of a section of *Radopholus similis* adult females, with **e** and **f** detailed pictures *Wolbachia* endosymbionts (indicated by arrowheads). Note the three membranes around the endosymbiont. cu, cuticle; sm, somatic musculature; int, intestine; ov, ovarium; ut, uterus; vu, vulva.

Table 4. Unigenes with BLASTx top-hits to *Wolbachia* sequences

ID	Wolbachia protein	Accession	E-value	%id/%sim*
CL342Contig1	N utilization substance protein A	ZP_00372417.1	5e-49	80/88
CL357Contig1	Hypothetical protein Wendoof_01000009	ZP_01315146.1	4e-19	58/70
CL458Contig1	RNA polymerase sigma factor RpoD	NP_967007.1	7e-41	61/72
CL561Contig1	Hypothetical protein WD0332	NP_966130.1	2e-14	31/51
CL622Contig1	cell division protein FtsZ	AAB38745.1	5e-26	65/77
CL882Contig1	cell division protein FtsZ	NP_966481.1	2e-31	92/98
aaa87b09.g1	Glyceraldehyde-3-phosphate dehydrogenase, GapA	YP_198129.1	4e-39	81/90
aaa88b06.g1	malate dehydrogenase	ZP_00372475.1	2e-53	75/91
aaa88b10.g1	hypothetical protein	NP_966393.1	8e-39	72/90
aaa96b05.g1	hypothetical protein	NP_966219.1	1e-12	35/59
aaa96g09.g1	ribosomal protein L3	NP_966445.1	1e-44	85/93
aab02h07.g1	ATP-dependent protease La	NP_966117.1	3e-39	65/87
aab05d02.g1	Ribosomal protein L14	YP_198162.1	2e-20	85/96
aab10b02.g1	adenylosuccinate lyase	NP_966540.1	4e-42	70/86
aab18b05.g1	hypothetical protein WD0631	NP_966396.1	2e-15	45/79
aab23e08.g1	50S ribosomal protein L20	NP_966615.1	3e-44	86/97
aab23f04.g1	Chaperonin GroEL (HSP60 family)	YP_198181.1	3e-78	73/84
aaa92c09.g1	Type IV secretory pathway, component VirB9	YP_198111.1	2e-38	63/80
aaa11f09.g1	hypothetical protein WD0474	NP_966260.1	3e-18	53/72
aaa16d02.g1	chaperonin, 60 kDa	NP_966107.1	3e-52	72/88
aaa16g01.g1	hypothetical protein WD1172	NP_966885.1	2e-13	37/59
aaa22f12.g1	30S ribosomal protein S12	NP_965847.1	5e-17	93/100
aaa23e06.g1	Integral membrane protein, interacts with FtsH	YP_198320.1	7e-24	68/91
aaa23e09.g1	bifunctional GMP synthase/glutamine amidotransferase protein	NP_966007.1	4e-81	71/88
aaa52b11.g1	50S ribosomal protein L16	NP_966438.1	8e-36	83/90
aaa57g11.g1	translation elongation factor Tu	ZP_01314396.1	4e-12	91/94
aaa96e08.g1	DNA-directed RNA polymerase, RpoB/RpoC	YP_198477.1	4e-58	84/94
aaa89c06.g1	ATP-dependent Clp protease, ATP-binding subunit ClpB	ZP_00372220.1	8e-12	85/92
aab16h05.g1	DNA-directed RNA polymerase, RpoB/RpoC	YP_198477.1	1e-60	75/88
aaa12f04.g1	Polyribonucleotide nucleotidyltransferase, pnp	YP_197853.1	3e-50	69/79

** %id/%sim: percentage identical and similar amino acid as reported by BLASTx search using bacterial codon table

that can neutralize reactive oxygen (ROS) species. ROS are produced by the host as a defense mechanism in response to infection by the nematode (Dubreuil et al. 2007; Jones et al. 2004a; Robertson et al. 2000). Furthermore, three unigenes show homology to fatty acid- and

retinol-binding proteins of parasitic nematodes. Fatty acids are compounds that play a role in the host defense signaling pathway (Kennedy et al. 1995; Prior et al. 2001). As a consequence, nematode proteins that bind such compounds could modulate the host defense to facilitate parasitism. One unigene of this group resembles a gene of the animal-parasitic nematode *Brugia malayi* (Kennedy et al. 1995). The other two unigenes (CL12Contig1 and CL70Contig1) show highest homology to SEC-2 proteins (also called FAR, fatty-acid and retinol-binding) of the PPN *Globodera pallida*. However, where CL12Contig1 shows also homology to a (hypothetical) protein of the free living nematode *C. elegans*, CL70Contig1 has only homology to SEC-2 proteins of parasitic nematodes. This could point to the existence of functionally distinct SEC-2 proteins, one with a general function and another with a function related to parasitism, as hypothesized previously (Garofalo et al. 2003; Prior et al. 2001). Two unigenes have homology to SXP/RAL-2 genes, a family of putative parasitism genes with expression in different tissues, such as the hypodermis, subventral glands and amphids (Jones et al. 2000; Tytgat et al. 2005). One SXP/RAL2 gene (51237540) is very similar to SXP/RAL-2 of *Globodera rostochiensis*, which is expressed in the amphids and most likely has a function in host localisation (Jones et al. 2000). In addition to those functionally known genes, some unigenes showed homology to putative parasitism genes without known function. To identify additional unknown parasitism gene candidates, we searched within the 1,316 unigenes without any BLASTx-homology for sequences predicted to be protein coding and with homology exclusively to ESTs of plant-parasitic nematodes (PPN). 212 such unigenes (5.8%) were retrieved, of which 18 (1.3%) were predicted to encode secreted proteins (Table 6). Five of them (CL26, CL546, CL793, 92h12, 23g06) were assigned as good parasitism genes candidate, since the homologous EST sequences originated exclusively from the parasitic stages of PPN (i.e. second stage juvenile to adults). *In situ* hybridisations succeeded for four of the parasitism gene candidates (Figure 18), and one of them (92h12) revealed expression in the pharyngeal gland cells. In the 341 unigenes with homology to PPN exclusively, 45 domains could be identified (Table 7). The domain with highest homology was glycosyl hydrolase family 5 domain, as identified in cellulases in numerous plant-parasitic nematodes. Interestingly, some of the identified domains (e.g. DUF1772) are until now only described in fungi and/or bacteria. Furthermore, the *R. similis* unigene containing the isochorismatase family domain has significant homology only to three ESTs in *Meloidogyne hapla* (1 originating of the female stage library, 2 of a second stage library), and to one of *Heterodera glycines* (of a fourth stage library). No homology is found to any of the numerous ESTs of *Caenorhabditis elegans* –

Table 5: Nematode specific unigenes and potential parasitism genes

Unigene ID	BLASTx homology	Accession	E-value	Species	Specificity*
aab12d02.g1 (EY194811)	SXP/RAL-2 protein	AAR35032.1	2e-16	*Meloidogyne incognita*	PPN
51237540 (CO897750)	SXP/RAL-2 protein	CAB66341.1	4e-22	*Globodera rostochiensis*	PPN
aaa92h12.g1 (EY195747)	gland-specific protein g4e02	AAO33473.1	1e-17	*Heterodera glycines*	PPN
51237561 (CO897771)	Unknown gene	AAW33662.1	8e-07	*Heterodera glycines*	PPN
51334228 (CO961044)	glutathione S-transferase	AAF81283.1	4e-17	*Haemonchus contortus*	MET
CL70 Contig1	SEC-2 protein	CAA70477.2	4e-12	*Globodera pallida*	PAR
aaa09c03.g1 (EY190991)	beta-1,4-endoglucanase	BAB68522.1	2e-52	*Pratylenchus penetrans*	GEN
aab06h02.g1 (EY194441)	xylanase D	AAB63573.1	2e-34	*Aeromonas punctata*	GEN
CL551 Contig1	thioredoxin peroxidase	AAF21097.1	1e-72	*Dirofilaria imminitis*	GEN
aab21a04.g1 (EY195324)	glutathione peroxidase	AAL09384.1	6e-09	*Haemonchus contortus*	GEN
aab18d05.g1 (EY195186)	Extracellular superoxide dismutase [Cu-Zn]	P51547	1e-23	*Haemonchus contortus*	EUK
CL152 Contig1	cathepsin L-like cysteine proteinase	AAY46196.1	4e-51	*Globodera pallida*	EUK
aaa90f12.g1 (EY193222)	SJCHGC01111 protein	AAW26476.1	4e-06	*Schistosoma japonicum*	MET
aaa78e10.g1 (EY192694)	major allergen	AAK18279.2	2e-14	*Brugia malayi*	PAR
51237728 (CO897938)	hypothetical protein L3ni51	AAT02162.1	9e-08	*Dictyocaulus viviparus*	PAR
CL308 Contig1	gp 15/400 antigen; Bm12	AAB32807.1	0.007	*Brugia malayi*	PAR
CL674 Contig1	class V aminotransferase	AAK26375.1	0.002	*Heterodera glycines*	PPN
CL919 Contig1	FMRFamide-related peptide 2	CAC32452.1	2e-13	*Globodera pallida*	PPN
CL939 Contig1	FMRFamide-related peptide	CAC36149.1	0.007	*Globodera pallida*	PPN
51237728 (CO897938)	hypothetical protein L3ni51	AAT02162.1	9e-08	*Dictyocaulus viviparus*	PAR
CL140 Contig1	galectin 3	AAD45606.1	8e-07	*Haemonchus contortus*	PAR
CL897Contig 1	Cyclophilin *Bm-cyp-2*	AAC47231	9e-06	*Brugia malayi*	PAR
aaa09f07.g1 (EY191024)	glycogen synthase	AAK18279.2	4e-06	*Steinernema feltiae*	PAR

* Classification of the unigenes according to BLASTx results for that unigene: 'PPN', only hits to plant-parasitic nematode proteins; 'PAR', only to (plant- and animal-)parasitic nematode proteins; 'MET', only to metazoan proteins; 'GEN', to proteins of pro- and eukaryotes.

although this domain is also reported in this species. The fact that the PPN ESTs originate solely from libraries made of parasitic stages (except for *R. similis* of course) adds to the possibility these ESTs are involved in the parasitism process. Interestingly, this domain is also found to be specific for phytopathogenic fungi as compared to non-phytopathogic fungi, and it is hypothesized to function in the downregulation of plant defenses through depletion of SA (Soanes et al. 2008). The identification of one *R. similis* unigene resembling the 'pepsin inhibitor-3-like repeated' domain of nematodes is also promising. This domain is present on some protease inhibitors of APNs and of *C. elegans*. The function for APNs was proposed to inhibit host proteases when passing through the intestinal tract, but since some of the APNs do not come into contact with the intestine, and homologues are found in *C. elegans*, recent hypotheses propose a regulatory function of endogenous proteinases (Girdwood et al. 2000). The *R. similis* unigene has strong homology (E-value 1E-15) to 11 ESTs of cyst nematodes (*Heterodera* and *Globodera*) exclusively, and is predicted to contain a signal peptide, indicative of an extracellular function.

Table 6: Selection of 18 parasitism genes candidates. Based on PPN specific expression, protein coding and the occurrence of signal peptide for secretion (see text), with reporting of the number of PPN-specific hits, the lowest E-value, and the distribution over different stage-specific libraries.

ID	Length (nt)	Nuc*	Size*	PPN-hits	E-value	Stages**							
						E/E	1	2	3	4	a	d	rest
CL14Contig1	523	YES a	22	4	4e-03	0	4	0	0	0	0	0	0
CL26Contig1	716	YES c	12	27	0e+00	7	0	19	0	0	1	0	0
CL277Contig1	439	YES c	3	1	20e-06	0	0	0	0	0	0	0	1
CL541Contig1	974	YES a	2	1	8e-02	0	0	0	0	0	0	0	1
CL546Contig1	640	YES d	2	1	9e-03	0	0	0	0	0	1	0	0
CL793Contig1	561	YES b	2	2	3e-03	0	0	0	0	0	2	0	0
aaa85g06.g1	545	YES b	1	4	6e-13	2	0	1	0	0	0	0	1
aaa97b02.g1	530	YES b	1	1	0e+00	0	0	0	0	0	0	0	1
aaa97e04.g1	568	YES c	1	1	6e-06	0	0	1	0	0	0	0	0
aaa92h12.g1	415	YES c	1	5	0e+00	0	0	0	0	1	0	0	4
aaa01c01.g1	276	YES a	1	1	1e-12	1	0	0	0	0	0	0	0
aaa17e11.g1	310	YES c	1	9	2e-12	2	0	4	0	0	2	0	1
aaa23g06.g1	576	YES b	1	9	0e+00	0	0	0	0	3	2	0	4
aaa24h12.g1	470	YES c	1	4	2e-03	1	0	1	0	0	0	0	2
aaa67g10.g1	274	YES a	1	3	1e-05	2	0	0	1	0	0	0	0
51334668	401	YES a	1	15	7e-07	3	0	11	0	0	1	0	0
51334619	457	YES a	1	15	4e-07	3	0	11	0	0	1	0	0
51334566	462	YES a	1	8	1e-17	4	0	3	0	0	0	0	1

* Nuc, Size: see Table 2. ** E/E: egg and embryo; 1 to 4: first to fourth juvenile stage; a: adult; d: dauer; rest: mixed stage libraries.

Table 7: Overview of domains identified in the PPN-specific unigenes, with E-value < 1e-04.

Domain name	Specificty*	E-value	No.**	Unigene
Cellulase (glycosyl hydrolase family 5) IPR001547	bact/fungi	1.30e-40	1	aaa09c03.g1
Protein of unknown function, DUF647 IPR006968	eukaryote	3.30e-36	1	CL751Contig1
Replication factor C IPR013748	general	5.70e-18	1	aaa52g07.g1
MA3 domain IPR003891	eukaryote	1.10e-17	2	CL441Contig1
Domain of unknown function (DUF1772) IPR013901	fungi	2.60e-16	1	CL281Contig1
Isochorismatase family IPR000868	bact/fungi	5.90e-16	1	CL627Contig1
Protein kinase domain IPR017442	general	7.90e-16	2	CL568Contig1
Bacterial transferase hexapeptide	general	1.70e-08	3	51334565
Glu-tRNAGln amidotransferase C subunit IPR003837	general	5.80e-08	1	aab08c06.g1
PUL domain IPR013535	eukaryote	9.70E-08	1	aab20b04.g1
EF hand IPR002048	general	1.20e-07	1	aab00c03.g1
Hsp20/alpha crystallin family IPR002068	general	6.40e-07	1	aab14a02.g1
HMG (high mobility group) box IPR000910	eukaryote	8.70e-07	1	CL382Contig1
Toprim domain IPR006171	general	1.20e-06	1	aab08h01.g1
Jumping translocation breakpoint protein (JTB) IPR008657	eukaryote	1.50e-06	1	CL771Contig1
NUC153 domain (nucleolar localisation) IPR012580	eukaryote	1.80e-06	1	aaa92e07.g1
RhoGAP domain IPR000198	eukaryote	2.90e-06	1	CL797Contig1
Ligand-binding domain of nuclear hormon IPR000536	eukaryote	6.70e-06	1	aaa91f05.g1
Ankyrin repeat IPR002110	general	1.80e-05	1	aab18e11.g1
Ground-like domain IPR007284	nematode	7.00e-05	1	aab11h03.g1
CD36 family IPR002159	eukaryote	7.40e-05	1	aab23d02.g1
Pepsin inhibitor-3-like repeated domain IPR010480	nematode	0.00019	1	CL533Contig1
Survival motor neuron (SMN) interacting prot IPR010304	metazoa	0.00089	1	CL916Contig1

* Specificity of the domain as indicated.
** No.: number of *R. similis* unigenes matching to the domain, one is reported in the last column.

4. Discussion

With the generation of thousands of new EST sequences from mixed stages of the plant-parasitic nematode *R. similis*, interesting research topics have been made possible. Based on our analysis, the *R. similis* ESTs are derived from three different sources. The majority of the unigenes are derived from the nuclear genome. A small fraction (~0.6%) has most likely a mitochondrial origin, corresponding to sequences with a very low GC content (~16% GC). Finally, a third subset (~1%) seems to be derived from a *Wolbachia* species. To our knowledge, this obligate intracellular endosymbiont has only previously been reported from a few filarial nematode species (Hise et al. 2004; Kramer et al. 2003; Taylor et al. 1999). In these nematode species, *Wolbachia* seems to be required for successful molting as well as for reproduction of the nematode. In only three genera of plant-parasitic nematodes (*Heterodera*, *Globodera* and *Xiphinema*), bacteria-like endosymbionts - other than *Wolbachia* - have been found (Noel et al.

2006; Vandekerckhove et al. 2002b). Localization studies by this thesis confirmed the presence of an endosymbiotic species in *R. similis*, which reside in the ovaria. If future investigations reveal that the bacterium is required for *R. similis* to develop, it could even open new ways for *R. similis* specific control.

The majority of the unigenes are likely to be derived from the nuclear genome of *R. similis*. Approximately one quarter of the unigenes code for proteins involved in general metabolic pathways. Other classifications based on BLASTx results can be found in figure 2, but can change slightly in the future as more sequence data become available. Besides the unigenes with clear homology, it was found that a relatively large part of our unigene dataset (30.8%)
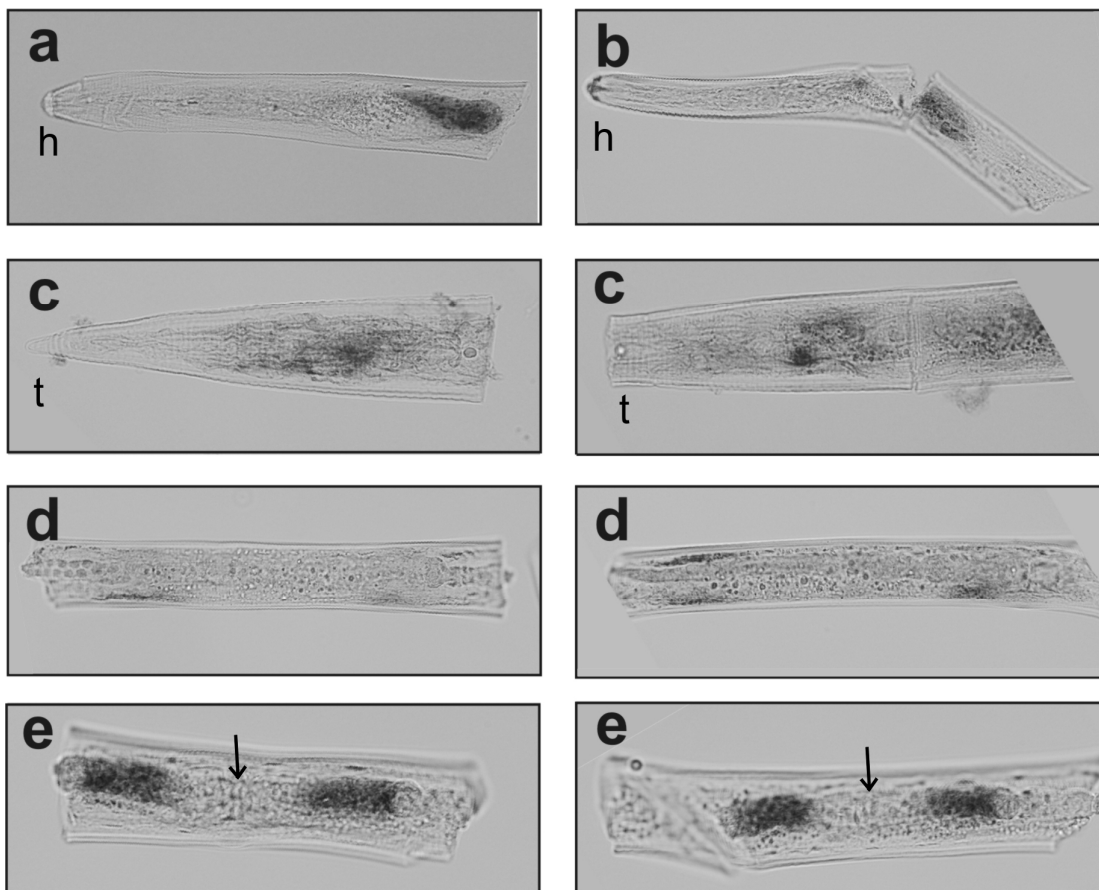

Figure 18: *in situ* hybridisation results of putative parasitism genes of *Radopholus similis*. **a,** positive control (endoglucanase 2, accession number EF693942.1). h: head. **b,** 92h12 shows expression in the pharyngeal gland cells. h: head. **c,** cluster 546 shows a strong signal in the tail region. This specific signal could be located in the retrovesicular ganglion. t: tail. **d,** cluster 26 shows signal in patches along the epidermis. **e,** 23g06 shows very strong signal in the region around the vulva (arrow), in the vicinity of the spermatheca. Negative controls did not shown any signals (data not shown).

2

lacked homology to any sequence in the database to date (called 'orphan' sequences). Multiple explanations can be found for these orphans: 1) the most 'preferred' and exciting one is that the unigene represents a genuine novel gene (estimated to be 40% of the orphans by the protein coding prediction program). However, 2), the length of the unigene plays also a role, as a correlation exists between the length of a unigene and its homology significance level. Thus homology can simply not be detected if the sequence is too short. Alternatively, 3), unigenes containing mainly untranslated region (UTR) will most likely lack homology, as UTRs are the most diverse regions of transcripts (Ko et al. 1994; McCarter et al. 2003). Moreover, 4), for unigenes derived from (non-coding) contaminating DNA, possibly no significant homology will be detected. But still another exciting possibility is, 5), that some unigenes of the orphans correspond to non-coding RNAs rather than mRNA, especially since evidence is accumulating on the ubiquitous role of these non-coding RNAs in translational regulation. Recent estimates in humans state that at least 20% of the genes are regulated by over 1,000 miRNAs (Bentwich et al. 2005; Berezikov et al. 2005; Lim et al. 2005; Xie et al. 2005) and in the model nematode *C. elegans*, 112 miRNA genes have been identified so far (Ohler et al. 2004; Ruby et al. 2006). But one of the most intriguing orphans is notably the largest cluster in our dataset, representing about 8% of the ESTs. In fact it is frequently reported in the literature that the largest clusters in EST analyses contain a disproportionate large number of ESTs compared to smaller clusters (the largest cluster contained 10% of the ESTs in Mitreva et al. (2004b), 5% in Mitreva et al. (2004a), 2% in Kikuchi et al. (2007), 7.36% in Abernathy et al. (2007), ~5% in Maizels et al. (2000), and 4.7% in Ranganathan et al. (2007)). Often these clusters do not show homology to any known sequence (Dubreuil et al. 2007; Maizels et al. 2000; Ranganathan et al. 2007). Unfortunately, not many attempts have been undertaken to clarify this. We suggest that cluster 1 has most likely a mitochondrial origin. If so, it should be a part of a transcriptionally active region with a very high expression level.

Another remarkable result from this analysis is the impossibility – despite the use of various approaches - to extract sequence fragments from the unigene dataset that resemble *trans*-spliced leader sequences, known to occur in other nematode species, such as *Meloidogyne incognita* and *Caenorhabditis elegans*. This could indicate that *R. similis* makes no use of *trans*-splicing. This seems unlikely based on the statement that *trans*-splicing is widely occurring throughout the phylum Nematoda. This process is already reported to take place in (amongst others) Caenorhabditis *elegans*, *Meloidogyne incognita*, *Pratylenchus penetrans*, *Globodera* and

Brugia malayi (Blaxter et al. 1996; Mitreva et al. 2004; Stratford et al. 1994). Therefore it seems more likely that *R. similis* makes use of a different set of spliced leader sequences. However, no common 5' sequence motifs could be extracted, which may be due to the large incidence of 5' end truncation observed in EST data. The lack of known *trans*-spliced leader sequences could also explain the difficulties encountered when constructing an oligo(dT)-SL1 PCR based cDNA library. This library turned out to have a low number of primary transformants, which could have been due to the unsuccessful amplification step. Remarkably, the same negative result was obtained with a similar preliminary search in the EST data from *Heterodera glycines* for known *trans*-spliced leaders, despite the various techniques used to construct the investigated libraries (data not shown). These results point to the need for a thorough investigation to validate the systematical occurrence of *trans*-spliced leader sequences (known and unknown) throughout the phylum Nematoda. Some comparisons between those four nematode species can be found in Table 8.

Unigenes correlated to the plant-parasitic life style of *R. similis* were found through homology with sequences of parasitic species, corresponding to genes with known and unknown function. An approach that has proven useful in discovering new parasitism genes is searching unigenes coding for secreted proteins. It is assumed that the majority of the parasitism genes are secreted into the host to modulate the nematode's environment. Most of them originate from the pharyngeal glands and are injected into the plant tissue through the stylet. We identified known parasitism genes based on homology, and additional novel parasitism genes through PPN-specific expression combined with *in silico* signal peptide identification.

Table 8: General comparison between 4 different nematode species

	Radopholus similis	*Heterodera glycines*	*Meloidogyne incognita*	*Caenorhabditis elegans*
Number of ESTs	7007	24,444	20,334	346,107
Genome size	~ 20 Mb (a)	92.5 Mb (b)	51 Mb (c)	100.2 Mb (d)
Spliced Leader	Not found	Not found	SL1/SL2	SL1/SL2
GC% / GC3%	49% / 65%	50% / 56% (e)	37% / 27% (e)	43% / 40% (e)
Trophic ecology	Endo migratory parasite	Endo sedentary parasite	Endo sedentary parasite	Free living

a) as inferred for *Pratylenchus coffeae*, a supposed close relative of *R. similis* (Leroy et al. 2007); b) Opperman et al. (1998); c) Hammond et al. (1992); d) Stein et al. (2003); e) Mitreva et al. (2006)

Interestingly, one of the putative novel parasitism genes (92h12) shows expression in the gland cells. A homologue was identified in *Heterodera glycines*, which is expressed in the subventral gland cells during the migratory and first sedentary stages of this species (*Heterodera glycines* gland-specific clone g4e02, accession number AAO33473.1, see Figure 19) (Gao et al. 2003). *gus* and *gfp*-fusion constructs in *Arabidopsis thaliana* confirmed that the *H. glycines* homologue is targeted to the plant-host nucleus (Elling et al. 2007a). Similarly, PSORTII predicts the 92h12 unigene to be localised in the nucleus. These results could indicate direct regulatory activity in the host nucleus of these alpha-helix rich secreted protein products. In addition, the identification of this gene may point to a more complex parasitism behaviour of *R. similis*, in which modulating the plant cell processes before withdrawal its contents plays an important role.
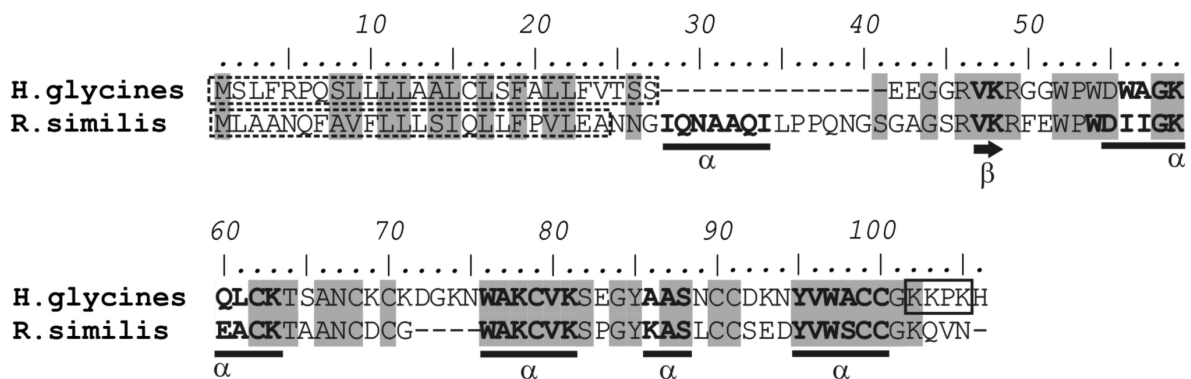

Figure 19: Alignment of EST 92h12 of *R. similis* with its homologue of *H. glycines* (g4e02). Identical residues are shaded. The peptide signal for secretion is boxed by dashed lines. Predicted secondary structures are indicated by a balk (α-helix) or an arrow (β-sheet) with the residues contributing to the corresponding structure in bold. The nuclear localisation signal on the *H. glycines* homologue is boxed.

Although GO classification brings much needed structure, it is not yet very suited to deal with symbiont-host interactions, and processes involved in parasitism. However, the recently established 'Plant-Associated Microbe Gene Ontology' (PAMGO) consortium is dedicated to fill this gap by expanding the GO with terms specifically applicable to the symbiont-host relationship, including terms specifying PPN-host interactions (PAMGO consortium 2009; Torto-Alalibo et al. 2009). Nevertheless, comparative results based on GO remain often difficult to interpret, especially when comparing heterogeneous small datasets (such as EST libraries), or when a high portion of novel genes with unknown function is present.

It is remarkable that the majority of the *R. similis* unigenes with homology exclusively to PPN lack a BLASTx-hit. This indicates a high potential for discovery of novel genes in PPN EST sequences. Besides the elucidation of the parasitism process as stepping stone to parasite control, GO and RNAi-phenotype data analysis suggest that suitable targets for controlling *R. similis* may also be found among genes involved in the regulation of processes. Recent experiments with RNA interference show that viability of nematodes can be severely affected when essential nematode genes are silenced (Kamath et al. 2003). A promising technique in this respect is the *in planta* generation of nematode specific double stranded RNA, leading to a decreased viability of the nematode when it ingests the inferring RNA molecules (Bakhetia et al. 2005; Gheysen et al. 2007b). The sequences delivered by this EST project can therefore aid efficient parasite control.

5. Material and Methods

a. Biological material

R. similis was cultured at 25°C on carrot disks in small parafilm sealed petri dishes (Huettel 1985; Jacob et al. 2007). Approximately 5000 mixed stages nematodes were collected in sterile demineralized water. After grinding these nematodes in liquid nitrogen, RNA was extracted with TRIzol® Reagent (Invitrogen, Carlsbad, USA), precipitated with isopropanol and washed with 70% ethanol. The pellet was redissolved in diethylpyrocarbonate(DEPC)-treated demineralized water. Integrity of the RNA was checked by electrophoresis on a 0.5X TAE 1% agarose gel. Concentration was determined with the ND-1000 spectrophotometer (Nanodrop, Wilmington, Delaware USA). This RNA served as a basis for cDNA library construction using the SMART™ cDNA Library Construction Kit, following the manufacturer's instructions (Clontech, Mountain View, USA). The resulting fragments were directionally cloned in the pDNR-Lib vector provided with the kit. The *Radopholus similis* mixed stage cDNA library contained over 10^5 primary transformants. Clones were sequenced using the M13 forward or reverse primer at the Genome Sequencing Center (GSC, Washington University, St.-Louis, USA). Sequences and quality files can be found on Nematode.net (Wylie et al. 2004), and sequences were submitted to the EST division of GenBank (dbEST) (Boguski et al. 1993).

b. Cleaning and clustering

The sequences were cleaned using Seqclean (www.tigr.org) with a locally downloaded vector database and default parameter settings, to remove vector, poly(A) and short (<100 bp) sequences. Next, the dataset was clustered using TIGR Gene Indices Clustering Tool (TGICL) (Pertea et al. 2003), and assembled sequences were constructed

by CAP3 (Huang et al. 1999) using default settings, generating contigs (clustered ESTs) and singletons (non-clustered ESTs), commonly referred to as 'unigenes'. Based on the clustering results, ESTstat was used to estimate the degree of fragmentation (Wang et al. 2006). To gather the ESTs with poly(A) sequences (and thus containing 3' untranslated regions), cleaning by Seqclean was performed also without poly(A) screening (option −A). The sequences differently cleaned compared to the first cleaning were gathered. The results of the clustering can be found in the online additional material of Jacob et al. (2008).

c. BLAST searches

The Basic Local Alignment Search Tool (BLAST) analyses (Altschul et al. 1990) were performed both locally and via netblast. The BLASTx results were parsed by in-house perl scripts: for each hit, the species and phylogenetic classification was obtained from its GenBank file and used for subsequent classification of the unigene query (as nematode-, animal-, eukaryote-specific, etc.). *R. similis* unigenes were also used for BLASTx against *C. elegans* sequences (E-value cut-off of 1E-05), and the top-hit sequences were used to estimate the degree of fragmentation (Mitreva et al. 2004). For this estimation, 384 of 1632 unigenes with *C. elegans* hits share the same top-hit with one or more other unigenes. Of these unigenes, 221 were 'redundant' as the *C. elegans* top-hit was already detected by (an)other unigene(s), from which the fragmentation can be estimated. Further, all nematode EST sequences were downloaded (June, 2007) and searched locally with tBLASTx for homology to the unigenes of *R. similis*. Of those tBLASTx hits, the developmental stage and nematode species used for the cDNA library construction, were parsed from its GenBank file. A tBLASTx search was performed with the *R. similis* unigenes as query against the coding mitochondrial sequences of all nematode species available in GenBank (September 2007). To address the nature of cluster 1, coding and non-coding classification was done for this cluster by ESTScan (trained with the invertebrate *Drosophila* sequences) (Iseli et al. 1999) and RNAz (Washietl et al. 2005).

d. Translation

The FrameD gene prediction and translation program (Schiex et al. 2003) was trained by manually selected full length coding open reading frames (ORFs) from the set of unigenes, based on the BLASTx results and CLUSTAL W alignment with the corresponding most homologous sequences (Thompson et al. 1994) (ORFs of cluster 4, 5, 8, 11, 12, 13, 15, 18, 23, 25, 28, 31, 35, 41, 42 and 43). This set was extended with 7 full length coding sequences of *R. similis* yet in GenBank (accession numbers AM691117.1, AM691118.1, EU190885, EF693940, EF693941, EF693942, EF693943), resulting in a total of 15,069 coding nucleotides. Using these sequences, the N resistant Markov model was build on the website of FrameD (http://bioinfo.genopole-toulouse.prd.fr/apps/FrameD/FDM.pl). Predictions were analyzed for the occurrence and position of coding sequence part of the unigene. GC content was calculated using an in-house perl script. Other unigenes which were predicted to be non-coding by FrameD, but with homologous proteins in other species according to BLASTx, were translated using Prot4EST trained with *H. glycines* sequences (Wasmuth et al. 2004). Signal peptide prediction on this set of translations was done by SignalP 3.0 (Emanuelsson et al. 2007) and a signal peptide was only assigned if both the neural network and the hidden Markov model predicted a signal peptide. The signal peptide was cleaved from the translated sequence and

subsequent transmembrane domain prediction was performed by TMHMM (http://www.cbs.dtu.dk/services/ TMHMM/) and SOSUI (Hirokawa et al. 1998). A protein was assigned to reside in the cell membrane, if both programs predicted a transmembrane domain. PSORTII (http://psort.nibb.ac.jp/) was used to predict the subcellular localisation.

e. *Trans*-spliced leader search

Searches were based on the spliced leader (SL) sequences reported in Guiliano and Blaxter (2006). In a first approach, a BLASTn search was set up with the SL sequences as query, a minimum matching length of 20, and a cut-off E-value of 0.1. Using these parameters, a search was done in the *R. similis* unigene dataset, and also in a locally downloaded set of EST sequences of the plant-parasitic nematode *Meloidogyne incognita* as a control, since SLs are shown to occur in *M. incognita* (Guiliano et al. 2006; Koltai et al. 1997; McCarter et al. 2003). To minimize influence of a technical nature, we selected an EST set of *M. incognita* of a similar size as ours and generated by the SMART technology was used. The chosen control EST set has 3,098 ESTs sequences from a SMART cDNA library, constructed from females of *M. incognita* (library "*Meloidogyne incognita* female SMART pGEM"). Since the BLASTn search yielded no results, a second approach was applied using a perl regular expression pattern, based on common features extracted from the SL sequences and based on the 5' position of the SL in the unigene. The resulting pattern /^[AGCT]{0,30}[GGT[^CG]{4,9}CCC[^C]\w{5,9}AG)/ was used to search the SL sequences set, the *R. similis* unigenes (both strands) and the M. incognita ESTs (both strands). As a confirmation, the sequences of *M. incognita* found to contain a SL sequence, were used for a tBLASTx search as a query (E-value cut-off of 1E-35), simultaneously against *C. elegans* EST sequences (from dbEST) and against *R. similis* unigenes, containing full length coding sequences and sequences with at least a 5' UTR sequence part (our data). When highly similar ESTs (ranging from 60% to 93% identity on the protein level) in both datasets were found, the corresponding ESTs were aligned (on the DNA level) using ClustalW.

f. Gene Ontology and KEGG biochemical pathway annotation

To map and annotate gene ontology (GO) terms, BLAST2GO was used (Conesa et al. 2005), with default parameters, except for an BLAST E-value cut-off of 1E-05, maximum number of 30 BLAST hits, E-value hit filter for annotation of 1E-05, the conversion of the annotation to GOSlim view, and a node scoring filter in the GO graph of 50 for biological process, 20 for molecular function and 20 for cellular component. Further, KOBAS (http://kobas.cbi.pku.edu.cn/index.jsp) was used to annotate KEGG biochemical pathways to the unigenes (Mao et al. 2005).

g. Annotating RNAi data to the unigenes

Using the RNAi data available of numerous *C. elegans* genes, we tried to assign an RNAi phenotype to the *R. similis* unigenes. A BLASTx search revealed the top-hit *C. elegans* sequence for a unigene (using cut-off E-value of 1e-05). Subsequently, the RNAi phenotype and GO terms (only of the *C. elegans* top-hits with observed RNAi

phenotypes) were retrieved via WormMart (Schwarz et al. 2006) and the results analyzed and visualized with WEGO (Ye et al. 2006).

h. Whole mount *in situ* hybridization

The whole mount *in situ* hybridization protocol was performed as described by Vanholme et al. (2002) with minor modifications. Nematodes were fixed in 3% paraformaldehyde for 15 hours at 4°C, followed by an additional incubation for 4 hours at RT. Hybridization was done overnight at 47°C. Digoxigenin(DIG)-labeled probes were generated by linear PCR with a single primer (rv-primers) and an appropriate cDNA fragment cloned into pGEM-T as template in the presence of DIG-labeled nucleotides. Used primers are listed in Table 9. As a positive control, an antisense probe against endoglucanase 2 (accession EF693942.1) was used.

Table 9: Primers used for cloning and probe generation.

Primer name	Primer sequence
CL26_fw	GACTGTTGGTCCCACTTTCC
CL26_rv	TGCTGCTTCTCATTGTCTGC
CL546_fw	GCGTCTGATCGTTGTTTCC
CL546_rv	CGGAGTTTTGTTTCGACAGG
CL793_fw	AGTGTCCACTGTTCGCTTCC
CL793_rv	CACTGTCGACAACCACAACC
92h12_fw	TTTTCTCCTGCTCTCGATCC
92h12_rv	CTTCAACTGATTTCCATTCTGC
23g06_fw	CCTGTCTGTTCCTCCTGTCC
23g06_rv	AGTCGGGCTGGTAGATGG
EG2_fw	ATGCTTAGCCTGTTCCTTTTGCC
EG2_rv	GTGAAGAAGGAGACCGCCTG

i. DAPI staining

R. similis nematodes were fixed in 50% acetic acid and DAPI (dissolved in methanol) was added drop by drop to a final concentration of 100 ng/ml. Nematodes were mounted on a glass slide, washed and embedded in Vectashield (Invitrogen) to reduce photobleaching. Nematodes and fluorescence signals were visualized with a Nikon TE 2000-E inverted microscope, equipped with a 100× oil objective (NA 1.2, Plan corrected) and a standard Nikon RGB camera. Excitation and detection was performed with filter cubes of the following composition (EX: excitation, DC: dichroic, EM: emission): EX 330-380; DC 400; EM 420 LP. Images were acquired using NIS-Elements software version 2.10.

j. TEM

Nematodes were fixed in Karnovsky solution at 60°C for 30 minutes, after which head and tail was removed. After additional fixation overnight at 4°C, the fixative was removed by washing three times in 0.134M sodium

cacodylate buffer (pH 7.2), followed by an overnight post-fixation in reduced osmium solution. After washing, samples were dehydrated in series of ethanol dilutions, transferred to absolute ethanol and CuSO4 bars were added to remove any remaining water. Specimens were infiltrated with a low-viscosity embedding medium and polymerized at 70°C for 8h. Ultrathin (70nm) longitudinal sections were cut, followed by poststaining with uranyl acetate and lead citrate stain (EM stain, Leica). Sections were visualized wit a Jeol JEM 1010 (Jeol Ltd., Tokyo, Japan) transmission electron microscope operating at 60 kV.

6. Acknowledgements

Chapter 3

Four transthyretin-like proteins of the migratory nematode *Radopholus similis*: members of an extensive nematode-specific family

3

This chapter is adopted from:

Four transthyretin-like genes of the migratory plant-parasitic nematode *Radopholus similis*: members of an extensive nematode-specific family

3

1. Abstract

Seven tags coding for proteins holding a transthyretin-like domain (PF01060) were identified in the ESTs of the plant-parasitic nematode *Radopholus similis*. The seven ESTs corresponded to four different genes, which were cloned from a cDNA library (accession numbers AM691117, AM691118, AM691119, AM691120). Transthyretin-like genes belong to a large family, different from the transthyretin and the transthyretin-related genes with whom they share some sequence similarity at the protein level. This similarity has caused an inconsistent use of different names and abbreviations in the past. To avoid further confusion, we introduce a standardized nomenclature for this gene family, and chose to name this barely characterized gene family TTL (as for transthyretin-like). Further examination of the identified genes, named *Rs-ttl-1 to -4*, showed that they are expressed in both juveniles and adults, but not in developing embryos. Whole mount *in situ* hybridization revealed a distinct spatial expression pattern for two of the genes: *Rs-ttl-1* is expressed in the tissues surrounding the vulva, whereas *Rs-ttl-2* is expressed in the ventral nerve cord. The deduced protein sequences contain a putative signal peptide for secretion, pointing to an extracellular function of the mature proteins. Database screens showed that the TTL family is restricted to nematodes. Moreover, a HMMER search revealed that ESTs derived from TTL genes are more abundant in parasitic nematode libraries, with a bias towards the parasitic stages. Despite their abundance in nematodes, including the extensively studied model organism *Caenorhabditis elegans*, the function of TTL proteins remains obscure. Our data suggest a role in the nervous system.

However, even without insight into their biological function, the nematode-specific nature of this gene family makes it a promising target for nematicides or RNAi mediated control strategies against parasitic nematodes.

2. Introduction

It has been postulated that future control techniques against parasitic nematode may rely on disruption of gene expression by means of promoting post-transcriptional gene silencing through the application of antisense RNA molecules (RNAi) (Gheysen et al. 2007; Huang et al. 2006a). Due to several practical concerns, this approach preferentially requires the use of nematode specific genes. Exploration of the vast amount of available nematode EST sequences revealed that many gene families meet those requirements, as an estimated 8.6% of the conserved gene families are specific for the phylum (Parkinson et al. 2004). One such conserved family is the transthyretin-like gene family. The corresponding proteins of these genes are characterized by the transthyretin-like domain (PF01060; IPR001534; DUF290), which is one of the most abundant nematode-specific domains (McCarter et al. 2003). Transthyretin-like (TTL) proteins were first described in *Caenorhabditis elegans* by Sonnhammer and Durbin (1997). However, despite their extensive occurrence throughout the phylum Nematoda, barely any studies have been performed to investigate this gene family and to elucidate their function.

As can be deduced from its name, transthyretin-like proteins have some similarity to transthyretins (Sonnhammer et al. 1997), which are characterized by the transthyretin domain (PF00576; not to be confused with the transthyretin-like domain). Transthyretins are specific to vertebrates and transport thyroid hormones (as a di- or tetramer) as well as vitamin A complexed with retinol binding protein in extracellular fluids of vertebrates (Palha et al. 2002). Besides TTL proteins, a second group of proteins with similarity towards transthyretin exists, namely the transthyretin-related proteins (TRP) (Eneqvist et al. 2003). These proteins contain the transthyretin domain (PF00576) and possess a characteristic Y-[RK]-G-[ST] peptide at the C-terminal end. In contrast to transthyretin and transthyretin-like proteins, TRPs are found in a much broader range of species, including bacteria, plants, vertebrates and invertebrates (Eneqvist et al. 2003; Lundberg et al. 2006). TRPs seem to perform very different functions in

different species: in plants they have a role in brassinosteroid signaling, whereas in mouse they apparently function in the uricase reaction pathway as 5-hydroxyisourate hydrolase (Lee et al. 2006; Li 2005).

In this chapter, analysis of 1,154 ESTs of *R. similis* revealed four different genes coding for transthyretin-like proteins. This analysis was performed in parallel to generating the additional cDNA library described in chapter 1. The 4 TTL genes were cloned and their spatial and temporal expression patterns were characterized. The distribution of the gene family in the phylum Nematoda was examined, and the relationship with transthyretin and TRP was determined. Finally, we discuss the unclear nomenclature used by previous papers and suggest some standardized nomenclature to annotate the different gene families.

3

3. Results

a. Identification of TTL proteins of *R. similis*

A small scale EST-sequencing project on mixed developmental stages of *R. similis* resulted in 1,154 ESTs. Clustering the ESTs revealed 127 contigs and 419 singletons, corresponding to 546 putative unigenes. A HMMER search on the translated putative unigenes using the transthyretin-like domain profile from the Pfam database reported the presence of four proteins containing the transthyretin-like domain. The four corresponding unigenes consisted of three contigs (Cl102; Cl113; Cl138) and one singleton (51334193), clustered from 7 ESTs (GenBank accession numbers CO897702, CO961414, CO961279, CO897942, CO961009, CO961234, CO961285). BLASTP searches with the four protein sequences against the non-redundant protein database resulted in numerous matches with proteins of *C. elegans* and *C. briggsae*. Besides these matches, the transthyretin-like proteins of *R. similis* showed similarity to Xi-TTR-1 (CAH89266) and Xi-TTR-2 (CAH89267) of the migratory PPN *Xiphinema index* and a 'hypothetical gland cell secretory protein 11' (HSP11; AAF76925) of the sedentary PPN *Heterodera glycines*. All retained matches were of nematode origin (E-value cut-off 0.1).

b. Full length coding sequences and corresponding proteins

Full length cDNA could be obtained for Cl102 and Cl113. For the singleton 51334193 and contig Cl138 we were not able to clone the 5'-end regions of the coding sequences. However, based on homology to *C. elegans* proteins, we estimate that only 3 (51334193) and 9 (Cl138) amino acids (AA) were missing at the N-terminal end of the corresponding proteins. A signal peptide for secretion was predicted in the protein sequences derived from Cl102, Cl113 and 51334193. A signal peptide was not predicted for Cl138, most likely due to the absence of AAs at the N-terminus of the predicted protein sequence, since the most homologous *C. elegans* transthyretin-like protein (NP_505561) does posses a signal peptide. To avoid confusion with transthyretin-related proteins (see section e), the genes were named *Rs-ttl-1* (Cl102), *Rs-ttl-2* (Cl113), *Rs-ttl-3* (51334193) and *Rs-ttl-4* (Cl138) as for t̲r̲ans̲t̲h̲yretin-l̲ike genes (with GenBank accession numbers AM691117, AM691118, AM691119, AM691120 respectively). The mature protein sequence of *Rs-ttl-1* differs from the other proteins in having a spacer of 10 AAs preceding the transthyretin-like domain. The mature protein sequence of *Rs-ttl-4* has a basic C-terminal extension of 22 amino acids. The secondary structure of the Rs-TTL proteins was predicted to consist of beta-strands, connected to each other by short loops. No alpha helices were predicted. The four proteins have 8 putative beta-strands in common. One additional beta-strand was predicted in the short N-terminal extension of Rs-TTL-1, whereas three additional predicted strands were found in the C-terminal extension of Rs-TTL-4. Some cysteines predicted to be involved in the formation of disulfide bridges are conserved among the Rs-TTLs. Rs-TTL-3 and -4 have 2 cysteines, whereas Rs-TTL-1 and -2 have one additional cysteine pair. These covalent bridges can provide inter-chain stabilization and can be important for extracellular proteins. Other predictions on the processed proteins are summarized in Table 10 and Figure 20.

Table 10: Characteristics of mature Rs-TTL proteins.

TTL	Length (AA)	MW (kD)	pI
Rs-TTL-1	131	14.5	4.93
Rs-TTL-2	115	13.0	5.50
Rs-TTL-3	121	13.3	4.97
Rs-TTL-4	144	16.1	8.87

c. Temporal and spatial expression pattern

To study the expression level of the four identified TTL genes in different life stages of *R. similis*, a semiquantitative RT-PCR was performed. Expression of the genes could not be detected in developing embryos. In juveniles, *Rs-ttl-1* and *Rs-ttl-4* have higher expression levels than *Rs-ttl-2* and *Rs-ttl-3*. Interestingly, this pattern is reversed in adults, where *Rs-ttl-2* and *Rs-ttl-3* are more abundantly expressed compared to *Rs-ttl-1* and *Rs-ttl-4* (Figure 21). To reveal the spatial expression pattern of the cloned genes, a whole mount *in situ* hybridization was performed. Since such a technique was not yet available for *R. similis*, the protocol was optimized with *Rs-sec-2* as a positive control. The spatial expression pattern of this gene was restricted to the epidermis (Figure 22), which is in agreement with the expression pattern of *sec-2* in other nematode species (Prior et al. 2001; Vanholme et al. 2002). Using the same conditions we were able to demonstrate a specific expression of *Rs-ttl-1* around the vulva, whereas the *Rs-ttl-2* expression was restricted to the ventral nerve cord (Figure 22). For *Rs-ttl-3* and *Rs-ttl-4*, variable staining could be observed in multiple parts of the nematode (data not shown). Searching Wormbase returns only two TTL genes of *C. elegans* with known expression patterns: *ttr-8* (R13A5.6, NM_066256) is expressed in the hypodermis, tail neurons and head neurons, and *ttr-51* (JC8.8, NM_070148) in the spermatheca.

3

Figure 20 (see next page): Alignment of parasitic nematode TTL proteins with one TRP (*Agrobacterium tumefaciens*) and one transthyretin sequence (*Gallus gallus*). Red boxes delineate predicted signal peptides. Black triangles above the alignment point to cysteines in the Rs-TTL sequences predicted to be involved in disulfide bonds. Rs-TTL-5* (CL41Contig1) and -6* (CL111Contig1) are two additional TTL genes identified in additional EST sequences (see chapter 2), and for completeness presented in this alignment. *Green /light green* bars represent resp. conserved/similar residues in the TTL sequences; *orange /light orange* bars (also indicated by dots) represent resp. conserved/similar residues in TTL, TRP and transthyretin sequences; *Blue /light blue* bars represent resp. conserved/similar residues in TRP and transthyretin sequences, and together with the *orange* bars, these residues are found to be conserved in TRPs and transthyretins as defined by Eneqvist et al. (2003). *Blue* stars mark residues aligning the binding channel in TRPs and transthyretins. The AAs in *blue* delineate predicted (for Rs-TTLs) and confirmed (for TRP and transthyretin) beta-strands; the AAs enclosed by a *rectangle* delineate the confirmed alpha-helix in TRPs and transthyretins. The AAs in red indicated for Rs-TTL-5* and -6* represent an alpha-helix structure. *Grey arrows* and the *grey box* indicate the position of beta-strands and the alpha-helix respectively: above the alignment as predicted for Rs-TTLs, under the alignment as confirmed for the 3D structure of TRPs and transthyretins (Blake et al. 1978; Hennebry et al. 2006; Lundberg et al. 2006; Sunde et al. 1996; Wojtczak 1997).

Common secondary structure elements as predicted for Rs-TTL:

pfam01060 domain   rkgsvavkgrlicggkp--aenvkvklyekdtgidpddlldstytde

Rs-TTL1	MAPMFLPTVSVFLLLLVVVQQSLLVLA	SPK**PLFGIGI**GRTQ**SAGVEGTLLC**EGKP--MAD**VLVKLYD**DDRGVDTDDLMAEGKTDS
Rs-TTL2	----MRLLFCAFLLLAVLCAFALC	-----------DV**QNITVRGQTIC**RRRS--VKA**LAVELRE**HDTF-DPDDALGS**TTT**GP
Rs-TTL3	---...LFAVVLFAIITIVPSPCGA	-----------MRQQ**AVAVKGQLMC**GSRP--ASG**VKVKLWD**EDDGPDPDDVLDE**AFT**DS
Rs-TTL4	-------...CFCSLGSA	-----------IRTQ**SAAVKGRLMC**GAAP--ARN**VRVKLF**DEDSGPDPDDLDAGY**YT**DS
Rs-TTL5*	----MHPVLVVCLGLSLAALGEA	-------GVGLLGS**EQSAAVRGVLMC**NGRP SVG**TKVRLYD**ED-DTDM**DDLM**AEG--VTDR
Rs-TTL6*	----MESLLPTLLLFLAVFSCSAA	-----------RFQQ**SVGIRGRLMC**GGQP LK**ETQVKLW-N**KNTLGTD**DQL**AAG--KTDN
Xi-TTR1	-MEKAVLVVCVFFCLLPVTLS	-----------SLGQRQRVIVKGRLLCGNAP--ASNIRVKLVDEDDGPDPDDMDDGYTND
Xi-TTR2	--MTSSLSVFAVLSVICMMAGEADS	WEGTLAVVGKKSVRRSYSVGRLLCGEKP TNAKDTEITLLDKKGGLDKDITMAKIVVKN
Hg-HSP11	-----MFRFSLLLLLVPLVRC	-VPNPKPLFGIGIGRKQSAGAEGKLTCAGEP--LADVKVKLYDDGHVHMGETRTDS
Ac-TTL	-------MQHIIFLLATVVCCLA	-----------VREQAVGVTGRLMCGNKP--AVGVKLWEEDGPDPDDLLDQGFTDS
At-TRP	--------------------------MT**GLTTHVLD**AAHGVP--AE**GLTIELYR**LS--GDR**REKLKTVKT**NS
Gg-TTR	------MAFHSTLLVFLAGLVFLSEA	-APLVSHGSVDSKC**PLMVKVLD**AVRGSP--AA**NVAVKVFK**KAADGTW**QD-FATGKT**TE
TTR domain		gpltthvldtasgkp aagvkvelfrleepgqgwellatgktna

Secondary structure as determined for TRP and transthyretin:

A   B   C

---

ngefrlsgstteitaifdpylniyhdcndvd----kepcyrkftitipdeyitsgktpkktfdigtlnlansfpgetrdcin

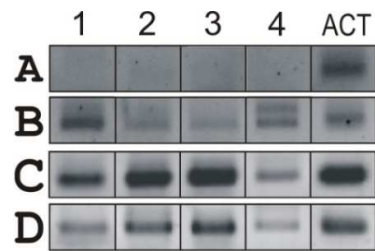| |
|---|
| KG**RFSLEGYTHEFTTI**-DP**KINIYHD**CND-----LLPCQ**RKISIMI**PDK**YI**SSGKTPER**YCNAGEVELE**GKFKGEERDCLH--------- |
| DGT**FEVRG**TENE**VGSI**R-P**YLRITEK**CDVKD----ELK**RRVTEIDI**PKD**KVN**HGVYE---**MNFINS**NVPGHR--DTETCD------- |
| SGA**FHLKGSERELTNI**-DP**VLKIYHD**CDDG-----IMPG**RKIKLRI**PGS**YI**TGGGVAKV**VFDVGLLN**LETKFPGEERDLL------- |
| NGN**FQLKGSTTEATNI**-DVV**FKIYHD**CDDG-----IKPGS**RKIKFKI**PNS**YV**SEGATPKK**TFDIGTLNLETI**FANEERELI**VS**KK**RRHLHH**YRGGEKF**NFFVR**L |
| QGG**RFQLSG**HVK**EVTTI**-DP**KLNVYH**DCNDVHV-----PCK**LKFSIKL**PDG**YV**SKGTVA**RKVFDAGTLNLEG**AFSGQSRDCVN------- |
| NGG**NFELQGGVGQLTKMS**-VH**VVYHD**CDDG-----VKPCQ**RKVDLGV**PDQ**YV**QRADRVS**KWFEAGTMNMEF**KFPDEDRSCIN------- |
| NGEFLLDGQQTEISPI-DP**VLKIYHD**CNDG-----LPCQRKWRFKIPNKY**IV**DPEDSSTVMMGTWNLEPIADDEERDCIH------- |
| DGTFRIDGISTGFLDIN-PRMY**IYTSSNKG**-----MNPCQKMWKMPLPKL**YQ**NTPEG----YHLGTWNLPLEMD------- |
| EGRFRLEGYTHE**ITTI**-DP**VLKIYHD**CDDG-----LKPGQ**RKISIMI**PDK**YI**ASGEHPNTYYDAGTVELEGKFSGETRDCLH------- |
| NGN**FNLKG**SERELTNI-DP**VFKVYHD**CDDG----FY----------------------RD------- |
| DG**RV**DGGPLLVGDSFK-**GEYELVFH**AGDYLRGRGVQ LAEPAFLDI**IPIRFGIA**DE-SGH**YHVPLLLSP**YSY**STYRGS**------- |
| FG**EI**H--ELTTEEQFV-E**GVYRVEFD**TSSYWKG---LGLSPFHEY**ADVVFTAN**DSGHR**HYTIAALLSP**FSYS**TTAVV**SDPQE------- |
| dgrid aplltgetla pgiyrlefdtgdyfkargvaladppflpyvpvrfgiadaghqhyhvplllspygystyrg |

D   E   α   F   G   H

Figure 2I: Semiquantitative RT-PCR. Amplification was performed on cDNA of embryos (A), juveniles (B), females (C) and males (D). Lanes 1 – 4: *Rs-ttl-1* to *-4*; ACT: actin.

Figure 22: *In situ* hybridization. **a**, *Rs-ttl-1* showed expression in the tissues around the vulva (arrow), inset shows top view. **c**, *Rs-ttl-2* showed expression in the ventral nerve cord. **b** and **d**, negative control of *Rs-ttl-1* and *Rs-ttl-2* respectively showed no staining. **e**, Positive control showed staining in the epidermis.

To examine the occurrence of the TTL proteins in other species, multiple database screens were performed. From the protein database, 192 protein sequences (hypothetical included) containing a 'transthyretin-like' domain were retrieved. All the proteins were of nematode origin. Most of them belong to *C. elegans* (n=128) and *C. briggsae* (n=60). The domain was also reported in proteins from the plant-parasitic nematodes *X. index* (n=2) and *H. glycines* (n=1) and the animal-parasitic nematode *Ancylostoma ceylanicum* (n=1). Most of these proteins (n=188) start with a methionine, and 164 of this subset (87%) contain a predicted signal peptide for secretion preceding the transthyretin-like domain. Values concerning identity and similarity between some TTL proteins are summarized in Table 11.

Table 11: Identity and similarity (grey cells) matrix for some mature parasitic nematode TTL proteins. Rs: *Radopholus similis*; Xi: *Xiphinema index*; Hg: *Heterodera glycines*; Ac: *Ancylostoma caninum*. In bold, the highest level of identity for Rs-TTLs with TTL proteins of other nematodes.

	Rs-TTL-1	Rs-TTL-2	Rs-TTL-3	Rs-TTL-4	Xi-TTR-1	Xi-TTR-2	Hg-HSP11	Ac-TTL
Rs-TTL-1		24.1	43.9	31.6	43.5	27.0	**78.4**	30.5
Rs-TTL-2	38.9		30.4	18.6	27.8	18.2	25.0	23.1
Rs-TTL-3	59.5	49.6		43.5	**48.8**	23.7	43.3	**49.6**
Rs-TTL-4	42.4	35.0	53.7		36.7	18.8	30.2	30.5
Xi-TTR-1	55.0	41.1	68.5	46.9		**33.1**	41.8	37.1
Xi-TTR-2	48.1	38.7	44.4	35.0	50.0		25.2	19.4
Hg-HSP11	84.3	41.0	56.7	42.9	56.0	47.0		29.9
Ac-TTL	39.7	36.5	59.5	35.6	49.2	37.1	40.3	

The majority of the current molecular data of nematodes consists of DNA sequence information (ESTs) rather than protein sequences. Since each EST entry holds a description of the library from which it was sequenced, this enabled us to trace back for instance the stage from which the EST was obtained. To use this vast amount of information and sensitively search the occurrence and distribution of the TTL genes within the phylum Nematoda, a HMMER search was performed on translated nematode ESTs. The profile used to search was constructed from a manually optimized alignment of TTL proteins of parasitic and free-living nematodes. This profile was preferred over the profile of the PFAM database which is constructed mainly of free-living nematodes TTL protein sequences. The search of the nematode ESTs resulted in 2,385 hits (cut-off E-value 0.1). From now on we will call these retained ESTs '*ttl*-ESTs'. The classification of *ttl*-ESTs revealed a significantly higher abundance in parasitic nematode ESTs (PPN: 569 hits in 147,867 ESTs (0.38%); APN: 1,260 hits in 221,543

ESTS (0.57%)] than in the ESTs of FLN (FLN; 556 hits in 384,052 ESTs (0.14%)) (p<0.05). However, this could be the result of a bias in sampled nematode stages, since ESTs have a higher chance to be sequenced (and retained by our approach) if their corresponding gene is expressed in those stages that are abundantly sampled for sequencing. For example, currently 19.9% of the ESTs of PPN are sampled from the adult stage, whereas for APN 40.9% of the ESTs are derived from this stage. If the gene of interest is expressed at a higher level in adults, it will be more abundant among APN ESTs. To clarify this possible bias, the distribution of *ttl*-ESTs was checked over the different stages (Figure 23). This revealed that the representation of *ttl*-ESTs was lower in all stages in FLN compared to parasitic nematodes. Moreover, within the parasitic nematodes, we found a significantly higher *ttl*-EST occurrence in parasitic stages (p<0.05). This is from the second stage on in PPN and from the third stage in most APN (Blaxter 2003). To get an idea of the diversity in the TTL family, *ttl*-ESTs of the genera with the highest number of *ttl*-ESTs were clustered. This revealed 45 unigenes clustered from 155 *ttl*-ESTs of *Meloidogyne* (PPN), 41 unigenes clustered from 370 *ttl*-ESTs of *Ascaris* (APN), and 62 unigenes clustered from 494 *ttl*-ESTs of *Caenorhabditis* (FLN). The large number of unigenes illustrates that the genes belong to an extensive gene family. After pooling these 148 unigenes together and clustering this set using TribeMCL with stringent parameters (to obtain clusters of similar sequences), most of the unigenes (n=145) formed one cluster, besides two very small clusters of one and two unigenes. The *ttl*-unigenes of the parasitic genera do not cluster separately from the free living genus.

e. TTLs, TRPs and Transsthyretins: similarities and differences

TTL proteins of parasitic nematodes were aligned to TRP and transthyretin protein sequences. The sequence similarity between TTLs and TRPs and between TTLs and transthyretins is for both on average ~30% (identity is ~16%), whereas the similarity between TRP and transthyretin is higher (similarity is ~38% and identity is ~23%). Furthermore, the non-redundant set of TTL protein sequences of *C. elegans* has an identity percentage of 26%, while this figure for TRPs reaches 47% and for transthyretin sequences even 69%. The alignment sequences of the three groups revealed several conserved residues between the proteins of the different families (Figure 20). The positions of most of the β-strands as predicted for TTLs correspond to the localization of the beta strands as experimentally determined in TRP and transthyretins. The region that forms an α-helix in TRPs and transthyretins is not preserved in

TTL proteins. In addition, the last β-strand which is needed for di- and tetramerisation in transthyretins, is lacking on TTLs (Naylor et al. 1999). To visualize the similarity between TTLs, TRPs and transthyretins, an evolutionary tree was constructed. The alignment used for the construction of the tree consisted of 555 characters, of which 358 were parsimony-informative. 382 maximum parsimony trees, with a consistency index of 0.3795 and a retention index of 0.7055, were retained. The 50% Majority Rule Consensus tree is shown in Figure 24. The tree illustrates the distinct branch leading to the TTL family, although this branch has a low bootstrap value (bootstrap value of 19). TTLs from parasitic nematodes do not cluster together within the TTL branch. The TRP and transthyretin sequences are clustered with high bootstrap support (bootstrap value of 85); moreover transthyretin proteins form a separate cluster (bootstrap value of 100) nested within the TRP branch.
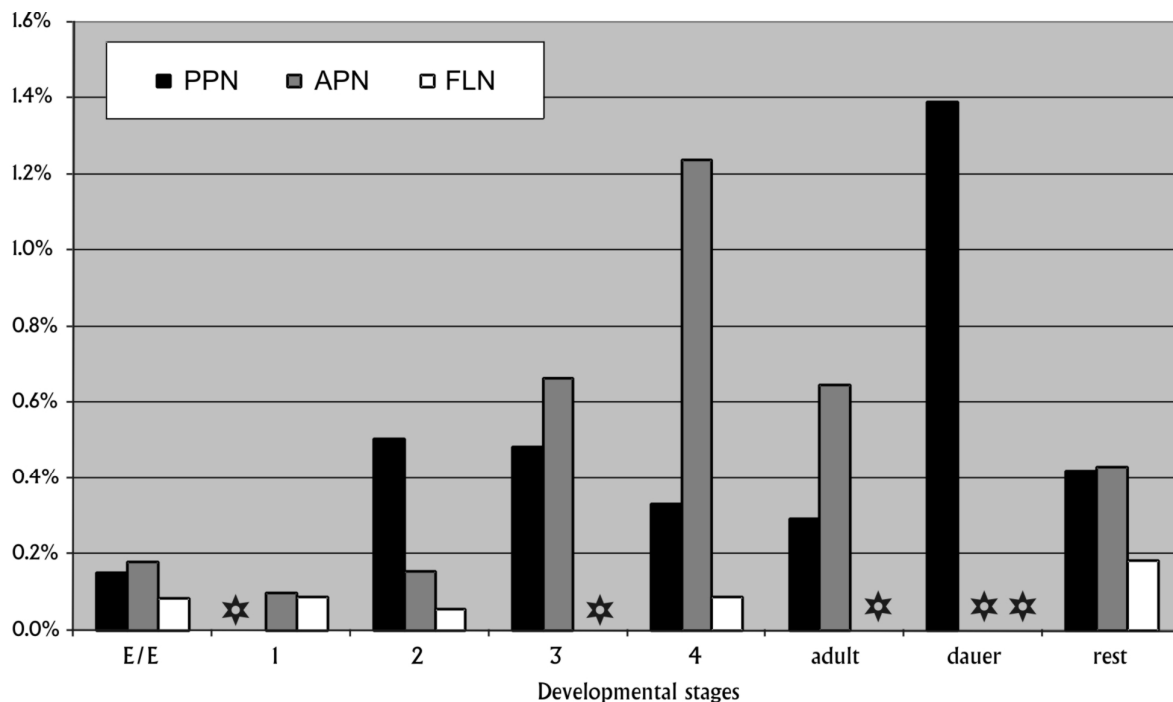
Figure 23: Distribution of *ttl*-ESTs over the developmental stages for PPN, APN and FLN. 'E/E': egg and embryonal stages; '1' to '4': first to fourth larval stages; 'adult': adult stages; 'rest': ESTs from dauer, L5 or mixed stages libraries or for which no developmental stage information was given. Stars indicate that there are no, or less than 20 ESTs from that stage. Y-axis: percentage of ESTs with a (whole or partially) transthyretin-like domain.
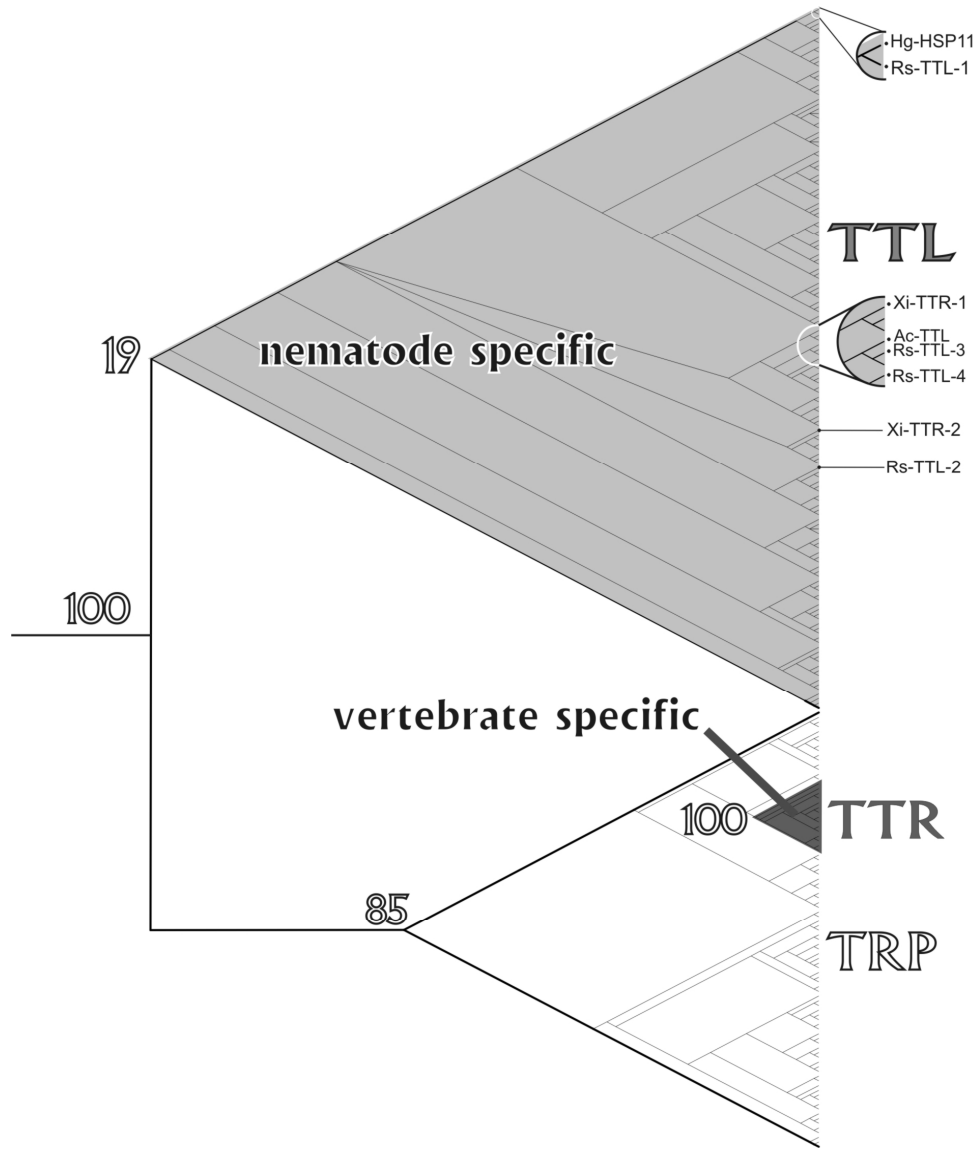
Figure 24: Phylogenetic tree showing the relationship between TTL, TRP en transthyretin (TTR) proteins with indication of some bootstrap values. The TTLs of parasitic nematodes are indicated by their names. The area in *light gray* indicates the nematode-specific TTL genes, the area in *dark gray* the vertebrate-specific TTR genes, and the remainder are TRP genes.

4. Discussion

Proteins that contain a transthyretin-like domain share some sequence similarity with transthyretin-related proteins and transthyretins. The similarity between these different families resulted many times in an inconsistent use of names and abbreviations in publications and annotated sequence databases. For example, TRPs were first described by Eneqvist et al. in

2003, however, several later papers used abbreviations such as TTL (Nam et al. 2004) or TLP (Hennebry et al. 2006) for naming TRP genes. In addition, many TRP sequences submitted to the database are described as 'transthyretin-like' (e.g. CAL34852), which can cause confusion because these proteins lack the transthyretin-like domain. Furthermore, when denoting a protein with a transthyretin-like domain, several authors used the abbreviation TTR (Bellafiore et al. 2008; Furlanetto et al. 2005; Hansen et al. 2005). We would like to clarify these inconsistencies, and strongly recommend reserving the use of TTR for transthyretin and restricting the use of TTL and TRP to point to transthyretin-like and transthyretin-related proteins, respectively.

In this study, we cloned four genes from *R. similis* whose corresponding proteins contain the transthyretin-like domain. These genes were named *Rs-ttl-1* to −*4*, following the proposed convention for naming (parasitic) nematode genes (Bird et al. 1994; Blaxter et al. 1997). Considering the sequence similarity between TTLs, TRPs and transthyretins at the protein level, our evolutionary tree placed the TTLs in a distinct cluster (bootstrap value of 100), separated from the transthyretins and TRPs. With a bootstrap value of 85%, TRP and transthyretins cluster together in the same group. However, within this group, transthyretins form a highly supported separate subcluster nested within the TRPs (Figure 24). This confirms previous studies suggesting that the TRPs are likely ancestors of the transthyretins (Eneqvist et al. 2003; Hennebry et al. 2006). Based on the observation that TTLs and TRPs form separate clusters in the tree, and the fact that the TTL proteins are only found in nematodes, whereas the TRPs occur in a wide range of species (including nematodes), an evolutionary model can be hypothesized. In this model TTL and TRP genes evolved independently from a common ancestral gene, which was probably a TRP-like gene. A duplication event leading to the current TLLs and TRPs most likely occurred in an early nematode species. The low sequence identity within the TTL family compared to TRPs can point to an accelerated evolution after the duplication event, which caused a divergence of the TTL family.

The proteins corresponding to the four cloned TTL genes of *R. similis* possess a signal peptide for secretion preceding the transthyretin-like domain. In addition, the mature Rs-TTL-1 contains a short N-terminal extension, and Rs-TTL-4 is characterized by a basic C-terminal extension (Figure 20). For each of the four TTL proteins of *R. similis* similar proteins are found in *C. elegans*, strongly suggesting that these extensions are not the result of PCR or sequencing errors. Database searches revealed no TTL proteins outside the phylum Nematoda,

pointing to a nematode-specific nature for this gene family. Furthermore, an estimation of the occurrence of TTL genes in the phylum Nematoda was obtained by exploring all nematode ESTs to date in the database. It revealed that the fraction of *ttl*-ESTs is greater in parasitic nematodes compared to free-living nematodes. In addition, in the parasitic nematodes, most *ttl*-ESTs are found in libraries constructed from the parasitic stages. The higher *ttl*-EST abundance in parasitic nematodes and more specifically in parasitic stages, can point to an involvement of TTL proteins in parasitism (Figure 23). TTLs have been detected in the excretory/secretory products of animal parasitic nematodes, to such an extent that they are the most immunogenic proteins, and good vaccine candidates (Hotez et al. 2003; Saverwyns et al. 2008; Vercauteren et al. 2003). In addition, Bellafiore et al. (2008) reported a TTL gene of *Meloidogyne incognita* to be expressed in the subventral gland cells. On the other hand, the four identified *R. similis* TTL genes are expressed in males of *R. similis* as well. This stage is considered non-parasitic since it possesses a degenerate stylet and smaller pharyngeal glands (Luc 1987). This fact seems contradictory to a putative role in the parasitism process for (at least) these four members of this large gene family.

Different functions have already been suggested for TTL proteins. One study revealed the expression of a TTL gene of the plant-parasitic nematode *X. index* (*Xi-ttr-2*) in the gland cell region. Since the corresponding protein contains a predicted signal peptide for secretion, it was suggested that the protein can be secreted by the nematode to fulfil a function in the establishment of a successful parasitic interaction (Furlanetto et al. 2005). However, there is no experimental evidence for this hypothesis. Based on their homology to transthyretin, both Sonnhammer and Durbin (1997) and Parkinson et al. (2004) suggested a role for TTL proteins in hormonal transport in nematodes. Also Hansen et al. (2005) suggested a ligand binding function for TTLs based on an RNAi experiment in *C. elegans* targeting *Ce-ttr-1* (NM_066803; a TTL gene despite its name). In TRP and transthyretin proteins, conserved beta-strands align a binding channel. Both proteins have a similar secondary structure, but the residues involved in ligand binding within these beta-strands are not conserved. The channel in TRP is characterized by positive charges, whereas corresponding residues are rather hydrophobic in transthyretins (Hörnberg et al. 2000; Lundberg et al. 2006). These differences determine the specificity of the interaction with the ligand. In this view, the secondary structure prediction of TTL proteins is remarkable, since the overall structure of β-strands is conserved in TTL proteins, suggesting a similar β-sandwich structure (see Figure 20, PDB

identifier 1DVQ) (Blake et al. 1978). The first, fourth and last two β-strands contain the residues which are crucial for ligand binding in TTR (Figure 20). This suggests that the cloned TTLs of *R. similis* fulfil a similar role as binding partner for a yet unknown ligand. Most interestingly, later additional searches in the extra *R. similis* EST data revealed two more TTL genes for which the secondary structure slightly diverges, in that an α-helix is located between the second and third β-strand. But all hitherto identified TTL genes lack the last beta-strand from TTRs needed for tetramerisation, indicating that TTL proteins do not from tetramers for their function.

The function of TTL proteins could be related to the nervous system, since expression of *Rs-ttl-2* was clearly restricted to the ventral nerve (Figure 22). The expression of *Rs-ttl-1* was restricted towards the region of the vulva, a region containing different tissues, which makes it difficult to pinpoint the expression to a specific cell type. However, the expression could be limited to the nerve cells extending from the ventral nerve cord in the vulval region (Decraemer et al. 2006). The putative role in the nervous system could also explain the absence of expression of any of the four studied TTL genes in developing embryos, since a nervous system is not yet formed at this point - in contrast to later stages, in which all Rs-TTL genes were expressed as determined by semiquantitative RT-PCR. Of the known expression pattern of 2 *C. elegans* TTL genes, one is found to be expressed in its nervous system and hypodermis. Interestingly, one 'neuropeptide-like protein', *nlp-4* (NM_060349), of *C. elegans* shows a weak similarity to the transthyretin-like domain (PF01060; bit score of -37.90). This NLP-4 protein has characteristic features of neuropeptide preproproteins and was reported by Nathoo et al. (2001) upon exploring the *C. elegans* genome. The expression in the nervous system, the prevalence of secretion signals, the large gene family and different expression patterns for several members, are all characteristics comparable to the features of neuropeptides, such as the FMRFamide-like neuropeptide (FLP) gene family in Nematoda (McVeigh et al. 2005). In addition, the existing RNAi phenotypic data of the *C. elegans* TTL genes reveals that the majority (49 genes of 53 genes tested) does not have an observable phenotype. Only 4 TTL genes show phenotypes after RNAi application (*ttr-1* - extended life span; R13A5.3 (NM_066257) - hyperactivity; T05A10.3 (NM_077400) - maternal sterility and larval death; ZC64.2 (NM_076265) - increased fat content). The lack of RNAi phenotypes for the majority of the *C. elegans* TTL genes is in agreement with expression in the nervous system, which is recalcitrant to RNAi in *C. elegans*. In contrast, the large TTL family might be

functionally redundant, causing one member to take over the function of another when silenced. Nevertheless, in an RNAi experiment on *R. similis* to silence *Rs-ttl-1* and *Rs-ttl-2*, transcript levels were slightly reduced, but no phenotypic difference was observed between the control *gfp* soaked nematodes (data not shown). Further identification, classification and characterization of additional members of the TTL gene family is needed to elucidate their function and to verify the possible link with parasitism.

Despite its uncharacterized biological role, the restricted occurrence of the TTL proteins in nematodes makes it a very promising target for nematicide development with minimal chance for off-target effects. If the TTL gene family has a role in the nervous system, this can be an additional advantage since drugs that target parts of the nervous system of nematodes are very effective (Sandstrom et al. 2004). Additionally, experiments targeting FMRFamide-like neuropeptides in *Globodera pallida* revealed an unusual neuronal sensitivity for RNAi, in contrast to RNAi experiments targeting the same neuropeptide in *C. elegans* (Kimber et al. 2007). The *in planta* generation of double-stranded RNA could lead to crop resistance against plant-parasitic nematodes by inducing RNAi silencing in the feeding nematodes (Gheysen et al. 2007). The choice of the target gene(s) is of great importance for success and sustainability of this approach. In our opinion the conserved and nematode-specific TTL gene family could be a suitable candidate for this approach.

5. Materials and Methods

a. *R. similis* culture and collection

See chapter 2 for a description of *R. similis* culture. Nematodes were either used freshly, or stored as a pellet at -20°C. Specific stages for RNA extraction (developing embryos, juveniles, adult males and females) were manually isolated from a fresh batch of nematodes, and directly transferred to a tube with 100μL TRIzol® Reagent (Invitrogen, Carlsbad, USA).

b. RNA extraction, cDNA synthesis, library construction and EST sequencing

The cDNA library used for sequencing of ESTs was prepared before the construction of the library reported in chapter 2. For the first library, constructed by Irena Ronko and Makedonka Mitreva, mixed stages of *R. similis* were ground in liquid nitrogen and RNA was extracted using $MgCl_2$ as described by Sambrook et al. (1989). The

library was constructed as described in Mitreva et al. (2004). Briefly, cDNA was synthesized and amplified using oligo(dT) and spliced leader 1 primers. PCR products were non-directionally cloned into the pCRII-TOPO vector (Invitrogen) following the TOPO-TA cloning protocol. EST sequencing was performed on this library as described in Mitreva et al. (2004). For the RT-PCR experiments, RNA was extracted using TRIzol® Reagent after grinding juveniles, females or males in liquid nitrogen or after sonication of developing embryos. For the latter, single cell eggs or eggs which had undergone few cell divisions were chosen. cDNA was synthesized using the SMART™ PCR cDNA system (Clontech Mountain View, USA).

c. Full length cDNA generation

Genes of *R. similis* coding for TTL proteins (for identification of the genes: see section g) were amplified from a mixed stage cDNA pool by a nested PCR under standard PCR conditions using primers based on the corresponding EST sequences (_fw1 and _rv1 primers for the first PCR (15 cycles) and _fw2 and _rv2 primers for the second pcr (35 cycles); see Table 12 for primer sequences). The resulting PCR fragments were separated on a 0.5X TAE 1.5% agarose gel and specific bands were purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Purified DNA fragments were cloned into pGEM-T (Promega, Madison, USA) and sequenced at the VIB Gene Sequencing Facility (Antwerp, Belgium). Missing 5'- and 3'-fragments were cloned from the SMART™ cDNA library using gene specific primers in combination with vector primers (M13F and M13R) (Table 12). The obtained fragments were gel excised, cloned into the pGEM-T vector and sequenced as described above. 5'- and 3'-fragments were assembled *in silico* to obtain full length coding sequences. Sequences were deposited in GenBank.

d. Sequence analysis

Sequences were aligned by ClustalW (Chenna et al. 2003; Thompson et al. 1994) and alignments were manually corrected using BioEdit (Version 7.0.5.2) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Nucleotide sequences were translated using the Translate program on the ExPASy proteomics server (http://www.expasy.ch/). The ProtParam tool on the same server was used to determine additional protein characteristics. To decrease the redundancy of a list of sequences, the Decrease Redundancy tool was used on ExPaSy with the maximum similarity set to 99%. Identity/similarity matrices were constructed by MATGAT2.02 (Campanella et al. 2003), which aligned the sequences using the BLOSUM62 matrix. Signal peptides for secretion were predicted on the SignalP3.0 Server (http://www.cbs.dtu.dk/services/SignalP/). Secondary structure prediction was performed by PROF (Rost et al. 1996), and predictions of the disulfide bonding state of cysteines were done by DISULFIND (Vullo et al. 2004) on the PredictProtein Server (Rost et al. 2004). Clustering of translated EST seqs was done by TribeMCL (Enright et al. 2002). Wormbase was searched for known expression patterns and RNAi phenotypes via WormMart (www.wormbase.org/Multi/Martview) (Bieri et al. 2007).

Table 12: Oligonucleotides used for amplification of *Rs-ttl* genes and actin.

Primer name	Sequence
RsTTL1_fw1	ATTGTTATTTTACTATGGCGCC
RsTTL1_fw2	TATGGCGCCTATGTTTCTGCC
RsTTL1_rv1	TCTTCCCCTCTGCCATCAGG
RsTTL1_rv2	TCAGGTCGTCGGTATCAACGCC
RsTTL2_fw1	CTCTGTTCCCAGTGCTTCCC
RsTTL2_ fw2	TGCTTCCCGCCAAAATGCGGC
RsTTL2_ rv1	ATGTCTCGGTGTCGCGATGGC
RsTTL2_ rv2	ATGAAGTTCATTTCGTACACGC
RsTTL3_ fw1	TCACCATTGTGCCGTCCCCG
RsTTL3_ fw2	AACAGGCGGTCGCAGTCAAGGG
RsTTL3_ rv1	AGCAAATCTCGCTCCTCGCC
RsTTL3_ rv2	TTTGCTGTTTGAGCGCGC
RsTTL4_ fw1	TCTGCTGCTGTAAAAGGTCGC
RsTTL4_ fw2	TGACGACCAATTAGACGCAGGC
RsTTL4_ rv1	TATTGTTTTGTTTGTTTTGGC
RsTTL4_ rv2	TTTGCCACCTAGAGACGAACG
RsSEC2_fw	AACGCCTACTACTTCTTCAGCTC
RsSEC2_rv	TGTCCCCTTCCCTTCTTCTTCTGC
RsACT_F	GAAAGAGGGCCGGAAGAG
RsACT_R	AGATCGTCCGCGACATAAAG
M13F	GTAAAACGACGGCCAGT
M13R	GAAACAGCTATGACCATGTT

e. Whole mount in situ hybridization

The whole mount in situ hybridization protocol was performed as in chapter 2. Hybridization was done overnight at 47°C. DIG-labeled probes were generated by linear PCR with a single primer and an appropriate cDNA fragment cloned into pGEM-T as template in the presence of DIG-labeled nucleotides. For the antisense probe against Rs-ttl-1 to -4, the primers RsTTL1_rv2, RsTTL2_rv2, RsTTL3_rv2 and RsTTL4_rv2 were used, respectively; for the sense probe generation the primers RsTTL1_fw2, RsTTL2_fw2, RsTTL3_fw2 and RsTTL4_fw2 were used, respectively. As a positive control an antisense probe against Rs-sec-2 was used. A cDNA fragment of Rs-sec-2 was cloned from a mixed stage cDNA pool using the primers RsSEC2_fw and RsSEC2_rv. The primer sequences were based on an EST of Rs-sec-2 (CO897936).

f. Semiquantitative reverse-transcriptase PCR (RT-PCR)

The different TTL genes of *R. similis* were amplified from the cDNA pools of developing embryos, juveniles, adult females and adult males via a standard PCR using gene specific primers (Rs-ttl-1: RsTTL1_fw1 & RsTTL1_rv1 / Rs-ttl-2: RsTTL2_fw1 & RsTTL2_rv1 / Rs-ttl-3: RsTTL3_fw1 & RsTTL3_rv1 / Rs-ttl-4: RsTTL4_fw1 & RsTTL4_rv1 / Rs-act: RsACT_F & RsACT_R) (Table 12). Cycling conditions were as follows: 2' 94°C; (25" 94°C; 25" 52°C; 40"

72°C)x40 for the embryo and the juvenile cDNA pool, x35 for the male cDNA pool and x30 for the female cDNA pool. The different number of cycles resulted in a similar amplicon level of the internal control (actin) between the different pools, so that comparing the signals over the cDNA pools was possible. 5 µl of each PCR product was separated on a 0.5x TAE 1.5% agarose gel.

g. Hidden Markov model search in *R. similis* ESTs and in nematode ESTs

The *R. similis* EST sequences (n=1,154) were clustered with the TIGR Gene Indices clustering tool (Pertea et al. 2003) resulting in CAP3 assemblies (contigs) (Huang et al. 1999) and singletons. Translation of the clustered ESTs was done according to Vanholme et al. (2006) resulting in putative proteins. TTL proteins are identified by searching the list of putative proteins with a 'hidden Markov model' using the transthyretin-like profile from the PFAM database (profile of PF01060 domain). Further, a manually optimized alignment was constructed by ClustalW (Thompson et al. 1994) using the identified Rs-TTLs, together with 2 TTL proteins of *Xiphinema index* and 12 C. elegans TTL protein sequences with highest homology to Rs-TTLs. This alignment was used to construct a new HMM profile (hmmerbuild of the HMMER software), optimized for finding local alignments in translated EST sequences (Durbin et al. 1998; Eddy 2004). The profile was used to screen all nematode sequences from dbEST (n=754045; January 2007) translated in 6 reading frames. For each hit (cut-off E-value of 0.1), the species and developmental stage information was parsed from the GenBank entry. In addition, we checked GenBank entries of all nematode ESTs and parsed the developmental stage from which each EST was obtained. We also classified the EST as derived from plant-parasitic, animal-parasitic or free-living nematodes (respectively PPN, APN, and FLN). For each class we calculated the fraction of ESTs belonging to a certain stage. FLN contained the genera *Caenorhabditis*, *Pristionchus*, and *Zeldia*. APN contained the genera *Ancylostoma*, *Angiostrongylus*, *Anisakis*, *Ascaris*, *Brugia*, *Dirofilaria*, *Haemonchus*, *Heterorhabditis*, *Litomosoides*, *Necator*, *Nippostrongylus*, *Onchocerca*, *Ostertagia*, *Parastrongyloides*, *Strongyloides*, *Teladorsagia*, *Toxocara*, *Trichinella*, *Trichostrongylus*, *Trichuris*, and *Wuchereria*. PPN contained the genera *Bursaphelenchus*, *Globodera*, *Heterodera*, *Meloidogyne*, *Pratylenchus*, *Radopholus*, and *Xiphinema*. Statistical comparison of the fractions of ESTs was performed as described in Audic and al. (1997) using the web interface (http://www.igs.cnrs-mrs.fr/Winflat/winflat.cgi).

h. Evolutionary analysis

TRP sequences (n=82) from bacteria, plants, invertebrates and vertebrates were selected from the Swissprot and TrEMBL databases (February 2007) based on the presence of their characteristic Y-[RK]-G-[ST] peptide at the C-terminal end using the ScanProsite tool (http://www.expasy.ch/tools/scanprosite/) (Sigrist et al. 2002). A HMMER search verified the presence of the transthyretin domain in all TRP sequences. A search for 'transthyretin' was done via the search field on the Entrez website (http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi). From the resulting list, 14 transthyretins originating from different vertebrate groups (mammalian, avian, amphibian and reptilian) were manually selected. Similarly, via the Entrez search field, the protein sequences containing the DUF290 (transthyretin-like) domain were retrieved (n=192), and manually checked for redundancy (n=147). The gathered transthyretin-like, TRP and transthyretin protein sequences (n=243) were aligned using

ClustalW (Thompson et al. 1994) (outgroup sequence: EAU94136). The optimized alignment was used to generate an evolutionary tree based on maximum parsimony using PAUP* 4.0 (http://paup.csit.fsu.edu/) (Swofford 2003). A heuristic search with 100 replications was done and 100 bootstraps were performed.

6. Acknowledgements

3

Chapter 4

The mitochondrial genome of the plant-parasitic nematode *Radopholus similis*

4

"Most claims of originality are testimony to ignorance and most claims of magic are testimony to arrogance."
James March, Stanford University

The mitochondrial genome of the plant-parasitic nematode *Radopholus similis*

1. Abstract

We present the annotation of the mitochondrial (mt) genome sequence of the plant-parasitic nematode Radopholus similis (accession number FN313571). Expressed sequence tags derived from mt genes (see chapter 2) served as a starting point to amplify and sequence the genome by long-distance PCR. At first sight, the resulting circular genome of 16,791 bp resembles other nematode mt genomes, in that it could encode for 12 proteins, 2 rRNA genes and 22 tRNA genes, all transcribed from the same strand. However, other features deviate strongly from other nematode mt genomes. The large size of the genome is due to the presence of a large non-coding region consisting of two repeat regions. The gene arrangement is not comparable to the mtDNA of any other nematode, and the protein-coding sequences - most of them lacking a Stop codon - are the highly diverged from other nematode sequences. Furthermore, secondary structure prediction of the large (16S) rRNA gene revealed a reduction in structural features. Most interesting, this very AT-rich (85.4%) genome uses a peculiar genetic code, in which UAA is reassigned from Stop to Tyrosine. These unique features reinforce the monophyletic position of *Radopholus similis*. Analysis of the large non-coding region revealed several secondary structures and motifs which appear in other nematode mt genomes, indicative of common regulation.

4

2. Introduction

In the second chapter, novel transcriptomic data on *R. similis* was analyzed (Jacob et al. 2008). An estimated 0.6% of the unigenes comprised transcripts derived from the mitochondrial DNA. Mitochondria are found in all eukaryotic cells. The main function of these organelles is to provide energy to the cell through the process of oxidative phosphorylation. They are

thought to originate from an ancestral endosymbiotic α-proteobacterial species. In the course of evolution practically all the endosymbiont genes have been transferred to the nucleus, and now approximately 1% of the nuclear genes have a function in the mitochondria. Still, few genes were not transferred, and mitochondria contain their own haploid autonomously replicating genome of relatively short length, in nematodes ranging from 12.5 kb to 26 kb, with a mean of 15 kb (Azevedo et al. 1993; He et al. 2005; Hu et al. 2006). Typically, a metazoan mt genome is divided into two regions: a non-coding region controlling transcription through the action of nuclear-encoded factors (Gissi et al. 2008; Scarpulla 2008) and a large coding region containing all the mt genes, whose gene products fulfil essential roles in mitochondrial processes. Almost all genes from the mt genome have transferred to the nuclear genome or have been lost in the course of evolution and it is not clear why some genes (an estimated 1% of the orlgal gene content) remain encoded by the mt genome. These gene products are usually 2 ribosomal RNAs, 22 transfer RNAs, which both play an essential role in transcription of mt encoded genes, and 12-13 intronless protein-encoding genes (PCGs), which encode for crucial subunits in the mt membrane-located respiratory complexes I, III, IV and V. Transcription is thought to be initiated in the non-coding regions, generating a long polycistronic transcript covering the complete coding part. Subsequent cleavage of the transcript to release mono- or dicistronic transcripts and RNA genes is guided by the tRNA structures, known as the "tRNA-punctuation" model (Ojala et al. 1981; Scarpulla 2008). Mt mRNA is polyadenylated in most metazoa by a mitochondrial poly(A) polymerase (Scarpulla 2008). However, many of the mechanisms of mitochondrial transcription and translation remain unresolved.

Today, only 27 mitochondrial genomes of nematodes are sequenced and available in the GenBank database (Hu et al. 2006; Jex et al. 2008; Kang et al. 2008; Li et al. 2008). In addition, some mt genomes have been partially sequenced (Armstrong et al. 2000; Gibson et al. 2007; Okimoto et al. 1991). Based on mt genomes, generally two different groups of species can be recognized. Species belonging to the Enoplea class contain mt genomes that are on average 20.0 kb ± 4.0 and resemble the average metazoan genome in gene content and arrangement. In contrast, mt genomes of the class Chromadorea are usually extremely compact (13.8 kb ± 0.8) and differ in several aspects from other typical metazoan mt genomes. Most remarkably, all genes are unidirectionally transcribed from one strand. Furthermore, they lack an ATPase subunit 8, the rRNA molecules are shorter, and tRNAs are truncated.

Most nematode mt tRNAs lack one arm, either the TΨC arm which is replaced by the 'TV-replacement loop', or the DHU arm which is replaced by 'D-replacement loops' (Hu et al. 2006; Montiel et al. 2006). The mt genome discovered in *Globodera* species deviates from other Chromadorean nematodes. Their mt genome is found to consist of 6 to 9 fragments, each containing part of the mt protein content and with indications of homologous recombination taking place (Armstrong et al. 2000; Riepsamen et al. 2008).

In this study, mitochondrial transcriptomic data served as a starting point to obtain the complete sequence of the mitochondrial genome of *R. similis*. The peculiar features of this nematode mt genome are investigated and compared to other nematode mt genomes.

3. Results

a. Genome size and characteristics

The complete mitochondrial genome of *R. similis* was amplified by long-distance PCR in three overlapping fragments, with the length of the complete assembled sequence being 16,791 bp (accession number FN313571). This length corroborated with Southern blot hybridization, indicating that the assembled sequence corresponds to the complete mt genome, which is the largest Chromadorean mt genome to date (Figure 25). With an AT content of 85.4% it is the most AT-rich nematode mt genome sequenced to date.
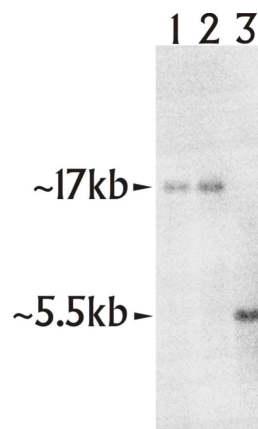
Figure 25: Southern blot on digested *R. similis* mt DNA. Lane 1 and 2 are digested with SpeI and XmnI respectively, lane 3 is digested with both enzymes. Expected lengths are 16.8 kb for lane 1 and lane 2 and 5.4 kb for lane 3.

The 36 mitochondrial genes comprise 76.6% of the genome and are unidirectionally transcribed from the coding strand, which has an asymmetrical nucleotide composition of 52.9% T, 32.5% A, 10.2% G and 4.4% C and is also referred to as the heavy strand, based on centrifugal separation on density gradients. Unusually for parasitic nematodes, a quarter of the mt genome of *R. similis* is non-coding, comprising two large repeat regions which contribute to the large size of the genome.

b. tRNAs

As in other described nematode mt genomes, 22 tRNAs have been predicted ranging from 51 nt to 59 nt in length (Figure 26). While most amino acids are encoded by one tRNA, leucine and serine are each encoded by two tRNAs (Leu-CUN and Leu-UUR, Ser-AGN and Ser-UCN). With the exception of tRNA^{Lys}, all tRNAs are adjacent to at least one other gene. Most of the anticodons are conserved in nematodes, except for the UCG anticodon of tRNA^{Arg}, which is common in other metazoans, but deviates from ACG anticodon used by most Chromadorean nematode mt genomes. All tRNAs are characterized by a 7 bp amino-acyl arm and a 5 bp anticodon arm with a uracil preceding the anticodon (Figure 26). In contrast to most tRNAs which have a DHU (dihydrouridine) arm and lack a TΨC (T-pseudouridine-C) arm, the tRNA^{Ser} has a TΨC arm and lacks a DHU arm. The predictions of the tRNA genes are likely to be accurate, since alignment with the corresponding genes of other nematodes revealed similarity on both primary sequence and predicted secondary structure. Six identified tRNAs (coding for Arg, Asp, Cys, His, Ile and Lys) lacked the conserved adenine on position 14 (first nucleotide of the DHU loop) identified on the majority of nematode mt tRNAs (Sakurai et al. 2006). Seemingly typical for *Radopholus* is the occurrence of 2 nucleotides - instead of one - positioned between the DHU arm and the anticodon arm. In those cases, the DHU arm is on both sides bordered by uracil (occurring in 8 tRNAs ; Figure 26). Another feature characteristic for the described tRNAs is the occurrence of three instead of two nucleotides between the amino-acyl acceptor stem and DHU arm (occurring in 10 tRNAs).

c. Protein-encoding genes

To our surprise, we discovered that the UAA codon, which encodes translation termination (together with UAG) according to the standard invertebrate mitochondrial genetic code, was abundant in all 12 PCGs, both in less and more conserved sequence parts. Based on alignments

with other nematode mt proteins, the program GenDecoder predicts with a high probability that UAA encodes for Tyr in *R. similis* (Figure 27). This codon reassignment UAA[Stop] to UAA[Tyr] is until now only identified in *R. similis*, and is thoroughly discussed in the next chapter. As in other Chromadorean nematodes, the mt genome of *R. similis* contains 12 PCGs and lacks the ATPase subunit 8 gene commonly found in other metazoan mt genomes. The cytochrome oxidase subunit proteins (COX1, 2 and 3) are the most similar compared to other nematodes, while NADH dehydrogenase subunit 6 (NAD6) is the least conserved (Figure 28).
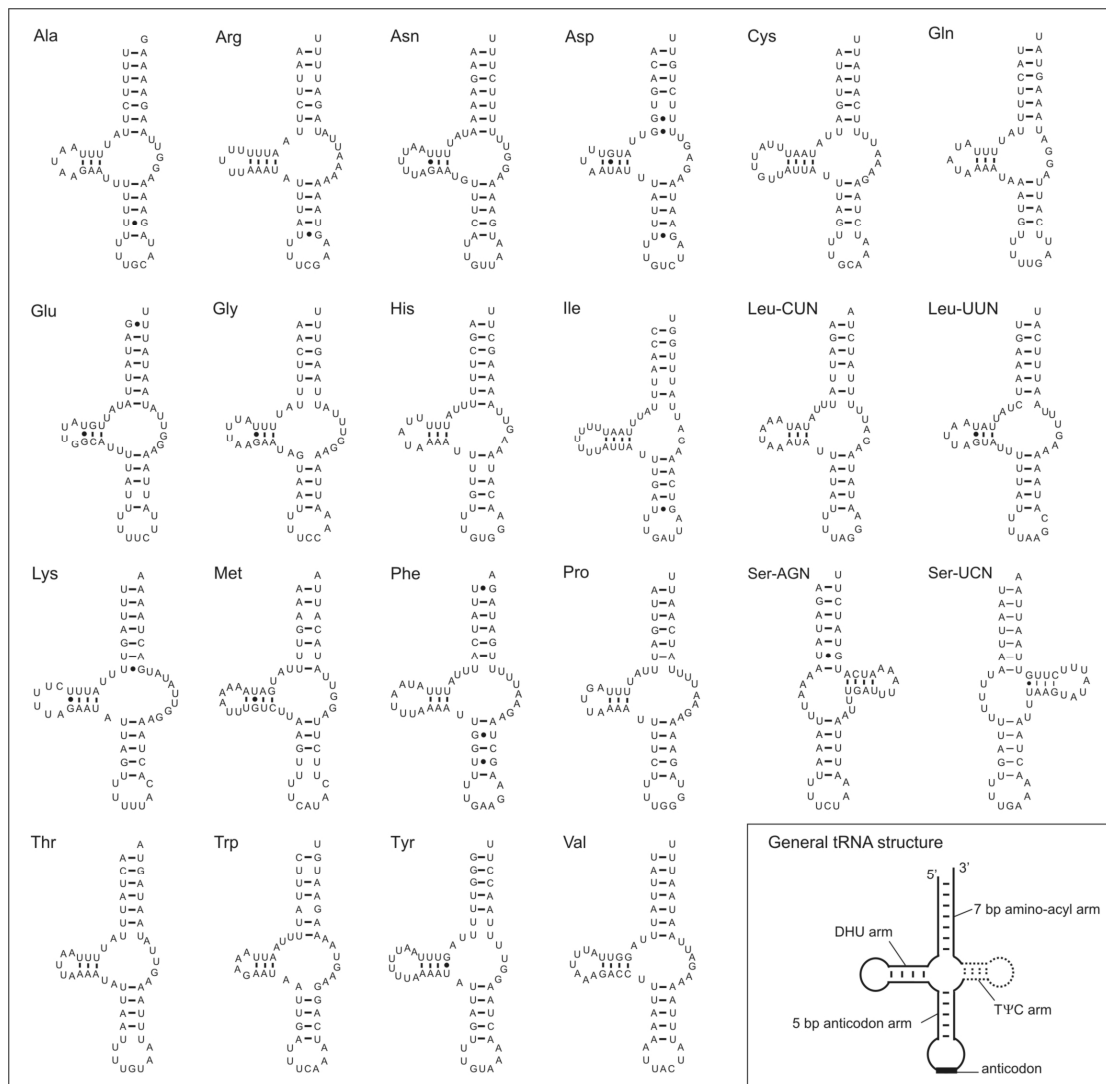
Figure 26: Secondary structures predicted for the 22 tRNA genes of the *R. similis* mitochondrial genome. In the downward right corner, the general tRNA structure is depicted with naming of the different stem-loops and features. The TΨC loop is indicated in dashed line since it is lacking on most tRNAs.

```
        --------------------------------------------------------------------
                        MAIN RESULT OF GENDECODER
        --------------------------------------------------------------------

     Species  : Radopholus similis
a    Expected : FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSSSSVVVVAAAADDEEGGGG
b    Predicted: FlLLS-SsY-Y*C-WWL-i?P-PpHhQQR-r-IvMMT-TtN?KKScSsVVVVAaA-D-EEGgGG
c    First    : TTTTTTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG
d    Second   : TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG
e    Third    : TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
                 ..  .  .   ..   .   .  ..   .  .  .  .   ..   .  .   .  ..   .

f    Codon-imp: +1+++0+3+0+5+0+6+030+0+2+1+9+020+1+++091+0+++1+5+4+8+170+0+++1+6
g    Codon-num: +1+++0+3+0+5+0+6+041+0+2+1+++020+1+++0+1+2+++1+5+4+++170+0+++1+6
h    Freq-aa  : 55455-594-4+5-886-3-9-9+8+89+-+-35336-684-8565674743778-8-787+89
i    Diff-freq: o2ooo-ooo-4oo-ooo-o-o-ooooooo-o-o3ooo-ooo-ooo5oooooooooo-o-oooooo
     Total number of codons: 3326

     Nucleotide usage: Total  : A: 30 %T: 55 %G: 11 %C: 5%    GC: 16%
                       1st pos: A: 31 %T: 49 %G: 15 %C: 5%    GC: 20%
                       2nd pos: A: 22 %T: 58 %G: 11 %C: 8%    GC: 19%
                       3rd pos: A: 36 %T: 58 %G: 6  %C: 0%    GC: 6%

     Effective number of codons: Nc=28.75
```

Figure 27: Results of GenDecoder. **a,** expected invertebrate mitochondrial genetic code. **b,** predicted mitochondrial genetic code in *R. similis*. Lowercase indicated cases in which the prediction is based on less than four codons or cases in which the difference between the frequency of the predicted and expected amino acids is lower than 0.25. A dash ("-") indicates the absence of a codon. A question mark ("?") indicates that the meaning of the codon cannot be predicted despite its presence, due to the lack of conserved amino acids at that position. **c, d** and **e** show respectively the first, second and third nucleotide of the corresponding codon. Dots below certain codons indicate the presence of tRNAs containing the corresponding anticodon. **f,** the number of times that the codon is observed at highly conserved positions, with '+' indicating more than 9 times, and **g,** the number of times that the codon is observed including less conserved positions. **h,** the frequency that the predicted amino acid is observed for the corresponding codon by indication of the first decimal, with '+' indicating more than 0.9, and **i,** the difference in frequency observed for the expected amino acids and the observed amino acids.

Remarkably, the mitochondrial encoded proteins of *R. similis* are one of the most dissimilar compared to those of other nematodes. The mean GC content of the protein-coding gene sequences is 13.9%, with the GC content of the nucleotides at the third codon position (GC3) markedly lower (6.4%) than that of GC1 and GC2 (18.3% and 17.0% respectively) (Figure 28). Due to the codon reassignment, the only Stop codon in use is UAG, which could intriguingly only be inferred for 5 PCGs: *nad4L*, *cox3*, *nad1*, *atp6* and *nad5*. The 3 latter genes are consecutively positioned on the only region lacking intergenic tRNAs. Of the remaining genes, *nad6* has a single uracil as translation termination signal. In other mt genomes this so-called truncated stop forms a complete UAA stop codon after polyadenylation (Bobrowicz et al. 2008) – but obviously this could not be the case for *Radopholus* since UAA encodes Tyr. Attempts to determine gene boundaries by CT-RT-PCR (see Materials and methods) failed for

PCGs – in contrast to 12S rRNA. The inferred start codons are UUG (2), AUA (5), AUG, AUU (2) and UUA (2), all of which have been reported before in nematodes (Jex et al. 2008). Although most genes are very tightly arranged, a few gaps and even overlaps with neighbouring tRNA genes are observed, as summarized in Figure 29.
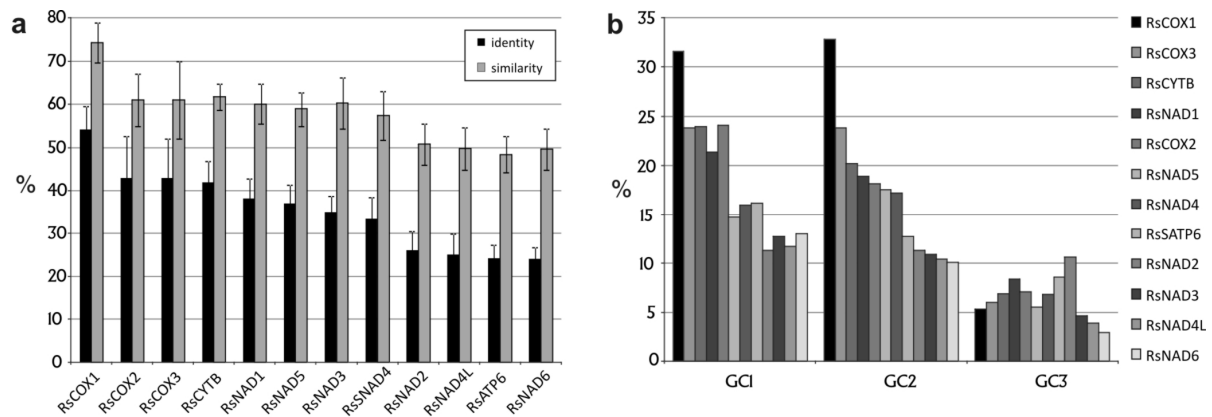

Figure 28: Characteristics of *R. similis* mitochondrial protein-coding genes. **a,** mean percentage identity and similarity with other nematode mt proteins. **b,** GC percentages of protein-coding genes.

d. Ribosomal RNA genes

The two rRNA genes (12S and 16S) of the *R. similis* mt genome were identified based on sequence similarities to rRNA genes of other nematodes. The boundaries of the genes were extended to the neighboring tRNA genes (Okimoto et al. 1994). The 12S rRNA gene (or *rrnS*) is 692 nt long, comparable to other nematode mitochondrial 12S rRNA genes (698 nt ± 33 nt). The predicted boundaries of this gene were experimentally validated through CT-RT-PCR (see Materials and Methods). The secondary structure of this rRNA was predicted based on 12S rRNA structures described by Gutell et al. (1994) (see Figure 30) and revealed a similar topology as other published nematode 12S rRNA structures (Hu et al. 2002; Jex et al. 2008; Keddie et al. 1998; Lavrov et al. 2001). The large subunit rRNA (16S rRNA or *rrnL*) is 840 nt in length, which is considerable smaller than the average of 943 nt (± 57 nt) for other nematodes. Despite this compactness, the predicted structure revealed an overall comparable topology. The most remarkable differences due to the deletions are indicated by arrows in Figure 31. The region after stem D11 is shortened and lacks stem loops. Helix D21 is shorter and
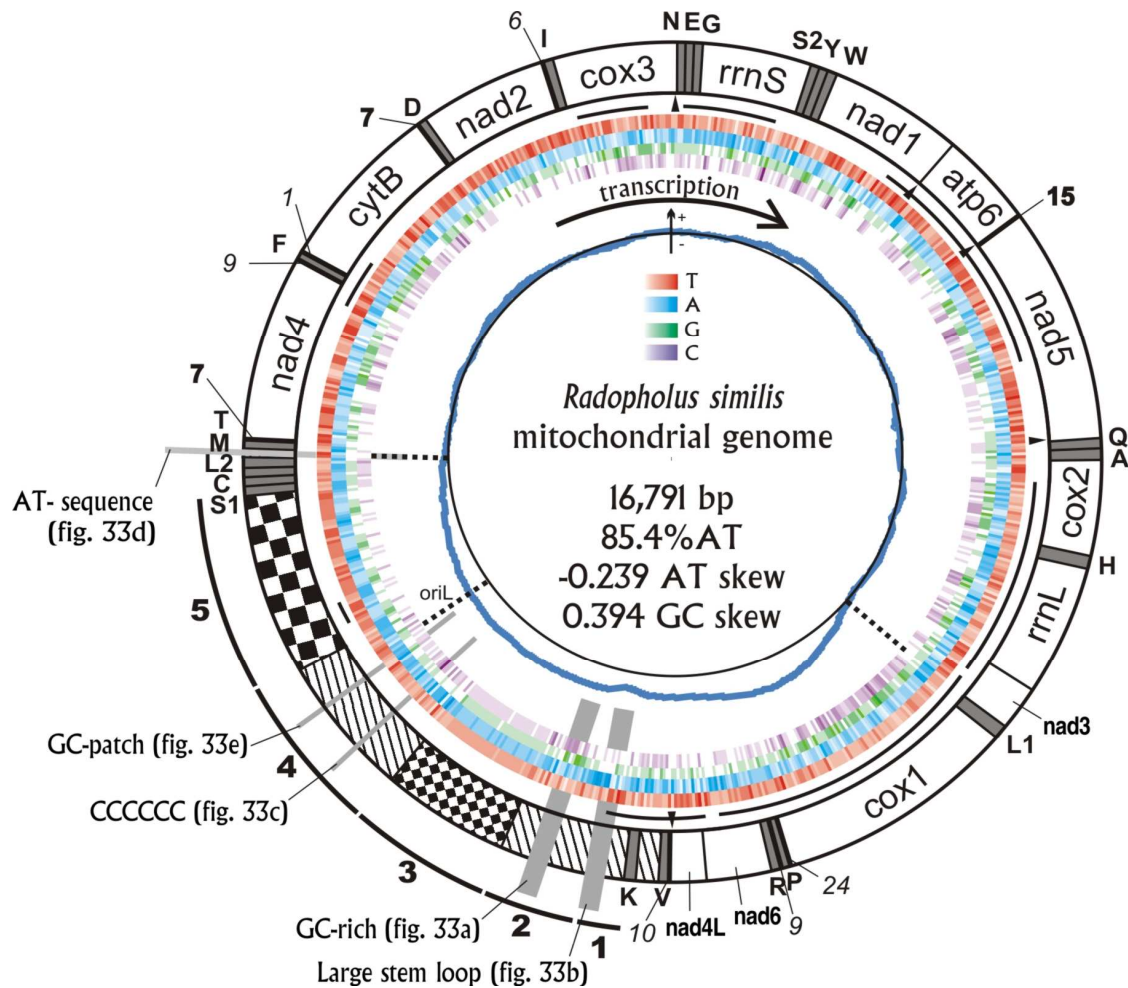
Figure 29: Overview of the organization of the circular mt DNA of *R. similis*. The arrow indicates the direction of transcription. Genes and non-coding regions are indicated on the outer ring: in white, the protein-coding and rRNA genes, in gray, the tRNA genes called by their amino acid symbol (S1: Ser-AGN, S2: Ser-UCN, L1:Leu-CUN, L2: Leu-UUR). Numbers in bold at gene boundaries indicate non-coding nucleotides between genes, and numbers in italic indicate overlapping nucleotides between neighboring genes. The motif-filled part of the ring represents the large non-coding region, divided in five regions as explained in the text. The repeat region of 302 bp is filled with large checkers and the 26 bp repeat region is filled with small checkers. The black lines at the inner periphery represent EST sequences, with UAG stop codons indicated by little black triangles. The colored bar-code-like circles represent the nucleotide content of the coding strand. The positions of secondary structures depicted in Figure 33 are indicated by light-gray balks, crossing the ring. The dashed line indicated by oriL marks the predicted origin of replication of light strand (oriL) (Grigoriev 1998, see Figure 32). The inner blue line indicates the cumulative GC count, mapped to the black circle representing a constant increase, with indication of an above average increase by a '+' on a little arrow at the top. An offset enriched in GC nucleotides, marked by dashed lines, is clearly visible. The reason for this accumulation is uncertain, but could be due assymetrical mutations caused by the single-strandness of the DNA strand during replication.

helix E22 is completely absent. Similarly, the sequence between D1 and E21 is severely shortened and lacks secondary structures. Although nematode mitochondrial 16S rRNAs are

severely shortened, this is even more apparent in *R. similis* (Montiel et al. 2006). Note that naming helical elements must be considered tentative, as similarity to *Escherichia coli* 23S rRNA model used as template is often not clear, especially in the 3' region (see De Rijk et al. 1999). Most conserved nucleotides in nematode mitochondrial rRNAs are generally found in loops and not in stems (Figure 30 and Figure 31), pointing to an importance in maintaining catalytic nucleotides which typically reside in loops, while more compensatory substitutions occur in stems to preserve the structure. Most of the nucleotides at the binding sites of the amino-acyl stem of tRNAs and/or peptidyl-transferase are conserved. However, the G3 stem, which has been implicated as part of the exit site, contains no conserved residues (Keddie et al. 1998).
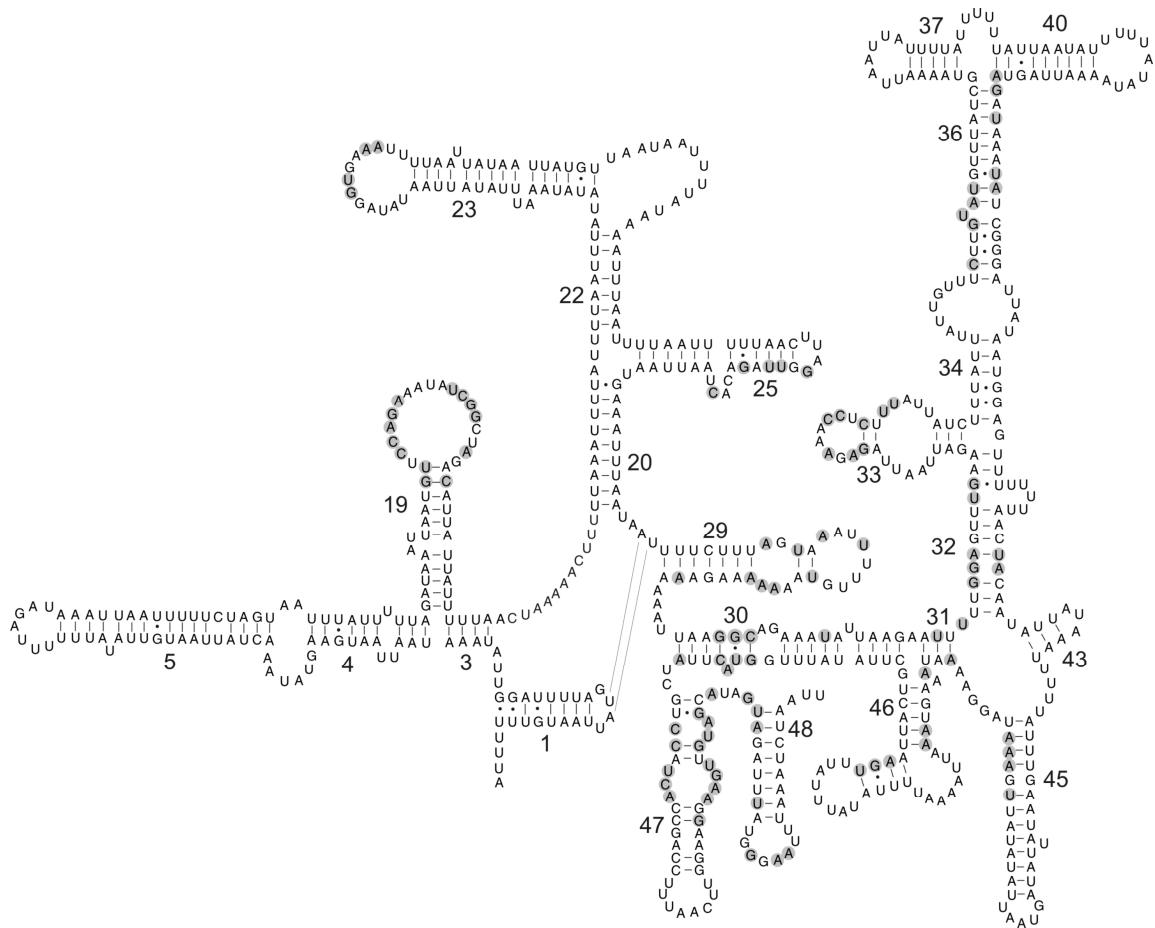
Figure 30: Predicted secondary structure of the 12S rRNA gene of *R. similis*. Watson-Crick base pairing is indicated by a line, whereas a G:U base pairing is indicated by a dot. Proposed tertiary interactions are represented by long, straight lines. Numbers identify the conserved secondary structure elements defined by Dams et al. (1988). Shaded nucleotides are conserved in at least 90% of the currently available nematode mt 12S rRNA sequences.
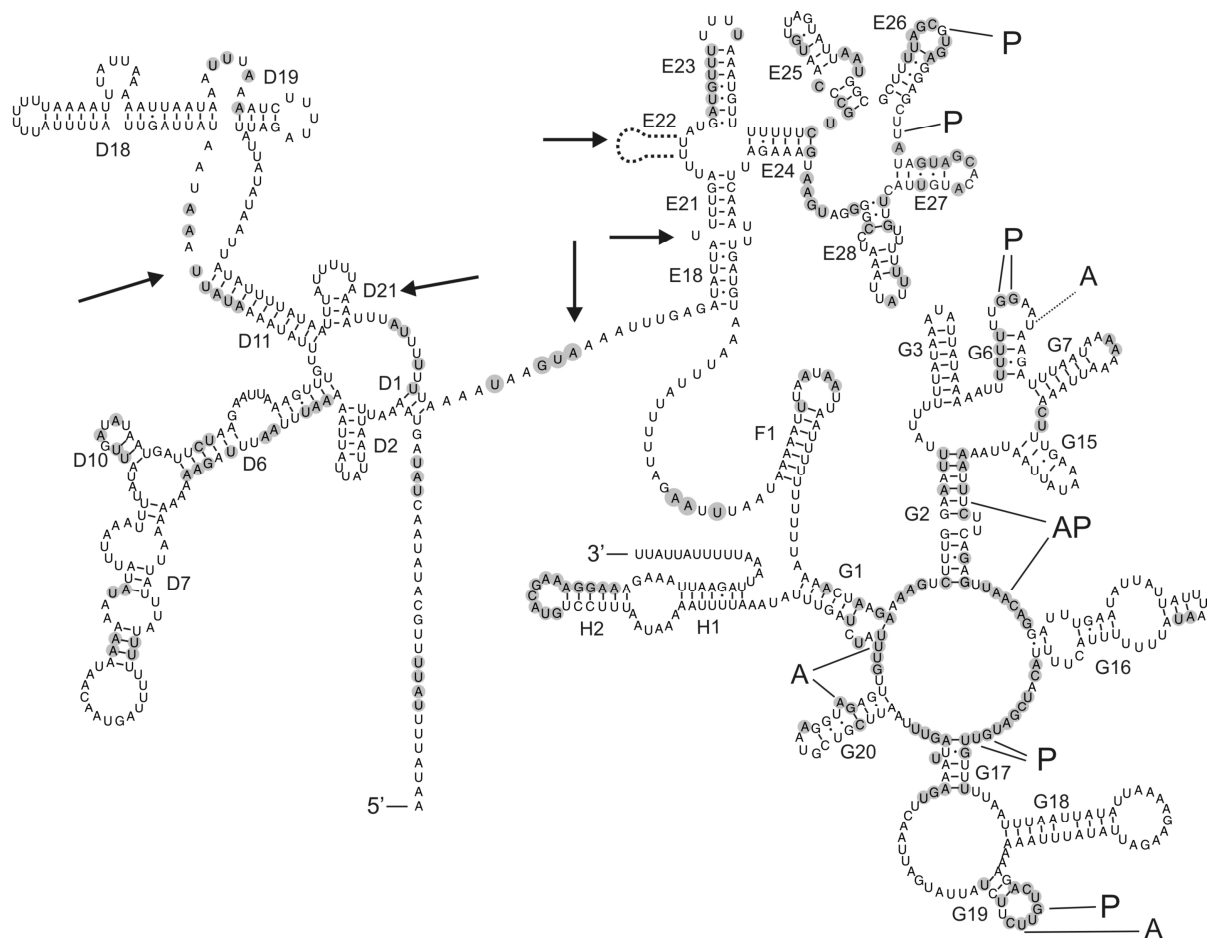
Figure 31: Predicted secondary structure of the 16S rRNA gene of *R. similis*. Watson-Crick base pairing is indicated by a line, whereas G:U base pairing is indicated by a dot. Helices are numbered according to De Rijk et al. (1999) and other published rRNA secondary structures of nematodes. Putative binding sites for the amino-acyl stem of tRNAs (A), peptidyl-transferase (P), or both (AP) as defined by Noller et al. (1986) and Hu et al. (2002) are indicated. The A-site at stemloop G6 is indicated by a dashed line since this nucleotide is not conserved. Shaded nucleotides are conserved in at least 90% of the currently available nematode mt 16S rRNA sequences based on sequence alignments. Arrows indicate marked alterations as compared to other nematode 16S rRNAs. The E22 stem loop present in 16S rRNA of other nematodes is completely absent in *R. similis*.

e. Transcriptomic data analysis

In an attempt to ascertain the gene boundaries we aligned EST data to the mitochondrial genome (Figure 29). This approach can reveal the start site of polyadenylation, and hence the 3′ end of the gene as no untranslated regions are present in mt genes (Gissi et al. 2003). 623 ESTs (8.9% of total available ESTs of *R. similis*) could be aligned to the mitochondrial genome, including ESTs previously predicted to be mitochondrion-derived (Jacob et al. 2008). The

different ESTs cover 7,971 nt or 47.5% of the mitochondrial genome. Remarkably, the majority (n=581) of the mt ESTs are derived from 12S rRNA, pointing to a higher expression level and /or higher stability compared to the other genes. None of the mt ESTs contain a poly(A) tail and none match what is undoubtedly a PCG boundary. Only one EST cluster aligned to a gene boundary by starting exactly at the 5' end of tRNA^{Glu}. In general, very few sequence differences between the mt DNA and ESTs were present. Most differences were indels, preferentially comprising one or two thymidines occurring in poly(T) tracts. In contrast to this overall sequence similarity, 3 different EST clusters — containing 549, 18 and 2 ESTs respectively — match the genomic 12S rRNA with sequence differences of 0, 2.4 and 7.5%. Some transcriptomic data matched parts of the non-coding regions: one EST sequence aligned to a part of the large repeat region, and few ESTs aligned to mt DNA sequence up to 323 nt after the last identified coding gene tRNAL^{Lys}. Surprisingly, the putative non-coding sequence between tRNA^{Val} and tRNA^{Lys} (155 nt) is completely covered by ESTs. Combined with the fact that most other tRNAs flank protein-coding genes, it may indicate that a small gene, encoding a protein of approximately 51 aa resides here. However no homology in databases was found with this 155 nt sequence as a query.

f. Non-coding regions

The large non-coding region in nematode mt genomes is called the control or AT-region and varies considerably both in length (from 94 nt for *Xiphinema americanum* to 2154 nt for *Heterorhabditis bacteriophora*) and nucleotide composition (from 71.8% AT for *X. americanum* to 93.1% AT for *C. elegans*). In the mt DNA of *R. similis* the control region has an AT-content of 87.2% and occupies 3,383 nt or 20% of the genome making it the largest nematode control region described to date. Based on sequence features, the region was divided into five sub regions (see Figure 29). The first part covers 362 nt and starts after the last gene (tRNA^{Lys}) and runs until a large stem loop structure. Characteristic for this region are some poly-thymine stretches (4 x T{5}, 2 x T{7}, 2 x T{10} and 1 x T{11}), and some poly-adenine stretches (2 x A{5} and 1 x A{6}), resulting in an AT content of 87.6%, which is slightly higher than the rest of the genome (Figure 29). Some of the poly-T stretches are separated by a single G nucleotide, a motif that can be found in several other nematode control regions (Li et al. 2008; Montiel et al. 2006; Okimoto et al. 1991). The large and nearly perfect stem loop of 126 nt at the end of the first region consists for 99% of AT resides, with a 'bulge' of A

nucleotides located in the stem (Figure 33 b). The second region of 460 nt runs until the small repeat region. In contrast to the first region, the second region is enriched in GC nucleotides (81.2% AT) and no ESTs were found corresponding to this region. The GC nucleotides are concentrated in a 225 nt region, and can adapt a stable secondary structure (Figure 33 a). The third region (84.5 AT%) contains a repeat region of 806 nt in which a 26 nt motif is repeated 31 times. The fourth region part is an AT-rich sequence part (87.5% AT) of 944 nt and two unique features distinguish this region from the rest of the genome. Firstly, the G-content is comparable to the C-content (6.9% versus 5.6%), indicating possible secondary structures, especially since GC-rich patches are surrounded by AT-rich sequences (Figure 33 c and e). Secondly, a stretch of 6 C nucleotides is present, which is remarkable in light of the very low C-content of the complete mitochondrial genome (4.4%). The sequence surrounding this C-stretch can fold into stable secondary structures (Figure 33 c). A pattern search identified similar C-rich stretches in all nematode mt control regions. Interestingly, the origin of replication of the light strand is predicted to be located in this region (Figure 32).

The figure shows a plot with "Cumulative GC-skew" on the Y-axis (0 to 800) and "Position" on the X-axis. A label "oriL" appears near position 11470. The trendline equation is:

$$y = -7E\text{-}07x^2 + 0.0473x - 13.768$$
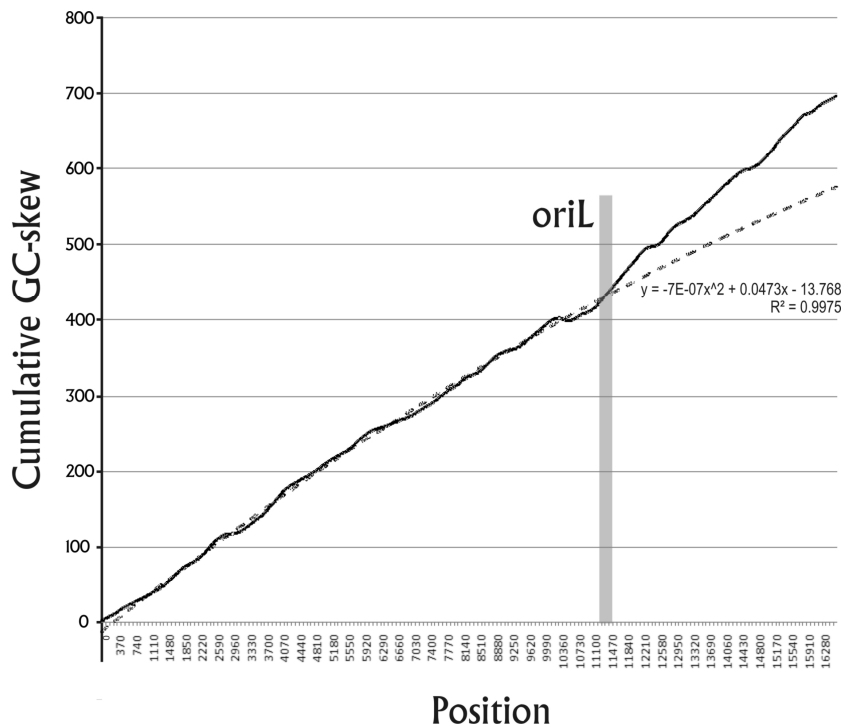$$R^2 = 0.9975$$

Figure 32: Cumulative GC-skew analysis of the mt genome of *R. similis* according to Grigoriev (1998). The X-axis represents the position on the genome (starting from tRNA[Asn]) and the Y-axis shows the cumulative GC-skew value. The nucleotide composition is biased due to the single stranded situation during DNA replication, which is visible as a GC-skew offset from the second order polynomial trendline. This offset corresponds to the origin of replication of the light strand (oriL) (see also Figure 29).

The last part of the control region is another repeat region of 1,208 nt, containing a 302 nt sequence directly repeated 4 times, with an AT content of 84.4%. One EST sequence is derived from this repeat. Surprisingly, this repeat region partly overlaps with tRNA$^{Ser}_{AGN}$. Consequently, the last 15 nt of the repeat unit are identical to the first 15 nt of this tRNA gene. The two repeat regions (the third and fifth region) contribute most to the size of the control region. In contrast to *C. elegans* and *A. suum*, no runs of AT dinucleotides are found in the control region (Okimoto et al. 1994). Besides the large non-coding region, 14 smaller non-coding regions were found interspersed in between the genes of the coding region. Although most are small (ranging from 1 to 15 nt), one is 65 nt in length and found between tRNA^{Met} and tRNA^{Thr}. This region is solely composed out of AT nucleotides and can potentially form a secondary structure (Figure 33 d).

4. Discussion

4

Comparative analysis of mt genomes, both at the sequence level and the gene arrangement level, can be used as a tool for resolving phylogenetic relationships on different levels (Hu et al. 2006; Hwang et al. 2001; Larget et al. 2005). This technique is often used based on assumed benefits over other techniques (Kang et al. 2008; Montiel et al. 2006). However, the lack of sequenced nematode genomes in combination with the extreme diversity of the phylum makes the validity of this approach for nematode phylogeny uncertain. The mt genome of *Radopholus* differs from any known nematode mt genome in both overall sequence similarity as in gene order. The mt DNA gene arrangement of many Chromadorean nematodes is relatively conserved between many lineages. Currently Chromadorean nematodes are divided into five 'gene arrangement' (GA) classes (Gissi et al. 2008; Hu et al. 2002; Kang et al. 2008). Comparison of the gene arrangement in *R. similis* mt DNA with other nematode mt genomes reveals very few conserved gene boundaries with known GA classes (Kang et al. 2008; Kim et al. 2006a). Only *nad1-atp6* (as found in GA2), *cox2*-tRNA^{His}-16SrRNA-*nad3* (as in GA2 and GA3), *nad6-nad4L* (as in GA2) are conserved. Hence, the *Radopholus* mt genome belongs to another novel GA class. Due to the lack of mitochondrial data from close relatives, resolving the phylogenetic position of *R. similis* with the current nematode mitochondrial genomic data is not useful.
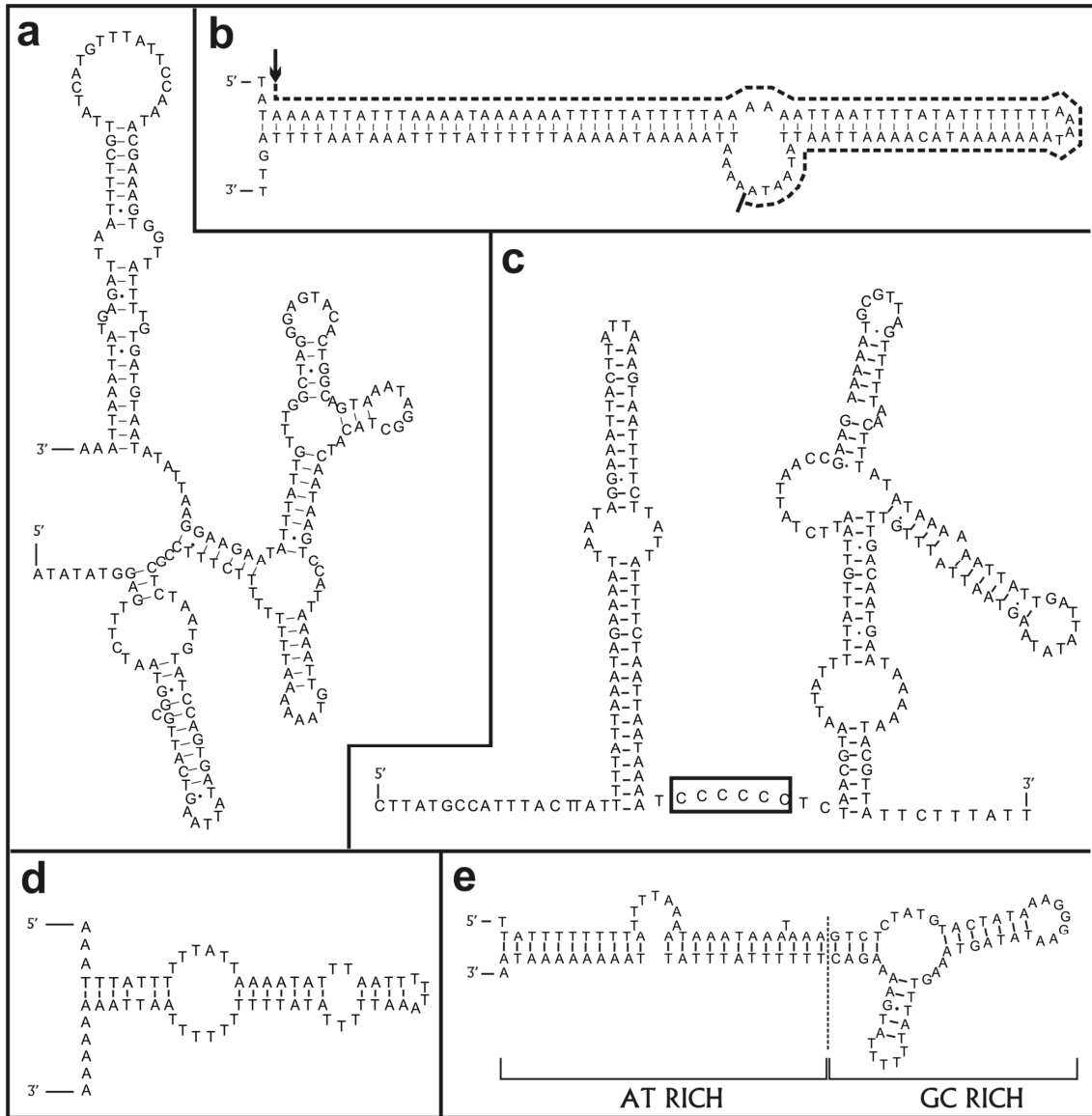
Figure 33: Summary of potential secondary structures in the non-coding regions. **a,** GC-rich region located before the short repeat region. **b,** large stem loop with A-rich 'bulge', with EST data indicated by dashed line (arrow: 5' start). **c,** The six C nucleotide motif located in the sequence between the two repeat regions; **d,** AT-rich non-coding region between tRNA^Met and tRNA^Thr. **e,** local GC-rich region, surrounded by AT-rich sequences, located in the sequence between the two repeat regions. Note the stretch of 8 A nucleotides at the 3' end, which is similar to a sequence described in few filarial parasites by Hu et al. (2002) and Jex et al. (2008).

The arrangement of three PCGs (*nad1-atp6-nad5*) as consecutive genes lacking intergenic tRNAs is rather unique in Chromadorean nematodes. Only in *Enterobius vermicularis* such arrangement of three genes is identified (*cox1-nad1-atp6*). In contrast, a consecutive arrangement of two (small) PCGs is often observed in many other species, which has been

explained by the properties of the small subunit of the ribosome, which needs at least 400 nt on an (mono- or dicistronic) RNA template for efficient binding (Taanman 1999).

Additional features characteristic for the *Radopholus* mt genome are the unique genetic code (see next chapter), the truncated 16S rRNA, the unusual transcriptomic data and the large sequence repeats in the control region. Two sorts of repeat units can be distinguished, namely large repeat units and short repeat units, of which the latter's tandem repetitions are commonly associated with very high frequencies of heteroplasmy (Zhang et al. 1997). Whether heteroplasmy occurs in *R. similis* remains to be investigated. Divergent mt genomes provide an opportunity to detect recurring regulatory patterns and so to reveal common features (Schäfer 2005). Insect and nematode mt genomes often posses control regions with long palindromes (Arunkumar et al. 2006). Indeed, an extremely long palindrome is located in the control region of the *Radopholus* mt genome, and this was also found in some filarial nematodes (Hu et al. 2002). This pattern is located in nematodes at the 5' end of the control region and in *Radopholus* coincides with the end of transcription as revealed by EST data. This secondary structure could therefore be involved in transcription termination. Notably, more or less conserved sequence motifs between nematode control regions are repeatedly located in GC-rich patches. Despite its alternative name (AT region), all nematode control regions contain locally GC-rich sequence parts. Remarkably, in a sliding window of 20 nucleotides, the most GC-rich regions in all nematode AT regions ranged from 30% GC up to 80% GC, with a mean of 51% GC (data not shown). It is clear that these GC-rich patches are of utmost importance for forming stable secondary structures in the control region, structures which may play roles in guiding nuclear-encoded regulatory factors, for example to guide the DNA replication. Based on GC-skew analyses, oriL is predicted to be located in the vicinity of potential secondary structures in the non-coding region. The observed increase in GC content in a large part of the mt genome can not be explained readily, but this nucleotide bias could be a result of the DNA replication processes (Grigoriev 1998, Figure 29).

The genomic sequence located between tRNA^{Val} and tRNA^{Lys} is classified as non-coding. However, some evidence suggests that this sequence encodes – or has encoded –a PCG. Firstly, most PCGs are bordered by tRNAs, and secondly, the sequence is covered by ESTs. On the other hand, no homology in any reading frame to known genes was detected. Intriguingly, the mt genome of *Meloidogyne javanica* contains a similar unidentified open reading frame, which was regrettably not further investigated (Okimoto et al. 1991). In both

cases the intergenic space encompasses only a very short nucleotide sequence, encoding a protein of approximately 50 amino acids. If this sequence truly encodes a protein, it could be an *atp8* gene, the shortest and most highly diverse gene found in most metazoan mt genomes. ATP8 is absent in nematode mt genomes, except for *Trichinella spiralis*. In this species, an ATP8 protein is characterized having a similarity with insect *atp8* genes (n = 77) of on average 40.7% (±5.3). The similarity to the unknown sequence of *R. similis* reaches a comparable 36.1% (±5.7). However, the sequence with highest homology contains an in frame UAG stop codon. It should be stressed that the ATP8 protein is characterized by a higher conservation of the secondary structure compared to the primary sequence, hampering correct annotation. Nevertheless, the annotation of the ATP8 gene in *T. spiralis* is being questioned, since it lacks the N-terminal MPQ motif frequently observed of other metazoa (Gissi et al. 2008). Similarly, no such motif is present in any of the reading frames of the sequence between tRNA[Val] and tRNA[Lys]. Maybe this region is not translated into a protein product, and could be a molecular evolutionary relic in the mt genome of *R. similis*.

Besides the unique features of the sequenced genome, some elements identified in other nematode mt genomes were not found. For example, the putative promoter element identified in the mt genome of *X. americanum* by He et al. (2005) could not be detected. Similarly, a non-coding region forming a hairpin structure with a stretch of thymidines in its loop - as described for *Caenorhabditis elegans*, *Ascaris suum*, and *Steinernema carpocapsae* (Montiel et al. 2006) - could not be identified in the mt genome of *R. similis*. This motif is thought to guide mtDNA replication. However, in *Radopholus* another small non-coding region located between tRNA[Leu]uuR and tRNA[Met] could have a stable secondary structure, which could be involved in replication regulation.

In conclusion, we determined the complete sequence of the mitochondrial genome sequence for *R. similis*. Although at first sight the sequence has all characteristics of a typical nematode mt genome, numerous idiosyncratic features confirm its uniqueness. The gene sequences and their arrangement are the most diverged of any nematode mt genome known to date. Furthermore, the unique genetic code and the complications for transcription and translation offer an opportunity to further elucidate the processes of the mt decoding machinery, which are currently poorly understood.

5. Material and methods

a. DNA extraction and LD-PCR

Total DNA of 10,000 *R. similis* nematodes was phenol/chloroform extracted. On this DNA, LD-PCR using a combination of Expand Long Range dNTP Pack (Roche, Mannheim, Germany) and Phusion High-Fidelity DNA Polymerase (Finnzymes) was performed with primers based on EST sequences of *coxI* and *nad5* (Jacob et al. 2008) and degenerate primers based on *cytB*. Resulting fragments were sequenced by AGOWA (Berlin, Germany). The PCRs were done following the manufacturer's instructions in a 50 μL reaction mixture, containing 6% DMSO on 500 ng of total DNA. After sequencing the fragments, additional primers were developed for LD-PCR. Finally, amplification of the complete genome in three overlapping fragments succeeded using the primers listed in Table 13. Fragments were directly sequenced by AGOWA.

b. Collection of mitochondria

Approximately 400 μL nematodes (corresponding to approximately 200,000 nematodes) were collected from carrot discs, washed once with demineralized water and placed into 1 mL sucrose buffer (300 mM sucrose, 30 mM Tris, and 10 mM EDTA, pH7.9). With the aid of a teflon pestel and some sand the nematodes were crushed in a glass tube. The nematode homogenate was transferred to centrifuge tubes, and centrifuged for 5 minutes at 500G, after which the supernatant was subsequently centrifuged for 5 minutes at 1000G. Finally, the supernatant was centrifuged at 15,000G for 45 minutes, the pellet redissolved in sucrose buffer, and again centrifuged at 15,000G for 45 minutes. The resulting purified mitochondria were immediately used.

Table 13 Primers used for amplifying the mitochondrial genome of *R. similis*

Primer name	Primer sequence	Amplicon
Rsrad5kb_3pr1	TAATACCTCCAGCCAAAACTGG	5kb piece
Rsrad5kb_5pr1	CGTAGTGTTTTACCTCGTTATCGTT	
Rsmit5kb_3pr2	GCTGATATAGCTTGACCTCGTG	10 kb piece
Rscox3_fw1	ACCTAAAGGCACTCAATTATCTCC	
Rsmit5kb_5pr2	AAGCAATTGGGCGAATTAAA	3 kb piece
Rscox3_rv1	AGTCCCTTTAGGTGATATTGGAGA	
RsNADHD_2	CCTAAAGCTATGAGAGCACCTACACC	3.5 kb piece
RsCOXI_7	CCAAATGACCAGATGATCTTAAAGG	

c. Mitochondrial RNA extraction

For RNA extraction, the mitochondria were dissolved in 200 μL Trireagent (Sigma, St. Louis, MO, USA), and incubated for 1h at RT. The solution was exposed to sonication at the lowest setting for three times two seconds

(Branson Sonifier 250, Danbury, USA). After incubation for 15 minutes at RT, 40 µL chloroform was added and the tube vigorously shaken by hand. Further RNA extraction steps were performed following the manufacturer's instructions.

d. CT-RT-PCR

Mitochondrial RNA was circularized using T4 RNA ligase (Fermentas, St. Leon-Rot, Germany), using 1 µg mitochondrial RNA in a total volume of 20 µL, following the standard manufacturers instructions. Additionally, we tried to optimize the reaction for the PCGs by using different concentrations of RNA (up to 10 µg), by adding 10% DMSO which may eliminate stable secondary structures, and by adding BSA to a final concentration of 0.06 mg/ml. After incubation for 1h at 37°C, the volume was adjusted to 200µL with demineralized water, and a standard phenol-chloroform extraction was performed. The reaction was also performed at 17°C for 10 hours for the PCGs. The resulting pellet was redissolved in 20 µL demineralized water, and used for the reverse transcriptase reaction. The reaction was performed, using SuperScript II (Invitrogen, Carlsbad, CA, USA) following manufacturer's instructions and a reverse-oriented gene-specific primer located at the 5' end of the gene. The resulting cDNA was subjected to nested PCR using gene-specific primers. After checking by agarose gelelectrophoresis, the PCR products were ligated into the pGEM-T vector and cloned into *Escherichia coli* DH5α cells. Selected inserts were sequenced.

e. Southern blot

1.2 µg total DNA was digested with SpeI and XmnI restriction enzymes, and separated by inversed field electrophoresis in 0.5XTBE 1% agarose gel. After capillary blot, hybridization was performed using the High Prime DIG labeling Starter Kit (Roche) following manufacturer's instructions with a probe constructed using primers RsNADHD_2 and RsCOXI_7 (Table 13) covering 3,504 bp of the mitochondrial genome. The chemiluminiscent signal was visualized on a fluor imager FLA-5100 (Fujifilm, Tokyo, Japan).

f. Annotation

Combinations of BLAST algorithm searches (Altschul et al. 1997) and ClustalW (Thompson et al. 1994) alignments were used to detect protein gene and rRNA sequences. Comparisons to other nematode mt genes were done after gathering the sequences from GenBank using the MitoBank2.1 program (Abascal et al. 2007). G+C percentages were calculated by an in-house perl program. Folding of rRNA sequences based on previously published structures was done with the aid of Mfold (Zuker 2003), and structures were edited with RNAviz (De Rijk et al. 1997). To detect tRNA genes, SE-tRNAScan (Lowe et al. 1997) with parameters set for 'nematode mitochondrial tRNAs' was used and for confirmation, tRNA genes were aligned using LARA (Bauer et al. 2007) to other nematode mt-tRNAs. MATGAT2.02 (accessible at http://bitincka.com/ledion/matgat/) (Campanella et al. 2003) was used for constructing identity and similarity matrices of nematode mitochondrial proteins, using the BLOSUM62 matrix. Stable secondary structures in the non-coding region (NCR) were sought with mfold.

Repeats in the NCR were detected by an in-house perl program, and shared motifs with other nematode NCRs were analyzed by the Pattern Search program on the Softberry Server (www.softberry.com).

g. Replication start site determination

In general, replication of both strands of the mt DNA is initiated on different locations. The heavy strand is synthesized first starting at the oriH site, until it reaches the oriL site, leading to initiation of replication of the light strand. Due to this mechanism part of the mt DNA spends more time as a single stranded, which affects the nucleotide distribution along the genome (Grigoriev 1998). The deviation from the remainder of the genome is the biggest at oriH and oriL, since at these site the strands spend the longest time being single-stranded. Through cumulative GC-skew analyses of mt genomes the oriL site can be predicted. With an in-house perl program, the GC-skew values (G-C/G+C) were calculated for each 5th position in the genome using a window of 300 nt and the cumulative values were obtained. In addition, the cumulative GC-content was counted for every 10th position using a window of 100 nt. This was incorporated in the overview figure (Figure 29).

Chapter 5

The evolutionary path to
the unique mitochondrial genetic code
of *Radopholus similis*

"The opposite of a correct statement is a false statement.
But the opposite of a profound truth may well be another profound truth."
Niels Bohr

5

The evolutionary path to the unique genetic code in *Radopholus* mitochondria

1. Abstract

The conserved nature of the genetic code over different kingdoms was an astonishing finding, but one that could be explained in the light of evolution. Later, it was the unexpected exceptions to the standard code – mostly identified in small nuclear genomes and mitochondrial genomes – that drew attention. Although different models were hypothesized to deal with the likely detrimental effects of genetic code changes, recent research recognizes that each case should be looked at separately to construct its most likely evolutionary model. Here we analyse the identified codon change in the mitochondrial genome of *Radopholus*, in which the UAA codon was reassigned from Stop to Tyrosine (Tyr). We gained insight into the evolutionary process by which this codon reassignment appeared through statistical analyses of nematode mitochondrial (mt) genomes. The evolutionary path towards this unique codon reassignment in this species can be explained by a novel synergistic action of two previously described codon reassignment hypotheses. A key player in our model is the significant influence of the nucleotide composition - partly shaped by situation of one coding strand – on the amino acid (AA) content of the encoded proteins, which probably has led to an initial non-optimal AA composition of the mt proteome in *Radopholus*. This pressure disappeared by extending the codon family of Tyr. We expected this unique stop:sense reassignment to be reflected in a unique DNA decoding machinery, especially since the majority of the protein-coding genes lacks a Stop codon. However, northern blotting revealed that transcription occurs from one strand, with transcript cleavage taking place as expected. Additional examination of mitochondrial transcriptomic data (ESTs) showed some aberrant features, such as a complete lack of polyadenylation but also a lack of transcript processing, strongly contrasting the northern blot results.

5

2. Introduction

Mt genomes use genetic codes that are slightly different from the standard nuclear code. Whereas some codon changes are widespread and evolved independently, (e.g. UGA[Stop] to UGA[Trp] occurred at least 12 times independently in the mt genomes of different eukaryotic lineages), other deviations are more restricted (e.g. AGR[Arg] to AGR[Stop] for vertebrate mt genomes) (Knight et al. 2001a). The mitochondrial codons are recognized by tRNAs *via* extended wobble rules for base pairing, by which G is allowed to pair with T. In this way, the 22 mt-derived tRNAs are able to recognize the 62 codons encoding all 20 amino acids (Osawa et al. 1992). Each amino acid is encoded by the codons of its 'codon family', which contains on average three codons (called synonymous codons). The relative use of these codons within a codon family is expressed as the relative synonymous codon usage (RSCU). RSCU differs between genes within one organism, depending on many parameters such as expression level, AA composition, gene length and others (Chen et al. 2004). In addition, RSCU differs between different genomes, correlated with GC content (Chen et al. 2004). Species-specific codon biases are partly determined by the DNA decoding system (favouring certain codons in codon families), and partly by genome-wide mutational pressures acting on all DNA present in a given organism.

Several mechanisms have been proposed as to how codon reassignments in genetic codes can occur, based on a simple 'gain'-'loss' framework (Sengupta et al. 2005). **1)** The extinction of a codon can lead to an altered decoding upon its reappearance ("Codon capture" or "Codon disappearance" (CD) mechanism) (Osawa et al. 1992; Sengupta et al. 2007). **2)** The loss of codon recognition can precede the gain of it by another tRNA, leaving the codon temporarily unassigned. **3)** The gain of recognition by another tRNA can occur while the original tRNA is still functional, creating a temporarily ambiguous decoding state ("Ambiguous intermediate" (AI) model) (Schultz et al. 1996). **4)** Finally, two simultaneous mutations can compensate each others lethal effect, by which the reassignment occurs at once (Sengupta et al. 2007). Regardless of the mechanism, evolutionary pressure must drive this change. We noted that cause (pressure) and effect (reassignment) are very often mixed up, although both are very different aspects of genetic code evolution (Knight et al. 2001a). An often cited evolutionary pressure is the strive for compactness of mt genomes, affecting the genetic code to minimize

the number of tRNAs needed to decode all codons (Andersson et al. 1995). However, no link between genome size and number of deviations of the canonical genetic code was observed (Knight et al. 2001b). Alternatively, most reassignments have lead to an increased use of amino acids with lower metabolic cost, making the proteomes less expensive to synthesize (Swire et al. 2005). In addition, the harsh oxidative environment of mitochondria is generally recognized as another driving force. Adaptation of the mitochondrial proteome to deal with oxidative stresses can occur through reassignment of the genetic code. For example, the reassignment of AUA from Ile in the standard code to Met in the mt code has lead to the accumulation of Met residues in mt encoded proteins. Since Met residues are preferably oxidized over other AA and can be reduced at low metabolic cost, vulnerable AA at catalytic cores are shielded from oxidative stress (Bender et al. 2008). Perhaps the fact that guanine is vulnerable to oxidation (Ravanat et al. 2000) can explain why mitochondrial genomes are driven towards AT enrichment, and subsequently favour the use of AT-rich codons. The single-strandness during replication favours C to T mutations, a phenomenon which could also contribute to the increased AT richness (Krishnan et al. 2004). Additionally, most genetic code changes are found in mitochondria, probably because of its small proteome and the occurrence of heteroplasmy (i.e. the presence of different mt genomes in the mitochondria of one cell), which may increase the chance of successful codon reassignment.

3. Results and Discussion

a. The unique UAA codon reassignment is conserved at the genus level in *Radopholus*

Upon sequencing of the mt genome of the nematode *Radopholus similis,* we discovered that the UAA codon, which encodes translation termination (together with UAG) according to the invertebrate mt genetic code, was abundant in all 12 PCGs (see previous chapter). Based on alignments with other nematode mt proteins, we deduced that this codon encodes Tyr in *R. similis*. Strikingly, the newly assigned codon is preferred over the commonly used UAU codon for Tyr (4.65% and 3.76% of the total number of codons respectively). A third Tyr-codon (UAC) is not used, as is the case for virtually all codons with a cytosine at the wobble position in this highly AT-rich genome (85.4 %AT) (Figure 34). The unique mitochondrial genetic code was confirmed at the genus level, since a sequenced fragment (NAD3 - tRNA$^{Leu}_{UCN}$– COXI) of

the mt genome of *R. arabocoffeae* also contained the UAA[Stop] to UAA[Tyr] reassignment (data not shown). Although a similar codon reassignment was once reported for the planarian *Dugesia japonica* (Bessho et al. 1992), subsequent studies could not confirm this finding (Jeon et al. 2007; Le et al. 2004; Nakao et al. 2000; Telford et al. 2000). Thus until now, the UAA[Stop] to UAA[Tyr] reassignment is unique for the mt genome of *Radopholus similis* and *R. arabocoffeae*.
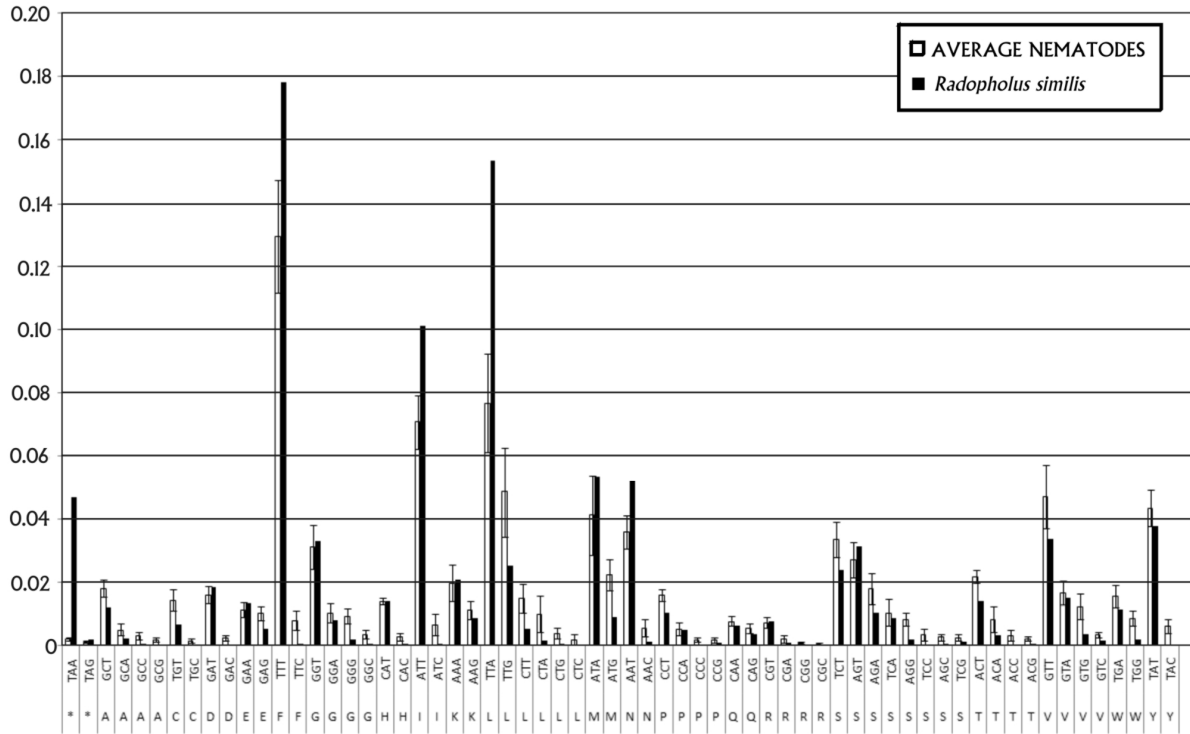

Figure 34: Frequency of codons in the mitochondrial genome of *Radopholus similis*, compared to the average in nematodes.

b. The nucleotide composition influences amino acid content in nematode mt genomes

To compare the mt genetic code of *Radopholus* with other nematodes, we performed a principal component analysis (PCA) on available nematode mitochondrial data. The first principal component (PC) resulting from a PCA based on codon usage data, revealed no difference between *Radopholus similis* and other nematodes. However, *Radopholus* was clustered separately from other nematodes based on the second PC (Figure 35). Similarly, PCA based on amino acid composition yielded a PC1 which is not discriminative for *Radopholus*,

however, again the PC2 discriminates strongly. Interestingly, this PC2 was strongly correlated with the average AT content (Figure 36), demonstrating that nucleotide composition has an influence on the AA content of the proteome (Boore et al. 2000). In other words, differing AT content does not only change RSCU, but changes also AA composition. However, the observed incidence for Tyr of 8.41% in *Radopholus* is lower than expected from the chances
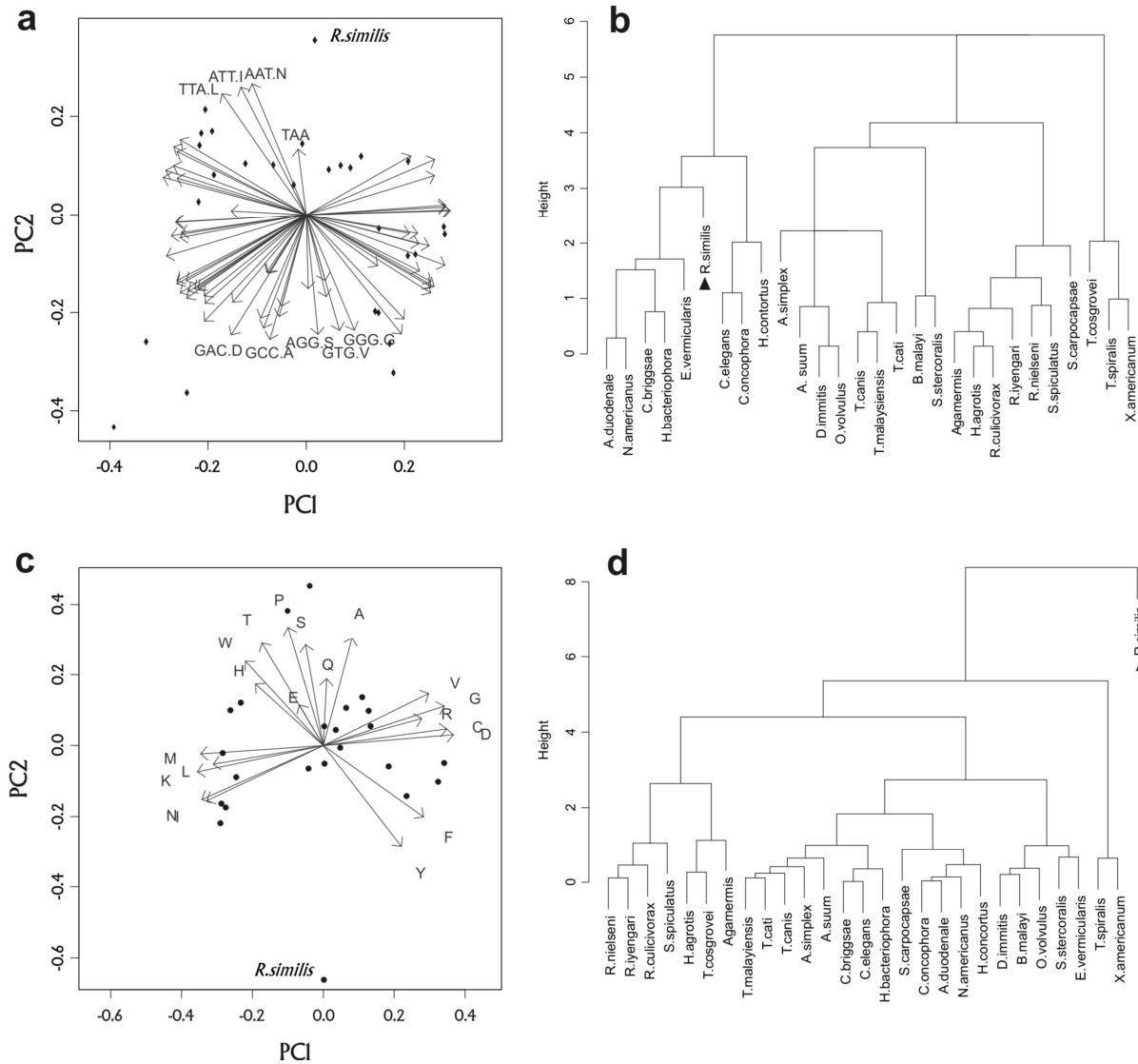

Figure 35: **a** and **c**, scatter plot of the two first principal components of the PCA, respectively on the codon usage and on the amino acid content of all nematode mitochondrial genomes. **b** and **d**, divisive hierarchical clustering using the first 4 principal components of the codon usage PCA and the amino acid composition PCA respectively. For the codon usage PCA, the first two components explain 66.6% of the variation in the data. For the amino acid PCA, the first two components explain 68.8 % of the variation. *R. similis* is indicated by an arrowhead.

of occurrence of UAU and UAA codons (15.02%) based on plain nucleotide composition, but is still higher than the nematode mean of 4.95%. An overview of AT content in nematode genomes on AA contents is given in Figure 37. The amount of some AAs encoded in the genome is barely correlated with the genomic GC content (see Figure 37, H, R and Q), but some amino acids respond heavily (see Figure 37, V, M and I).

c. Pressure on the codon family of Tyr led to the UAA codon reassignment

Although the 'CD' model (see Introduction) was used to explain other stop:sense reassignments (Swire et al. 2005), this mechanism can not explain the observed reassignment in *R. similis*. In all other nematode mt genomes UAA codon is the preferred Stop codon (42% vs. 24% for the alternative UAG), and despite being a Stop codon, it is unlikely that UAA has disappeared in an ancestral genome, prior to reassignment to Tyr in the mt genome of *Radopholus*. Indeed, other large scale studies found that low codon frequencies are neither necessary nor sufficient for reassignment (Knight et al. 2001b; Yokobori et al. 2001). Instead, a more complicated scenario is needed to explain this unique codon change. The decrease of GC content in the mitochondrial genome of *Radopholus* drove the UAC codon to extinction (as a first step of the CD model), leaving Tyr to be encoded by UAU only, a situation which generates a pressure on the encoding amino acid Tyr. This pressure could manifest itself as a too low incidence of Tyr and as multiple slightly deleterious mutations in the mt genome.

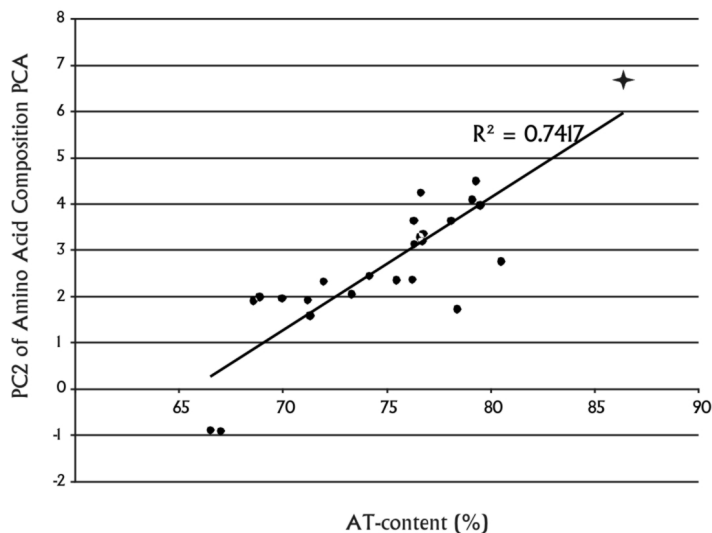
Figure 36: Influence of AT-content on the nematode mt proteomes, with the *Radopholus similis* indicated by a star. The differences in amino acid (AA) composition of nematode mt genomes (Y-axis) correlate well with their AT-content (X-axis).

Consequently, a gain emerges when the tRNA of Tyr expands its codon recognition capacity via ambiguous decoding of the UAA codon (following the 'AI' model). This codon could have occurred maximally 12 times in the ancestral mt genome as a Stop. Since tRNA[Tyr] (with anticodon GUA) is much preserved in nematode mt genomes, we hypothesize that an additional modification causes the extension of the recognition capacity. By extending the coding repertoire of Tyr, some AT-pressure related mutations in the mt genome may be fixed, leading to an increased use of the newly reassigned UAA codon and its corresponding AA Tyr. This should have created an advantage, despite the ambiguous coding situation having a severe effect on the correct translation of the mitochondrial transcriptome due to competition between the release factor and tRNA[Tyr] for recognition of the UAA codon. Truncated proteins or conversely, extension of proteins, would have been the consequences of such a compete-
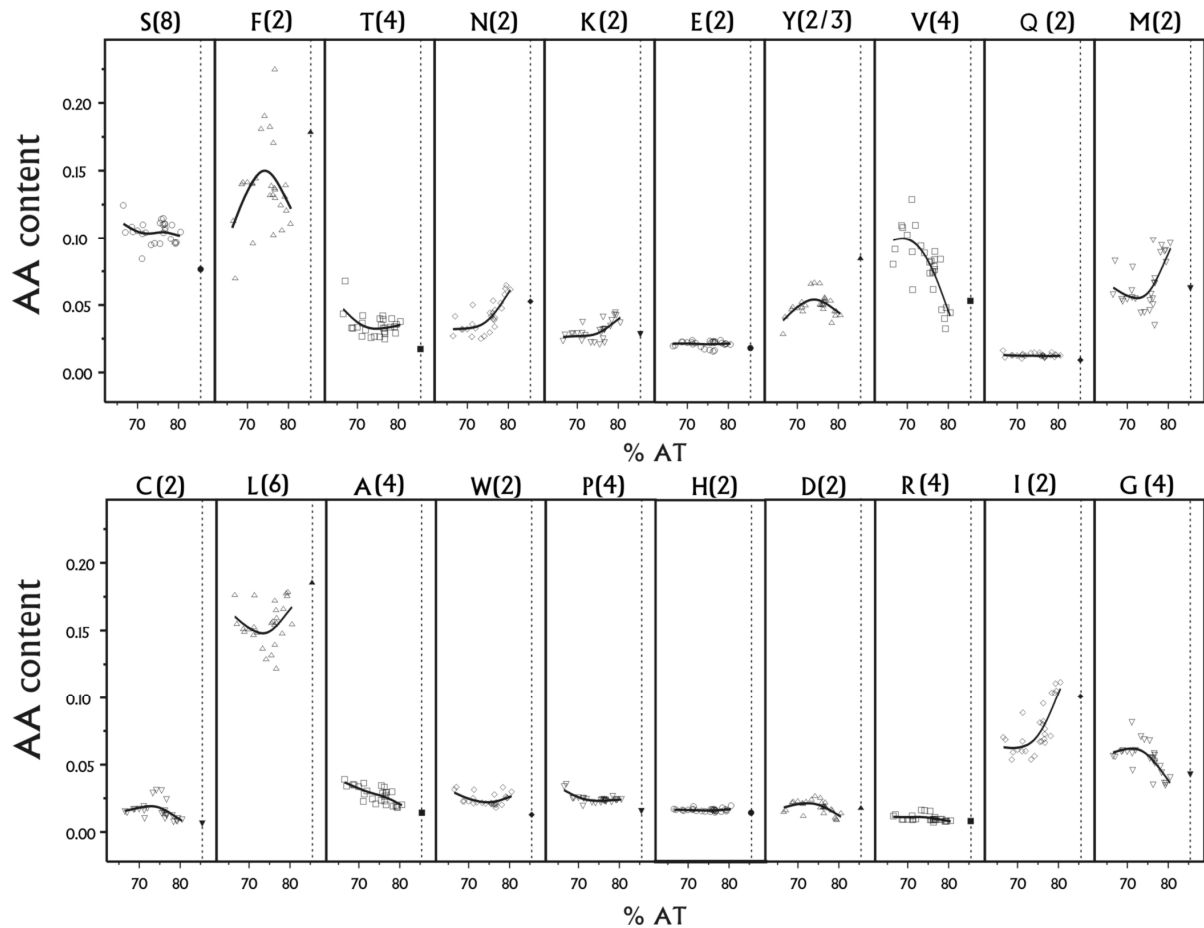
Figure 37: The influence of AT content on the amino acid content in nematode mt genomes, with the *R. similis* data point indicated by a vertical dashed line at 85.4%. A smoothing spline curve with 4 degrees of freedom is fitted to the data points – excluding the *R. similis*. The number between brackets shows the number of codons encoding the residue.
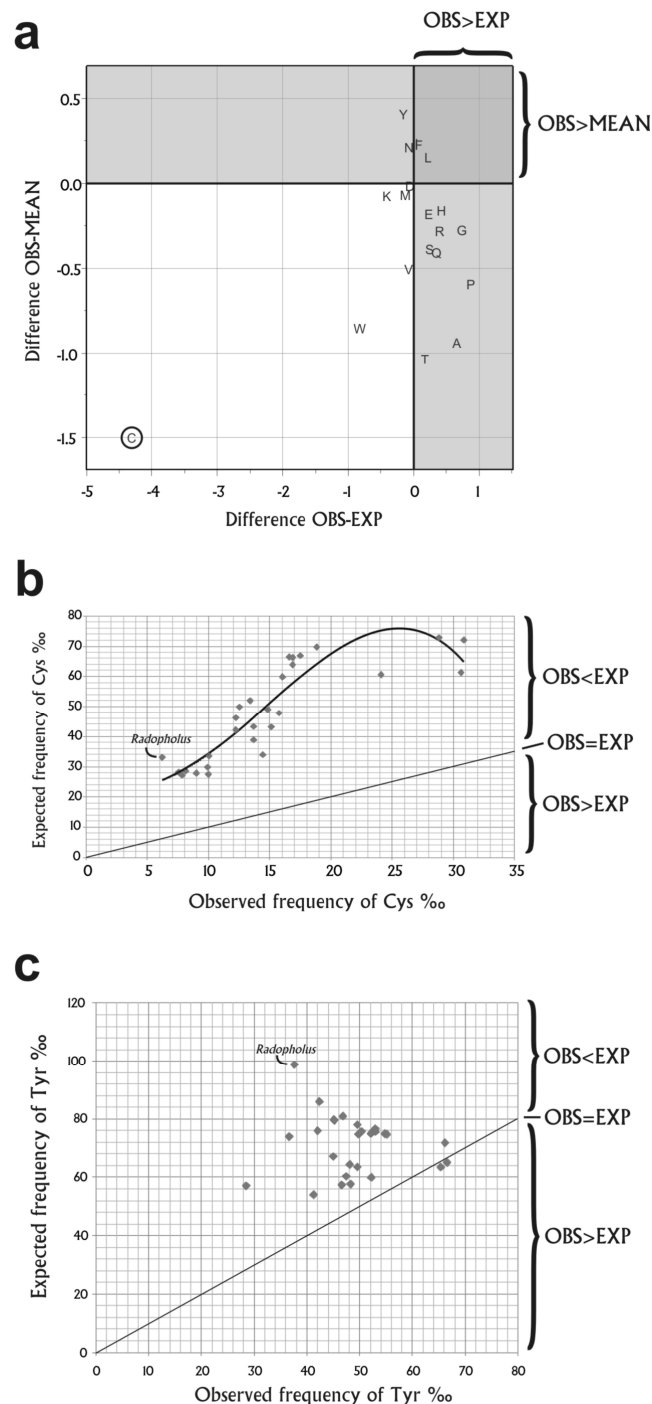
Figure 38: **a,** overview of the AA composition in *R. similis*. The Y-axis represents the relative difference of the observed incidence (OBS) of an AA minus the nematode mean (MEAN) of that AA divided by OBS. A value close to zero indicates a similar content as the nematode mean. Similarly, the X-axis shows the relative difference of OBS with the expected incidence (EXP), based on the nucleotide composition of the mt genome of *R. similis*. From this, it is clear that Cysteine (C, encircled) is encoded much less as the nematode mean, and is encoded less than expected from nucleotide composition. **b,** Scatterplot of OBS (X) versus EXP (Y-axis) of Cys content in all nematode mitochondrial genome. In all nematodes, Cys is encoded less than expected, and in *R. similis* it is encoded the least. **c,** Similar scatterplot for Tyr as explained in b.

tion. Nevertheless, even these events could have been not entirely negative, since an elevated stress response may be triggered, giving the organism a benefit in an environment under strong selective pressure (Miranda et al. 2006). In addition, high numbers of Tyr (since *Radopholus* mt proteins contain above average Tyr residues) could exert a beneficial effect as Tyr has a powerful antioxidant capacity, providing potentially an additional benefit to mitochondrial protein in the oxidative environment (van Overveld et al. 2000). Furthermore, heteroplasmy could have helped mitochondria with the ambiguous situation during the early phases of the reassignment to persist in the cell. Modification of translation termination mechanisms, such as a mutation of the release factor to recognize solely the UAG stop codon constitute the last phase of the codon reassignment.

Note that our model does not obey to the simple 'gain'-'loss' framework proposed previously, since the lost codon is not the same as the gained codon (Sengupta et al. 2007). Tryptophan (Trp) could have followed a similar scenario, being encoded before its reassignment by UGG only, after which it took over the UGA stop codon. Due to increasing AT content, UGG could have disappeared. For this case, often the 'CD' model is supposed as the correct one: UGA disappeared (leaving two other stop codons UAA and UAG) from the mt genome, and upon its recurrence the meaning was changed from Stop to Trp. The tRNA^Trp has UCA as anticodon in nematodes, which needs to be modified or edited to recognize the appropriate codons – in contrast to e.g. fungal mt genomes in which tRNA^Trp has changed its anticodon to UCA (Alfonzo 2009; Carullo et al. 2008). But apparently the situation of 'one amino acid – one codon' suffices not to force a change. Analysis of all amino acids in *Radopholus* revealed that Cysteine (Cys) is the most deviating, as it is 2.5 times less often used than the nematode average (Figure 38). Remarkably, in all nematodes – including *Radopholus* – Cys is encoded less than expected from the coding strand nucleotide composition. In *Radopholus* however, it is encoded the least compared to all nematodes, and represents the only case in which Cys is encoded by only one codon (UGU) in the mt genome, since UGC does not appear. However, its codon family can not be expanded readily, since in this case it should hijack a codon of Trp, shifting the problem.

The majority of codon reassignments involve tRNA alterations by post-transcriptional modification (Knight et al. 2001b). Not many tRNA modifications are described today that can comply with the extended recognition capacity of *R. similis* mt tRNA^Tyr, although tRNAs that recognize stop codons have been identified (so called suppressor tRNAs) (Beier et al. 2001).

5

Methylation of G at the first anticodon position, as has been found in Echinoderm and squid mt genomes, is known to enable the tRNA to recognize the complete codon family (Knight et al. 2001a). However, in this case the UAG stop codon would also be recognized. Because the UAG Stop codon is present on 3 consecutively positioned PCGs without being separated by
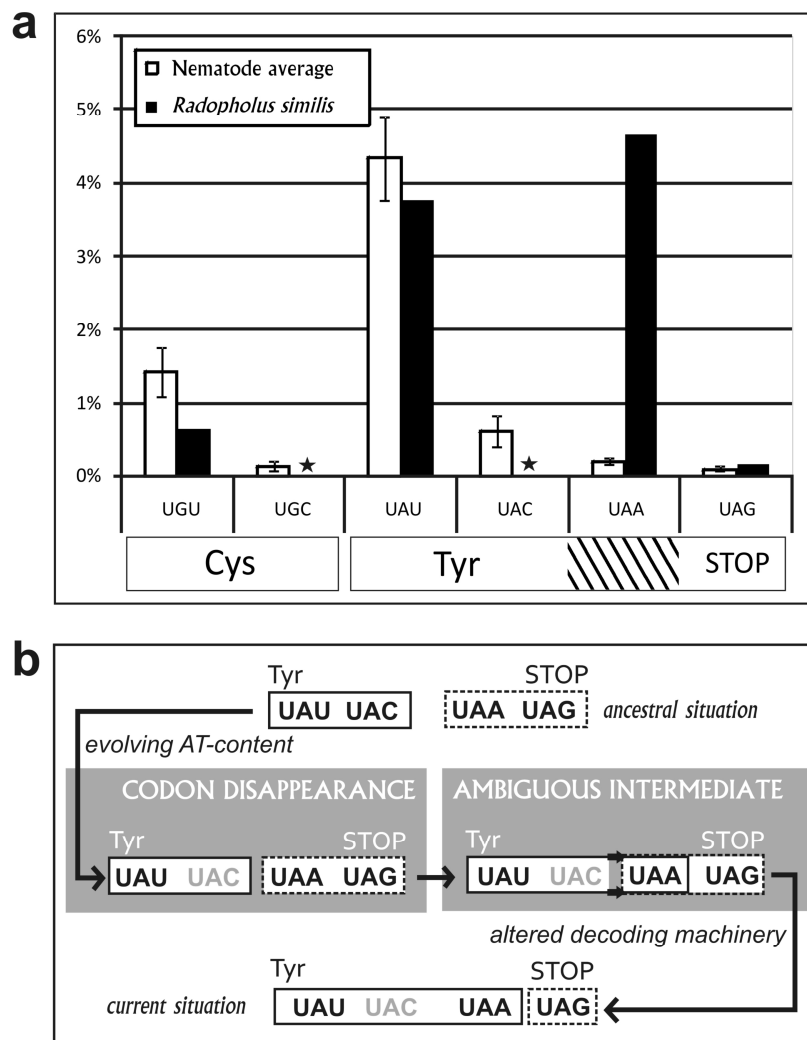

Figure 39: Overview of the influence of the AT-content on Cysteine (Cys) and Tyrosine (Tyr), and the putative evolutionary model behind the observed genetic code change. **a**, The high AT-content of the mt genome of *R. similis* has greatest influence on codon family of amino acids Cys and Tyr (unused codons are indicated by a star). While Tyr has expanded its codon repertoire with UAA (hatched rectangle: in *Radopholus* coding for Tyr, in all other nematode for Stop), Cys has not expanded. **b**, Schematic representation of the evolutionary path to reassignment of the UAA codon, combining two previously presented models (at the right side 'CD' and at the left side 'AI' model, see text). Due to the increasing AT-content (and concomitant strong decrease in C nucleotides) (upper arrow), the UAC codon disappeared, leaving just one codon (UAU) for Tyr. This pressure forced tRNA^{Tyr} to expand its recognition capacity to include the UAA codon (middle arrow), causing an ambiguous decoding situation, which initially could be tolerated in stressful environments. The ambiguity was resolved by modifying the decoding machinery to restore its efficiency and functionality (lower arrow).

tRNAs, seemingly this codon still functions as a Stop, making an existing ambiguous situation highly unlikely. Otherwise, the ciliate *Euplotes* has a conserved tRNA[Cys] (anticodon GCA), which is able to recognize UGU, UGC and UGA. No modification is (yet) found on the corresponding tRNA species to enable such recognition. Alternatively, the authors explain the decoding by suggesting an altered release factor with diminished binding capacity, and an increased abundance of a (nuclear encoded) tRNA species (Grimm et al. 1998). In *R. similis* mitochondria we do not expect a higher abundance of tRNA[Tyr] compared to other tRNAs, as tRNAs ought to be produced in equal quantities based on the proposed transcription process. To investigate possible mutations in *R. similis* release factors, ESTs were explored for them. Unfortunately, no corresponding tags were identified in EST data, so no comparison with other nematodes was made. Note however, as an alternative to tRNA modification or mutation, a cytosolic tRNA specifically recognizing the UAA codon may also be imported from the cytosol through the RNA import pathway in mitochondria, although this pathway has not yet been investigated in nematodes (Kamenski et al. 2007).

d. Inconclusive results for an altered mitochondrial transcription in *Radopholus*

Clearly, for the codon reassignment to be completed, alterations to the DNA decoding machinery should have taken place in response to the ambiguous coding situation. Due to the shift in codon usage, the codon table contains only a single stop codon (UAG) which remarkably could be inferred for only 5 PCGs, namely *nad4L*, *cox3*, *nad1*, *atp6* and *nad5*. Previous chapter reported that nearly all EST sequences run over gene boundaries. In addition, closer investigation showed that all mt ESTs lack 3' polyadenylation tails, although mitochondrial ESTs in other metazoa contain 3' polyadenylation tails (Gissi et al. 2003). Whether this result is specific for *Radopholus* or holds also true for other nematodes remains to be investigated.

These two observations suggested no processing (cleaving and polyadenylation) of mt transcripts is taking place. To gain more insight into this issue, CT-RT-PCR and northern blot analysis were performed. In short, for CT-RT-PCR, extracted RNA of mitochondria was self-ligated to obtain circularized RNA, after which cDNA was synthesized over the region of the ligated ends which was subsequently amplified and sequenced by PCR. In this way, information can be obtained of both 5' and 3' ends of an RNA molecule. However, the CT-RT-PCR did not yield any results for protein-coding genes in contrast to the small rRNA gene, despite the use

5

of extensive approaches (see previous chapter). This could confirm the observed lack of cleavage in the EST data, as the expected gene ends, on which the primers were based, are lacking. To confirm this, a northern blot analysis on mitochondrial RNA using radio-labeled probes for *cytB* (no stop codon), *coxI* (no stop codon) and *nad5* (UAG stop codon) was performed. Surprisingly, the results indicated one major transcript for each gene (Figure 40). Based on the RNA marker, all three transcripts are approximately 500 bp longer than expected from the open reading frame of the gene. However, this aberrant length might be a result of the high AT-content of the sequences, as was reported before (Zhu et al. 2005). From this we conclude cleavage of the large polycistronic transcript is taking place to set the mRNA free. The sense probes of the analyzed genes did not give any result, indicating that only the coding strand is transcribed in *Radopholus*. Even the *nad5* gene showed cleavage, even though no tRNA gene is positioned before this gene. Since tRNAs were thought indispensable to guide transcript cleavage, other factors must assist in determining the end of these genes. We searched for secondary structures at the 3' end of the genes, which could act as a signal for transcription cleavage (Figure 41). Remarkably, for all but two protein-coding genes such structures could be identified. Thus genes followed by a tRNA contained also these features, which would lead to redundant signal as both could be guiding the cleavage machinery. Nevertheless, with the confirmation of transcript processing, the reason for failure of the CT-RT-PCR is unknown. It is unlikely that secondary structures have hampered the ligation or amplification step, as the CT-RT-PCR of 12S rRNA succeeded. Alternatively, a modification of the gene ends may be present. However, a cap structure as 5'-terminal m7GpppG occurring on nuclear mRNA is not present on mitochondrial RNA (Grohmann et al. 1978). The aberrant EST data remains an enigma. One explanation is that the EST data could be derived from mtDNA rather than RNA, to which the oligo(dT) primers used for cDNA synthesis occasionally bind due to the high AT content. However, no preferential binding sites from the primers could be detected.

From extrapolation of the northern blot results, it seems that the majority of the mt genes in *R. similis* are translated from cleaved transcripts lacking Stop codons. Absence of stop codons is a very uncommon feature in mitochondrial genes but it has been shown that mitochondria of humans, plants and dinoflagellates are capable of translating templates lacking Stop codons (Chrzanowska-Lightowlers et al. 2004; Jackson et al. 2007; Raczynska et al. 2006). How this is achieved remains a matter of speculation since release factors are deemed essential for

disassembly of the ribosomes of the RNA templates. It could be that tmRNA-like RNA species assist in this process. tmRNAs act as tRNA and mRNA at the same time, and are capable of restarting protein synthesis by providing a terminal mRNA section that encodes a functional stop codon (Muto et al. 1998). The stable secondary structures identified at the 3' end of the PCGs could assist in the translation machinery in terminating as opposed to guiding the transcript cleavage (Fenn et al. 2007). However, these structures were not identified on all protein-coding genes lacking a UAG stop codon (*nad4* and *nad6* lack them), and in contrast, could be identified on genes with a UAG stop codon. In this respect, the identified structures seem also not to be a full replacement for a Stop codon. In conclusion, further work on the mitochondrial transcription needs to be conducted in order to gain insight into mitochondrial DNA decoding in *R. similis*. But also research in other Chromadorean nematode mitochondria genetics is required, since it is expected to deviate from most other metazoa due to the unidirectional transcription, the unique initiation codons, unconventional tRNA structures, the large protein content of the ribosomes and remarkably small rRNAs (Gissi et al. 2008; Ohtsuki et al. 2007). For example, the majority of metazoa have one EF(elongation factor)-Tu(thermo-unstable) protein to recognize the standard cloverleaf structure of mt tRNAs. In contrast, Chromadorean nematodes have two mt EF-Tu
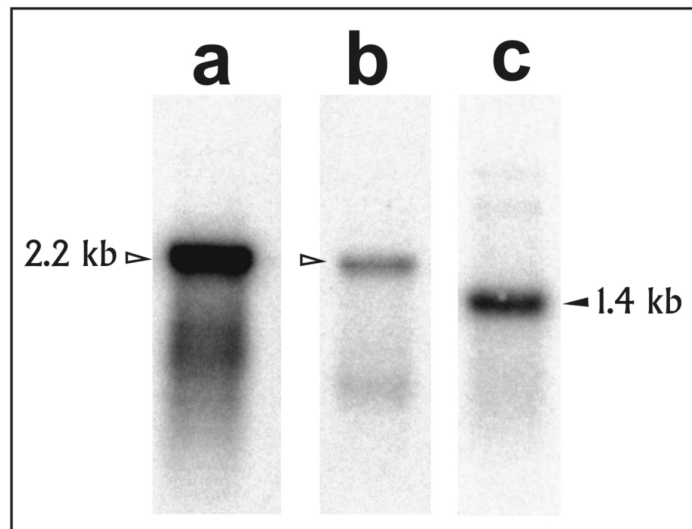

Figure 40: Northern blot on mitochondrial RNA of *Radopholus similis*. **a**, *coxI* antisense probe, with the hollow arrowhead indicating 2.2 kb. Expected length is 1.6 kb. **b**, *nad5* antisense probe, with the arrowhead indicating 2.2 kb. Expected length is 1.5 kb. **c**, *cytB* antisense probe, with the black arrowhead indicating 1.4 kb. Expected length is 1.0 kb. Probing with the sense probes for the three genes failed to produce signals (data not shown).
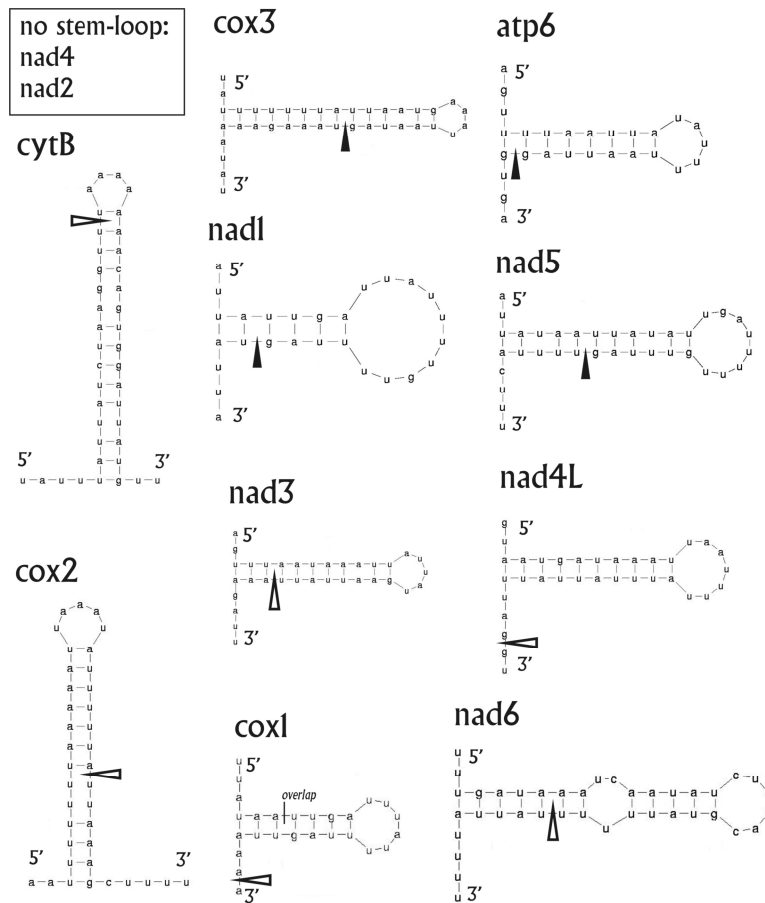
Figure 41: Identified secondary structures at the 3' ends of the protein-coding genes. Black arrowheads indicate the 3' end of the UAG stop. For the other genes, the hollow arrowheads indicate the inferred 3' end of the gene based on protein homology.

proteins, one to bind specifically to TΨC-armless tRNAs by recognition of the DHU-arm, and the other to bind the 2 tRNA^Ser which lack the DHU arm (Sakurai et al. 2006). The EF-Tus have a C-terminal protein extension to deal with the altered tRNAs. In this respect, the truncated 16S rRNA in *R. similis* may indicate complementation by altered proteins in the ribosome, perhaps forced by the codon reassignment.

e. Other AT-rich mitochondrial genomes

Regarding the proposed model, the question rises why this codon reassignment did not happen in other AT-rich mitochondrial genomes, e.g. in the mt genome of the bee *Melipona bicolor* which has an AT content of 86.2%.

A first observation is that insect mt genomes are bidirectionally transcribed, leading to a fundamentally different relationship between the four nucleotides compared to Chromadorean nematodes (Figure 42) – although they generally share the same mt genetic code. The mt genomes of Enoplean nematodes, which are also bidirectionally transcribed, often follow a similar trend as insects, indicating that the differences are partly due to the unidirectional transcription (Figure 40). In insects, C content is strongly and negatively correlated with A and T content. In contrast, in all nematodes, C content has barely any correlation with A and T content. The C content of the mt genome of *Radopholus* makes it unique: it is the lowest (4.4%) of all Ecdysozoa (the moulting animals) mt genomes sequenced to date (Aguinaldo et al. 1997). For Chromadorean nematodes, T is positively correlated (linear regression R^2 = 0.74), and C negatively correlated (R^2 =0.59) with Tyr content. A and G are both barely correlated (R^2 = 0.04 and 0.05 repectivitly) (Figure 43). Clearly, nucleotides at the wobble position determine the amount of Tyr encoded. However, the negative correlation between C nucleotide content and Tyr content is somewhat odd, especially since the amount of C is barely correlated with A and T content (Figure 42). Nevertheless, the codon change in *Rado-*
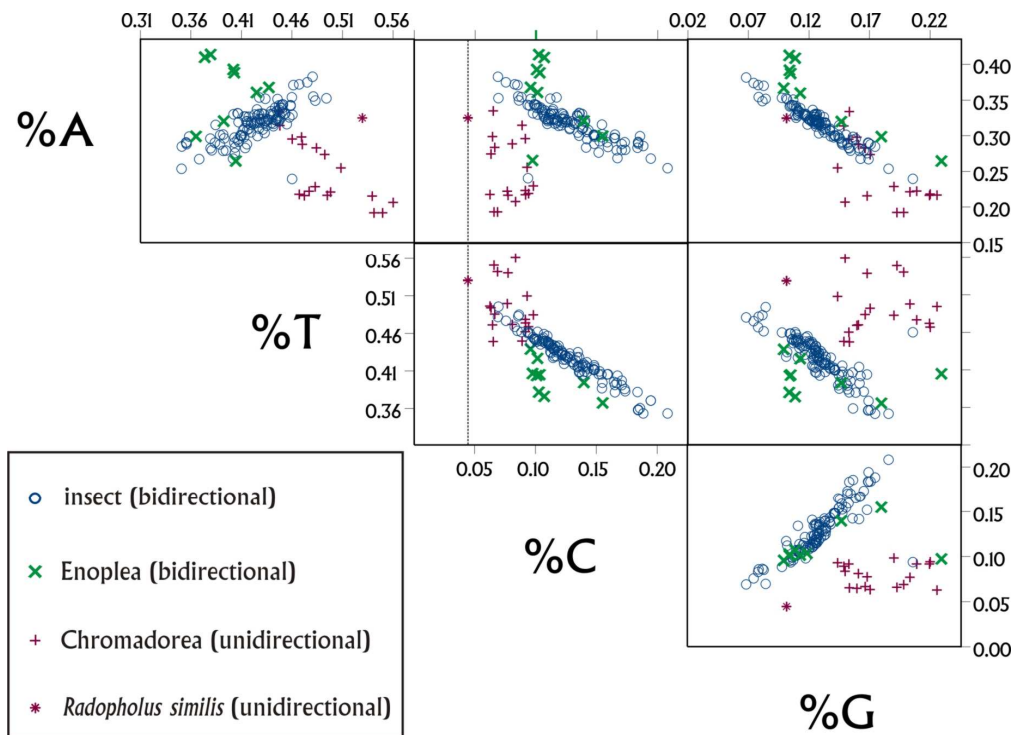

Figure 42: Visualisation of correlations between nucleotide contents in insect and nematode (Enoplean and Chromadorean) mitochondrial genomes, with differing directionality of the transcription. The mt genome of *R. similis* is indicated by a star, and has the lowest C content of all Ecdysozoa (nematode + arthropoda) sequenced to date.
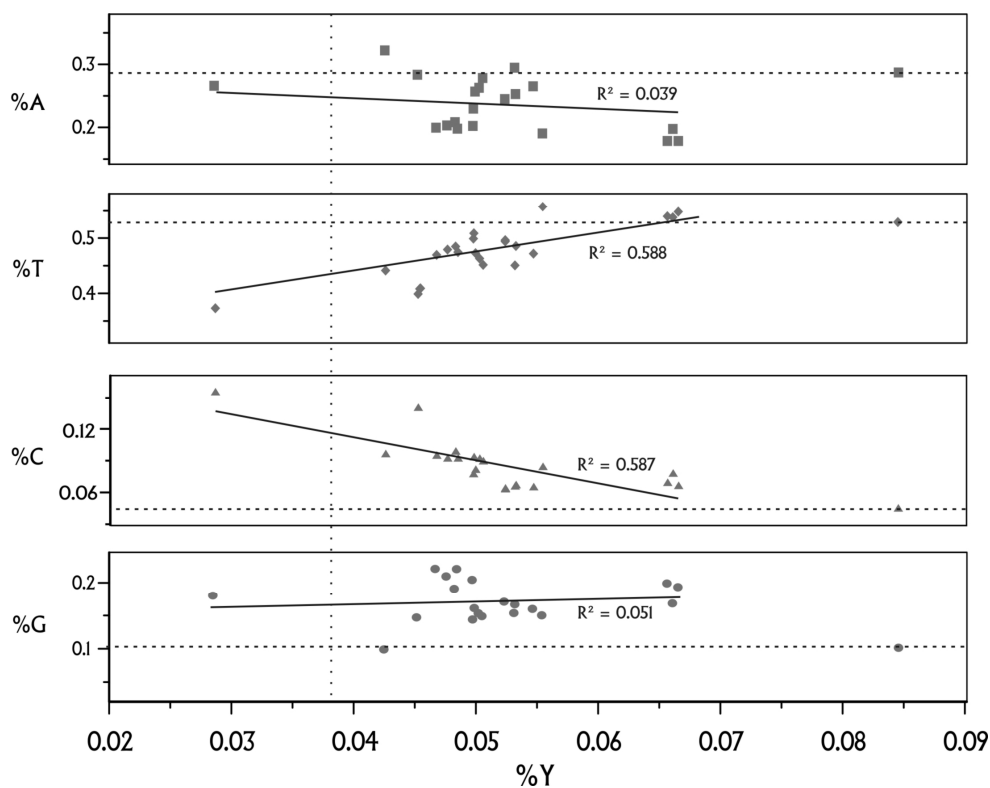
Figure 43: Correlations of nematode mitochondrial genome nucleotide compositions on the total Tyrosine (Tyr) content. The *Radopholus* data point is at utmost right, crossed by the horizontal dashed line. The dotted line indicates the amount of Tyr when only the codons UAU are counted, approximating the ancestral situation before the codon reassignment.

pholus mt genomes caused an increase in Tyr, which leads to a relationship between T and C content more in line as observed in other Chromadorean nematodes (Figure 42).

In summary, the unidirectional transcription in mt genomes of Chromadorean nematodes has to some extent influenced the nucleotide composition, causing in the long term the codon reassignment now observed in *Radopholus similis*. In addition, perhaps the unique mtDNA decoding processes of Chromadorean nematodes are more easily adapted to deal with the effects of such a codon change.

4. Conclusion

The mitochondrial genome of *Radopholus* is clearly unique in its kind, even compared to other nematode mt genomes. The unidirectional translation has shaped the nucleotide content

(driven to AT-richness by a pressure apparently acting on all mt genomes) in such way that the Tyr was forced to expand its codon family. This event caused ambiguous coding for the Stop codon, leading to a disturbance of the mt DNA decoding process, which needed to further differentiate to deal with the new situation. This scenario gives an explanation for the conclusion made by Knight et al. (2001b) based on the analysis of 213 mt genomes, which state that codon disappearance and coding ambiguity act in concert most of the cases. A closer investigation of the codon reassignment in *R. similis* mt DNA shows that this can be explained by the fact that the codon reassignment always acts on a codon which is part of a codon family encoding for one amino acid. A combination of several forces has driven this codon change, the major force being the genome wide GC content (Chen et al. 2004), combined with a weak selection on amino acid content (as proposed by Swire et al. 2005) and a weak selection by the DNA decoding machinery (as proposed by Jia et al. 2008). Clearly, the mt genome of *R. similis* is a very exciting case to study evolution of codon reassignments. Some features in the genome deviate strongly from other metazoan animals and even Chromadorean nematodes, which could be related to the codon reassignment. But many details of the mt DNA decoding processes in *R. similis* remain yet unresolved, leading to a myriad of questions. How does translation termination occur without Stop codons on the RNA templates? Does polyadenylation occur on mRNAs? Why do some genes still have a UAG stop codons, unlike the majority? Is the release factor for this recognition still present? What is the nature of the modification of the tRNA[Tyr]? Answers to these questions can add to our better understanding of nematode mitochondrial genomes and may the in the long run even provide a basis for parasitic nematode-specific control through disruption of mitochondrial function. For example, RNA interference showed that *C. elegans* could not survive when mt EF-Tu1 was suppressed (Sakurai et al. 2006). Since the properties of nematode mt DNA decoding processes differ substantially from those of animal or plant hosts, these processes may indeed be a safe and suitable target.

5. Materials and methods

a. Northern blot

RNA of purified mitochondria was extracted by short sonication of the mitochondria followed extraction with TriReagent (Sigma) (see previous chapter). RNA was loaded in an ethidium bromide containing loading buffer on a 1 x MOPS 1% formaldehyde agarose gel and run at 50V for 4 hours. Downward capillary blotting using a slightly alkaline buffer (5X SSC, 10 mM NaOH) transferred the RNA out of the gel onto an H-bond N+ membrane in about 2 hours (Amersham Bioscience, Uppsala, Sweden). This method of blotting is much faster than the classic upward capillary blotting, and involves placing the gel on a blotting membrane which is put on a pile of tissues. A bridge of whatman paper connects the buffer tank with the top of the gel. When the transfer was finished, the membrane was baked for 20' at 80°C. The membrane was prehybridized at 56°C in a hybridization buffer containing 50% formamide, 5x SSC, 0.5% SDS, 5X Denhardt's solution and a final concentration of 4 µg/ml salmon sperm DNA. RNA probes were generated by cloning the *coxI*, *cytB* and *nad5* genes into pGEM-T. Clones containing inserts in reverse orientation with respect to the T7 promoter were selected and purified. After cutting with restriction enzyme NotI for linearization and subsequent purification of the template, RNA probe was generated with the Riboprobe kit (Promega, Madison, WI, USA) using T7 RNA polymerase in presence of ^{32}P-labeled nucleotides (Perkin-Elmer, Waltham, Massachusetts, USA) for 2 hours at 38°C. After denaturation at 95°C for 5 minutes, probes were cooled on ice and added to 10 ml hybridization buffer, after which hybridization occurred overnight at 56°C. The membranes were washed at 56°C for 5 minutes with 3XSSC/0.1%SDS, and 30' with 3XSSC/0.1%SDS, and 20' with 0.1XSSC/0.1%SDS. The membranes were exposed for four hours to a phosphor imaging plate and the plate scanned using a FLA-5100 imager (Fujifilm, Tokyo, Japan).

b. Bioinformatic and statistical analyses

With the aid of Mitobank2.1 (Abascal et al. 2007) and in-house perl scripts, all nematode (n=28) and insect (n=138) mt nucleotide and protein sequences were extracted from GenBank (March 2009). Complete mitochondrial genomes were downloaded from NCBI. Codon usage data for the nematode genomes was extracted using CUSP program of the emboss package (Rice et al. 2000). Nucleotide content and amino acid content was calculated by in-house perl scripts, and correlations examined through scatter plots. Principal component analysis was done on codon usage data and nucleic acid data and divisive hierarchical clustering was performed on the four first principle components. Statistical calculations were done with S-Plus 8.0 (Insightful, Palo Alto, USA) and Microsoft Excel (Microsoft, Redmond, Washington, USA). See chapter 4 for EST data analysis.

6. Acknowledgements

5

Chapter 6

NEXT, a web-based Nematode EST Exploration Tool

"Whether you think you can or you think you can't, you're right."
Henry Ford

6

NEXT, a web-based Nematode EST Exploration Tool

1. Abstract

Although genomic sequences of few nematodes have recently become available, the majority of nematode molecular data still consists of expressed sequence tag (ESTs) sequences. While genomic data delivers the blueprint of what an organism is capable of, EST data is derived from the transcriptome and provides insight on which part of the genome is actually active. As such, they can contain valuable information when derived from specific developmental stages, or specific tissues, giving a temporal and spatial overview of gene expression within one species. Unfortunately, this valuable information about the ESTs lies buried deep in the GenBank files, so that it cannot be easily accessed. To unlock this wealth of information we present in this chapter an on-line tool, called "nematode EST exploration tool" (NEXT, accessible at http://zion.ugent.be/joachim/next). This tool allows users to search homologues of their gene in all publicly available nematode EST libraries, which are grouped according to life stage or tissue. The library information in NEXT database is manually curated and updated on a regularly basis. Furthermore, we present a novel approach to compare EST libraries, which is based on normalization using newly identified and frequently used internal reference genes. This feature is implemented in NEXT, and explained through some case studies. In addition, information of the libraries, such as length and clustering statistics can be accessed, allowing assessment of the library quality. In summary, NEXT enables scientists to obtain a spatial and temporal gene expression profile of homologues in different nematode species based on available EST data.

6

2. Introduction

'Expressed sequence tags' (ESTs) are sequences derived from mRNA. Being complementary to genomic sequences, they provide insight into which part of the genome is transcriptionally active in certain selected tissues. For the phylum of nematodes – one of the major phyla on earth – ESTs are the largest source of molecular data. Of the more than 60 million sequences currently available in dbEST, 1,080,521 are derived from nematodes (Boguski et al. 1993). The average nematode EST library, which is a collection of ESTs derived from a single experiment, contains approximately 4,000 sequences (Figure 44). ESTs serve multiple purposes. They play a key role in identifying coding regions in genomes. However, this practice is currently of value for a limited number of nematode species, as genome sequences are currently only publicly available for *Brugia malayi* (Williams et al. 2000), *Meloidogyne incognita* (Abad et al. 2008), *Meloidogyne hapla* (Opperman et al. 2008), *Caenorhabditis elegans* (C. elegans Genome Sequencing Consortium 1998) and *Caenorhabditis briggsae* (Stein et al. 2003). Nonetheless, other useful information can be extracted from EST data (Kim et al. 2008; Munoz et al. 2004). For many species, EST data provide a first glimpse of the transcriptome, which can subsequently be exploited for identification of genes and for evolutionary analyses (Jacob et al. 2008; Mitreva et al. 2005). Furthermore, a collection of EST libraries of carefully selected tissues and developmental stages can provide a spatial and temporal gene expression profile of a species. In addition, since ESTs are generated through random sequencing of cDNA clones, sequence redundancy in EST libraries can give an estimation of the level of gene expression. This is because the sampling frequency of a tag is assumed to be proportional to the abundance of the corresponding transcript in a given cellular mRNA pool (Audic et al. 1997).

The construction of an EST library is a very tedious job. Although the term EST is used for any sequence obtained from mRNA in a high-throughput manner, the characteristics of EST libraries can vary considerably, due to the used protocol and inclusion of controlled and uncontrolled systemic biases. Depending on the purpose of the library, several applied techniques add to a biased mRNA representation in EST libraries, such as amplification steps, primers used for cDNA synthesis, the system used for cDNA synthesis, cDNA size fractionation, techniques such as subtraction and normalization, and the choice of cloning

6

vectors. In addition to these 'intentionally' introduced variations, further deviations arise from differing quality and quantity of starting material, reverse transcriptase reaction efficiency, ligation efficiency, difference in cloning 'capacity' of sequences (e.g. toxins) and other unforeseen events along the library construction process. Therefore, it is clear that comparison of different EST libraries – if they are not constructed using identical protocols – should be done with necessary precaution and appropriate controls. It must be noted that with the upcoming 'next generation' sequencing platforms, some of these biases will be dealt with leading to more consistent EST data sets, since a number of steps in the library construction process can be excluded (Simon et al. 2009).
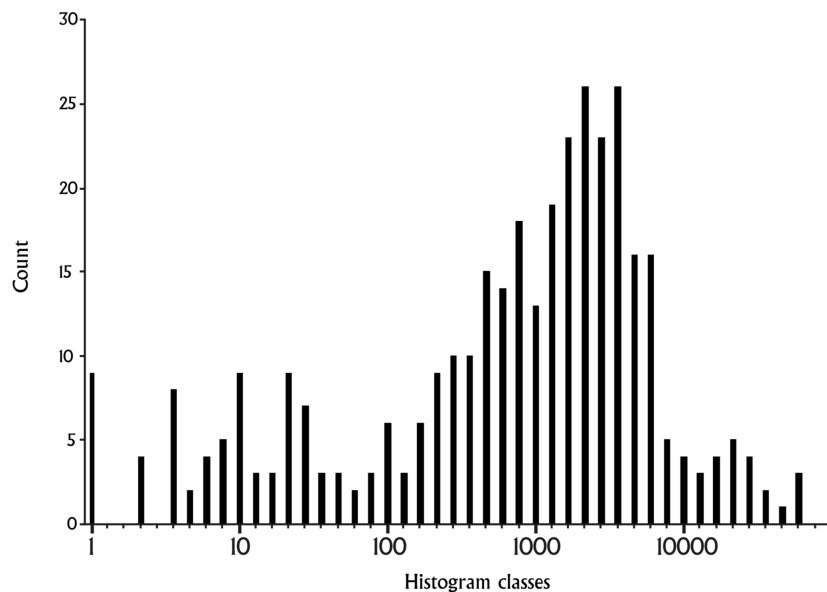

Figure 44: Logarithmic histogram of sizes of current available nematode libraries. The Y-axis shows the EST library counts. The X-axis shows classes with bin range x – 1.259 x (on a logarithmic scale)

Interestingly, many systemic biases in cDNA libraries can be detected through visualisation of some statistics. For example, enrichment of sequences with a certain length (visible in length distribution graphs) can point to inclusion of normalization steps (Liu et al. 2006). Plotting histograms of GC content can differentiate between sequences of different sources (Jacob et al. 2008). And the observation of aberrantly large clusters (a cluster is a collection of identical ESTs representing one gene) in a cluster size distribution may point to overrepresentation of certain genes, mostly due to artefacts during library construction (Kikuchi et al. 2007).

Despite the usefulness of EST data analysis, and their value in parasitic nematode research (Dubreuil et al. 2007), essential information of EST libraries is still not easy to access and remains therefore largely unexplored. A major obstacle is the complete lack of standard nomenclature for specifying library characteristics (such as tissue and stage information) and library construction details, making it difficult to extract this valuable information. The two major online nematode ESTs repositories, which also offer EST analysis tools are www.nematode.net (Wylie et al. 2004) and Nembase3 on www.nematodes.org (Parkinson et al. 2004). These two websites provide both a tool for searching nematode ESTs using BLAST, but the results are difficult to interpret in terms of library specifications, and it is not possible to retrieve the resulting sequences in batch. In addition, Nembase3 provides a tool to compare EST expression levels in different life stages, of different species. This tool is based on precalculated clusters of different nematode libraries, and reports clusters which match a user-defined set of criteria for raw EST counts. This tool requires knowledge of the size and characteristics of the libraries, both which are cumbersome to find. Currently no tool exists for BLASTing a query sequence to different nematode EST libraries, with visualization of the BLAST results to allow straightforward interpretation of the results. To comply with a potential need, we developed a web-based tool called 'Nematode EST exploration tool' (NEXT) to easily explore the wealth of information contained in nematode EST data. This tool complements the other tools available for nematode EST analysis. Furthermore, several issues when identifying homologous sequences of a diverse range of species are considered. A method based on internal reference genes to deal with these issues is implemented. Finally, we demonstrate the usefulness of NEXT through some case studies.

3. The NEXT web tool

a. The NEXT interface

The NEXT website aids the researcher in exploring the occurrence of his gene in all nematode libraries to date. Upon receiving a query, NEXT decides to use tBLASTn or tBLASTx, in the case of a nucleotide or a protein sequence query, respectively. The query is blasted against all nematode ESTs and information of the EST hits is fetched from the NEXT database. Through a web interface (Figure 45), various options can be set to deal with normalized libraries (in- or

excluded) and aberrant clusters, which often are observed in EST libraries and are likely to represent constructional artefacts (Jacob et al. 2008; Mitreva et al. 2004). The latter's default setting is 2%, i.e. clusters containing more than 2% of the total sequences of an EST library are subtracted of the total. For graphical representation of the results, libraries are pooled together according to user-defined settings, based on two criteria. The 'grouping' criterion is based on nematode species. Up to three groups of species can be compared. Predefined groups are available as plant-parasitic (PPN), animal-parasitic (APN) and free-living (FLN) nematodes. The 'contrast' criterion divides the results of the grouping criterion further into developmental stage or selected tissue. At the bottom of the interface page the option to access directly library statistics information is provided.

Figure 45: The NEXT interface (zion.ugent.be/joachim/next)

b. Processing and output

After receiving the request, a comprehensible extensive report is displayed. A link is provided to download all EST hits, with essential library information in the header of the sequence. The main output consists of several graphs. The first graph are based on simple BLAST hit counts, reported as 'ESTs per million' (EPM), which is the total number of hits divided by the total

number of sequences of the pooled libraries, multiplied by 1,000. The first graph groups the results following the grouping term. The second graph shows the distribution of the E-value of the hits according to the grouping term. The third graph reports the further subdivision of the results following the contrast term. The second part of graphs shows estimated relative expression, normalized to internal reference genes (see further). Clicking on the graphs provide the data on which construction of the graphs is based. This data can then be copied to spreadsheet software to allow maximum flexibility. The last part of the output reports form a detailed output of all EST hits in tabular, grouped per library. Library statistics can be visualized by clicking on the library name.

4. Development of a novel approach to EST library normalization

Comparing EST libraries for differential expression of homologues produces results of two different kinds. The first type reveals the presence or absence of the homologue in the different libraries. This can be due to the fact that too little EST sequences are available, or that the homologue is absent in the genome. For example, cellulase enzymes are only identified in ESTs of plant-parasitic nematodes, in contrast to other nematode species. Genome analysis has revealed that these genes are indeed lacking in non-PPN genomes. A second type of result emerges if homologous genes are present in all EST datasets, but at different levels. In this case, some degree of quantification of gene expression is needed to draw conclusions.

In contrast to the two most popular 'analog' methods for gene expression quantification (micro-arrays and quantitative PCR (Q-PCR)) – which are specifically designed for this purpose - , 'digital' measurement of gene expression, based on tag count, is not yet widely used. However, little by little it receives more attention due to the ever growing collection of transcriptomic data. The digital method for gene expression quantification can be divided in two approaches. A first approach is applied by 'serial analysis of gene expression' (SAGE) (Velculescu et al. 1995) and 'massively parallel signature sequencing' (MPSS) (Brenner et al. 2000), and comprises generation of small (10-22 nucleotide) unique tags of expressed genes, which are subsequently mapped to genomic sequences for identification of the tag. The tag count correlates with the expression level of the corresponding gene. A second approach is

based on EST libraries (Stekel et al. 2000). This approach confers advantages over the other methods. The most important is that prior genomic sequence information is not needed, and is applicable to any organism for which enough starting material can be obtained. In addition, sufficient sequence information of potentially every gene is generated, which can directly be used for design of further experiments. However, a critical requirement in the case of gene expression quantification is that the analyzed EST library is an as accurate as possible reflection of the initial mRNA pool of which it was constructed. Unfortunately, due to the laborious protocol used to construct EST libraries, the composition often deviates strongly from the original mRNA pool, an effect that is regularly selected for, at the cost of losing quantitative information. As for all digital methods, the accurate assignment of each tag to just one gene is required. Digital gene expression quantification is intrinsically noisy (a property shared with microarrays), but — in contrast to microarrays — this noise can be reduced simply by increasing sampling size. This practice leads simultaneously to increased detection of low abundance transcripts.

Some statistical methods for digital gene expression quantification have been developed, based on an ideal cDNA library. These methods start from a Poisson distribution of EST counts and have been developed and used for comparing gene expression in two or more cDNA libraries (Audic et al. 1997; Herbert et al. 2008; Murray et al. 2007; Stekel et al. 2000). A p-value is assigned to differences between counts for assessing its significance. Recently, the relative abundance of all transcripts in a transcriptome — and by extrapolation, in an ideal cDNA library — was shown to follow a generalized inverse Gaussian distribution (Zhu et al. 2008). Based on this complex distribution, it is estimated that 5 million tags need to be generated to cover 90% of the human transcriptome. Although the distribution was determined from SAGE data, we found that combining clustered EST data of nematode EST libraries leads to the same distribution (Figure 46).

Currently, the statistical methods outlined above are implemented by several tools to estimate gene expression based on EST counts, such as TissueInfo, BodyMap, SOURCE, exQuest and TEPSS (Aguilar et al. 2008; Brown et al. 2004; Diehn et al. 2003; Hishiki et al. 2000; Skrabanek et al. 2001). These tools are only available for human and mouse, and make use of the genomic information of these species to specifically assign each EST sequence to one gene. They allow comparing expression levels of different genes in different tissues and developmental stages of one species.

NEXT has a profoundly different starting point: it allows the analysis of one query gene by identification of homologues in EST libraries of 66 different nematode species. Comparing whether homologous genes are differentially present in EST libraries of different species is not a trivial task. As a first remark, a BLAST search typically not only identifies orthologous sequences. Second, the construction characteristics of EST libraries tend to differ, leading to a possible biased gene representation, and hence a different representation of tag count for orthologous sequences. With this in mind, the EPM count based on a BLAST search depends on the query gene, the E-value cut-off used, and the EST libraries searched.
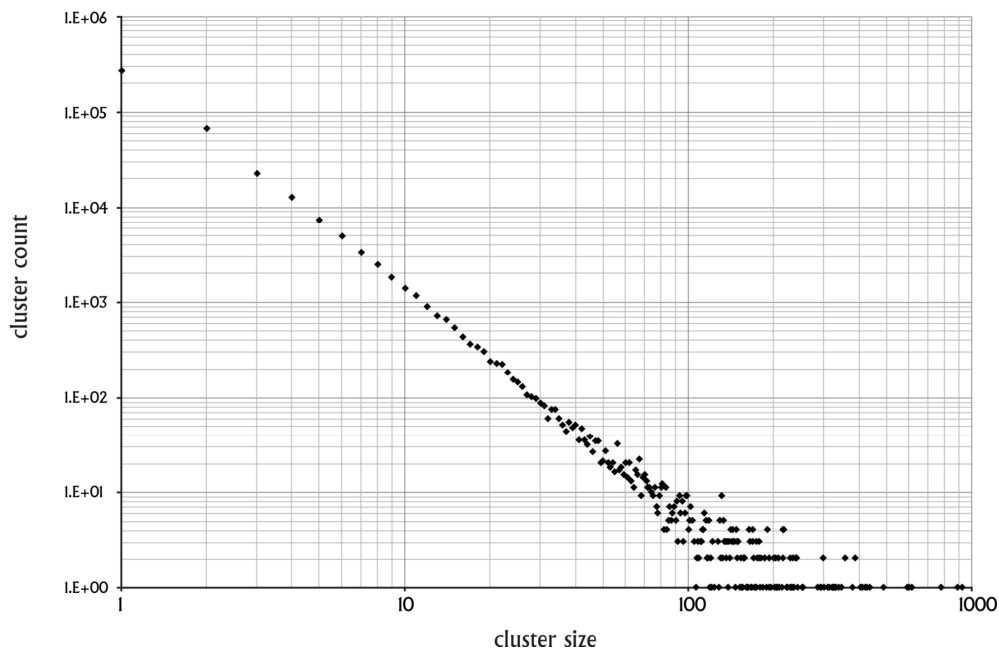
Figure 46: The cluster size distribution of all nematode ESTs, clustered per library. The distribution follows a generalized inverse Gaussian distribution as described in Zhu et al. (2008).

As a novel approach to allow comparison between EST libraries of different species, we searched whether the application of internal reference genes can be used. Rather than comparing gene expression by library density (i.e. expression as EPM), the expression could be compared to an established index of reference genes. The purpose of this index is to capture a large part of the variability in EST libraries. The members of the index need to fulfill some requirements. They need to be present in the majority of the libraries, and their relative

ranking may not fluctuate to a large extent. The latter requirement is inspired by the use of internal reference genes in Q-PCR experiments.
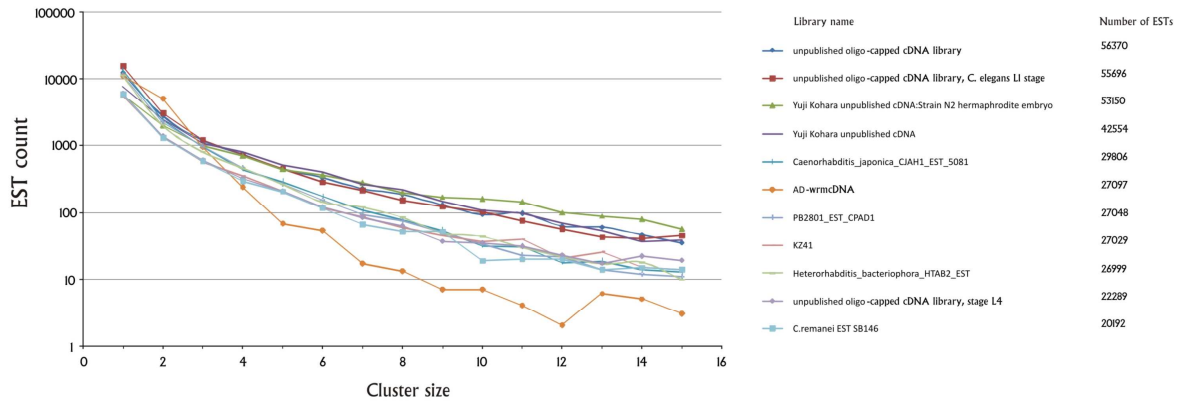

Figure 47: Clustering statistics for the nematode EST libraries with more than 20,000 sequences. Shown are clusters containing up to 15 ESTs. Note the remarkable different clustering of the different libraries.

To obtain such a reference index, genes were identified of which ESTs appeared in almost all libraries (see Materials and Methods). The corresponding genes of *Caenorhabditis elegans*, rather than the EST sequences, were further used for analysis. The expression stability of each gene, as the standard deviation of the EPM count at a BLAST E-value cut-off of 1e-20, was analyzed in the libraries containing over 10,000 sequences. Internal control genes of *C. elegans* identified by Hoogewijs et al. using Q-PCR (2008) were added to the analysis and their expression stability determined in the same way. Nine best scoring genes were retrieved, exhibiting a stable expression level ratio over a large range of EST libraries (Table 14, Figure 48). Analysis of the EPM counts revealed relatively stable ranking in the four biggest nematode EST libraries (all of *C. elegans*) (Figure 50). Pooling the results from EST libraries of different species also resulted in similar ranking of the internal reference genes. However, *tax-6*, *cpz-2* and *pmp-3* were excluded from further analyses, because the E-value distribution of the hits differed over the three major nematode groups FLN, PPN and APN, indicative of too divergent sequences (Figure 49). Analysis of the remaining *C. elegans* reference genes in the four biggest nematode EST libraries revealed that the maximum BLAST E-value of the hits was largely determined by the mean sequence length of the library (Figure 51).

6

Table 14: Used reference genes of *Caenorhabditis elegans*.

Gene	Accession	Length (nt)	Description	Reference*
act-1	NP_505819.1	1128	Actin family member	H
tba-1	NP_492600.1	1347	Tubulin alpha family member	H
T05D4.1	NP_741281	1095	Hypothetical protein	This study
cpz-2	NP_506318	1401	Cathepsin Z family member	This study
tax-6	NP_001076658	1791	abnormal chemotaxis family member	This study
csq-1	NP_510438.1	1251	Calsequestrin family member	H
mbk-2	NP_001040951	1503	Minibrain kinase (*Drosophila*) homolog family member	This study
pmp-3	NP_506620.1	1980	Peroxisomal membrane protein related family member	H
eif-3.C	NP_492638.1	2694	Eukaryotic initiation factor family member	H

* H: Hoogewijs et al., 2008
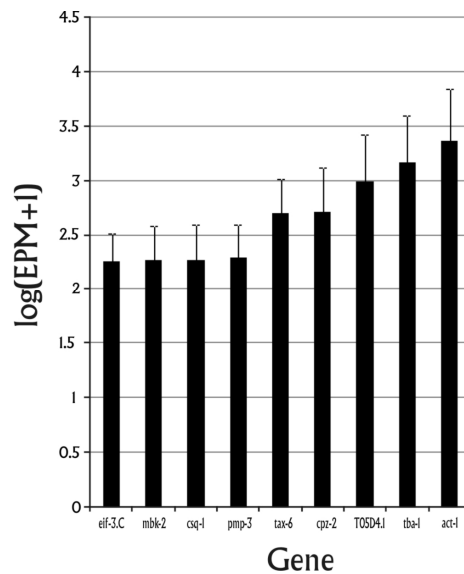
Figure 48: Average EST abundance of nine internal reference genes in all nematode EST libraries containing more than 3,000 sequences. For each library, counts, as 'ESTs per million' (EPM) of the library size, are log10(x+1) transformed.
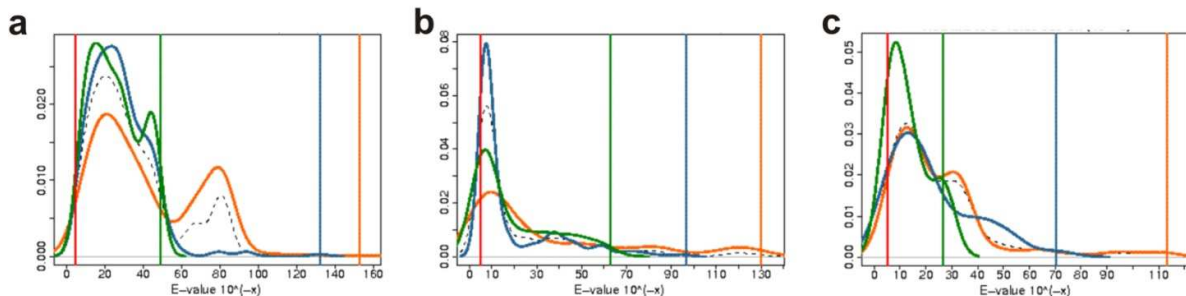

Figure 49: Distribution of the E-value of the hits of three internal reference genes visualised by a density line. **a**. *tax-6*, **b**. *cpz-2* **c**. *pmp-3*. The green vertical line represents the maximum E-value observed for plant-parasitic species, the orange for free-living, and the blue for animal parasitic species.
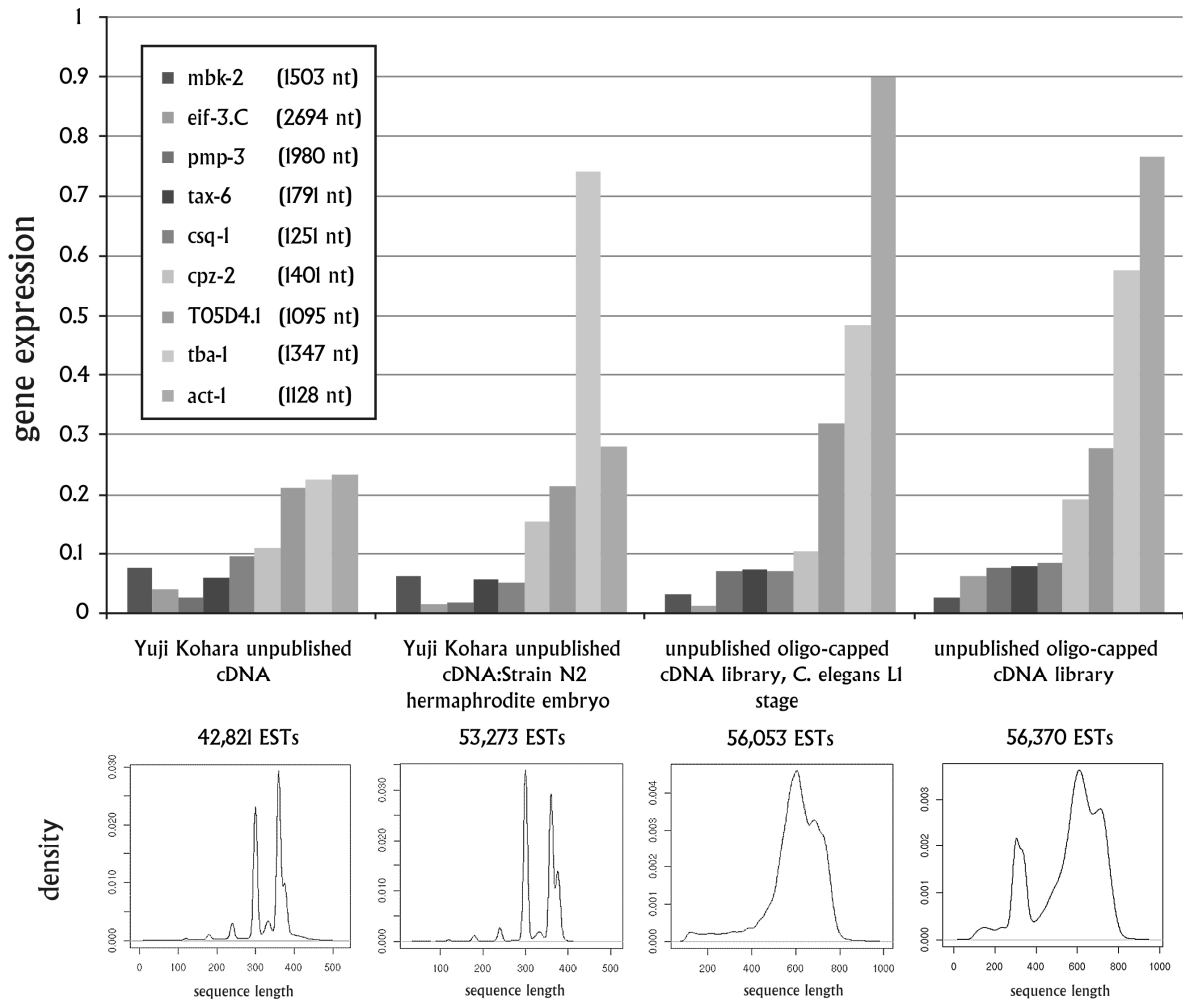
Figure 50: Analysis of BLAST hit counts of the nine selected reference genes in the four biggest nematode EST libraries. The E-value cut-off was set to 1E-20. In Y-axis, EPM counts are divided by gene length, since gene length has been found to be inversely correlated with tag count in EST datasets (Munoz et al. 2004), and the result is $\log_{10}(x+1)$ transformed. The graphs at the bottom show the EST length distribution of the libraries (taken from NEXT), and show that despite the large differences in sequence length distribution, the ranking of the reference genes is not disturbed to the same extent.

Therefore, comparison of BLAST hit counts of a query (reference) gene between different EST libraries requires usage of library-specific E-values. Determining these E-values to achieve comparable hit counts between libraries may prove being a non-trivial task. In addition, when comparing different species, additional variability on the E-value will arise due to sequence differences. One of the solutions to minimize the library-dependent part of the E-value variability can be reprocessing of EST sequences, so that each library has a similar sequence length distribution. Another solution, which also deals with sequence differences, is clustering of each EST library. In the latter approach, upon BLASTing a (reference) gene, the size of the

most homologous EST cluster (i.e. the number of EST sequences contained in the cluster) can serve as a reference point to compare different libraries. Investigating the remaining 6 reference genes applying this method shows that ranking of most reference genes is preserved in most of libraries. However, *tba-1* was found to change its position in the ranking more than the other reference genes. A closer look on this gene revealed a marked increase in expression level in embryo-specific libraries, when the counts were normalized to the other reference genes (Figure 51, but also Figure 50). This increase in tubulin expression has been observed in
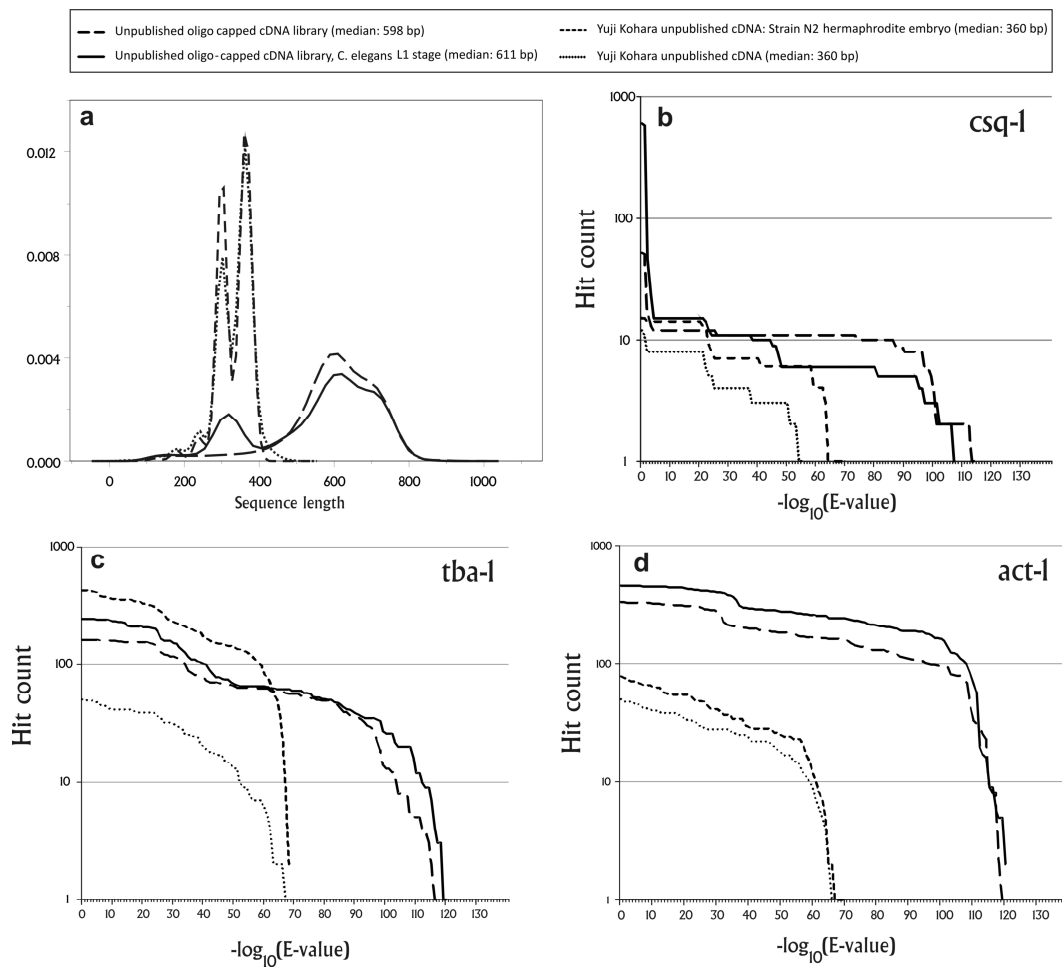

Figure 51: Number of BLAST hit count of 3 *C. elegans* genes in 4 *C. elegans* libraries against the E-value cut-off (X-axis). **a,** the sequence length distribution of the four analyzed libraries. **b,** *csq-1* has relatively low counts in all libraries. **c,** *tba-1*. Remark the much higher counts in the embryo library (short dashes). **d,** *act-1* has the highest counts of the internal reference genes. As expected, the number of hits decreases with increasing homology level. Each library has its maximum E-value, largely depending on the mean sequence length.

embryos of *C. elegans* (Wright et al. 2003). Because of this behaviour, we decided to remove *tba-1* of the reference list, and use the remaining 5 reference genes to include in NEXT (Figure 52). To normalize gene expression of the homologues occurring in EST libraries, NEXT determines the most homologous cluster in each library for a query gene, taking into account the user-defined E-value cut-off. The size of this cluster is determined, and this value is normalized against the cluster size for the reference genes in that library. The library results are pooled together following the user-defined settings, and the output graphs are constructed.

The normalization results are presented in the second part of the output of NEXT. Clicking on the graphs shows detailed information on the occurrence of the reference genes in the libraries. Following case studies will demonstrate the use of NEXT.
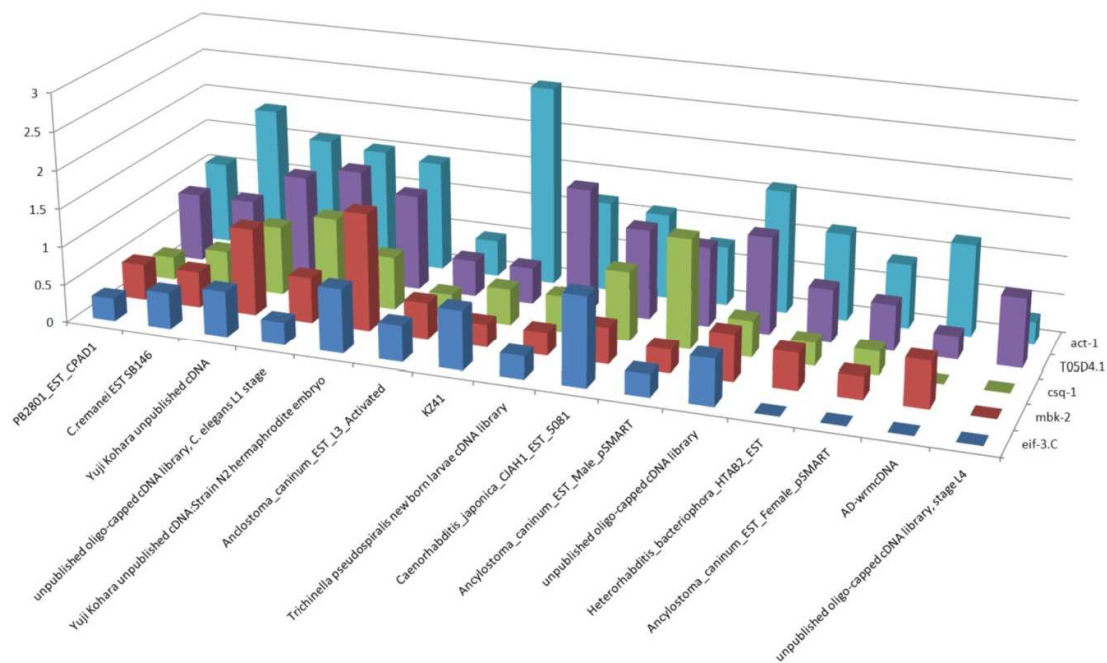

Figure 52: The abundance of 5 reference genes in nematode EST libraries containing over 20,000 sequences. The number of ESTs contained in the cluster most homologous to the reference gene is counted, and this count is $\log_{10}(x+1)$ transformed. Such a reference gene index partly captures the deviating gene representation characteristic for each EST library, and can serve as a starting point to compare heterogeneous libraries.

5. Case studies

a. tba-1

As a first case study, the alpha-tubulin 1 gene (*tba-1*) of *C. elegans* was used as query, and its occurrence investigated in different developmental stages. With an E-value cut-off of 1E-05, 3,202 EST hits were identified in 155 libraries (which is 42.8% of the total number of libraries). The EST hit counts normalized to library density (reported as EPM) reveals two stages with a higher abundance of *tba-1*: the egg/embryo stages for FLN and APN, and the fourth stage for FLN. Remarkably, for PPN no upregulation for *tba-1* homologues is observed in any of the stages. The normalized expression graph shows a similar pattern, but with better differentiation between the stages with higher expression. Having a closer look to the E-value distribution reveals several peaks coinciding with different *tba-1* homologues. By setting the E-value cut-off to 1E-70, the peak in the egg/embryo stage of FLN disappears, while only the fourth stage shows upregulation for FLNs. In addition, when querying the best homologue of *tba-1* in the egg/embryo stage of FLN (gi:5582329), a sole strong peak is seen in the egg/embryo stage. It is important to realize that the normalized expression graph reports the result of one cluster, which represents in the ideal case all ESTs of one gene in a library. This count is therefore less influenced by the applied E-value cut-off, in contrast to the EPM graph. Restricting the analysis to *C. elegans*, two different homologous tubulin genes are observed, one with a strong expression in fourth stage, and one which is upregulated in egg/embryo stage. BLASTing against *C. elegans* protein sequences reveals that the latter ESTs are derived from *tba-1*, while the former are derived from *tba-2*. These results are in agreement with Wormbase (Bieri et al. 2007). Thus by carefully adjusting E-values, it is possible to discriminate closely related homologues. Note, however, that a query gene identifies all close homologues, and that ESTs with best E-values correspond not necessarily to orthologous sequences.
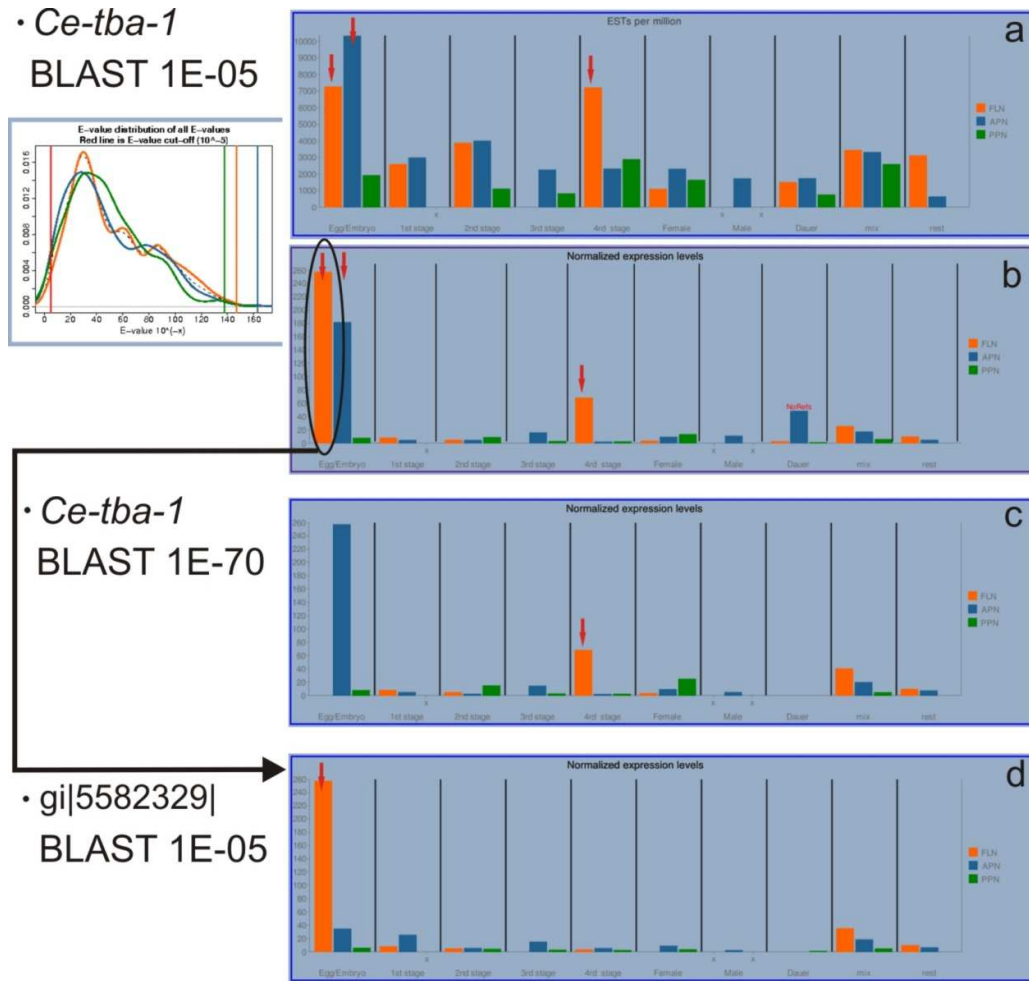
Figure 53: Analysis of *tba-1* using NEXT. **a** and **b**, results of the BLAST with E-value 1E-05. The homologues covered in this search are upregulated in egg/embryo and the fourth stage (red arrows). Note in **b**, that for the APN dauer stage none of the reference genes are present, but homologous ESTs above the E-value cut-off are withheld. The graph differentiates more between the different stages than the EPM graph. **c**, setting the E-value threshold more stringent to 1E-70, causes the upregulation in egg/embryo stage for FLN to disappear completely. In contrast, the peak in the fourth stage for FLN persists (red arrow). The ESTs corresponding to this peak are derived from *tba-2*. We note also that homologues in APN are upregulated in egg/embryo stage. **d**, querying the most homologous EST of the egg/embryo stage of FLN identified in the first search (the sequence of which can be retrieved by clicking on the gi number in the NEXT output), reveals a strong upregulation in the egg/embryo. The ESTs corresponding to this peak are derived from *tba-1*.

b. Tissue inhibitor of metalloproteinase homologue

Identification of parasitism genes in parasitic nematodes has hugely profited of EST analyses. Often, candidates are identified by screening for genes encoding secreted proteins (Elling et al. 2009; Vanholme et al. 2006). One of the candidates we have recently identified by screening

ESTs of the plant-parasite *Radopholus similis* was a homologue to a 'tissue inhibitor of metalloproteinase' (TIMP). The complete open reading frame was retrieved from the ESTs and encodes a protein of 153 amino acids. Through NEXT, a better view on the distribution of this homologue in all nematodes was obtained. Using an E-value cut-off of 1E-05, we retrieved 103 homologous ESTs, the majority of which were derived from plant-parasitic species. Since the E-value of the most homologous EST of outside the group of plant-parasitic nematodes was 1E-19, we repeated the search with an E-value cut-off of 1E-20. The E-value graph of the resulting PPN-specific ESTs reveals two peaks, with upon closer investigation, the first peak corresponding to nearly all egg/embryo ESTs, and the second with the ESTs of other stages.
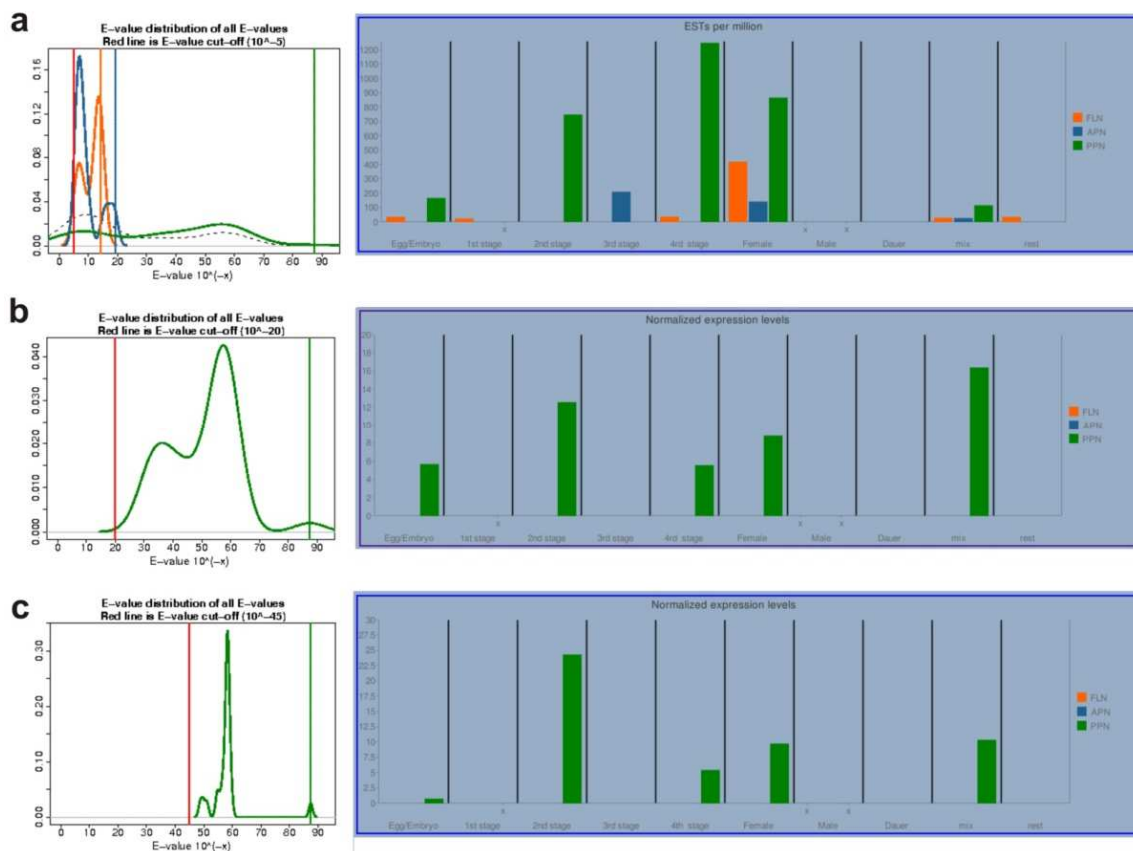

Figure 54: Analysis of a 'tissue inhibitor of metalloproteinase' homologue of the plant-parasite *Radopholus similis* using NEXT. Using gradually more stringent BLAST criteria (a − c), the expression is predicted to be at its maximum during the second stage in plant-parasitic nematodes.

We repeated the search with an E-value cut-off of 1E-45. This search shows that the highest expression is observed in the second stage of plant-parasitic nematodes, with a lower expression in the fourth and female stage. These stages are all parasitic stages in PPNs. The homologues originate of different kinds of plant-parasitic nematodes, strongly suggesting an important role of TIMPs in the parasitism process of PPNs. The corresponding sequences can be easily downloaded from NEXT for further investigation. Interestingly, a BLAST search with the query sequence against the nr database reveals homology to a 'hypothetical esophageal gland cell secretory protein 12' of *Heterodera glycines*, the soybean cyst nematode (accession AAF76926) (Wang et al. 2001). However, using 'tissue' as contrast did not reveal any homologous ESTs originating from gland-specific libraries.

6. Discussion

The NEXT website provides a useful tool for accessing relevant information of nematode EST libraries. It allows searching homologues in nematode EST libraries. By carefully analyzing the occurrence over different life stages and tissues in different species, clues about their function may emerge. Furthermore, statistics of all libraries are presented in graphs, allowing the researcher to investigate several features, such as sequence length distribution and clustering statistics.

Digital gene expression quantification bases its estimation on analysis of redundancy observed in cDNA libraries. Although identification of tags is an issue in any tag-based gene expression method (Zhu et al. 2008), extra care has to be taken when using the results from the NEXT database for gene quantification. The sequence query must preferentially be full-length, since most ESTs are not full-length. The search uses BLAST with a user-defined homology threshold, in which the difference between gene family members and orthologous sequences is sometimes fuzzy. Further, attempts to compare pure EST counts between libraries should include corrections for observed and expected biases. Independent of the technique used for gene expression measurement (microarray, Q-PCR, digital northern), it is often recognized that in contrast to absolute quantification, comparing gene expression ratios over different samples produces more accurate results. This approach might even be suited to compare different gene expression quantification methods (Chen et al. 2007; Vandesompele et al. 2002). We identified

6

internal reference genes, whose ranking based on gene expression as estimated from EST counts, remains the same in most nematode EST libraries. Rather than genuine gene expression differences, the variations in such a set arise from a combination of random tag count fluctuations and biases in EST libraries. This set may assist in comparing different heterogeneous EST libraries. The internal reference gene approach presented in this study can be transferred to other systems requiring comparison of EST libraries. Note that the ideal internal reference gene index does not exist, and best reference genes differ for each combination of EST libraries to be compared. In contrast to Q-PCR and micro-arrays, reference genes in EST libraries come 'for free', and as many reference genes as possible may be included in a reference index. Future analysis to optimize such indices is needed, to asses which – and how many – reference genes they should include. Combined with a better understanding of the influence of EST library characteristics on BLAST searches, this may further improve comparing EST libraries. However, accurate differential gene expression based on EST data currently requires still sufficient large non-normalized libraries, created with the same protocol. When these conditions are not met, the results from the NEXT website can help to provide indications for differential gene expression, which can be further validated through additional gene quantification experiments.

7.   Materials and methods

a.   The NEXT database

All nematode ESTs were downloaded from NCBI with the aid of Entrez-utilities, and stored in a MySQL database. Subsequent processing involved cleaning by Seqclean and clustering by TGICL without assembling cluster sequences (Pertea et al. 2003). The clustering results for each EST were stored as additional information. Repeated rounds of clustering were performed if initial clustering did not give satisfying results (Pertea et al. 2003). Library information was parsed from the GenBank files of the ESTs, and stored into the database. Manual correction of library information is indispensable, since a uniform nomenclature of library specifications is completely lacking. Finally, library specificities as length, cluster and GC percentage distribution were calculated and presented with the aid of R (http://www.r-project.org/). A local blastable database of the ESTs is constructed (Altschul et al. 1990). The database is semi-automatically updated once a month.
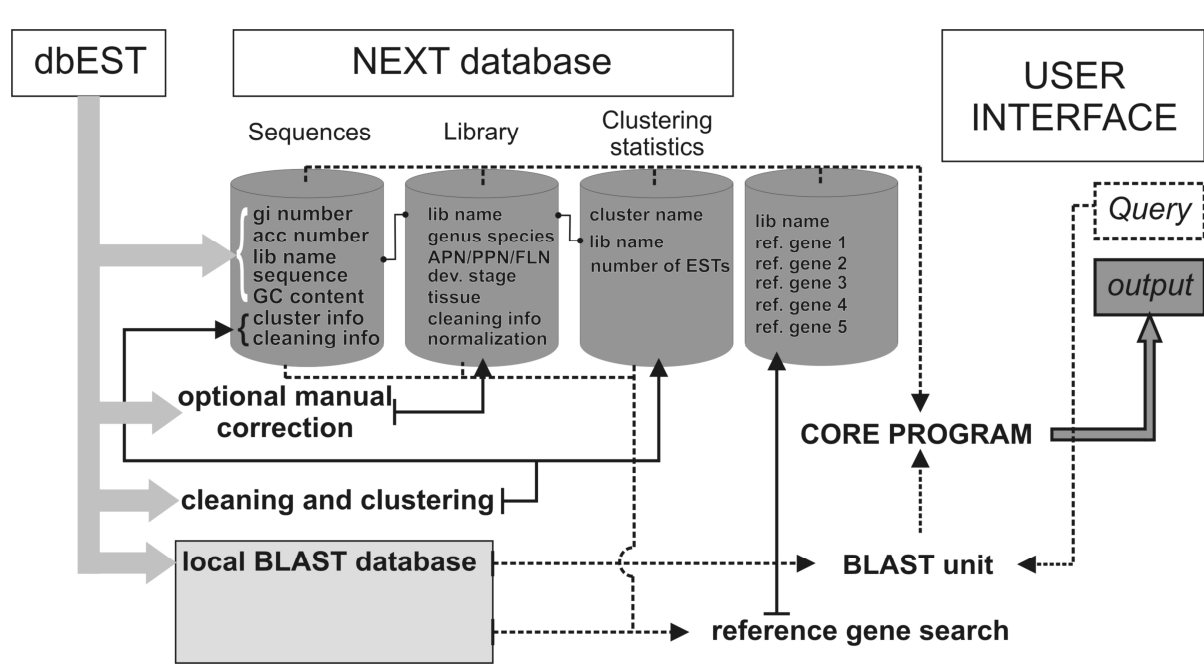
Figure 55: Scheme of the NEXT database. The four tables of the MySQL database are updated once a month by fetching sequences from dbEST and manual correction of the automatically parsed library information. The ESTs are clustered and cleaned, and the results stored. A local blastable database of the EST sequences is constructed. Finally, homologues of the 5 reference genes are identified in each library and their cluster sizes stored. Upon receiving a query, the 'BLAST unit' blasts the query against the blastable database, and the resulting hits are passed to the 'core program', which collects the information of the hits from the NEXT database following user-defined specifications and passes them to the output file.

b. Candidate reference gene search

With the aid of the NEXT database, we attempted to search for candidate reference genes in the EST libraries by identifying widely occurring clusters. For this we iteratively blasted (with E-value 1e-05 and effective database length fixed at 15,761,056) the clusters of the biggest nematode library ("unpublished oligo-capped cDNA library" (*C. elegans*), n=56,370) to all libraries in order of decreasing size (Figure 56), and proceeded with the next library solely with homologous clusters. The resulting graph of this procedure shows a steep decrease from libraries of size 10,000 on, pointing to strongly biased gene representation in libraries with smaller sizes. We analyzed the corresponding (*C. elegans*) protein sequences of 67 clusters present in all nematode libraries with sizes larger than 10,000, by searching and counting homologues in the four largest nematode libraries (all *C. elegans*). The BLAST search was performed with 5 different E-values: 1E-01, 1E-05, 1E-20, 1E-45, 1E-75. Taking known biases in EST data into consideration (such as aberrant clusters and gene length dependence (Munoz et al. 2004)), the resulting counts were transformed by

$$\log_{10}\left(\frac{\text{EST count}*10^6}{(\text{library size}-\text{aberrant clusters})*\text{gene length}} + 1\right)$$

with clusters containing more than 2% of the ESTs marked as aberrant. After ranking the genes by their standard deviation, four genes that showed similar expression levels in the four libraries were selected. The same procedure was applied for 10 reference genes identified by Q-PCR in *C. elegans* by Hoogewijs et al. (2008), and the five most stable were selected. Searching all nematode libraries larger than 3,000 sequences with this gene set resulted in similar rankings of the genes for most libraries, although some genes were missing in some nematode libraries, and some genes were not ranked in some libraries as in the majority of libraries. Unexpectedly, using 'cluster size counts' instead of the 'EST counts' increased the standard deviations on the reference genes on average with 70% (data not shown). This is probably due to the mean length differences of EST libraries, which influences the maximum E-value obtainable for each EST library. Since we used full length sequences as a query, we decided to implement only EST counts in NEXT. Further analysis on the reference genes is reported in section 4 of this chapter.
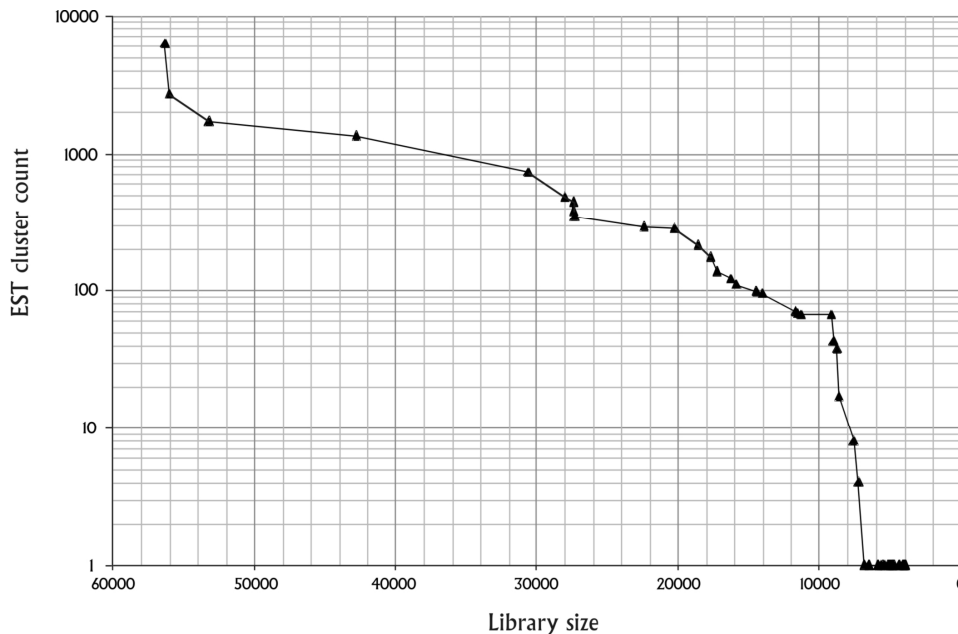
Figure 56: Number of homologous clusters shared by nematode EST libraries. The cumulative number of homologous clusters (Y-axis) is depicted against the library sizes (X-axis). For every subsequent library (according to in decreasing size), the number of common clusters with the bigger EST libraries decreases, with a steeper decrease from libraries containing 10,000 sequences.

8. Acknowledgements

Chapter 7

Cyst nematode chorismate mutase
and early plant responses
to cyst nematode infection

"One of the most painful circumstances of recent advances in science
is that each one makes us know less than we thought we did"
Bertrand Russell

Cyst nematode chorismate mutase and early plant responses to cyst nematode infection

1. Abstract

Sedentary plant-parasitic nematodes have sophisticated molecular tools to manipulate plant cells in order to accomplish the formation of specialized nematode feeding sites. One of the proteins secreted during the early parasitic stages of the nematode is chorismate mutase. This enzyme metabolizes one of the most important key compounds of the plant as the first synthesis step to a wide array of important primary and secondary metabolites. In this chapter, different hypotheses about the putative function of chorismate mutase in the parasitism process are considered. We found that the flavonoid pathway remains unaltered during early nematode infection. This was reflected in the unaltered susceptibility of different flavonoid mutants of *Arabidopsis thaliana* to the beet cyst nematode (*Heterodera schachtii*). In addition, detailed microscopic investigations of the interaction revealed no upregulation of flavonoids in the nematode feeding site. *Arabidopsis* overexpressing *Hs-cm-1* (a secreted chorismate mutase of *H. schachtii*) did not show an altered flavonoid content. Therefore, we conclude that nematode chorismate mutase is not capable to alter flavonoid levels into the plant cell. In addition, the *Arabidopsis* transformants lacked clear visible phenotypic effects and the auxin distribution seems not to be disturbed, rendering the involvement of nematode chorismate mutase in auxin production unlikely. Finally, with the aid of quantitative PCR we measured a strong increase in JA in *Arabidopsis* roots 1 day post inoculation with *H. schachtii*, but this response was completely vanished 2 days post inoculation at the time most nematodes had initiated a nematode feeding site. In contrast, no change in SA response in the roots compared to uninoculated *Arabidopsis* was detected at both time points. However, a strong increase in SA response was observed in shoots at 2 days post inoculation. We discuss these results extensively in the light of the possible roles of the secreted chorismate mutase in the parasitism process.

7

2. Introduction

Plant-parasitic nematodes (PPNs) secrete dozens of proteins through the stylet, at least part of which is presumably injected into a plant cell to induce a nematode feeding site. One of the first identified 'parasitism' proteins shedding light on the unique ways how PPNs alter the plant's metabolism to their benefit, was a chorismate mutase (CM) enzyme (Lambert et al. 1999). Chorismate – the substrate for CM – is the end-product of the shikimate pathway, which is only found in plants, bacteria, fungi and some protozoa (Herrmann et al. 1999; Roberts et al. 1998). The fact that animals do not possess this metabolic pathway suggests that PPNs have acquired this gene solely to assist in parasitism. Root-knot as well as cyst nematodes apparently obtained CM from prokaryotic species, through a process called 'horizontal gene transfer', and they seem to have gained the gene purely to assist in the parasitism process – a hypothesis which could be extended to other parasitism genes (Jones et al. 2003).

Based on the presence of the CM gene only in sedentary nematodes (as opposed to migratory nematodes), this gene may play a role in feeding site induction and/or maintenance. CM has been found in a number of nematodes in which it is part of a large multi-gene family: *Heterodera glycines* (Bekal et al. 2003), *Heterodera schachtii* (Vanholme et al. 2008), *Globodera pallida* (Jones et al. 2003), and *Meloidogyne* species (Lambert et al. 1999; Long et al. 2006). CM is expressed during the early phases of parasitism in the pharyngeal gland cells (both subventral and dorsal) (Painter et al. 2003; Vanholme et al. 2008). The nematode CM genes have a modular structure: a first part encodes a signal peptide for secretion. A second part is cyst-nematode specific and not necessary for enzymatic activity. It contains 6 conserved cysteine residues that probably form stabilizing cysteine bridges, which are often found in secreted proteins. The third domain consists of two parts previously considered distinct domains separated by a proline-rich motif. This third domain contains a single catalytic site (Vanholme et al. 2008) and is found in all CMs of the AroQ$_\gamma$ class, the class of secreted CM present in plant-parasitic nematodes and also some bacteria.

The substrate for CM is the key branch-point compound chorismate, which is the end product of a very important plastid-located pathway, the shikimate pathway. This pathway, through which an estimated 20% of fixed carbon flows under normal growth conditions, generates chorismate through 7 enzymatic steps starting from erythrose-4-phosphate and

phosphoenolpyruvate (see Figure 57). Chorismate serves as a starting point to synthesize both important primary metabolites and a diverse array of specialized secondary metabolites through different metabolic pathways (Herrmann et al. 1999). Besides being a substrate for anthranilate synthase (first step in the synthesis of tryptophan (W)), aminodeoxychorismate synthase (synthesis of tetrahydrofolate), and isochorismate synthase (synthesis of salicylic acid in the plastids), chorismate is also utilized by the plant's CM enzymes. Through a pericyclic Claisen rearrangement of chorismate, phrephenate is formed by CM, leading to the synthesis of tyrosine (Y) and phenylalanine (F) through 2 enzymatic steps. Besides being used in protein synthesis, these two amino acids serve on their turn as precursors for a lot of important growth and defense-related compounds, such as lignin, salicylic acid (synthesis *via* the cytosolic pathway), alkaloids, glucosinolates, flavonols, anthocyanins and condensed tannins (also called proanthocyanidins) (Dixon 2005). The model plant *Arabidopsis* contains 3 genes encoding CM enzymes, of which 2 (At-CM-1 and At-CM-3) are targeted to the plastids, while a third (At-CM-2) is cytosolic. Unlike the At-CM-2, the plastidic CMs are regulated by the levels of the amino acids F, Y and W and are pathogen-inducible (Eberhard et al. 1996). But evidence points to additional feed-back mechanisms controlled by other chorismate-derived compounds (Romero et al. 1995). The plastids seem to be the sole location for the chorismate pool, so the finding of a cytosolic CM is enigmatic since the cytoplasm lacks a shikimate pathway. It is largely unknown how transport of chorismate between different subcellular compartments occurs. It is clear however that the metabolic network originating from the shikimate pathway is finely regulated by numerous mechanisms in order to respond most efficiently to internal metabolic demands and to external influences, which require the synthesis of secondary metabolites (Weaver et al. 1997).

Hence, the secretion of a nematode CM enzyme into a plant cell seems to be a molecular tool allowing the nematode to directly influence the plant's metabolic core machinery (Lambert et al. 1999). The lack of a plastidic localisation signal in the secreted nematode CMs indicates that this enzyme most likely functions in the cytoplasm. Due to the wide metabolic connections of chorismate, the function of the nematode CM is still a matter of debate. Nevertheless all hypotheses assume that it encompasses adjusting the concentrations of the chorismate-derived compounds for the nematode's own benefit.

Doyle and Lambert (2003) generated soybean hairy roots expressing *Mj-cm-1* (of the root knot nematode *Meloidogyne javanica*), which showed an altered phenotype with decreased

7

numbers of lateral roots and inhibition of vascular formation. Interestingly, this phenotype could be rescued by applying indole acetic acid (IAA), indicating that Mj-CM-1 reduced IAA levels in the hairy roots. This led to the "chorismate competition" model: the nematode CM utilizes chorismate in the cytoplasmic pool, causing the chorismate to be drained from the plastids, in which the IAA synthesis is subsequently lowered due to lack of precursors.

Although based on experimental evidence, this hypothesis has never received wide acceptance. First, upon nematode feeding site initiation, the IAA level in the NFS increases rather than the opposite (Karczmarek et al. 2004). Second, the major site of IAA synthesis is located in the shoots, and IAA is transported acropetally through the phloem with increase of IAA levels being accomplished by regulation of transport proteins or flavonoids (mainly quercitin and kaempferol) (Brown et al. 2001). Thirdly, hairy roots are an artificial system, sensitive to disturbance by hormone level alterations. Drawing conclusions from such a system must be done carefully (Plovie et al. 2003).

Another hypothesis states that CM could increase the cytosolic synthesis of flavonoids – that on their turn alter IAA levels by inhibiting auxin transport (Gheysen et al. 2002). In favour of this hypothesis, one study noted an increase of flavonoids in and around syncytia (nematode feeding site of cyst nematodes) compared to control roots (Jones et al. 2007). Flavonoids are known to regulate auxin transport (Brown et al. 2001), which led to the hypothesis flavonoids might be involved in the auxin increase in syncytia during onset of its formation (Hutangura et al. 1999). However, no flavonoid accumulation has been observed in giant cells, even though a similar initial auxin increase is observed as in syncytia (Hutangura et al. 1999; Wuyts 2006). Instead, increased levels of flavonoids are observed in the gall cells surrounding the giant cells. Indeed, it seems quite inefficient for the nematode to influence auxin levels through alteration of flavonoid levels (as an effect of injected nematode CM), compared to alternatives as stimulating auxin production, or providing auxin in the secretions. Interestingly, different plant hormones are found in PPN extracts and in pharyngeal gland secretions, although the amounts seem to be very low (Meutter et al. 2003; Vanholme et al. 2004). In addition, *Heterodera schachtii* infection of several *Arabidopsis* mutants deficient in the first steps of flavonoid biosynthesis (*tt4, tt5* and *tt6*, and double mutants hereof) revealed no difference or – in some cases – even an increase in susceptibility compared to the wild type (Jones et al. 2007).
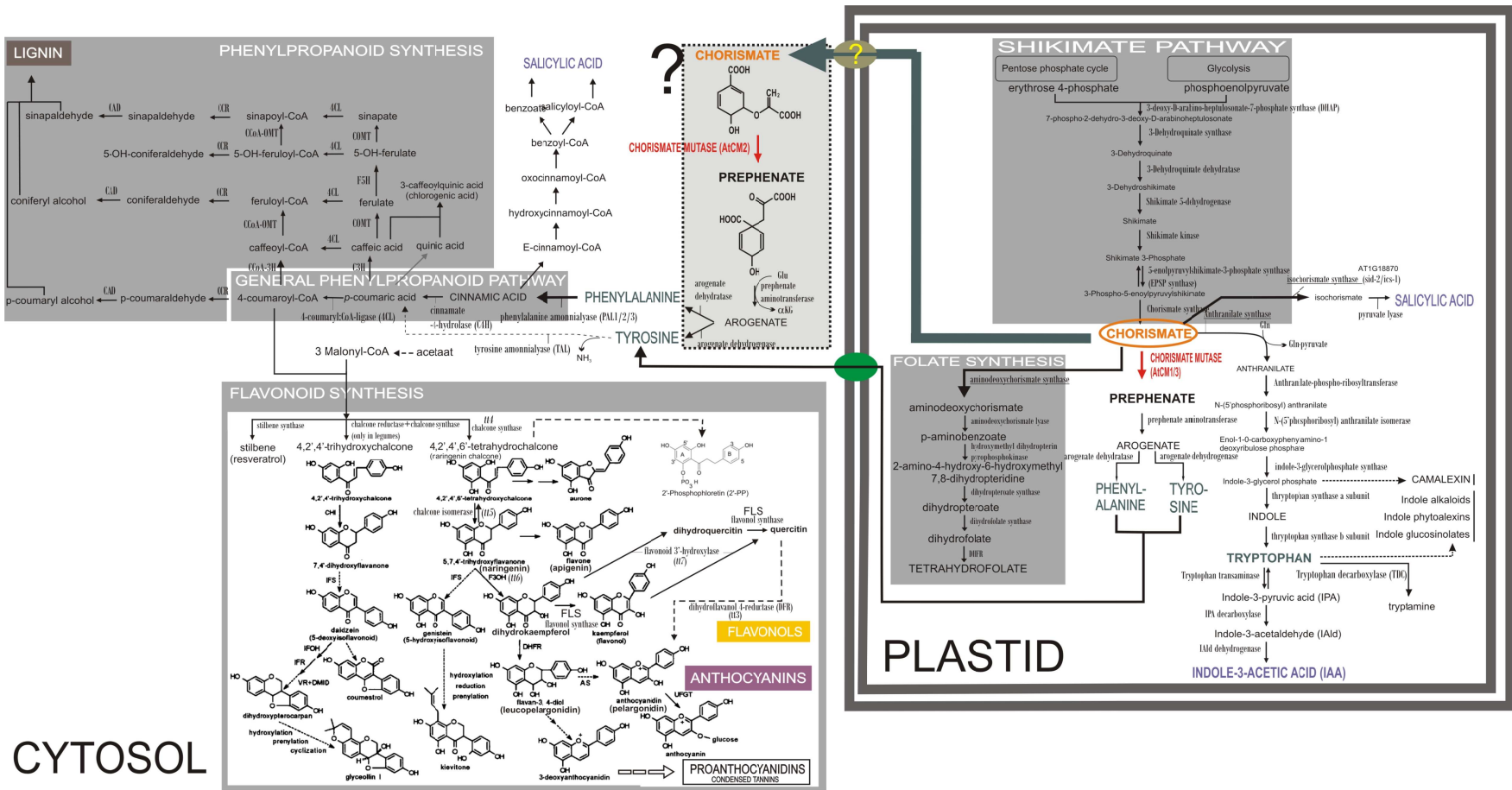
Figure 57: Schematic representation of the molecular pathways in plants linked to chorismate, with indication of some enzymes and gene names occurring in *Arabidopsis thaliana*. Chorismate synthesis is taking place in the plastids. The occurrence of chrorismate utilizing enzymes in the cytosol suggests that chorismate is transported to this compartment. Different pathways producing aromatic ring containing metabolites are located in the cytosol. Many of these pathways are organized as 'metabolons', chains of enzymes anchored to the rough endoplasmatic reticulum.

Based on the nematostatic effect of some phenylpropanoid derived compounds (phytoalexins), a third hypothesis states that CM influences phenylpropanoid levels in order to regulate plant defense against the nematode. Phenylpropanoid-derived phytoalexins, such as pyrocatechol, coumestrol, glyceollin and medicarpin accumulate in some plant hosts upon nematode infection, and have the capacity to impede nematode development (Giebel 1973; Veech 1981; Wuyts et al. 2006). In fact, nematode resistance is often associated with increased phytoalexin and lignin production and is accompanied by upregulation of the general phenylpropanoid pathway (phenylalanine ammonia lyase (PAL) and tyrosine ammonia lyase (TAL) increase) (Edens et al. 1995; Wuyts et al. 2006). The regulation of phytoalexin biosynthesis is not known, but it might be that signalling hormones such as JA play an important role, as an increase of flavones is detected as well upon application of JA to oat seedlings as after nematode infection (Soriano et al. 2004). Could it be that CM causes an alteration of phenylpropanoid blends in the NFS, so to avoid accumulation of harmful phytoalexins and counter any plant defenses? Pointing in this direction, *Pectobacterium atrosepticum* mutants deficient in genes required for normal suppression of basal defenses could suppress basal plant defense again after transformation with CM (J. Jones, personal communication). However, plant defense is tightly regulated by complex signalling pathways. And one of the major defense related signalling molecule, salicylic acid, is metabolically linked to CM. Hence, the link with plant defense could be established by alteration of the signalling, instead of altered phenylpropanoid compounds (see section c).

Another straightforward possibility is that the extra CM causes a better nutritional content of the feeding site. Nutritional studies on free-living and non-obligate parasitic nematodes show that aromatic amino acids are an essential dietary requirement for nematode growth and reproduction (Vanfleteren 1978). Through stimulation of the synthesis of aromatic amino acids, it could be that the nutritional value of the NFS is boosted for the nematode, allowing completion of the life cycle in most optimal conditions. Infection experiments with nematodes soaked in dsRNA targeting nematode CM results only slightly less nematode infection (J. Jones, pers. comm.). But fewer nematodes developed to adults, with the largest effect observed in female nematodes. This could point to an establishment and maintenance of a healthy and nutrient-rich NFS as a crucial function for chorismate mutase. Finally, recently the beneficial effects of dietary addition of flavonoles (e.g. quercitin) on the life-span of the nematode

7

Caenorhabditis elegans is reported, an effect which could be mediated by CM if it stimulates flavonoid synthesis (Kampkötter et al. 2008).

Increasing evidence points to a possible correlation of secreted chorismate mutase with pathogenicity, arguing for an essential role in the parasitism process. Firstly, nematode CM is an avirulence gene, since different CM isoforms are expressed in nematode lines with different virulence (Bekal et al. 2003). Furthermore, CMs are also secreted by some pathogenic bacteria. It is believed that secreted CM functions in the periplasm of these gram-negative bacteria (*Erwinia herbicola, Pseudomonas aeruginosa*) (Calhoun et al. 2001). However, the discovery that *Mycobacterium tuberculosis*, a gram-positive bacterium that lacks a periplasmic space, secretes highly active CM in the extracellular space within its mammalian host, could indicate a more fundamental and yet unknown role of secreted CM in pathogenicity (Kim et al. 2006b; Sasso et al. 2005).

It is clear from all the above examples that yet no conclusion can be drawn about the functional role of CM in the nematode feeding site. Our laboratory has cloned and characterized a secreted chorismate mutase of *Heterodera schachtii* (*Hs-cm-l*) (Vanholme et al. 2008). In this chapter, a further understanding of the role of CM was obtained through the analysis of *Arabidopsis* transformants, infection tests on flavonoid mutants, visualizing flavonoid content and performing quantitative PCR (Q-PCR).

3. Results

a. Analysis of *Arabidopsis* transformants overexpressing Hs-CM-1

Arabidopsis Columbia 0 was transformed with *Hs-cm-l* under control of the 35S promoter, and 4 different homozygous lines were generated. On those lines, the presence of the construct was checked by PCR, and expression was analyzed by RT-PCR (Figure 58). In addition, 2 homozygous lines expressing *Hs-cm-l* as an N-terminal green fluorescent protein (*gfp*) fusion were also generated. Remarkably, none of the obtained transformants differed phenotypically from the non-transformed control plants (data not shown). Furthermore, the flavonoid content appeared not to be altered, since diphenyl boric acid (DPBA) staining (which

7

Figure 58: Confirmation of transgene expression on the homozygous lines (F3 generation). RT-PCR (35 cycles) on cDNA of transformed *Arabidopsis* overexpressing 1, *Hs-cm-l* (line 54), 2, *Hs-cm-l* (line 11), 3, *Hs-cm-l* (line 5), 4, *Hs-cm-l* (line 4), 5, *gfp-Hs-cm-l* (line 3), 6, *gfp-Hs-cm-l* (line 6). Note that lines with different levels of expression are selected.
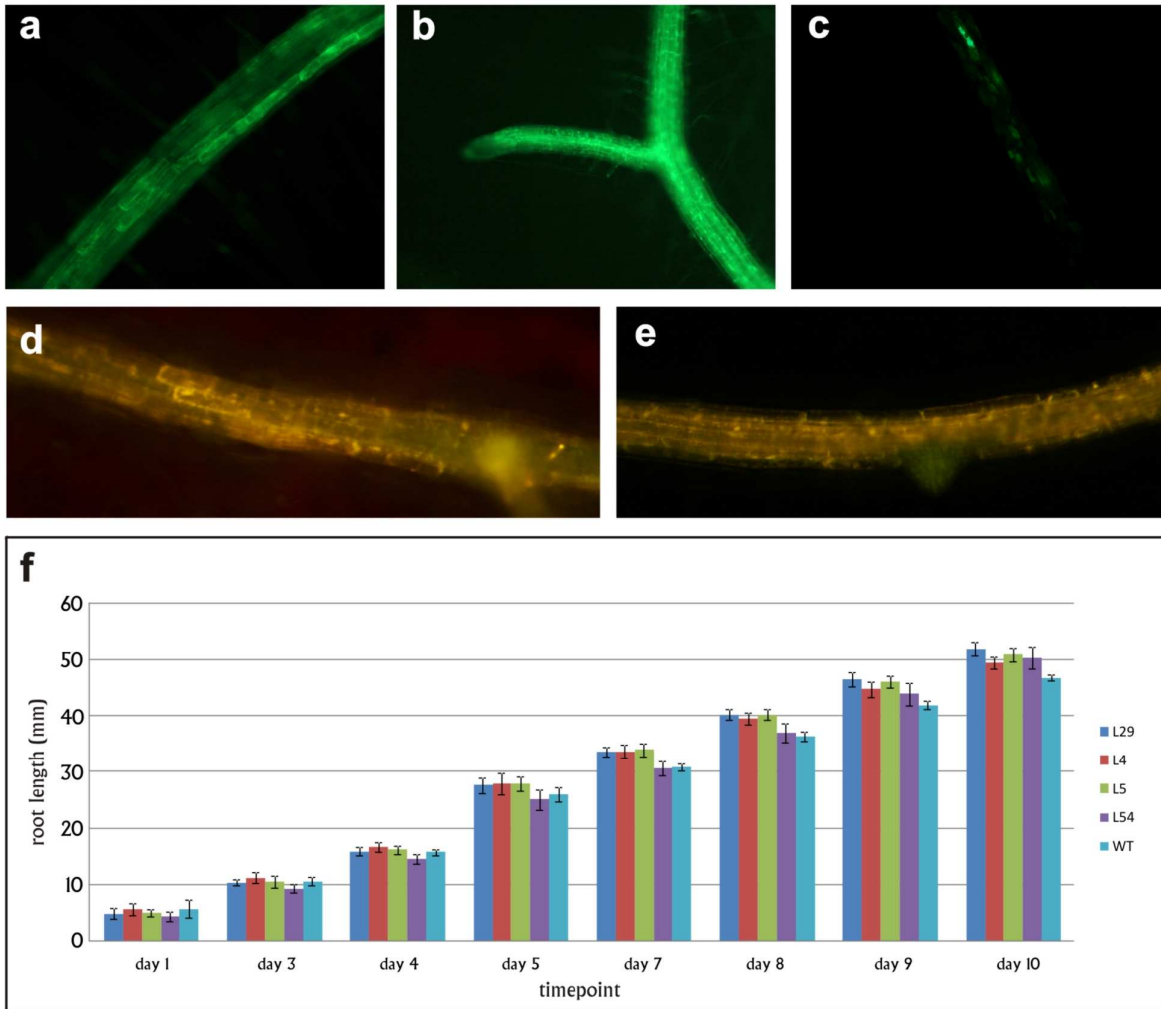
Figure 59: Phenotypic analysis of *Arabidopsis* transformed with *Hs-cm-l*. **a** and **b**, *gfp-Hs-cm-l* fusion constructs showed a cytoplasmic localisation signal. **c,** silencing occurred already in plants of the F3 generation. **d** and **e**, flavonoid staining of WT roots (d) and overexpressing *Hs-cm-l* (e). **f,** root growth assay on four homozygous *Hs-cm-l* overexpression lines. No significant difference between WT and the transgenic lines was measured.

differentially stains flavonoids) revealed no differences with WT *Arabidopsis* (Figure 59). To detect more subtle phenotypic differences through a possibly altered auxin homeostasis, we monitored root-growth of 1 week old seedlings during 1 week (Figure 59). No significant differences in mean root length and mean root growth could be detected. All analyzed *gfp*-fusion plants displayed silencing of *gfp* expression (patchy GFP signal with complete disappearance in the later generations), but any fluorescent signal clearly showed a cytoplasmic localization of the Hs-CM-1-GFP fusion protein (Figure 59). Several western blot attempts using *gfp*-antibodies on protein extracts of the GFP-fusion lines (F2 generation) were not successful in detecting the fusion protein (data not shown), indicating strong silencing.

b. No evidence for the involvement of flavonoids neither in nematode infection nor in nematode feeding site development

To investigate the effect of different flavonoid compounds on nematode infection, several *Arabidopsis* mutants deficient in certain enzymes of the flavonoid biosynthesis pathway were assessed in infection tests (Table 15 and Figure 60). Compared to infection of wild type *Arabidopsis* (Wassilewskija ecotype) none of the tested flavonoid mutants had a significantly different infection level totalled over all infection tests (Figure 61). Nevertheless, mutant *tt10* seems to support higher infection levels compared to WT, but this was only significant in one of the four repeated experiments (Tuckey Method, $p<0.05$).

In contrast to the reported accumulation of flavonoids at the feeding site of cyst nematodes (Grunewald 2008; Jones et al. 2007), no difference in susceptibility was observed for flavonoid mutants (Jones et al. 2007 and this study). Therefore, accumulation of flavonoids in the mutant and WT roots after nematode infection was analyzed by DPBA staining. The different flavonoid mutants were investigated at 2, 5 and 12 days post infection (dpi). The local green

7

Table 15: Used *Arabidopsis* flavonoid mutants.

Arabidopsis mutant	Affected enzyme	AGI	Described flavonoid profile	Reference
tt4	Chalcone synthase	AT5G13930	Absence of flavonoids	Shirley et al. (1995)
tt3	Dihydroflavonol-4-reductase	AT5G42800	Accumulation of flavonols	Shirley et al. (1995)
fls-1	Flavonolsynthase	AT5G08640	Accumulation of quercitin	Routaboul et al. (2006)
tt10	polyphenoloxidase	AT5G48100	Accumulation of epicatechin	Pourcel et al. (2005)
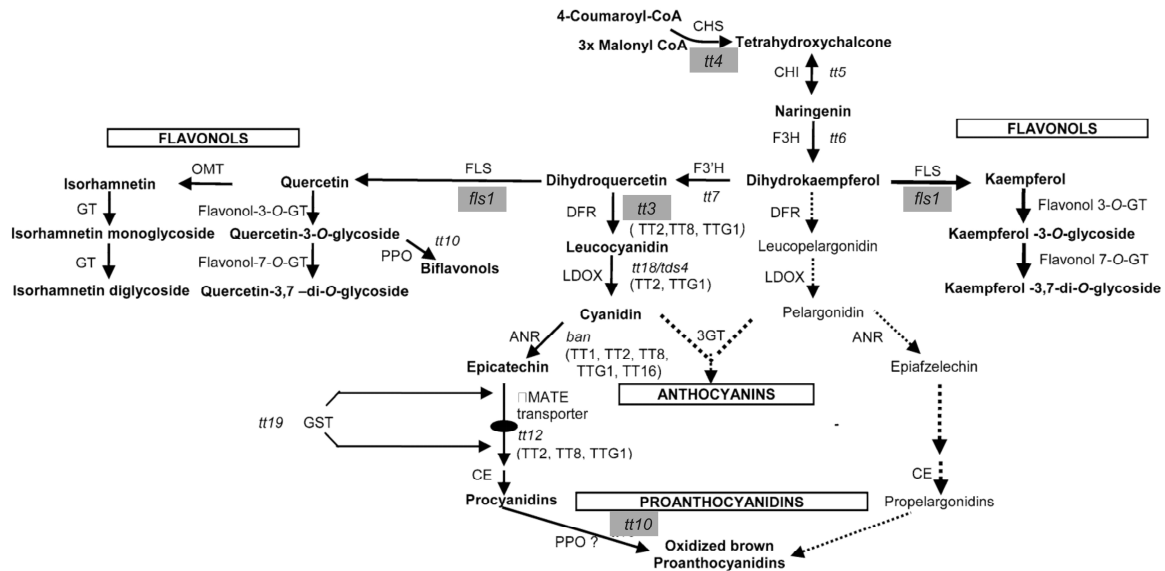
Figure 60: The flavonoid pathway, with indication of the used *Arabidopsis* flavonoid mutants.
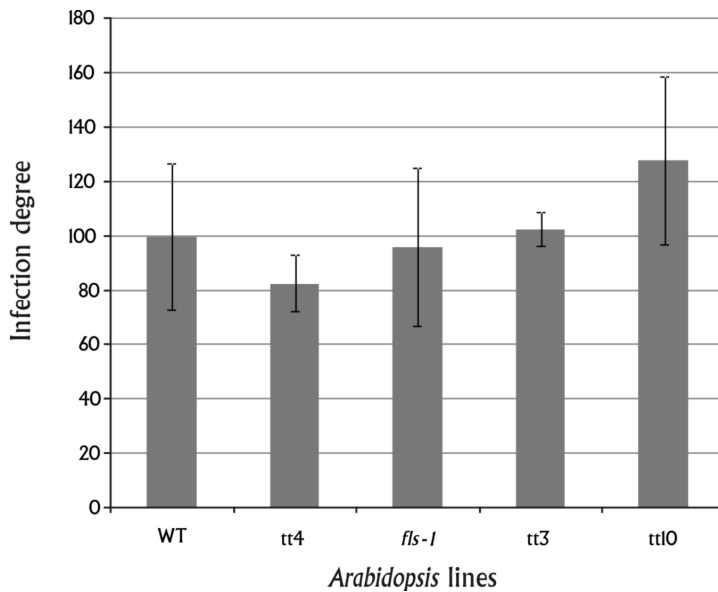
Figure 61: Results of infection experiments of *Arabidopsis* flavonoid mutants relative to WT infection (set to 100). Shown is the mean of 4 experiments, except for *tt4* and *tt3* representing the results of one experiment.

fluorescence observed at the site of nematode infection in all mutants seems not to originate from flavonoids, because it was also present in *tt4* mutants who are completely devoid of flavonoids (Figure 62, a and c). This green fluorescence is localised at the interface of the nematode with the plant root cells. The source of this flavonoid-independent signal is unknown. In the other mutants and WT *Arabidopsis*, to our surprise, flavonoids were not

consistently upregulated in the vicinity of the site of nematode infection. In a few cases, a bright gold fluorescence accumulated at some distance of the nematode infection site (Figure 62, b). However, in the vicinity of the parasitizing nematode generally no change and in most cases even a decrease in flavonoid content was observed. This was observed in all different mutants and in the WT (Figure 62).

The results of the flavonoid staining on cyst nematode feeding sites were clearly in contrast with previous reports, and prompted us to investigate in more detail. *wrky23::gus* plants infected with nematodes produce GUS in the feeding site (Grunewald et al. 2008), and additional staining of flavonoids on these plants enabled us to precisely localize flavonoids with respect to the feeding site. The pictures emerging from this experiment were clear: feeding sites formed by cyst nematodes do not accumulate flavonoids at 5 dpi (Figure 63). Most fluorescence was observed around the migratory path of the nematode and at the site of penetration around the nematode's head.

c. Roots and shoots respond very differently to nematode infection

Another pathway closely linked to CM is salicylic acid (SA) biosynthesis (see Figure 57). The phytohormone SA acts together with jasmonic acid (JA) and ethylene (ET) in a complex signalling network regulating plant defense responses through transcriptional induction of numerous defense related genes (Bostock 2005). Biosynthesis of SA can occur through two separate pathways, both starting from chorismate. One pathway is located in the cytoplasm and runs via the conversion of phenylalanine to trans-cinnamic acid by PAL, followed by multiple enzymatic steps to form benzoic acid, a precursor for SA. A second pathway is located in the plastids, where SA is produced out of chorismate via isochorismate through the action of isochorismate synthase, isochorismate pyruvate lyase and additional uncharacterized enzymes (Strawn et al. 2007). Recent results suggest that both biosynthesis pathways are differently regulated in response to biotic and abiotic stresses. The plastid biosynthesis pathway responds to biotic stresses, while the cytoplasmic pathway is more responsive to abiotic stresses (Catinot et al. 2008; Strawn et al. 2007; Wildermuth et al. 2001). Biotic stresses evoking a SA response involve mostly biotrophic diseases and some insects such as aphids. In the defense regulatory network, the SA response is known to work antagonistically to the JA response, which is mostly elicited by necrotrophic diseases and most chewing and
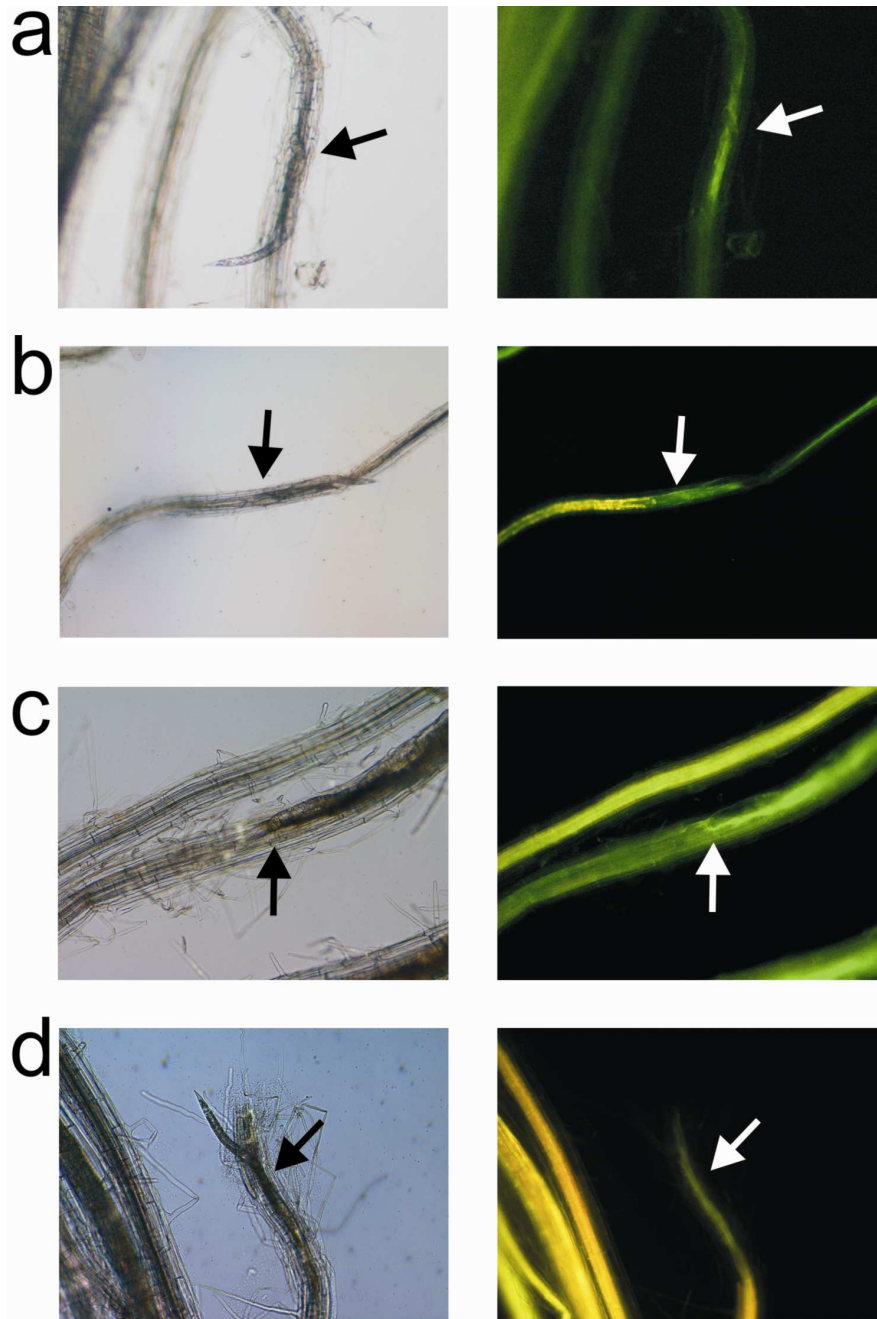
7

Figure 62: Some selected examples of DPBA staining on *Heterodera schachtii* (arrow) infected transparent testa mutants and WT *Arabidopsis*. A faint green background is caused by autofluorescence of sinapate esters (Wuyts 2006). Kaempferol fluoresces yellow-green, quercitin gives a gold fluorescence, and its glycoside quercitrin gets a yellow-orange colour (Peer et al. 2001). **a,** As observed in all mutants, the *tt4* mutant showed a bright green fluorescence around the nematode. **b,** In WT *Arabidopsis*, often – but not in all cases – accumulation of quercitin (gold) at some distance of the infecting nematode is observed. **c,** In *H.s.* infected roots of the *tt10* mutant, the same localised green fluorescence as observed in a is seen. No upregulation of flavonoids occurs. Instead, often a decrease in flavonoid content is seen (compare the upper root with the lower infected root). **d,** As here shown in *fls-1* mutants, similar local decrease in flavonoid content were often observed in all mutants.

sucking insects (Zhu-Salzman et al. 2005). Remarkably, this fine-tuned trade-off between defense responses is often abused by pests to their own benefit: by stimulating one response that is not harmful for the invader (e.g. SA in the case of aphids), the most effective plant defense response is attenuated (e.g. JA elicits effective responses against aphids) (Zhu-Salzman et al. 2005).
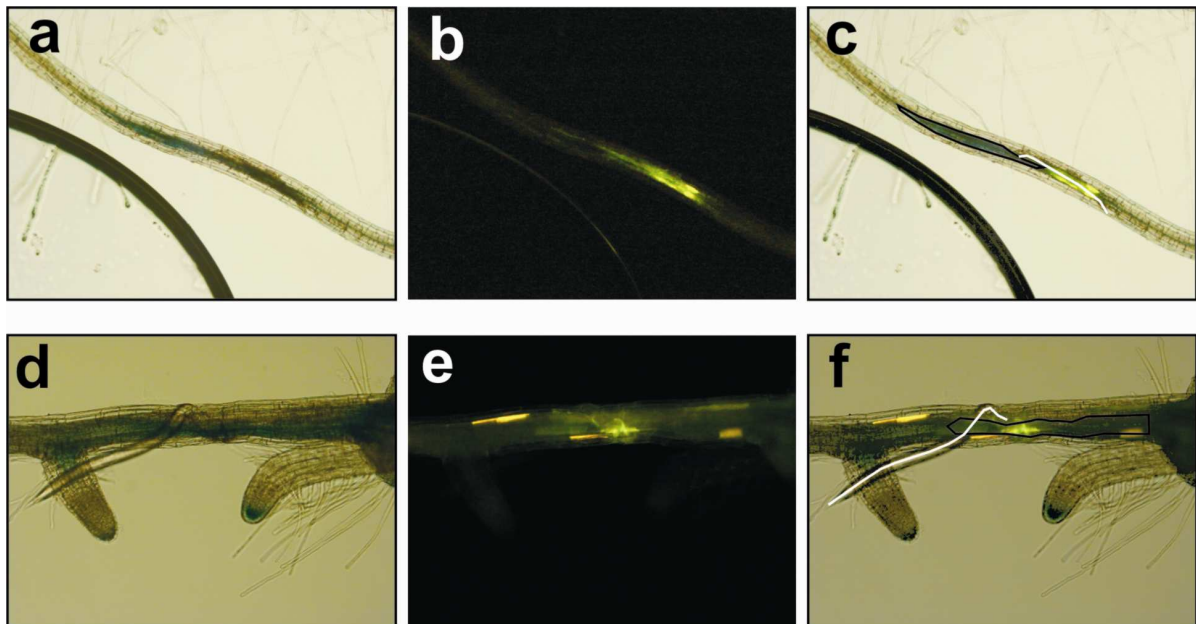
Figure 63: Flavonoid accumulation in *wrky23::gus* transformed *Arabidopsis* roots upon *Heterodera schachtii* infection, 5 dpi. The syncytium is stained blue (**a** and **d**), and the flavonoids are stained with DPBA (**b** and **e**). In no case, flavonoid staining was found colocalising with the syncytium (merged figures **c** and **f**, with the nematode indicated by a white line and the syncytium delineated with a black line). Instead, flavonoids accumulate in the migratory path of the nematode. Occasionally, root cells of the cortex accumulate quercitin (gold fluorescence). The bright green fluorescence seen at the head of the nematode is flavonoid independent and is also observed in *tt4* mutants.

Different scenarios can be thought of in which CM functions to alter the SA-regulated response (as well as for plant-pathogenic bacteria as for PPN). Attenuation of plastidic SA-production in nematode feeding sites can be achieved by competition for the common precursor chorismate. Otherwise, cytosolic SA biosynthesis can be stimulated by increased prephenate levels caused by CM activity. Unfortunately, the general plant responses to nematode infection are not yet fully resolved. Typically, the experiments performed in this research lack uniform set-up, further hampering a thorough understanding (see Table 16; for a good review, see Gheysen et al. (2008)). Nevertheless, the majority of the mutants or

7

transformants with altered hormone levels show a significantly altered susceptibility towards nematodes. Although this may point to an involvement of all investigated phytohormones in a successful parasitic interaction, the results can also be explained in light of the interaction between different hormones in defense-signalling (see above).

To shed more light on the plant response during early nematode infection, a Q-PCR experiment was set up to monitor SA and JA responsive genes (respectively *pr-1* and *pdf-1.2*), chalcone synthase (*chs*) and phenylammonia lyase (*pal-1*), both in roots and in leaves of infected plants (Figure 64). The results from this Q-PCR experiment show that the overall *pal-1* levels

Table 16: Summary of reported effects of altered ET, JA and SA hormone levels in susceptible plants upon nematode infection.

Mut/transf/appl.*	Description	Nematode**	Host	Result***	Reference
ein	ET insensitive	H.s.	Arabidopsis	-50%	Wubben et al. 2001
eto	ET overproducer	H.s.	Arabidopsis	+80%	Wubben et al. 2001
ACC supplementation	Increased ET level	H.s.	Arabidopsis	+25%	Wubben et al. 2001
Foliar JA application	Induced JA response	M.j.	Tomato	-75%	Cooper et al. 2005
Root JA application	Induced JA response	M.j.	Soybean	-90%	Shimizu et al. 2007
JA application	Induced JA response	H.a.	Oat	-70%	Soriano et al. 2004
def-1	JA deficient	M.i.	Tomato	=	Bhattarai et al. 2008
jai-1	JA insensitive	M.i.	Tomato	-75%	Bhattarai et al. 2008
Foliar SA application	Induced SA response	M.j.	Pineapple	=	Chinnasri et al. 2006
Foliar SA application	Induced SA response	Rot.	Pineapple	=	Chinnasri et al. 2006
Foliar SA application	Induced SA response	M.i.	Okra	-70%	Nandi et al. 2003
Foliar SA application	Induced SA response	M.i.	Cowpea	-75%	Nandi et al. 2003
Medium SA application	Induced SA response	H.s.	Arabidopsis	-66%	Wubben et al. 2008
NahG	SA deficient	M.i.	Tomato	=	Bhattarai et al. 2008
sid	SA deficient	H.s.	Arabidopsis	+20%	Wubben et al. 2008
NahG	SA deficient	H.s.	Arabidopsis	+20%	Wubben et al. 2008
npr1	SA insensitive	H.s.	Arabidopsis	+20%	Wubben et al. 2008

* name of the mutant or transformant, or indication of the phytohormone application.
** H.s.: *Heterodera schachtii*, H.a.: *Heterodera avenae*, M.j.: *Meloidogyne javanica*; M.i.: *Meloidogyne incognita*; Rot.: *Rotylenchulus*.
*** average effect on susceptibility compared to control plants.

remain relatively constant in both tissues at both time-points. Similarly, the expression level of *chs* does not alter significantly. These two observations point towards a minimal influence of the nematode infection on the regulation of the general phenylpropanoid pathway (through *pal-l*) and flavonoid synthesis (through *chs*) on the level of the complete plant. In contrast, we see drastic changes in the SA and JA response. In roots, *pr-l* expression levels stay low at both time-points, indicative of the lack of SA signalling in the roots during early infection. However, in the above ground parts, a significant SA response is visible through the 100 times upregulation of *pr-l* at 48 hours after infection. The scenario is inversed for the JA response. Initially a very strong response is visible in the roots 24 hours after infection, but this response could not be detected anymore 48 hours after infection. No sign of JA response is observed in the above ground parts at both time points. Infecting *Arabidopsis* transformed with *prl::gus* and *pdfl.2::gus* and subsequent GUS staining revealed no staining at the level of the syncytium at 2, 5 and 12 dpi (data not shown).
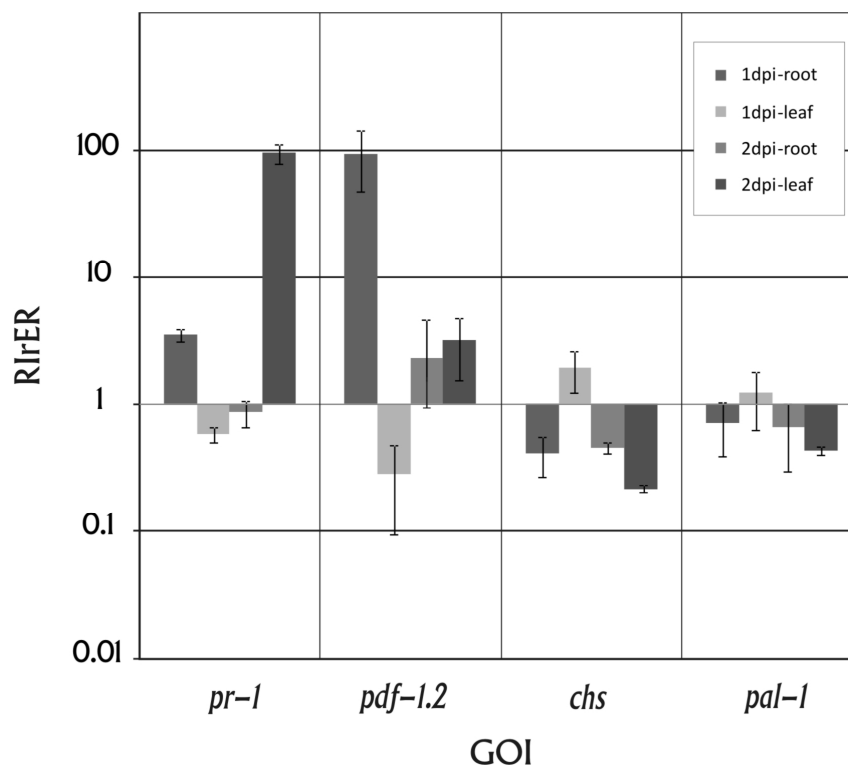
Figure 64: Relative expression levels (RirER) of four genes-of-interest (GOI), *pr-l* (SA responsive), *pdfl.2* (JA responsive), *chs* (flavonoid biosynthesis pathway) and *pal-l* (general phenylpropanoid pathway). The value is the Reference Index relative Expression Ratio (RIrER, see Materials and Methods), which shows the relative expression of the GOIs compared to the WT and normalized to three internal control genes. Each measurement was done in technical triplicate and biological duplicate.

7

4. Discussion

The reason for secretion of chorismate mutase by sedentary plant-parasitic nematodes into a developing nematode feeding site remains largely unknown. Due to the recognition of chorismate as a precursor for many different pathways involved in plant growth, development and defense, equally many hypotheses about its function have been proposed. In a series of experiments, we gathered further insights to direct future experiments dedicated to unravel the role of CM.

The hypothesized role of CM to alter auxin levels in NFS seems unlikely to be genuine. *Arabidopsis* overexpressing *Hs-cm-1* showed no signs of altered auxin levels. The transformants were phenotypically not distinguishable from WT *Arabidopsis*, and exhibited no altered root growth. In agreement with this, a crossing of some overexpressors with *dr5::gus* lines revealed no difference in GUS staining patterns compared to WT (data not shown). Therefore, CM seems not to directly modify the auxin levels to a biological relevant extent. This result is in contrast to previous experiments overexpressing CM in soybean hairy roots (Doyle et al. 2003), and can be explained by the difference between the two systems in sensitivity to hormone levels.

The hypothesis that CM alters flavonoid content in NFS in order to influence auxin accumulation was also not supported by experimental data. Moreover, in contrast to previous reports (Grunewald et al. 2008; Jones et al. 2007), no flavonoid accumulation could even be detected in NFS initiated by cyst nematodes. Thus auxin accumulation in NFS can not be established by increased flavonoid content. In addition, the CM transformants showed no different flavonoid content based on microscopic observations of DPBA stained flavonoid. Recent results point rather to an involvement of auxin transporters in auxin accumulation (Grunewald et al. 2009). In addition, the role of flavonoids regarding nematode infection seems to be neutral based on similar cyst nematode infection levels in mutants compared to WT *Arabidopsis*, except for an insignificantly slightly higher infection with *tt10* mutants. *In vitro* nematode infection tests suffer from high variability (Sijmons et al. 1991). Therefore, only repeated infection tests on *tt10* mutants could give an indication if the observed difference is significant. If this difference is genuine, the reason for it can be dual. The *tt10* mutants exhibit lower levels of oxidized procyanidins (condensed tannins), and higher levels of quercetin

7

derivatives, epicatechin and procyanidins (Pourcel et al. 2005). Both differences in flavonoid content could potentially negatively influence nematode reproduction (Brunet et al. 2006; Molan et al. 2004). Exposing *Heterodera schachtii* to the different flavonoids altered in the *ttl0* mutant could clarify this issue. Nevertheless, it will remain difficult to explain results of any infection test on flavonoid mutants with simultaneously lower and increased levels of certain flavonoids. Wuyts and colleagues exposed three different PPN species (not including *Heterodera* species) to numerous different flavonoids and phenolic compounds. They found that the majority of the compounds had no effect, with the few effective compounds showing many species-specific effects (Wuyts et al. 2006).

Although flavonoids seem to participate to only a lesser extent in the plant-nematode interaction, evidence exists that the general phenylpropanoid pathway does play a role in plant resistance against diseases and nematode infection (Mauch-Mani et al. 1996). Nematode resistant potato expresses *pal* and *tal* at a higher level compared to non-resistant plants (Giebel 1973). Furthermore, *pal* overexpression can cause resistance to certain diseases, most likely through the action of SA (Mauch-Mani et al. 1996). Our data shows that *pal-1* expression remains unaltered in roots and leaves in susceptible *Arabidopsis* infected with *H. schachtii* at 1 and 2 dpi. Interestingly, micro-array studies on syncytia of soybean infected with *Heterodera glycines* revealed upregulation of many genes of the phenylpropanoid pathway, and the subsequent pathway of lignin biosynthesis, at 2 dpi, reaching back normal levels at 5 and 10 dpi (Ithal et al. 2007). Coincidentally, in *Heterodera cm-1* expression peaked in the early parasitic stages, while expression was barely undetectable in adult stages using RT-PCR (Vanholme et al. 2008). Nematode CM seems a good candidate to be the mediator of the observed changes in syncytia, although this is again a hypothesis. On the level of the complete plant, also the flavonoid pathway seems not to be altered, since no upregulation is measured of *chs* during early nematode infection. This is in agreement with the only report we could find on this, in which *chs* expression stays equal - besides a slight upregulation 6 hours after infection - as measured in alfalfa infected with root-lesion nematodes (*Pratylenchus*) (Baldridge et al. 1998). In contrast to nematode infection, increased expression of chalcone synthase has been reported after infection with fungi (Fofana et al. 2005; Nagy et al. 2004). In addition, alfalfa resistant against root-lesion nematode showed an almost 2-fold higher constitutive *chs* expression level (Baldridge et al. 1998). In summary, there is evidence for a role of the general phenyl-propanoid and derived pathways in plants resistant to PPNs, but evidence is lacking for

7

the hypothesis that PPNs actively influence these pathways in susceptible plants. The limited accumulation of flavonoids observed around some nematode feeding sites is most likely a plant response.

The hypothesis that nematode CM stimulates a SA response through fueling the cytoplasmic biosynthesis pathway in order to depress a more harmful JA response seems very unlikely with the data we have gathered. No SA response is detected, nor at the level of the NFS, nor at the level of the complete root system. In contrast, above-ground parts seem to accumulate SA from 48 hours after infection on. Wubben et al. (2008) measured at 3 dpi an 25-fold increase of *pr-1* transcripts in leaves compared to the controls, but interestingly the transcript levels decreased at 8 and 13 dpi to a barely 5-fold increase. Upon application of SA to plant roots, most of the SA is transported to the leaves where it can initiate *pr-1* expression (Molinari et al. 2006). So in the case CM increases SA production at the site of infection, it could be immediately translocated to above-ground parts causing no SA-related response in the roots. Interestingly, SA seems to be an attractant for *Meloidogyne incognita* (Wuyts et al. 2006), even though SA-induced defense response hampers nematode infection (Table 16). It must be noted that the plastidic biosynthesis pathway of SA could be downregulated due to the action of nematode CM, although no conclusion can be drawn from our results.

The JA response tells a different story. The plant appears to react to the entrance of the nematode into its root system, since a very strong JA response is observed 24 hours after infection, a phase in which the majority of the nematodes is migrating (~65%) or still outside the plant (~25%) (Tytgat 2003). Very remarkably, 48 hours after infection no JA response can be detected anymore in the roots. At this time point, over 80% of the nematodes have established a feeding site (Tytgat 2003). It could be that the disappearance of JA response is due to downregulation of JA biosynthesis genes, as revealed by micro-array data obtained 2 dpi from soybean infected with the cyst nematode *Heterodera glycines* (Ithal et al. 2007). In addition, a fatty-acid binding protein has been detected in the nematodes secretions of the hypoderm (FAR-1) that has the ability to bind linoleic and linolenic acids, which are involved in plant defense, either as signalling molecules or as precursors of JA biosynthesis (Prior et al. 2001). These observations suggest that a JA response is elicited early in plant-nematode interaction probably due to the severe wounding during nematode migration (McConn et al. 1997), and that the nematode probably succeeds in actively reducing this defense signalling after 2 dpi.

7

In conclusion, several lines of evidence indicate that the nematode does not upregulate the SA or JA response in order to suppress other defense responses. Induction of both SA or JA responses in plants generally decreases susceptibility to nematode infection (table 2). Due to lack of a direct metabolic link between JA production and cytosolic CM, the reduction in JA observed at 2 dpi is in all probability not directly achieved by action of nematode CM. Since no SA response in the root is observed, the JA reduction seems unlikely to be accomplished by nematode CM stimulating cytosolic SA production. From these data, no link between plant defense and nematode CM can easily be concluded. Still, besides PPNs and plant-pathogenic bacteria, a secreted CM has also been found within *Mycobacterium tuberculosis*. Based on these discoveries a more general – but yet undefined – role of secreted CM in pathogenesis is proposed, unrelated to SA biosynthesis (Sasso et al. 2005). Interestingly, in secretomes of pathogenic fungi, isochorismatase domains (PF00857, IPR000868) were identified which were not present in secretomes of non-pathogenic fungi (Soanes et al. 2008). Isochorismatase catalyses the conversion of isochorismate to 2,3-dihydroxybenzoate and pyruvate, but its motif is also found in a number of hydrolases such as nicotinamidase. Most remarkably, chapter 2 of this thesis reports on the identification of an isochorismatase domain specific for plant-parasitic nematodes. It is tempting to speculate that these isochorismatases exhibit a similar function in both plant-pathogen systems. Isochorismate, which is produced out of chorismate, is a precursor of SA. One of the hypothesized roles of the secreted phytopathogenic isochorismatase, which is most likely functioning in the cytosol, is the depletion isochorismate out of the plastids so that the plastidic SA production is reduced due to lack of precursors (similar as the plastidic biosynthesis pathway reduction hypothesis of chorismate) (Soanes et al. 2008). As such, no SA can accumulate in response to pathogen attack. These hypotheses still need experimental confirmation, but the parallel between isochorismatase as a secreted enzyme of plant pathogens acting on a derivative of chorismate, and CM as an established parasitism gene in PPN is striking. Both enzymes are homologues of plastidic enzymes of the plant, but are predicted to function in the cytosol. And both metabolize chorismate or a direct product of chorismate, separated by one enzymatic step.

So what is the function of nematode CM in the NFS? No confident answer can be given to this question yet. Since CM is a key enzyme for the plant, many possibilities were initially proposed, although a few of them have become less likely with new data arriving. Further extensive analysis (e.g. infection tests, metabolite screening) of the HsCM1 transformed

7

Arabidopsis plants may provide more insights. However, due to the large number of possible outcomes when influencing the shikimate pathway's products, probably the most promising way to shed light onto the function of nematode secreted CMs might come from high-throughput methods capable of measuring numerous responses simultaneously (e.g. micro-array combined with metabolomics) in a well-established model plant.

5. Materials and methods

a. Transformation of *Arabidopsis thaliana* Columbia 0

Starting from the *Hs-cm-I* gene contained in pGEM-T vector, the Gateway system was used to clone the gene into pDONR-221 vectors. Expression vectors were created using the Gateway system starting from vectors pK7WGF2 and pK7WG2 to obtain vectors containing *Hs-cm-I* under control of the 35S promoter, with and without an N-terminal GFP fusion. These were subsequently electroporated in *Escherichia coli* strain DH5α. With the aid of triparental mating, the vectors were transferred to *Agrobacterium tumefaciens* strain C58. *Arabidopsis thaliana Columbia 0* plants were subsequently transformed through floral dipping. Selection on kanamycin containing (75 µg/mL) K1 medium of primary transformants and analysis of segregation of following generations allowed us to select 4 and 2 homozygous lines of overexpression and GFP-fusion lines respectively. DNA of the plant lines was extracted using 2% CTAB extraction buffer on plant material grinded in liquid N_2. PCR on the subsequent DNA solution using construct-specific primers confirmed the presence on the DNA level. From the same plants, RNA was extracted using TRIzol (Invitrogen, Carlsbad, USA) following the manufacturer's instructions, and this RNA was subsequently used for first strand cDNA synthesis with the aid of Superscript II (Invitrogen) using oligodT primers. Expression of the constructs was checked by PCR on the first strand cDNA using gene-specific primers. Although on the protein level a strong silencing was observed in plant transformed with the GFP-fusion constructs through decreased GFP signal, we were able to confirm expression of the construct on cDNA level (data not shown). Solid K1 medium contains 0.5x Murashige and Skoog salts with vitamins, 1% sucrose and 0.8% agar set to pH 5.7, which is autoclaved before being poured in petri dishes.

b. Root growth assay

Seeds of the homozygous transformed lines (L4, L5, L29 and L54) were packed in miracloth and sterilized by 2 minutes of incubation in 70% EtOH (50 mL), immediately followed by 12 minutes incubation in NaOCl. The seeds were washed 5 times is sterile tap water and resuspend in 0.1% water agar, which was plated out on K1 medium for germination. After a week, three seedlings of each line were transferred to a single square petri dishes containing K1, with the seedling put in a row at the top of the plate in a random fashion. Four such plates were prepared. The plates were grown under a day/night regime of 16h/8h at 22°C. Starting from day 2 after

7

transfer, the end of the main root tip was marked with a dot each day, until day 8 after transfer. The root growth path was indicated on the petri dish and lengths were measured for each timepoint. The lengths were grouped per line over all plates. The root lengths were analyzed for homoscedasticity by a modified levene test. The normality criterion was not met for all lines at all timepoints. Only 6 days after transfer were all criteria for ANOVA met and with a p-value of 0.342 the null hypothesis (no difference in root length) was not rejected.

c. Infection tests

Arabidopsis plants were germinated on K1 medium and after 1 week transferred to KNOP medium, with 6 plants on one square plate. After two weeks of growth, plants were inoculated with infective *Heterodera schachtii* juveniles stage 2 (J2). In three experiments, 4 drops with 50 nematodes were placed in low melting point agarose (LMP, 0.3% agarose dissolved in bidest), in one experiment 4 drops with 125 nematodes in LMP agarose. For each mutant, 2 up to six plates were used per experiment, depending on the degree of contamination. The J2s were obtained an *in vitro* cyst nematode nematode stock on *Synapis alba* (mustard), of which the cysts were hatched in sterile condition in 3M ZnCl solution, after which the J2s were sterilized in 300 times diluted HAC solution (SSL Healthcare, London, United Kingdom). Two weeks after infection, acid fuchsin staining was performed and nematodes counted. Data per experiment were analyzed with S-Plus (Insightful, Palo Alto, USA): normality (Kolomogorov-Smirnov test) and homoscedasticity (Levene test) were validated. If these conditions were met (in all 4 experiments), ANOVA was performed with pairwise comparisons using the Tuckey method. If the conditions were not met, data were ln(x+1) transformed, and rechecked. If conditions for ANOVA were still not met, Wilcoxon-signed-rank test (a non-parametric test) was used for comparison. KNOP medium is a complex medium suited for *in vitro* nematode infection tests of *Arabidopsis*.

d. DPBA staining

Heterodera schachtii infected *Arabidopsis* plants were removed of KNOP medium and incubated in 0.025% DPBA solution with 0.02% Triton-X for 15 minutes. After 15 minutes of washing with tap water, plants were examined under a epifluorescence microscope (Nikon Corporation, Tokyo, Japan) using the long-pass FITC filter.

e. GUS staining

7

Arabidopsis plants were infected following the same scheme as in the above paragraph. Afterwards, plants were gathered and GUS staining performed: first plants were incubated in 90% aceton at 4°C and after 30 minutes washed with NT-buffer (0.1M Tris, 0.05M NaCl, pH 7.5). After incubation at 37°C for at least 30 minutes in 100 mM ferricyanide, GUS staining proceeds by incubation at 37°C in 100 mM ferricyanide containing 1 mM X-Gluc for at least 2 hours. The staining reaction can be stopped by washing with NT buffer. Incubation in lactic acid to clarify the root is recommended before visualization of the signals. Optionally, DPBA staining can precede GUS staining.

6. Q-PCR

Due to the inherent complexity of the Q-PCR technique, we describe into more detail in this paragraph the selection of internal control genes and analysis of the measured results.

a. Internal control gene selection

Since the goal was to monitor the response of the plant in below and above ground tissues, a validation of six control genes was performed. We decided not to use a rRNA gene as a control gene, since rRNA genes are more abundantly present in a total RNA extraction compared to mRNA (and thus not suited to normalize the latter's expression level). In addition, the used oligo(dT) primers amplify only mRNA. Three control genes were selected based on literature: ubiquitin-binding protein 22 (*ubp-22*) (Hofmann et al. 2007), elongation factor 1 alpha (*ef-1a*), and cyclophilin 5 (*cyp-5*) (Nicot et al. 2005). Three other genes were searched from micro-array data through the Genevestigator website (www.genevestigator.com) (Figure 65). For this, we selected all micro-array experiments of *Arabidopsis thaliana* on which biotic stress was applied. Using an iterative approach, we searched genes which were most up- or downregulated in one of the experiments, but not in the others. For these 17 selected genes selected from the micro-array data, together with the three control genes selected based on the literature, we compared the expression stability in micro-array data of nematode-infection experiments, as well as in all experiments with biotic and abiotic stresses, and in all experiments with different tissues (since we wanted to compare leaves and roots). Expression stability was measured as the relative difference of highest to lowest expression level. From these comparisons, we withheld 6 candidates which showed the greatest stability (Figure 65). EF-1a outperformed all other candidates in "all stresses" micro-array data and "all tissues" (Figure 65). Based on micro-array data of nematode-infection experiments, *agl-42* showed extreme stability.
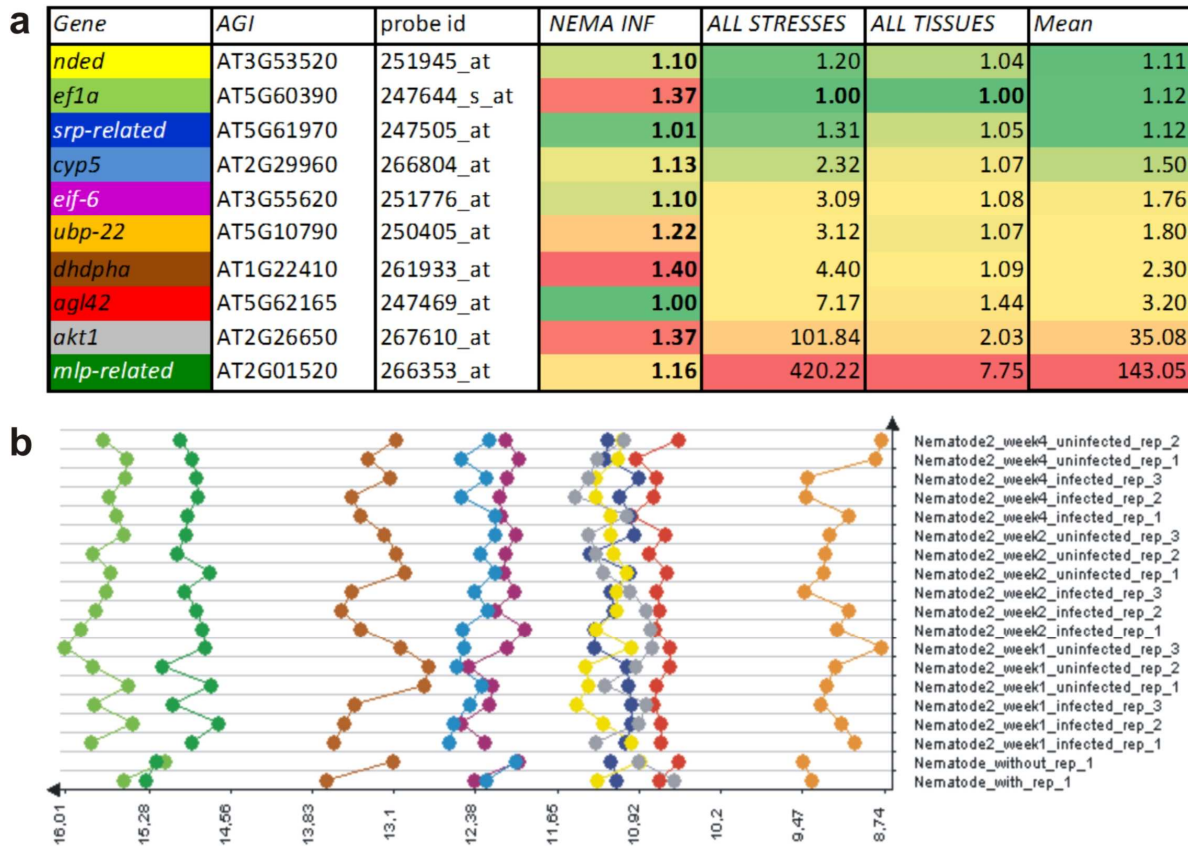
7

a

Gene	AGI	probe id	NEMA INF	ALL STRESSES	ALL TISSUES	Mean
nded	AT3G53520	251945_at	1.10	1.20	1.04	1.11
ef1a	AT5G60390	247644_s_at	1.37	1.00	1.00	1.12
srp-related	AT5G61970	247505_at	1.01	1.31	1.05	1.12
cyp5	AT2G29960	266804_at	1.13	2.32	1.07	1.50
eif-6	AT3G55620	251776_at	1.10	3.09	1.08	1.76
ubp-22	AT5G10790	250405_at	1.22	3.12	1.07	1.80
dhdpha	AT1G22410	261933_at	1.40	4.40	1.09	2.30
agl42	AT5G62165	247469_at	1.00	7.17	1.44	3.20
akt1	AT2G26650	267610_at	1.37	101.84	2.03	35.08
mlp-related	AT2G01520	266353_at	1.16	420.22	7.75	143.05

b

Figure 65: Overview Genevestigator data of selected internal control genes. **a,** overview of different stresses on internal control gene candidates based on micro-array data. Green means less responsive, red means highly responsive, with the value indicating the ratio of the maximum over the minimum expression value measured from micro-array data. "NEMA INF": nematode infection data; "ALL STRESSES": biotic and abiotic stresses data; "ALL TISSUES": specific tissues data; Mean: average value of all three grouping of micro-array data. **b,** relative expression of different internal control genes candidates in response to nematode infection.

Table 17: Primers used for Q-PCR with indication of specificities.

gene	colour	forward primer	Tm* (°C)	Intr**	reverse primer	Tm* (°C)	Intr**	a	b	c
srp-rel.	Orange	TAGCCGAATTTCTTCTCTCTGG	60	no	GCATTTCGTTTCAGAAGCTACG	61.3	no	143	131	−28.1
uxs-1	Green	GAACTTGCAGAGGTGGTTAAGG	60.2	yes	TCAACTGTTCCTTTGCTTTGC	60.4	no	121	350	−6.3
ef-1a	Blue	ATTAAGAGCGTGGACAAGAAGG	59.8	no	CCAAAGAAAACCATAAGAGAGTCG	60.5	no	117	357	−33.6
eif-6	Yellow	GAATGGTCCATCCTCACACC	60.2	yes	CCAATCATTAACGGTCATACCC	60.3	no	131	326	−11.4
ubp22	Red	TTGAAGAAGAAGTGGTGAGAGG	58.6	no	AAGAGTCAGCAATCAGCAAAGG	60.9	no	142	69	−8.2
cyp5	purple	GGACCAGGTGTACTTTCAATGG	60.6	yes	ATCTAACCAGCTTGTTGTCACC	58.3	no	90	282	−8.04

* Tm: melting temperature; ** Intr. sp.: over intron-exon boundary; **a** amplicon size; **b** distance to the poly-A tail; **c** folding energy as determined by Mfold on the complete amplicon
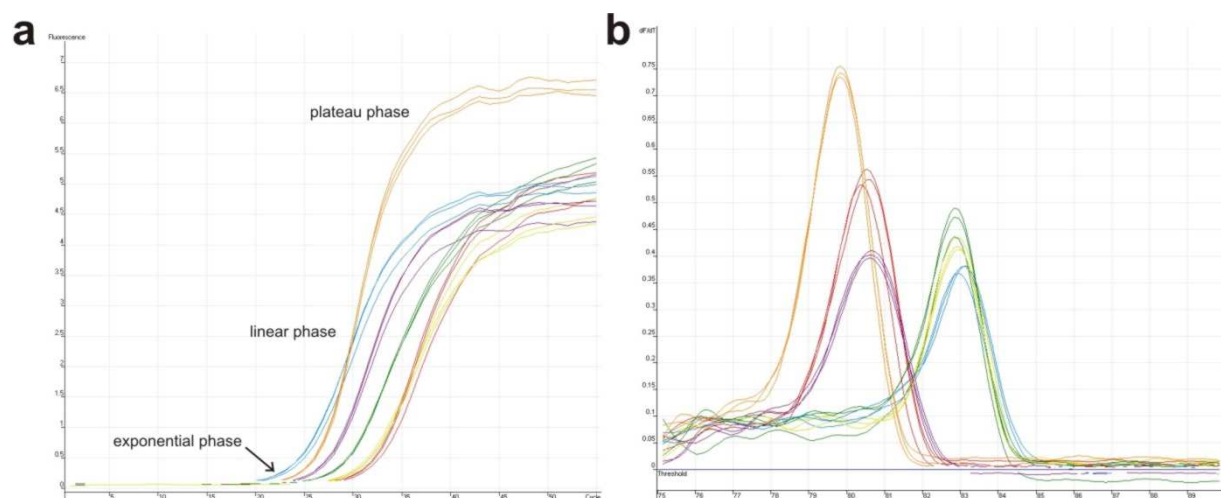
7

Figure 66: **a,** Q-PCR run with six control gene candidates on cDNA of *Arabidopsis Col O*. When the signal increases over the background fluorescence, the curve is in its exponential phase. In the linear phase and plateau phase of the curve, no reliable quantification is possible. **b,** melting curves.

For the six most stable genes (*uxs-1* (=*nded*), *ef-1a*, *srp-related*, *cyp-5*, *eif-6* and *ubp-22*) gene specific primers were developed, taking into account following specifications: the amplicon should be between 60 and 150 bp in length, the melting temperature of both primers should be 60±2°C, the primers should preferably be located over an intron-exon border, or otherwise should span an intron (so to distinguish possible DNA amplification). Furthermore, the amplicon should be located as close a possible at the 3' end of the cDNA and lack any stable secondary structures. Primers are listed in Table 17. A run on a test cDNA sample confirmed the performance of the primers.

b. Sample gathering, cDNA synthesis and PCR.

For evaluating internal control genes, complete roots, leaves, root tips (5 mm of the tip's end), and flowers were gathered from wild type *Arabidopsis* grown *in vitro* by immersion of the complete plants in liquid N_2 and collection of the tissues. RNA was extracted by grinding in liquid nitrogen and subsequent TRIzol extraction was performed (see above). The 1.5 µg RNA was DNAse I (Invitrogen) treated, and subsequently used for cDNA synthesis with an oligo(dT) primer using SuperScript II (Invitrogen) in a volume of 20 µL.

Using a pipetting robot CAS1200 (Corbett Research, Sydney, Australia), 20 µL of the Q-PCR mixture was made, containing 0.5 µL of the first strand cDNA mixture, 4 µM of each primer,

and reaction mixture mastermix from the 'Q-PCR Core kit SYBR-Green No ROX' kit (Eurogentec, Liège, Belgium), which was made following the manufacturer's instructions. The PCR was run in a Rotorgene 3000 machine (Corbett Research). The threshold fluorescence level was manually set to 0.015 for each run. Treshold cycle (Ct) and efficiency values were exported and further processed in Excel (Microsoft, Redmond, USA).
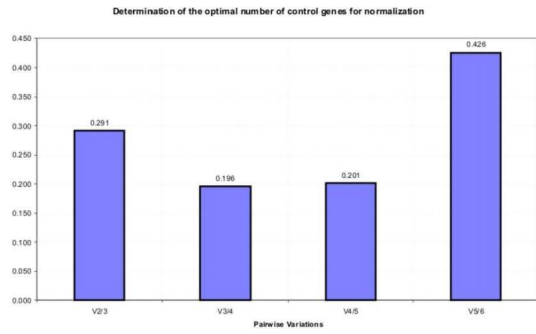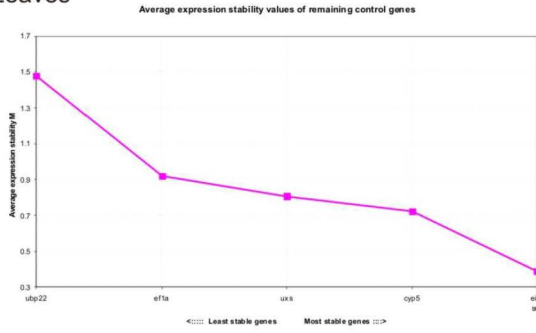
c. Internal control genes.

To establish the best internal control genes, we applied the method developed by Vandesompele and colleagues (2002). The principle is that the ratios of control gene expression in the investigated samples are equal. Therefore, pairwise comparing of all Ct-values over all samples reveals the most stable control gene combination, represented by M-values. In addition to the Ct-values (as originally proposed in the method), we included the efficiency of the PCR reaction, as efficiency between different samples and genes can vary considerably.

The PCR efficiency E (with $1<E<2$) can be determined by analyzing the fluorescence increase during the PCR reaction. When the fluorescence exceeds the background, the fluorescence curve is in the exponential phase of the PCR reaction. From this exponential phase, the efficiency of the PCR reaction can be calculated, by a logarithmic transformation of the data points and determining the highest possible slope of the now linear curve (Schefe et al. 2006). Including this data to transform the Ct-values of the control genes tested on different tissues (Table 18), we used the geNorm program to identify the most stable genes from the different tissues (Vandesompele et al. 2002). geNorm ranks the control genes by stability of expression over the different tissue samples. The optimal strategy with a large heterogeneity of tissues (roots and leaves in our case) is to eliminate the genes with the highest M-value (highest instability over the tissues). In all four tissues, *ubp-22* is the most unstable gene, and will not be used. *srp* is the second most unstable gene in two tissues, while *uxs* and *eif-6* are both the third most unstable gene in all tissues (Figure 67). From this ranking, we decide to use *ef-1a*, *cyp-5* and *eif-6* as internal control genes.

7

d. Measurement of expression levels of genes of interest

The goal is to evaluate the early plant response during cyst nematode infection (*H. schachtii*) in *Arabidopsis thaliana*, in comparison with a non-infected plant. Primers were developed using
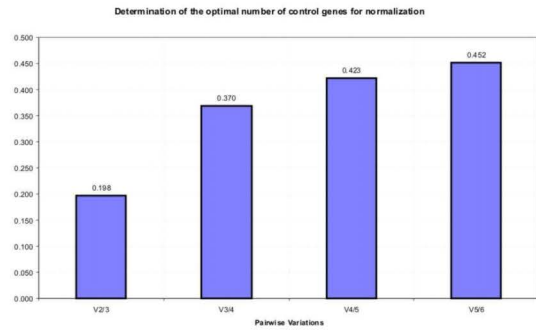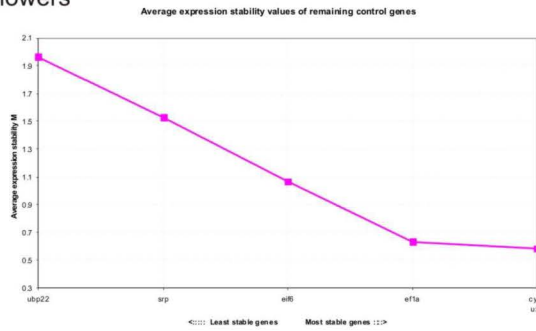
Figure 67: Output of geNorm showing the variability of 6 candidate internal control genes for four different analyzed tissues (a – d). From this analysis, we decided to use *ef-1a*, *cyp-5* and *eif-6* as reference genes in the final Q-PCR experiment.
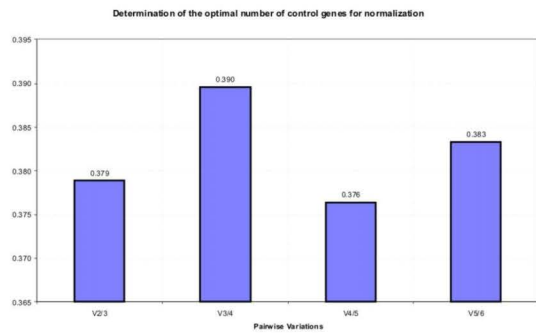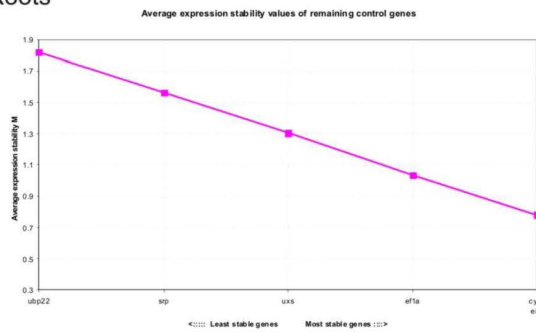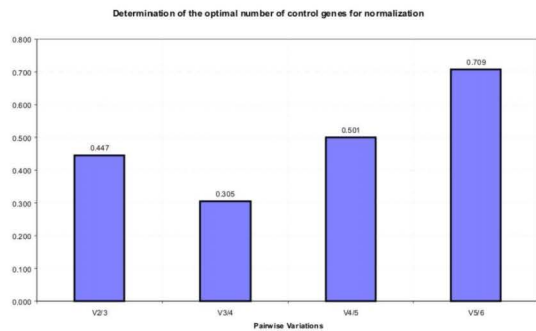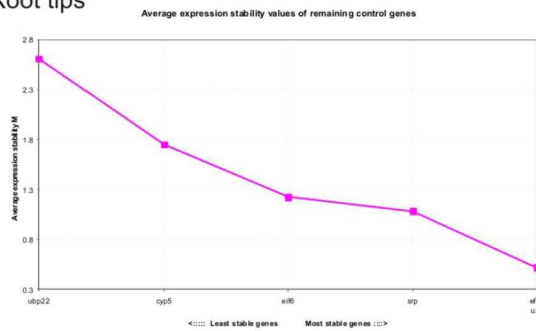
the same specifications as described above, for *pr-1*, for *pdf-1.2* for *chs* and for *pal-1*. The MgCl$_2$ concentration was optimized (3 mM) and the optimum primer concentration determined (0.5 µM), with final reaction efficiencies ranging from 61% to 71% based on the linear regression method (see above). The efficiency was more influenced by primer concentration than by MgCl$_2$ concentration. We infected *in vitro Arabidopsis* plants with infective J2 *Heterodera schachtii* juveniles as described above. 24 hour after infection, we washed the roots with sterile standard pore water, and collected 1 dpi root and leaf tissues. After another 24 hours we collected the 2 dpi root and leaf tissues. The Q-PCR was done on cDNA samples of these tissues, in biological duplicate and technical triplicate. Prior to relative quantitation, we examined the correlation of the efficiency with the samples (SOI) and genes-of-interest (GOI) (Figure 68). The Q-PCR efficiency was stronger correlated with the GOI (and thus the used primers), than with the SOI. ANOVA analysis indicated a significant interaction
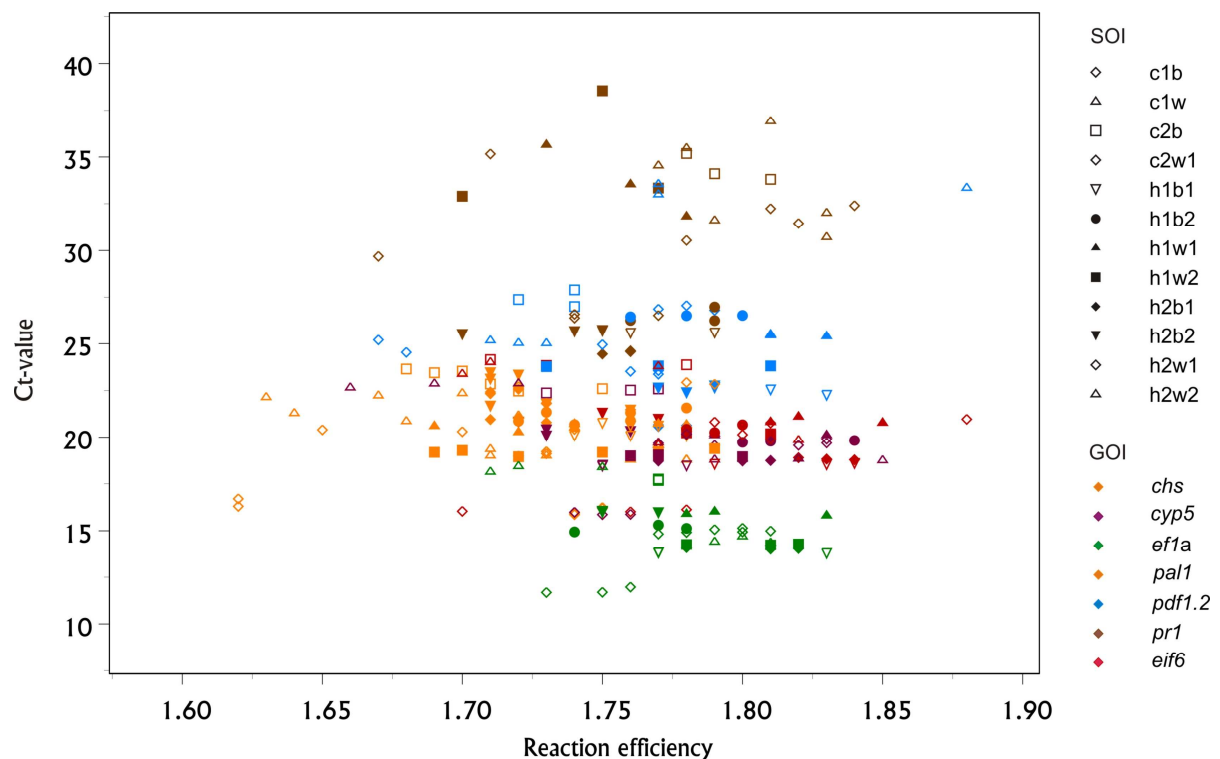

Figure 68: Scatter plot of efficiency values (X-axis) against Ct values (Y-axis), with indication per data point the sample-of-interest (SOI) and gene-of-interest (GOI). A wide variability exists on the efficiency values after optimization, which is correlated to both SOI and GOI. Therefore we decided to approximate the efficiency per SOI and per GOI.

between between SOI and GOI. Therefore, we decided to take the mean efficiency and Ct-values for each sample-primer pair, and use these for further calculations. Schefe and colleagues (2006) worked out a formula to relatively compare expression levels, which includes the Ct value and efficiency of the PCR reaction. Unfortunately, it is based on only one internal control gene.

We have used a reference index consisting of 3 internal control genes. Bearing the recommendations of Vandesompele and colleagues (2002) in mind, we deduced a formula based on Schefe's formula, which can be used for normalization against a 'reference index' of internal control genes (see Box 1). The resulting RIrER (Refence Index relative Expression Ratio)

is:
$$RIrER = \frac{RI(SOI)}{RI(ref)} * E(GOI, SOI)^{-Ct(GOI,SOI)} / E(GOI, ref)^{-Ct(GOI,ref)}$$

with 'ref' being the control sample and the 'Reference Index' RI(x) calculated by:

$$RI(x) = \sqrt[n]{\prod_{i=1}^{n} E_{i,x}^{Ct_{i,x}}}$$

where 'n' is the number of internal control genes, and 'x' the sample index.

Table 18: (page 170-172) Overview of transformation of the Ct values to match geNorm input. Ct values were determined using a threshold of 0.015 in all Q-PCR reactions through the use of RotorGene software 6.0.19 (Corbett Research). Efficiency was estimated as the mean of RotorGene software and LinRegPCR (Ramakers et al. 2003).

Measured Ct-values												
	Flowers2	Leaves3	Roottip1	Roots3	Flowers3	Roottip3	Leaves1	Flowers1	Roots1	Leaves2	Roottip2	Roots2
cyp5	23.18	21.40	27.73	31.42	23.80	30.64	20.50	19.55	19.59	18.80	25.23	20.45
ubp22	33.42	27.94	39.00	37.08	32.62	36.23	27.71	22.26	22.85	20.74	32.08	23.89

ef1	23.30	18.97	36.04	30.59	21.58	29.31	18.79	15.33	16.22	15.39	22.59	17.87
uxs	24.31	21.57	28.59	27.68	23.32	28.98	21.99	21.20	23.57	19.93	25.05	22.96
eif6	29.84	26.67	42.00	40.59	29.29	44.14	26.58	20.67	21.75	19.25	29.06	21.90
srp	27.31	22.72	36.61	35.96	25.98	34.38	23.75	20.32	20.87	18.80	27.46	21.83
Measured efficiency values												
cyp5	1.41	1.33	1.33	1.29	1.39	1.32	1.41	1.67	1.67	1.65	1.58	1.61
ubp22	1.17	1.27	1.19	1.21	1.17	1.16	1.15	1.66	1.61	1.62	1.67	1.59
Ef1a	1.21	1.30	1.20	1.28	1.23	1.20	1.22	1.62	1.58	1.58	1.63	1.57
uxs	1.30	1.33	1.30	1.36	1.34	1.22	1.33	1.53	1.37	1.53	1.57	1.48
eif6	1.20	1.25	1.22	1.26	1.23	1.17	1.23	1.61	1.61	1.59	1.64	1.58
srp	1.25	1.32	1.27	1.24	1.43	1.20	1.31	1.67	1.66	1.69	1.64	1.63
Absolute quantification based on PCR efficiency and Ct values (efficiency^Ct-values)												
cyp5	3106.87	446.73	2699.43	3331.94	2334.95	5072.85	1134.19	22628.52	24413.66	12886.22	106861.36	18072.58
ubp22	174.99	747.64	838.08	1218.31	190.35	247.83	43.37	77070.43	57176.86	22100.46	13099334.69	60827.68
Ef-1a	89.33	138.96	671.71	1772.33	86.81	235.41	43.63	1695.80	1685.32	1169.30	61913.51	3157.43
uxs	594.77	437.98	1687.39	5136.70	951.82	334.18	546.47	8114.17	1748.90	4647.51	79896.42	8488.64
eif6	206.92	357.23	4578.02	10364.52	474.19	1090.86	242.79	19526.26	29450.57	7261.30	1788690.64	22238.98
srp	463.04	547.86	6698.82	2106.58	11012.71	492.79	579.62	33254.89	37181.72	18182.44	752765.76	41448.75

7

Relative quantification to lowest expressed gene (ubp22)											
cyp5											
71.63	10.30	62.24	76.82	53.84	116.96	26.15	521.7	562.9	297.1	2463.9	416.69
ubp22											
4.03	17.24	19.32	28.09	4.39	5.71	1.00	1777.0	1318	509.6	302028	1402.4
Ef-1a											
2.06	3.20	15.49	40.86	2.00	5.43	1.01	39.10	38.86	26.96	1428	72.80
uxs											
13.71	10.10	38.91	118.4	21.95	7.71	12.60	187.09	40.32	107.2	1842	195.72
eif6											
4.77	8.24	105.6	239.0	10.93	25.15	5.60	450.21	679.0	167.4	41241	512.76
srp											
10.68	12.63	154.5	48.57	253.9	11.36	13.36	766.75	857.2	419.2	17356	955.67

Inverse of relative quantification to lowest expressed gene											
cyp5											
0.013	0.097	0.016	0.013	0.019	0.009	0.038	0.002	0.002	0.003	0.000	0.002
ubp22											
0.248	0.058	0.052	0.036	0.228	0.175	1.000	0.001	0.001	0.002	0.000	0.001
ef1a											
0.486	0.312	0.065	0.024	0.500	0.184	0.994	0.026	0.026	0.037	0.001	0.014
uxs											
0.073	0.099	0.026	0.008	0.046	0.130	0.079	0.005	0.025	0.009	0.001	0.005
eif6											
0.210	0.121	0.009	0.004	0.091	0.040	0.179	0.002	0.001	0.006	0.000	0.002
srp											
0.094	0.079	0.006	0.020	0.004	0.088	0.075	0.001	0.001	0.002	0.000	0.001

7. Acknowledgements

Box 1   Deduction of the Reference Index relative Expression Ratio (RIrER)

The RIrER of a gene-of-interest (GOI) in a sample-of-interest (SOI) is a value representing the expression level relative to a set of internal control genes (the reference index RI) and to a reference sample (the calibrator, C). We based the deduction on Vandesompele et al. (2002), who proposed to normalize expression level of a GOI to a geometric mean of expression levels of a reference index, consisting of at least three internal control genes. We combine this with the relative Expression Ratio formula (gene expression's Ct difference formula) suggested by Schefe et al. (2006), which includes the efficiency (E, $1<E<2$, ideally the E-value should be 2) of the PCR reaction to estimate the expression more accurately.

In any RNA sample, the relative expression of a GOI to one reference gene (RG) is the ratio of the initial copy number (R_0) of the GOI to the initial copy number of the RG:

$$\frac{R_0(GOI)}{R_0(RG)}$$

According to Vandesompele et al. (2002), the RG is replaced by a more accurate reference index, consisting of a small number of validated control genes:

$$Expression\ relative\ to\ reference\ index\ \mathrm{R}_{RI} = \frac{R_0(GOI)}{R_0(RI)}$$

with the (virtual) copy number $R_0(RI)$ being the geometric mean of the *n* control genes with index i. At any given cycle p in the exponential phase of the PCR reaction, the 'amount' R of the RI is equal to the geometric mean of the amplicon numbers of the control genes (measured by the fluorescence increase).

$$R_\mathrm{p}(RI) = \sqrt[n]{\prod_{i=1}^{n} R_p(RG_i)}$$

With the basic PCR kinetics, linking the amount of amplicons to the cycle number Ct at which fluorescence is detected above the background signal to the initial copy number:

$$R_{Ct}(RG) = R_0(RG) * E^{Ct}$$

Thus:
$$R_{Ct}(RI) = \sqrt[n]{\prod_{i=1}^{n} R_0(RG_i) * E_i^{Ct_i}}$$

$$\Downarrow$$

$$R_{Ct}(RI) = \sqrt[n]{\prod_{i=1}^{n} R_0(RG_i)} * \sqrt[n]{\prod_{i=1}^{n} E_i^{Ct_i}}$$

$$\Downarrow$$

$$R_0(RI) = \sqrt[n]{\prod_{i=1}^{n} R_0(RG_i)} = \frac{R_{Ct}(RI)}{\sqrt[n]{\prod_{i=1}^{n} E_i^{Ct_i}}} =$$

Expression in one sample SOI relative to RI:

$$R_{RI}(S) = \frac{R_0(GOI,S)}{R_0(RI,S)} = \frac{R_{Ct}(GOI,S)}{R_{Ct}(RI,S)} * \frac{\sqrt[n]{\prod_{i=1}^{n} E_i(S)^{Ct_i(S)}}}{E_{GOI}(S)^{Ct_{GOI}(S)}}$$

with $\frac{R_{Ct}(GOI,S)}{R_{Ct}(RI,S)}$ = constant K, because R_{Ct} is number of copies at detection, irrespective of gene, when primers are properly designed, and amplicon sizes do not differ largely.

7

Deduction of the Reference Index relative Expression Ratio (RIrER) (continued)

To obtain the expression level relative to the control sample C:

$$RIrER = \frac{R_{RI}(SOI)}{R_{RI}(C)} = \frac{K * \frac{\sqrt[n]{\prod_{i=1}^{n} E_i(SOI)^{Ct_i(SOI)}}}{E_{GOI}(SOI)^{Ct_{GOI}(SOI)}}}{K * \frac{\sqrt[n]{\prod_{i=1}^{n} E_i(C)^{Ct_i(C)}}}{E_{GOI}(C)^{Ct_{GOI}(C)}}}$$

$$\Downarrow$$

$$RIrER = \frac{\sqrt[n]{\prod_{i=1}^{n} E_i(SOI)^{Ct_i(SOI)}}}{\sqrt[n]{\prod_{i=1}^{n} E_i(C)^{Ct_i(C)}}} * \frac{E_{GOI}(SOI)^{-Ct_{GOI}(SOI)}}{E_{GOI}(C)^{-Ct_{GOI}(C)}}$$

with the 'Reference Index' RI of sample x: $RI = \sqrt[n]{\prod_{i=1}^{n} E_i(x)^{Ct_i(x)}}$

$$RIrER = \frac{RI(SOI)}{RI(C)} * \frac{E_{GOI}(SOI)^{-Ct_{GOI}(SOI)}}{E_{GOI}(C)^{-Ct_{GOI}(C)}}$$

Practical approach: one calculates the RI value of all samples, and divides this value with the RI value of the control sample, giving the correction factor (CF) for this sample.

$$CF(SOI) = \frac{RI(SOI)}{RI(C)}$$

This factor represents the correction applied to the relative expression of the GOI in the SOI relative to the control sample C.

When ANOVA confirms the lack of correlation between efficiencies of sample and GOIs, efficiencies of GOI can be averaged over all samples, giving a more accurate estimate of the efficiency of amplification of the amplicon (E'). In this case, the formula can be simplified to:

$$RIrER = CF(SOI) * E'_{GOI}{}^{(Ct_{GOI}(C) - Ct_{GOI}(S))}$$

In practice however, we have observed that the efficiency varies over different samples, with some samples having consistently a higher PCR efficiency over other samples.

7

Chapter 8

Perspectives

8

Perspectives

Migratory nematodes are generally thought of as adopting a more primitive way of parasiting plants compared to the sedentary plant-parasitic nematodes. However, to accomplish successful plant parasitism both share some parasitism genes. In this respect, identification of (pioneer) parasitism genes from nematodes of a diverse range of parasitic behaviour, even including animal-parasites, may shed more light on their function. The identification in *Radopholus* of homologues of sedentary nematode parasitism genes, which most likely function within the plant host nucleus, is an exciting finding. Also the identification of an isochorismatase domain containing protein that is shared by *Radopholus similis*, sedentary plant-parasitic nematodes and phytopathogenic fungi, points to shared molecular tools to establish a successful parasitic interaction. Most likely the products of these genes interfere with processes within the plant cell. Therefore they must act in the short salivation period observed in migratory nematodes, preceding the ingestion of cytoplasm. Indeed, detailed observation of parasitic behaviour of another migratory nematode (*Pratylenchus*) revealed that this nematode frequently revisited, up to three times, the same plant root cell to feed, before the cell eventually died (Zunke 1990). The components contained in the salivation of the migratory nematodes are as yet unknown. Since these components should function in such a small time frame, it could be argued that their role is to counter fast plant responses to nematode infection. The first plant defence responses, known to build up rapidly after infections, are reactive oxygen species and SA accumulation. It is tempting to speculate that the salivation of migratory nematodes mainly functions to prevent these responses so to keep the plant cell alive. Sedentary nematodes evolved from migratory species, so it is reasonable to assume that sedentary nematodes would have (and need) similar components (Holterman et al. 2009). Hence comparison between salivation of sedentary nematodes and migratory nematodes may lead to a better understanding of the factors crucial for establishing the sedentary nematode feeding site. Specific for migratory PPN research, the protein based approach for parasitism gene discovery in migratory PPNs has a big potential as migratory nematodes are more easily reared in large numbers compared to sedentary and salivation of parasitic stages can be readily induced and collected, at least for *Radopholus*. Identification of the secretion components through screening of cDNA libraries or ESTs and subsequently establishing expression in the gland cells, should allow constructing of extensive collections of parasitism genes. However, the *in*

8

situ hybridisation step often limits the speed of parasitism gene discovery: many nematodes have to be collected, during the protocol many nematodes are lost, and the overall sensitivity of the technique is very low. It would be beneficial if this approach could also be scaled up and improved. For example, constructing cDNA pools of micro-aspirated contents of gland cells, and performing large PCR screens on these pools for *in silico* selected candidates is one of the possible ways to accelerate parasitism gene discovery in migratory nematodes.

Assessing the function of parasitism genes, which is currently receiving increased attention in sedentary nematode research, can be extended to migratory nematodes only with continuous effort to further identify additional parasitism genes. Most research on parasitism gene function has focused on analysis of a single gene, as we have done for nematode chorismate mutase (CM). However, this approach may prove to reach its limits, since parasitism genes are expected to work in a concerted way, with products of parasitism genes functioning complementary to others in the plant. Indeed, if secreted nematode CM has an important role in establishing the nematode feeding site through manipulation of key pathways in the plant, this did not become clear by overexpressing this gene in *Arabidopsis thaliana*. Maybe the endogenous levels of chorismate mutase mask the additional action of the nematode CM. Or maybe nematode CM requires action of additional parasitism genes before reaching full functionality. Nevertheless, unravelling the role of parasitism genes can only succeed with a thorough understanding of processes within the plant tissues responding to nematode infection, starting from the initial contact with the plant, over the migratory phase and the feeding process, to – for sedentary nematodes at least – the establishment of the nematode feeding site. Our data show that during the migratory phase of cyst nematodes, a very strong JA response is being elicited in the plant roots, which eventually dies away. It will be interesting to analyze which factors cause the suppression of this response. For migratory nematodes such as *Radopholus similis*, no molecular characterization of plant responses has yet been performed. Measuring the plant root cell responses in the few minutes of salivation before feeding of the migratory nematodes may prove to be highly challenging. Nevertheless, common themes emerging in plant-nematode interactions, both at the nematode as on the plant side, can provide a basis for identification of nematode control targets, effective in a broad range of parasitic nematodes.

Unexpectedly, *Radopholus similis* seems to have many eccentricities as is clear from the results of this thesis. And as a consequence, an equal amount of questions arise. Perhaps these

8

aberrant features are reflected in the phylogenetic position of *Radopholus similis* as an isolated monophyletic sister clade of the Hoplolaimidae. Firstly, the mitochondrial genome contains a unique codon change, until now identified solely in this species. The reason for this seems to be the unidirectional transcription in combination with the very high AT content. However, the question now shifts to the reason for this high AT content, for which still no adequate answer is given. Since the unidirectional mitochondrial transcription is unique for Tylenchids within the Ecdysozoa, unraveling mitochondrial transcription mechanisms may even provide a means to develop parasitic nematode specific control measures.

A second remarkable feature is the identification of an endosymbiotic *Wolbachia* species residing in the ovaria of *R. similis*. This finding leads to exciting new research topics. It will be interesting to find out if this endosymbiont has a same mutualistic relationship with *R. similis* as reported for filarial nematodes. However, the basis of this relationship is still unknown. We do not know what the nematode gives and gets from *Wolbachia* and vice versa. Some facts may provide a clue though. *Wolbachia* reside mostly in the ovaria, and the ovaria of nematodes contain most mitochondria of all tissues. The mitochondria contained in oocytes of nematodes suffice to support development through the early larval stages. With sexual maturation, a 30-fold increase in total mtDNA copy number is observed in *Caenorhabditis elegans*, mostly associated with the production of oocytes (Tsang et al. 2003). Furthermore, mitochondria of some nematode species have the unique capability to function in the abscence of oxygen, due to the hypoxic environment they live in and the lack of a circulatory system, as investigated in many filarial nematodes (Iwata et al. 2008). However, during this anaerobic energy production (through fumarate reduction), acetate and propionate are formed (Tielens et al. 2002). Perhaps the colocalisation of *Wolbachia* with mitochondria in the ovaria is not coincidence: maybe *Wolbachia* enables the filarial nematodes, and *Radopholus*, to deal more easily with the end products of these mitochondrial processes. If *Wolbachia* is indispensable for *Radopholus* survival, targeting this endosymbiont may prove to be a valuable approach for *R. similis* control. Recently, attempts to identify novel drug targets in the genome of the *Wolbachia* endosymbiont from *Brugia malayi* have been undertaken using bioinformatic approaches (Foster et al. 2005). We are still far away from such an approach for *R. similis*, but with the increased use of the novel 'next generation' sequencing platforms, this may arrive sooner than expected.

8

Indeed, the current molecular research is increasingly moving into the direction of high-throughput techniques. However, with the increased use of high-throughput techniques such as the 'next generation' sequencing platforms and the availability of numerous micro-arrays, astronomical amounts of data will be generated. As a consequence, bioinformatics and computational biology are pushed to the limits to deal with and to analyse these data. The NEXT website aims to provide a convenient tool to explore the ever increasing amounts of EST data from nematodes, as it has been shown (by this thesis and others) that careful comparison of many different data sources may aid in identifying most interesting research topics. However, it takes time and a lot of effort to dig into the increasing pile of data to eventually find some nuggets. Nevertheless, the strength of EST data lies with the amount of data generated, and the 'next generation' sequencing techniques may ultimately prove to be of equal (or even more) value for transcriptome analysis as for genome sequencing (Weber et al. 2007). For example, the power to quantify gene expression with the aid of EST data increases with increasing data, providing the construction of cDNA libraries is accurately described. However, results obtained through bioinformatics can never stand on their own, and biological relevance needs to be confirmed through experimental procedures. The crucial step herein is the transfer of results between both research fields – as often the 'not discovered here' syndrome shows its face. Ideally, a scientist in molecular biotechnology should master both fields to an adequate extent to fully profit of the synergy between both fields. And this certainly holds true for the field of parasitic nematology. Due to the microscopic scale and the intrinsic complexity of the parasitic interaction, research in parasitic nematodes has profited the last years hugely of recent advances in molecular techniques. And undoubtedly, this will continue to be the case in the years to come. But one time or another, the scientist has to put aside his scientific appreciation for the fascinating interaction of the parasitic nematode with his host, and start concentrating on means to effectively combat these pests. But 'killing your darlings', as some would describe it, is apparently not easy.

8

Summary
Samenvatting

"The story so far: In the beginning the Universe was created. This has made a lot of people very angry and has been widely regarded as a bad move."
Douglas Adams, The Restaurant at the End of the Universe.

Summary

Nematodes are small roundworms, inhabiting numerous diverse ecological niches. Many of them parasitize plants in various ways. The migratory plant-parasites migrate through plant roots by feeding on and destroying plant root cells, causing substantial damage to the plant and hampering its development.

The first thorough molecular analysis of the migratory plant-parasitic nematode *Radopholus similis* is reported in chapter 2. For this, novel **expressed sequence tags** (ESTs) representing ca. 16% of the total gene content were generated from a mixed stage population. One third of them are novel sequences, which have a shorter mean length and are predicted to contain a large fraction of non-protein-coding sequences. The remainder had homology to sequences of a wide range of species but none of them had known *trans*-spliced leaders as common in other nematodes. The most striking discovery was tags corresponding to a *Wolbachia*-like endosymbiont, which was subsequently located in the ovaria of *R. similis*. Numerous tags corresponding to **parasitism genes** with potential roles in, amongst other things, host localisation, detoxification, cell wall modification, and even putative host transcriptional reprogramming were identified.

Different **transthyretin-like genes** were identified in the ESTs, which are part of large nematode-specific family and some members are described in chapter 3. All stages except developing embryos express the four analyzed genes. One of them is expressed in the **ventral nerve cord**, and another shows expression in tissues surrounding the vulva. All proteins of this family are predicted to be secreted. Predicted secondary structure shows resemblance to the structure of transthyretins and transthyretin-related proteins, suggestive of a similar **binding capacity with a yet unknown ligand**. Analysis of nematode ESTs shows that the TTL domain is more prevalent in parasitic stages of parasitic nematodes, although the functional link to parasitism remains unclear.

Chapter 4 presents the annotation of the complete **mitochondrial (mt) genome** of *R. similis*. Remarkably, a huge fraction (ca. 8%) of the EST data was derived from mt 12S rRNA. The mt genome, the largest of all Chromadorean nematodes known to date, has the expected gene content but shows many aberrant features such as: a considerably smaller 16S rRNA with reduced structures, two large repeat regions, the **lack of stop codons** on many genes and a **unique codon reassignment** UAA:Stop to UAA:Tyrosine. This latter phenomenon is seemingly the result of the unidirectional transcription in combination with the AT pressure in the mt genome, leading to extinction of nearly all NNC codons, which forced the amino acid Tyrosine to expand its codon family. The aberrant mt genome features

could be related to this codon reassignment but results on expected additional changes of the transcription/translation process are ambiguous.

The **NEXT database** is described in chapter 5. This tool extends the usefulness of **nematode EST data** by extracting and storing temporal and spatial information of each library available in public databases. The database is accessible through a website. Due to normalisation and uncontrolled biases in cDNA library construction the possibility to absolutely quantifying gene expression (in which redundancy of tags is related to expression level of the corresponding gene) is lost to a large extent. But analysis of **reference genes** reveals that relative expression levels are roughly preserved in most libraries and even in the case of combination of libraries. A novel approach to normalize heterogeneous EST libraries is implemented in the tool.

In addition to the research on *Radopholus similis* and EST data, the sixth and last chapter reports on plant responses in reaction to early **cyst nematode infection**. Cyst nematodes are sedentary plant-parasitic nematodes which, after a migratory phase through the plant root, become sedentary for the remainder of their life. In the sedentary phase, they select one root cell to transform with the aid of pharyngeal gland secretions into a nematode feeding site, which develops by cell wall dissolution and incorporation of increasing numbers of root cells. In regularly intervals, the nematode withdraws nutrients from this feeding site. During the migratory phase, a **strong JA induction** was observed in roots, which eventually vanished after the nematode became sedentary. From that moment on, a strong SA response in the shoots was observed. Cyst nematodes express a **chorismate mutase** (CM) enzyme in their pharyngeal gland cells, which is presumably secreted into a plant root cell to assist in the feeding site formation. However, the precise details of its role are currently unknown. *Arabidopsis* overexpressing cyst nematode CM did not show any phenotypic effects, did not show an altered root growth and did not show an altered flavonoid content. Detailed investigations showed that **flavonoids** were not upregulated in the nematode feeding site. Similarly, several flavonoid mutants did not show significant altered susceptibility towards cyst nematode infection. These results render the proposed roles of nematode CM in altering flavonoid and corresponding auxin levels highly unlikely.

By combination of thorough analysis of molecular data and diverse experimental techniques, both on the nematode and the plant side of the interaction, this thesis contributes to the efforts in expanding the knowledge on processes of the plant-parasitic nematodes and their interface with their host plant.

Samenvatting

Nematoden of rondwormen komen voor als talrijke soorten over de hele wereld. De meeste zijn microscopisch klein, en te vinden in zeer diverse ecologische niches. Een groot aantal soorten parasiteren planten, elk met hun eigen methode. Zo leven migratorische plantenparasitaire nematoden hoofdzakelijk in de wortels van planten, waarin ze wortelcellen aansteken met een speciale naaldachtige structuur om zich te voeden. De wortelcellen sterven uiteindelijk af en de nematode beweegt zich verder door de wortel. Na verloop van tijd ontstaan grote holtes in de plantenwortel met nefaste gevolgen voor de plant en zijn ontwikkeling.

In het tweede hoofdstuk van deze thesis wordt de eerste diepgaande analyse van moleculaire data van de migratorische plantenparasitaire nematode *Radopholus similis* (klasse Chromadorea) beschreven. Nieuwe sequentiedata ("**expressed sequence tags**" of ESTs) werd gegenereerd van transcriptioneel actieve genen, die naar schatting ongeveer 16% van het totaal aantal genen vertegenwoordigen. Eén derde blijken uit nieuwe sequenties te bestaan zonder homologe sequenties in alle tot nu toe onderzochte soorten. Opmerkelijk genoeg komt een groot deel van hen overeen met niet-eiwit-coderende sequenties. Voor de rest van de sequenties zijn wel homologen gevonden, maar geen van de sequenties bevatten *trans*-gesplicete leidersequenties, nochtans een frequent gerapporteerd verschijnsel in nematoden. De meest intrigerende ontdekking zijn sequenties die afkomstig blijken van *Wolbachia*. Het is de eerste keer dat deze endosymbiontische bacterie, die zich ophoudt in de ovaria van *R. similis*, gedetecteerd is in een plantenparasitaire nematode. Verder zijn verscheidene **parasitisme-genen** geïdentificeerd met diverse functies in plant lokalisatie, detoxificatie, plantencelwand modificatie en zelfs mogelijke transcriptionele herprogrammering van de gastheercel.

De EST data bevat verschillende '**transthyretin-gelijkende' genen**, die deel uitmaken van een grote nematode-specifieke genfamilie en waarvan enkele besproken worden in hoofdstuk 3. Alle stadia behalve ontwikkelende embryo's brengen de vier onderzochte genen tot expressie. Eén van hen komt tot expressie in de **ventrale zenuwstreng**, en een andere in de weefsels gelokaliseerd rond de vulva. Alle eiwitten worden hoogst waarschijnlijk uitgescheiden. De voorspelde secundaire structuur van de eiwitten gelijkt op de structuur van transthyretin and transthyretin-gerelateerde eiwitten, suggestief voor een gelijkaardige **bindingscapaciteit** van een nog onbekend ligand. De functie van deze familie blijft echter onbekend, maar het corresponderende eiwitmotief blijkt meer voor te komen in de sequentiedata van parasitaire stadia, een resultaat dat de mogelijke link van deze genfamilie met parasitisme bevestigt.

Het vierde hoofdstuk rapporteert de annotatie van het volledige **mitochondriale genoom** van *R. similis*, dat werd gesequeneerd met de hulp van EST data. Opmerkelijk genoeg blijkt een disproportioneel groot deel van de ESTs (ca. 8%) afkomstig te zijn van het mt 12S rRNA. Tot nu toe is dit mt genoom het grootste van alle Chromadorea nematoden. Het codeert voor de verwachte genen, hoewel ook vele merkwaardigheden in het genoom terug te vinden zijn, zoals: de sterk verkorte 16S rRNA, de twee grote regios met sequentieherhalingen, de afwezigheid van stop codons op de meeste genen, en een **unieke genetische code** waarin het codon UAA veranderd is van Stop naar Tyrosine. Deze unieke verandering is van alle eukaryoten tot nu toe enkel gevonden in het genus *Radopholus*, en blijkt het resultaat van de unidirectionele transcriptie in combinatie met de aanrijking aan AT nucleotiden kenmerkend voor mt genomen. Dit heeft uiteindelijk geleid tot de verdwijning van NNC codons, en de daaropvolgende expansie van de codonfamilie van Tyrosine. De afwijkende kenmerken in het mt genoom kunnen gerelateerd zijn aan deze codonverandering, hoewel verwachte bijkomende mutaties aan het transcriptie/translatie systeem niet éénduidig aan te wijzen zijn.

De **NEXT** databank wordt besproken in hoofstuk 5. Deze *tool* breidt de bruikbaarheid van nematode EST data uit door het extraheren en beschikbaar stellen van weefsel- en stadia-specifieke informatie van publiek beschikbare cDNA banken. De databank kan bevraagd worden via een uitgebreide interface beschikbaar op het internet. Door normalisatiepraktijken en ongecontroleerde aberraties gedurende het construeren van cDNA banken is de mogelijkheid tot absolute kwantificatie van genexpressie (gebaseerd op de redundantie van sequenties in cDNA banken) grotendeels verloren gegaan. Desondanks toont de analyse van **referentiegenen** aan dat relatieve genexpressie wel behouden blijft in de meeste banken voor de meeste referentiegenen, zelfs bij het combineren van verschillende banken. Gebaseerd op deze genen is een nieuwe methode in de *tool* geïntegreerd om heterogene EST banken te vergelijken.

Naast het onderzoek op *Radopholus similis* en EST data, handelt het zesde en laatste deel over reacties van de plant op de vroege stadia van cystennematode infectie. Cystenenmatoden zijn **sedentaire plantenparasitaire nematoden**, die na een migratorische fase door het plantenwortelsysteem stoppen met migreren en sedentair worden voor de rest van hun levenscyclus. In deze fase selecteren ze een wortelcel die onder impuls van nematodesecreties transformeert tot een nematodevoedingsplaats. Deze ontwikkelt zich door het oplossen van de celwanden van naburige cellen en het incorporeren van deze cellen in een aangroeiend syncytium. Op regelmatige tijdstippen onttrekt de nematode nutriënten van deze voedingsplaats. Gedurende de migratorische fase in de levenscyclus van deze nematode werd een sterke jasmijnzuurrespons opgemeten enkel in het wortelsysteem van een geïnfecteerde zandraket (*Arabidopsis thaliana*), een respons die echter volledig verdwenen was bij het inzetten van de sedentaire fase. Van dan af werd een verhoogde salicylzuurrespons opgemeten enkel in de bovengrondse

plantdelen. Eén van de eiwitten in de secreties van de faryngeale klieren van de cystennematoden is het enzym chorismate mutase (CM). Dit eiwit speelt een rol gedurende de fase van de syncytiumvorming, hoewel de precieze functie nog onbepaald is. We hebben getracht meer inzicht te verwerven in de functie van CM gedurende het infectieproces. Zandraket die nematode CM tot overexpressie brengt, vertoont geen veranderd fenotype of flavonoïedinhoud. Gedetailleerde observaties tonen aan dat flavonoïeden niet accumuleren in syncytia van een paar dagen oud. En testen op flavonoïedmutanten van *Arabidopsis* vertonen geen veranderde gevoeligheid voor cystennematode-infectie. Algemeen kunnen we stellen dat de voorgestelde rol van CM in het beïnvloeden van flavonoïedniveaus en de accumulatie van het plantenhormoon auxine zeer onwaarschijnlijk is.

Door combinatie van diepgaande analyse van moleculaire data en diverse experimenten, zowel aan nematode- als aan plantzijde, dragen de resultaten van dit proefschrift bij tot beter begrip van de processen in plantenparasitaire nematoden en de interactie met hun waardplanten.

References

Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin 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 (2008) Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nat Biotech 26:909-915

Abascal F, Posada D, Zardoya R (2007) MtArt: A new model of amino acid replacement for Arthropoda. Mol Biol Evol 24:1-5

Agrios GN (2005) Plant Pathology. 5th edition. Academic Press, San Diego, California, pp. 952

Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MG, Garey JR, Raff RA, Lake J (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489 - 493

Alfonzo JD (2009) Editing of tRNA for structure and function. In: Goringer H (Ed.), RNA editing Volume 20. Springer, Berlin/Heidelberg, Germany, pp. 33-50

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Andersson SGE, Kurland CG (1995) Genomic evolution drives the evolution of the translation system. Biochem Cell Biol 73:775-787

Armstrong MR, Blok VC, Phillips MS (2000) A multipartite mitochondrial genome in the potato cyst nematode Globodera pallida. Genetics 154:181-192

Arunkumar KP, Nagaraju J (2006) Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. PLoS ONE 1:e110

Atkinson H, Grimwood S, Johnston K, Green J (2004) Prototype demonstration of transgenic resistance to the nematode Radopholus similis conferred on banana by a cystatin. Transgenic Res 13:135-142

Audic S, Claverie J-M (1997) The significance of digital gene expression profiles. Genome Res 7:986-995

Azevedo JL, Hyman BC (1993) Molecular characterization of lengthy mitochondrial DNA duplications from the parasitic nematode Romanomermis culicivorax. Genetics 133:933-942

Bakhetia M (2005) RNA interference of dual oxidase in the plant nematode Meloidogyne incognita. MPMI 18:1099-1106

Bakhetia M, Charlton WL, Urwin PE, McPherson MJ, Atkinson HJ (2005) RNA interference and plant parasitic nematodes. Trends Plant Sci 10:362-367

Bakhetia M, Urwin PE, Atkinson HJ (2007) qPCR analysis and RNAi define pharyngeal gland cell-expressed genes of Heterodera glycines required for initial interactions with the host. MPMI 20:306-312

Bakhetia M, Urwin PE, Atkinson HJ (2008) Characterisation by RNAi of pioneer genes expressed in the dorsal pharyngeal gland cell of Heterodera glycines and the effects of combinatorial RNAi. Int J Parasitol 38:1589-1597

Baldridge GD, O'Neill NR, Samac DA (1998) Alfalfa (Medicago sativa L.) resistance to the root-lesion nematode, Pratylenchus penetrans: defense-response gene mRNA and isoflavonoid phytoalexin levels in roots. Plant Mol Biol 38:999-1010

Bauer M, Klau G, Reinert K (2007) Accurate multiple sequence-structure alignment of RNA sequences using combinatorial optimization. BMC Bioinformatics 8:271-271

Beier H, Grimm H (2001) Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res 29:4767-4782

Bekal S, Niblack TL, Lambert KN (2003) A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. MPMI 16:439-446

Bellafiore S, Shen Z, Rosso MN, Abad P, Shih P, Briggs SP (2008) Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. PLoS Pathog 4:e1000192

Bender A, Hajieva P, Moosmann B (2008) Adaptive antioxidant methionine accumulation in respiratory chain complexes explains the use of a deviant genetic code in mitochondria. Proc Natl Acad Sci U S A 105:16496-16501

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37:766-770

Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RHA, Cuppen E (2005) Phylogenetic shadowing and computational identification of human microRNA Genes. Cell 120:21-24

Bert W, Leliaert F, Vierstraete AR, Vanfleteren JR, Borgonie G (2008) Molecular phylogeny of the Tylenchina and evolution of the female gonoduct (Nematoda: Rhabditida). Mol Phylogenet Evol 48:728-744

Bessho Y, Ohama T, Osawa S (1992) Planarian mitochondria II. The unique genetic code as deduced from cytochrome c oxidase subunit I gene sequences. J Mol Evol 34:331-335

Bhattarai KK, Xie Q-G, Mantelin S, Bishnoi U, Girke T, Navarre DA, Kaloshian I (2008) Tomato susceptibility to root-knot nematodes requires an intact jasmonic acid signaling pathway. MPMI 21:1205-1214

Bieri T, Blasiar D, Ozersky P, Antoshechkin I, Bastiani C, Canaran P, Chan J, Chen N, Chen WJ, Davis P, Fiedler TJ, Girard L, Han M, Harris TW, Kishore R, Lee R, McKay S, Muller HM, Nakamura C, Petcherski A, Rangarajan A, Rogers A, Schindelman G, Schwarz EM, Spooner W, Tuli MA, Auken KV, Wang D, Wang X, Williams G, Durbin R, Stein LD, Sternberg PW, Spieth J (2007) WormBase: new content and better access. Nucleic Acids Res 35:D506-D510

Binder H, Preibisch S (2005) Specific and nonspecific hybridization of oligonucleotide probes on microarrays. Biophys J 89:337-352

Bird DM, Opperman CH (2009) The secret(ion) life of worms. Genome Biol 10:205

Bird DM, Riddle DL (1994) A genetic nomenclature for parasitic nematodes. J Nematol 26:138-143

Blake CC, Geisow MJ, Oatley SJ, Rérat B, Rérat C (1978) Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. J Mol Biol 121:339-356

Blaxter ML, Liu L (1996) Nematode spliced leaders – Ubiquity, evolution and utility. Int J Parasitol 26:1025-1033

Blaxter ML (2003) Nematoda: genes, genomes and the evolution of parasitism. Adv Parasitol 54:101-195

Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, Vida JT, Thomas WK (1998) A molecular evolutionary framework for the phylum Nematoda. Nature 392:71-75

Blaxter ML, Guiliano DB, Scott AL, Williams SA (1997) A unified nomenclature for filarial genes. Parasitol Today 13:416-417

Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK (2002) A global analysis of *Caenorhabditis elegans* operons. Nature 417:851-854

Bobrowicz AJ, Lightowlers RN, Chrzanowska-Lightowlers Z (2008) Polyadenylation and degradation of mRNA in mammalian mitochondria: a missing link? Biochem Soc Trans 36:517-519

Boguski MS, Lowe TMJ, Tolstoshev CM (1993) dbEST - database for "expressed sequence tags". Nat Genet 4:332-333

Boore JL, Brown WM (2000) Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol Biol Evol 17:87-106

Bostock RM (2005) Signal crosstalk and induced resistance: straddling the line between cost and benefit. Annu Rev Phytopathol 43:545-580

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotech 18:630-634

Brown AC, Kai K, May ME, Brown DC, Roopenian DC (2004) ExQuest, a novel method for displaying quantitative gene expression from ESTs. Genomics 83:528-539

Brown DE, Rashotte AM, Murphy AS, Normanly J, Tague BW, Peer WA, Taiz L, Muday GK (2001) Flavonoids act as negative regulators of auxin transport *in vivo* in *Arabidopsis*. Plant Physiol 126:524-535

Brunet S, Hoste H (2006) Monomers of condensed tannins affect the larval exsheathment of parasitic nematodes of ruminants. J Agric Food Chem 54:7481-7487

C. elegans Genome Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science, 282:2012-2018

Campanella J, Bitincka L, Smalley J (2003). MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4: 29

Calhoun DH, Bonner CA, Gu W, Xie G, Jones JT (2001) The emerging periplasm-localized subclass of AroQ chorismate mutases, exemplified by those from *Salmonella thyphimurium* and *Pseudomonas aeruginosa*. Genome Biol 2:0030.1-0030.16

Carullo M, Xia X (2008) An extensive study of mutation and selection on the wobble nucleotide in tRNA anticodons in fungal mitochondrial genomes. J Mol Evol 66:484-493

Catinot J, Buchala A, Abou-Mansour E, Métraux JP (2008) Salicylic acid production in response to biotic and abiotic stress depends on isochorismate in *Nicotiana benthamiana*. FEBS Letters 582:473-478

Chen J, Agrawal V, Rattray M, West MAL, St Clair DA, Michelmore RW, Coughlan SJ, Meyers BC (2007) A comparison of microarray and MPSS technology platforms for expression analysis of *Arabidopsis*. BMC Genomics 8:414-414

Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH (2004) Codon usage between genomes is constrained by genome-wide mutational processes. Proc Natl Acad Sci U S A 101: 3480-3485

Chen W-H, Wang X-X, Lin W, He X-W, Wu Z-Q, Lin Y, Hu S-N, Wang X-N (2006) Analysis of 10,000 ESTs from lymphocytes of the cynomolgus monkey to improve our understanding of its immune system. BMC Genomics 7:82

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497-3500

Chinnasri B, Sipes BS, Schmitt DP (2006) Effects of inducers of systemic acquired resistance on reproduction of *Meloidoygne javanica* and *Rotylenchulus reniformis* in pinapple. J Nematol 38:319-325

Chrzanowska-Lightowlers ZMA, Temperley RJ, Smith PM, Seneca SH, Lightowlers RN (2004) Functional polypeptides can be synthesized from human mitochondrial transcripts lacking termination codons. Biochem J 377:725-725

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676

Cooper WR, Jia L, Goggin L (2005) Effects of jasmonate-induced defenses on root-knot nematode infection of resistant and susceptible tomato cultivars. Journal of chemical ecology 31:1953-1965

Cutter AD, Wasmuth JD, Blaxter ML (2006) The evolution of biased codon and amino acid usage in nematode genomes. Mol Biol Evol 23:2303-2315

Dams E, Hendriks L, Van de Peer Y, Neefs JM, Smits G, Vandenbempt I, De Wachter R (1988) Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res 16 Suppl:r87-173

Davis EL, Hussey RS, Baum TJ, Bakker J, Schots A (2000) Nematode parasitism genes. Ann Rev Phytopathol 38:365-396

De Boer JM, McDermott JP, Davis EL, Hussey RS, Popeijus H, Smant G, Baum TJ (2002a) Cloning of a putative pectate lyase gene expressed in the subventral esophageal glands of *Heterodera glycines*. J Nematol 34:9-11

De Boer JM, McDermott JP, Wang XH, Maier T, Qu F, Hussey RS, Davis EL, Baum TJ (2002b) The use of DNA microarrays for the developmental expression analysis of cDNAs from the oesophageal gland cell region of *Heterodera glycines*. Mol Plant Pathol 3:261-270

De Boer JM, Overmars H, Pomp HR, Davis EL, Zilverentant JF, Goverse A, Smant G, Stokkermans J, Hussey RS, Gommers FJ, Bakker J, Schots A (1996a) Production an characterisation of monoclonal antibodies to antigens from second stage juveniles of the potato cyst nematode *Globodera rostochiensis*. Fundamental Applied Nematology 19:545-554

De Boer JM, Smant G, Goverse A, Davis EL, Overmars H, Pomp HR, van Gent-Pelzer M, Zilverentant JF, Stokkermans J, Hussey RS, Gommers FJ, Bakker J, Schots A (1996b) Secretory granule proteins from the subventral esophageal glands of the potato cyst nematode identified by monoclonal antibodies to a protein fraction from second-stage juveniles. MPMI 9:39-46

De Meutter J, Tytgat T, Witters E, Gheysen G, Van Onckelen H, Gheysen G (2003) Identification of cytokinins produced by the plant parasitic nematodes *Heterodera schachtii* and *Meloidogyne incognita*. Mol Plant Pathol 4:271-277

De Rijk P, De Wachter R (1997) RnaViz, a program for the visualisation of RNA secondary structure. Nucleic Acids Res 25:4679-4684

De Rijk P, Robbrecht E, de Hoog S, Caers A, Van de Peer Y, De Wachter R (1999) Database on the structure of large subunit ribosomal RNA. Nucleic Acids Res 27:174-178

Decraemer W, Hunt DJ (2006) Structure and Classification. In: Perry RN, Moens M (Eds.), Plant Nematology. CABI1, Oxfordshire, pp. 4-32

Ding X, Shields J, Allen R, Hussey RS (2000) Molecular cloning and characterisation of a venom allergen AG5-like cDNA from *Meloidogyne incognita*. Int J Parasitol 30:77-81

Dixon RA (2005) Engineering of plant natural product pathways. Curr Opin Plant Biol 8:329-336

Doyle EA, Lambert KN (2003) *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. MPMI 16:123-131

Dubreuil G, Magliano M, Deleury E, Abad P, Rosso MN (2007) Transcriptome analysis of root-knot nematode functions induced in the early stages of parasitism. New Phytol 176:426-436

Ducharme EP (1968) Burrowing nematode decline of citrus; a review. Tropical Nematology 20-37

Durbin R, Eddy S, Krogh A, Mitchison G (1998) Biological sequence analysis: probabilistic models of proteins and nucleic acids. University Press, Cambridge, UK, pp. 368

Eberhard J, Ehrler TT, Epple P, Felix G, Raesecke HR, Amrhein N, Schmid J (1996) Cytosolic and plastidic chorismate mutase isozymes from *Arabidopsis thaliana*: molecular characterization and enzymatic properties. Plant J 10:815-821

Eddy SR (2004) What is a hidden Markov model? Nature Biotechnology 22:1315-1316

Edens RM, Anand SC, Bolla RI (1995) Enzymes of the phenylpropanoid pathway in soybean infected with *Meloidogyne incognita* or *Heterodera glycines*. J Nematol 27:292-303

Elling A, Mitreva M, Gai X, Martin J, Recknor J, Davis EL, Hussey RS, Nettleton D, McCarter JP, Baum TJ (2009) Sequence mining and transcript profiling to explore cyst nematode parasitism. BMC Genomics 10:58

Elling AA, Davis EL, Hussey RS, Baum TJ (2007a) Active uptake of cyst nematode parasitism proteins into the plant cell nucleus. Int J Parasitol 37:1269-1279

Elling AA, Mitreva M, Recknor J, Gai X, Martin J, Maier TR, McDermott JP, Hewezi T, Bird DM, Davis EL, Hussey RS, Nettleton D, McCarter JP, Baum TJ (2007b) Divergent evolution of arrested development in the dauer stage of *Caenorhabditis elegans* and the infective stage of *Heterodera glycines*. Genome Biol 8:R211

Elsen A, Jain SM, Swennen R, De Waele D (2004) Recent developments in early in vitro screening for resistance against migratory endoparasitic nematodes in Musa. In: Jain SM, Swennen R (Eds.) Banana improvement: cellular, molecular biology, and induced mutations. Proceedings of a meeting held in Leuven, Belgium, 24-28 September 2001.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols 2:953-971

Eneqvist T, Lundberg E, Nilsson L, Abagyan R, Sauer-Eriksson AE (2003) The transthyretin-related protein family. Eur J Biochem 270:518-532

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30:1575-1584

Esser EP (1962) Life cycle stages of burrowing nematode. Nematology Circular 1.

Fenn JD, Cameron SL, Whiting MF (2007) The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Mol Biol 16:239-252

Ferraz LCCB, Brown DJF (2002) Parasitism of plants. Pensoft Publishers, Sofia, Bulgaria, pp. 221.

Fioretti L, Warry A, Porter A, Haydock P, Curtis R (2001) Isolation and localisation of an annexin gene (gp-nex) from the potato cyst nematode, *Globodera pallida*. Nematology 3:45-54

Fofana B, Benhamou N, McNally DJ, Labbé C, Séguin A, Bélanger RR (2005) Suppression of induced resistance in cucumber through disruption of the flavonoid pathway. Phytopathology 95:114-123

Fogain R, Gowen SR (1997) Damage to roots of Musa cultivars by *Radopholus similis* with and without protection of nematicides. Nematropica 27:27-32

Foster J, Ganatra M, Kama I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLOS Biology 3:0599-0614

Furlanetto C, Cardle L, Brown DJF, Jones JT (2005) Analysis of expressed sequence tags from the ectoparasitic nematode *Xiphinema index*. Nematology 7:95-104

Gao B, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS (2001a) Molecular characterization and expression of two venom allergen-like protein genes in *Heterodera glycines*. Int J Parasitol 31:1617-1625

Gao BL, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS (2001b) Identification of putative parasitism genes expressed in the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. MPMI 14:1247-1254

Gao BL, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS (2003) The parasitome of the phytonematode *Heterodera glycines*. MPMI 16:720-726

Garofalo A, Kennedy MW, Bradley JE (2003) The FAR proteins of parasitic nematodes: their possible involvement in the pathogenesis of infection and the use of *Caenorhabditis elegans* as a model system to evaluate their function. Med Microbiol Immunol 192:47-52

Gheysen G, Fenoll C (2002) Gene expression in nematode feeding sites. Annu Rev Phytopathol 40:191-219

Gheysen G, Mitchum M (2008) Molecular insights in the susceptible plant response to nematode infection. In: Berg H (Ed.), Cell Biology of Plant Nematode Parasitism, Springer, Berlin/Heidelberg, Germany, pp. 45-81

Gheysen G, Vanholme B (2007) RNAi from plants to nematodes. Trends Biotechnol 25:89-92

Gibson T, Blok V, Dowton M (2007) Sequence and characterization of six mitochondrial subgenomes from *Globodera rostochiensis* : multipartite structure Is conserved among close nematode relatives. J Mol Evol 65:308-315

Giebel J (1973) Biochemical mechanisms of plant resistance to nematodes: A review. J Nematol 6:175-184

Girdwood K, Berry C (2000) The disulphide bond arrangement in the major pepsin inhibitor PI-3 of *Ascaris suum*. FEBS Letters 474:253-255

Gissi C, Iannelli F, Pesole G (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101:301–320

Gissi C, Pesole G (2003) Transcript mapping and genome annotation of Ascidian mtDNA using EST data. Genome Res 13:2203-2212

Goellner M, Smant G, De Boer JM, Baum TJ, Davis EL (2000) Isolation of beta-1,4-endoglucanase genes from *Globodera tabacum* and their expression during parasitism. J Nematol 32:154-165

Gravato-Nobre MJ, McClure MA, Dolan L, Calder G, Davies KG, Mulligan B, Evans K, von Mende N (1999) *Meloidoygne incognita* surface antigen epitopes in infected *Arabidopsis thaliana*. J Nematol 31:212-223

Grigoriev A (1998) Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res 26:2286-2290

Grimm M, Brünen-Nieweler C, Junker V, Heckmann K, Beier H (1998) The hypotrichous ciliate *Euplotes octocarinatus* has only one type of tRNAcys with GCA anticodon encoded on a single macronuclear DNA molecule. Nucleic Acids Res 26:4557-4565

Grohmann K, Amairic F, Crews S, Attardi G (1978) Failure to detect "cap" structures in mitochondrial DNA-coded poly(A)-containing RNA from HeLa cells. Nucleic Acids Res 5:637-651

Grunewald W (2008) The quest for WRKY23's target genes. In: Functional analysis of *Arabidopsis thaliana* genes expressed during feeding site establishment of plant-parasitic nematode, PhD-thesis, Ghent University, Ghent, Belgium, pp. 136-148

Grunewald W, Cannoot B, Friml J, Gheysen G (2009) Parasitic nematodes modulate PIN-mediated auxin transport to facilitate infection. PLOS Pathogens 5:e1000266

Grunewald W, Karimi M, Wieczorek K, Van de Cappelle E, Wischnitzki E, Grundler F, Inzé D, Beeckman T, Gheysen G (2008) A role for AtWRKY23 in feeding site establishment of plant-parasitic nematodes. Plant Physiol 148:358-368

Guiliano DB, Blaxter ML (2006) Operon conservation and the evolution of trans-splicing in the phylum Nematoda. PLoS Genetics 2:e198

Gutell RR (1994) Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. Nucleic Acids Res 22:3502-3507

Haegeman A, Jacob J, Vanholme B, Kyndt T, Gheysen G (2007) A family of GHF5 endo-1,4-beta-glucanases in the migratory plant-parasitic nematode *Radopholus similis*. Plant Pathol 3: 581 - 590

Hajarnavis A, Korf I, Durbin R (2004) A probabilistic model of 3' end formation in *Caenorhabditis elegans*. Nucleic Acids Res 32:3392-3399

Hammond MP, Bianco AE (1992) Genes and genomes of parasitic nematodes. Parasitol Today 8:299-305

Hansen M, Hsu A-L, Dillen A, Kenyon C (2005) New genes tied to endocrine, metabolic and dietary regulation of lifespan from a *Caenorhabditis elegans* genomic RNAi screen. PLoS Genetics 1:0119-0128

He Y, Jones J, Armstrong M, Lamberti F, Moens M (2005) The mitochondrial genome of *Xiphinema americanum sensu stricto* (Nematoda: Enoplea): considerable economization in the length and structural features of encoded genes. J Mol Evol 61:819-833

Hennebry SC, Wright HM, Likic VA, Richardson SJ (2006) Structural and functional evolution of transthyretin and transthyretin-like proteins. Proteins: Structure, Function, and Bioinformatics 64:1024-1045

Herbert JMJ, Stekel D, Sanderson S, Heath VL, Bicknell R (2008) A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes. BMC Genomics 9:153-153

Herrmann KM, Weaver LM (1999) The shikimate pathway. Annual Review of Plant Physiology and Plant Molecular Biology 50:473-503

Hirokawa T, Boon-Chieng S, Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14:378-379

Hise AG, Gillette-Ferguson I, Pearlman E (2004) The role of endosymbiotic *Wolbachia* bacteria in filarial disease. Cell Microbiol 6:97-104

Hofmann J, Grundler FMW (2007) Identification of reference genes for qRT-PCR studies of gene expression in giant cells and syncytia induced in *Arabidopsis thaliana* by *Meloidogyne incognita* and *Heterodera schachtii*. Nematology 9:317-323

Holdeman QL (1986) The burrowing nematode *Radopholus similis sensu lato*. Nematology Publication, Sacramento, CA, USA, California Department of Food and Agriculture, Division of Plant Industry, pp. 52

Holterman M, Karssen G., van den Elsen S, van Megen H, Bakker J, Helder J (2008) Small subunit rDNA-based phylogeny of the Tylenchida sheds ligth on relationships among some high-impact plant-parasitic nematodes and the evolution of plant feeding. Nematology 99:227-235

Hoogewijs D, Houthoofd K, Matthijssens F, Vandesompele J, Vanfleteren JR (2008) Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in *C. elegans*. BMC Mol Biol 9:9

Hörnberg A, Eneqvist T, Olofsson A, Lundgren E, Sauer-Eriksson AE (2000) A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. J Mol Biol 302:649-669

Hotez PJ, Zhan B, Bethony JM, Loukas A, Williamson A, Goud GN, Hawdon JM, Dobardzic A, Dobardzic R, Ghosh K (2003) Progress in the development of a recombinant vaccine for human hookworm disease: the Human Hookworm Vaccine Initiative. Int J Parasitol 33:1245-1258

Hu M, Chilton NB, Gasser RB (2002) The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). Int J Parasitol 32:145-158

Hu M, Gasser RB (2006) Mitochondrial genomes of parasitic nematodes - progress and perspectives. Trends Parasitol 22:78-84

Huang G, Allen R, Davis EL, Baum TJ, Hussey RS (2006a) Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. Proc Natl Acad Sci U S A 103:14302-14306

Huang GZ, Dong RH, Allen R, Davis EL, Baum TJ, Hussey RS (2006b) A root-knot nematode secretory peptide functions as a ligand for a plant transcription factor. MPMI 19:463-470

Huang GZ, Dong RH, Maier T, Allen R, Davis EL, Baum TJ, Hussey RS (2004) Use of solid-phase subtractive hybridization for the identification of parasitism gene candidates from the root-knot nematode *Meloidogyne incognita*. Mol Plant Pathol 5:217-222

Huang GZ, Gao BL, 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*. MPMI 16:376-381

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868-877

Huettel RN (1985) Carrot disc culture. In: Zuckerman BM, Mai WF, Harison MB (Eds.) Plant nematology, laboratory manual. University of Massachussetts Agricultureal Experiment Station, University of Massachusetts, pp. 212

Hutangura P, Mathesius U, Jones MGK, Rolfe BG (1999) Auxin induction is a trigger for root gall formation caused by root-knot nematodes in white clover and is associated with the activation of the flavonoid pathway. Aust J Plant Physiol 26:221-231

Hwang UW, Friedrich M, Tautz D, Park CJ, Kim W (2001) Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature 413:154-157

Iseli C, Jongeneel V, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf intell Syst Mol Biol 138-148

Ithal N, Recknor J, Nettleton D, Maier T, Baum TJ, Mitchum MG (2007) Developmental transcript profiling of cyst nematode feeding cells in soybean roots. MPMI 20:510-525

Iwata F, Shinjyo N, Amino H, Sakamoto K, Islam MK, Tsuji N, Kita K (2008) Change of sununit composition of mitochondrial complex II (succinate ubiquinone reductatse/quinol-fumarate reductase) in *Ascaris suum* during the migration in the experimental host. Parasitol Int 57:54-61

Jackson C, Norman J, Schnare M, Gray M, Keeling P, Waller R (2007) Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. BMC Biology 5:41

Jacob J, Mitreva M, Vanholme B, Gheysen G (2008) Exploring the transcriptome of the burrowing nematode *Radopholus similis*. Mol Genet Genomics 280:1-17

Jacob J, Vanholme B, Haegeman A, Gheysen G (2007) Four transthyretin-like genes of the migratory plant-parasitic nematode *Radopholus similis*: Members of an extensive nematode-specific family. Gene 402:9-19

Jaubert S, Ledger TN, Laffaire JB, Piotte 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

Jeon HK, Kim KH, Eom KS (2007) Complete sequence of the mitochondrial genome of *Taenia saginata*: Comparison with T. solium and T. asiatica. Parasitology Int 56:243-246

Jex AR, Hu M, Littlewood DT, Waeschenbach A, Gasser RB (2008) Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). BMC Genomics 9:11

Jia W, Higgs P (2008) Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. Mol Biol Evol 25:339-351

Jones JT, Smant G, Blok V (2000) SXP/RAL-2 proteins of the potato cyst nematode *Globodera rostochiensis*: secreted proteins of the hypodermis and amphids. Nematology 2:887-893

Jones JT, Furlanetto C, Bakker E, Banks B, Blok V, Chen Q, Phillips M, Prior A (2003) Characterization of a chorismate mutase from the potato cyst nematode *Globodera pallida*. Mol Plant Pathol 4:43-50

Jones JT, Furlanetto C, Phillips MS (2007) The role of flavonoids produced in response to cyst nematode infection of *Arabidopsis thaliana*. Nematology 9:671-677

Jones JT, Reavy B, Smant G, Prior AE (2004) Glutathione peroxidases of the potato cyst nematode *Globodera Rostochiensis*. Gene 324:47-54

Jones JT, Smant G, Blok VC (2000) SXP/RAL2 proteins of the potato cyst nematode *Globodera rostochiensis*: secreted proteins of the hypodermis and amphids. Nematology 2:887-893

Kalorizou HA, Gowen SR, Wheeler TR (2006) Genotypic differences in the growth of bananas infected with migratory endoparasitic nematodes. 1. Roots. Experimental Agriculture 43:331-342

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P, Ahringer J (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature 421:231-237

Kamenski P, Kolesnikova O, Jubenot V, Entelis N, Krasheninnikov IA, Martin RP, Tarassov I (2007) Evidence for an adaptation mechanism of mitochondrial translation via tRNA import from the cytosol. Mol Cell 26:625-637

Kampkötter T, Timpel C, Zurawski RF, Ruhl S, Chovolou Y, Proksch P, Wätjen W (2008) Increase of stress resistance and lifespan of *Caenorhabditis elegans* by quercetin. Comp Biochem Physiol B Biochem Mol Biol 149:314-323

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32:D277-D280

Kang S, Sultana T, Eom KS, Park YC, Soonthornpong N, Nadler SA, Park JK (2008) The mitochondrial genome sequence of *Enterobius vermicularis* (Nematoda: Oxyurida)- An idiosyncratic gene order and phylogenetic information for chromadorean nematodes. Gene 429:87-97

Kaplan DT, Opperman CH (2000) Reproductive strategies and karyotype of the burrowing nematode *Radopholus similis*. J Nematol 32:126-133

Karczmarek A, Overmars H, Helder J, Goverse A (2004) Feeding cell development by cyst and root-knot nematodes involves a similar early, local and transient activation of a specific auxin-inducible promoter element. Mol Plant Pathol 5:343-346

Keddie EM, Higazi T, Unnasch TR (1998) The mitochondrial genome of *Onchocerca volvulus*: Sequence, structure and phylogenetic analysis. Mol Biochem Parasitol 95:111-127

Kennedy MW, Allen JE, Wright AS, McCruden AB, Cooper A (1995) The gp15/400 polyprotein antigen of *Brugia malayi* binds fatty acids and retinoids. Mol Biochem Parasitol 71:41-50

Kikuchi T, Aikawa T, Kosaka H, Pritchard L, Ogura N, Jones JT (2007) Expressed sequence tag (EST) analysis of the pine wood nematode *Bursaphelenchus xylophilus* and *B. mucronatus*. Mol Biochem Parasitol 115:9-17

Kikuchi T, Jones JT, Aikawa T, Kosaka H, Ogura N (2004) A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode *Bursaphelenchus xylophilus*. FEBS Letters 572:201-205

Kikuchi T, Shibuya H, Jones JT (2005) Molecular and biochemical characterization of an endo-beta-1,3-glucanase from the pinewood nematode *Bursaphelenchus xylophilus* acquired by horizontal gene transfer from bacteria. Biochem J 389:117-125

Kim KH, Eom KS, Park JK (2006a) The complete mitochondrial genome of *Anisakis simplex* (Ascaridida: Nematoda) and phylogenetic implications. Int J Parasitol 36:319-328

Kim N, Lee C (2008) Bioinformatics detection of alternative splicing. Methods Mol Biol 452:179-197

Kim SK, Reddy SK, Nelson BC, Vasquez GB, Davis A, Howard AJ, Patterson S, Gilliland GL, Ladner JE, Reddy PT (2006b) Biochemical and structural characterization of the secreted chorismate mutase (Rv1885c) from *Mycobacterium tuberculosis* H37Rv: an *AroQ enzyme not regulated by the aromatic amino acids. J Bacteriol 188:8638-8648

Kimber MJ, McKinney S, McMaster S, Day TA, Fleming CC, Maule AG (2007) *flp* gene disruption in a parasitic nematode reveals motor dysfunction and unusual neuronal sensitivity to RNA interference. FASEB J 21:1233-1243

Knight RD, Freeland SJ, Landweber LF (2001a) Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2:49-58

Knight RD, Landweber LF, Yarus M (2001b) How Mitochondria Redefine the Code. J Mol Evol 53:299-313

Ko MSH, Wang X, Horton JH, Hagen MD, Takahashi N, Maezaki Y, Nadeau JH (1994) Genetic mapping of 40 cDNA clones on the mouse genome by PCR. Mamm Genome 5:349-355

Koltai H, Spiegel Y, Blaxter ML (1997) Regulated use of an alternative spliced leader exon in the plant parasitic nematode *Meloidogyne javanica*. Mol Biochem Parasitol 86:107-110

Kramer, Kramer L, Passeri, Passeri B, Corona, Corona S, Simoncini, Simoncini L, Casiraghi, Casiraghi M (2003) Immunohistochemical/immunogold detection and distribution of the endosymbiont *Wolbachia* of *Dirofilaria immitis* and *Brugia pahangi* using a polyclonal antiserum raised against WSP (*Wolbachia* surface protein). Parasitol Res 89:381-386

Krishnan NM, Seligmann H, Raina SZ, Pollock DD (2004) Detecting gradients of asymmetry in site-specific substitutions in mitochondrial genomes. DNA Cell Biol 23:707-714

Kurppa S, Vrain TC (1985) Penetration and feeding behavior of *Pratylenchus penetrans* in strawberry roots. Revue Nématology 8:273-276

Lambert KN, Allen K, Sussex I (1999) Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*. MPMI 12:328-336

Larget B, Simon DL, Kadane JB, Sweet D (2005) A bayesian analysis of metazoan mitochondrial genome arrangements. Mol Biol Evol 22:486-495

Lavrov DV, Brown WM (2001) *Trichinella spiralis* mtDNA: A nematode mitochondrial genome that Encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of Coelomate metazoans. Genetics 157:621-637

Le TH, McManus DP, Blair D (2004) Codon usage and bias in mitochondrial genomes of parasitic platyhelminthes. Korean J Parasitol 42:159-167

Lee Y, Park BC, Lee DH, Bae K-H, Cho S, Lee CH, Lee JS, Myung PK, Park SG (2006) Mouse transthyretin-related protein is a hydrolase which degrades 5-hydroxyisourate, the end product of the uricase Reaction. Mol Cells 22:141-145

Leroy S, Bouamer S, Morand S, Fargette M (2007) Genome size of plant-parasitic nematodes. Nematology 9:449-450

Li J (2005) Brassinosteroid signaling: from receptor kinases to transcription factors. Current Opinion in Plant Biology 8:526-531

Li MW, Lin RQ, Song HQ, Wu XY, Zhu XQ (2008) The complete mitochondrial genomes for three *Toxocara* species of human and animal health significance. BMC Genomics 9:224-224

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433:769-773

Liu D, Graber J (2006a) Quantitative comparison of EST libraries requires compensation for systematic biases in cDNA generation. BMC Bioinformatics 7:77

Liu F, Lu J, Hu W, Wang SY, Cui SJ, Chi M, Yan Q, Wang XR, Song HD, Xu XN, Wang JJ, Zhang XL, Zhang X, Wang ZQ, Xue CL, Brindley PJ, McManus DP, Yang PY, Feng Z, Chen Z, Han ZG (2006b) New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. PLoS Pathog 2:e29

Long H, Wang X, Xu J (2006) Molecular cloning and life-stage expression pattern of a new chorismate mutase gene from the root-knot nematode *Meloidogyne arenaria*. Plant Pathol 55:559-563

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955-964

Luc M (1987) A reappraisal of Tylenchina (Nemata). 7. The family Pratylenchidae Thorne, 1949. Revue Nématologie 10:203-218

Lundberg E, Backstrom S, Sauer UH, Sauer-Eriksson AE (2006) The transthyretin-related protein: structural investigation of a novel protein family. Journal of Structural Biology 155:445-457

Maizels RM, Tetteh KKA, Loukas A (2000) Toxocara canis: genes expressed by the arrested infective larval stage of a parasitic nematode. Int J Parasitol 30:495-508

Mao X, Cai T, Olyarchuk JG, Wei L (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 21:3787-3793

Mauch-Mani B, Slusarenko AJ (1996) Production of salicylic acid precursors is a major function of phenylalanine ammonia-lyase in the resistance of *Arabidopsis* to *Peronospora parasitica*. Plant Cell 8:203-212

McCarter JP, Mitreva M, Martin J, Dante M, Wylie T, Rao U, Papa D, Bowers Y, Theising B, Murphy CV, Kloek AP, Chiapelli BJ, Clifton SW, Bird DM, Waterston RH (2003b) Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*. Genome Biol 4:R26

McClure MA, von Mende N (1987) Induced salivation in plant-parasitic nematodes. Phytopathology 77:1463-1469

McConn M, Creelman RA, Bell E, Mullet JE, Browse J (1997) Jasmonate is essential for insect defense in *Arabidopsis*. Proc Natl Acad Sci U S A 94:5473-5477

McVeigh P, Leech S, Mair GR, Marks NJ, Geary TG, Maule AG (2005) Analysis of FMRFamide-like peptide (FLP) diversity in phylum Nematoda. Int J Parasitol 35:1043-1060

Meutter JD, Tytgat T, Witters E, Gheysen G, Onckelen HV, Gheysen G (2003) Identification of cytokinins produced by the plant parasitic nematodes *Heterodera schachtii* and *Meloidogyne incognita*. Mol Plant Pathol 4:271-277

Miranda I, Silva R, Santos MAS (2006) Evolution of the genetic code in yeasts. Yeast 23:203-213

Mitreva M, Elling AA, Dante M, Kloek AP, Kalyanaraman A, Aluru S, Clifton SW, Bird DM, Baum TJ, McCarter JP (2004) A survey of SL1-spliced transcripts from the root-lesion nematode *Pratylenchus penetrans*. Mol Genet Genomics 272:138-148

Mitreva M, Wendl MC, Martin J, Wylie T, Yin Y, Larson A, Parkinson J, Waterston RH, McCarter JP (2006) Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. Genome Biol 7:R75

Molan AL, Sivakumaran S, Spencer PA, Meagher LP (2004) Green tea flavan-3-ols and oligomeric proanthocyanidins inhibit the motility of infective larvae of *Teladorsagia circumcincta* and *Trichostrongylus colubriformis* in vitro. Res Vet Sci 77:239-243

Molinari S, Loffredo E (2006) The role of salicylic acid in defense response of tomato to root-knot nematodes. Physiol Mol Plant Pathol 68:69-78

Montiel R, Lucena M, Medeiros J, Simões N (2006) The complete mitochondrial genome of the entomopathogenic nematode *Steinernema carpocapsae*: insights into nematode mitochondrial DNA evolution and phylogeny. J Mol Evol 62:211-225

Munoz E, Bogarad L, Deem M (2004) Microarray and EST database estimates of mRNA expression levels differ: The protein length versus expression curve for *C. elegans*. BMC Genomics 5:30

Murray D, Doran P, MacMathuna P, Moss AC (2007) *In silico* gene expression analysis: an overview. Mol Cancer 6:50

Muto A, Ushida C, Himeno H (1998) A bacterial RNA that functions as both a tRNA and an mRNA. Trends in Biochemical Sciences 23:25-29

Nagaraj SH, Gasser RB, Ranganathan S (2007) A hitchhiker's guide to expressed sequence tag (EST) analysis. Briefings in Bioinformatics 8:6

Nagy NE, Fossdal CG, Krokene P, Krekling T, Lönneborg A, Solheim H (2004) Induced responses to pathogen infection in Norway spruce phloem: changes in polyphenolic parenchyma cells, chalcone synthase transcript levels and peroxidase activity. Tree Physiol 24:505-515

Nakao M, Sako Y, Yokoyama N, Fukunaga M, Ito A (2000) Mitochondrial genetic code in cestodes. Mol Biochem Parasitol 111:415-424

Nam KH, Li J (2004) The *Arabidopsis* transthyretin-like protein is a potential substrate of BRASSINOSTEROID-INSENSITIVE 1. Plant Cell 16:2406-2417

Nandi B, Kundu K, Banerjee N, Babu SPS (2003) Salicylic acid-induced suppression of *Meloidogyne incognita* infestation of okra and cowpea. Nematology 5:747-752

Naylor HM, Newcomer ME (1999) The structure of human retinol-binding protein (RBP) with its carrier protein transthyretin reveals an interaction with the carboxy terminus of RBP. Biochemistry 38:2647-2653

Nicot N, Hausman JF, Hoffmann L, Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot 56:2907-2914

Noel GR, Atibalentja N (2006) *'Candidatus Paenicardinium endonii'*, an endosymbiont of the plant-parasitic nematode *Heterodera glycines* (Nemata: Tylenchida), affiliated to the phylum Bacteroidetes. Int J Syst Evol Microbiol 56:1697-1702

Noller H, Asire M, Barta A, Douthwaite S, Goldstein T, Gutell RR, Moazed D, Normanly J, Prince JB, Stern S, Triman K, Turner S, Van Stolk B, Wheaton V, Weiser B, Woese C (1986) Studies on the structure and function of ribosomal RNA. 143-163

O'Bannon JH (1977) Worldwide dissemination of *Radopholus similis* and its importance in crop production. J Nematol 9:16-25

Ohler U, Yekta S, Lim LP, Bartel DP, Burge CB (2004) Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. RNA 10:1309-1322

Ohtsuki T, Watanabe Yi (2007) T-armless tRNAs and elongated elongation factor Tu. IUBMB Life 59:68-75

Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470-474

Okimoto R, Chamberlin HM, Macfarlane JL, Wolstenholme DR (1991) Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host-race identification. Nucleic Acids Res 19:

Okimoto R, Macfarlane JL, Wolstenholme DR (1994) The mitochondrial ribosomal RNA genes of the nematodes *Caenorhabditis elegans* and *Ascaris suum*: consensus secondary-structure models and conserved nucleotide sets for phylogenetic analysis. J Mol Evol 39:598-613

Olsen AN, Skriver K (2003) Ligand mimicry? Plant-parasitic nematode polypeptide with similarity to CLAVATA3. Trends Plant Sci 8:55-57

Opperman CH, Bird DM (1998) The soybean cyst nematode, *Heterodera glycines*: a genetic model system for the study of plant-parasitic nematodes. Curr Opin Plant Biol 1:342-346

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 (2008) Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. Proc Nat Academy Sci of the U S A 105:14802-14807

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229-264

Painter JE, Lambert KN (2003) *Meloidogyne javanica* chorismate mutase transcript expression profile using real-time quantitative RT-PCR. J Nematol 35:82-87

Palha JA, Nissanov J, Fernandes R, Sousa JC, Bertrand L, Dratman MB, Morreale de Escobar G, Gottesman M, Saraiva MJ (2002) Thyroid hormone distribution in the mouse brain: the role of transthyretin. Neuroscience 113:837-847

PAMGO consortium (2009) Plant-Associated Microbe Gene Ontology - http://pamgo.vbi.vt.edu/.

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. Nature Genet 36:1259-1267

Parkinson J, Whitton C, Schmid R, Thomson M, Blaxter M (2004) NEMBASE: a resource for parasitic nematode ESTs. Nucl Acids Res 32:D427-D430

Peer WA, Brown DE, Tague BW, Muday GK, Taiz L, Murphy AS (2001) Flavonoid accumulation patterns of transparent testa mutants of *Arabidopsis*. Plant Physiol 126:536-548

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

Plovie E, De BS, Goeleven E, Tanghe M, Vercauteren I, Gheysen G (2003) Hairy roots to test for transgenic nematode resistance: think twice. Nematology 5:831-841

Popeijus H, Overmars H, Jones J, Blok V, Goverse A, Helder J, Schots A, Bakker J, Smant G (2000a) Degradation of plant cell walls by a nematode. Nature 406:36-37

Popeijus M, Blok VC, Cardle L, Bakker E, Phillips MS, Helder J, Smant G, Jones JT (2000b) Analysis of genes expressed in second stage juveniles of the potato cyst nematodes Globodera rostochiensis and G-pallida using the expressed sequence tag approach. Nematology 2:567-574

Pourcel L, Routaboul J-M, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I (2005) TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerizaton of flavonoids in *Arabidopsis* seed coat. Plant Cell 17:2966-2980

Price NS (2006a) The banana burrowing nematode, *Radopholus similis* (Cobb) Thorne, in the Lake Victoria region of East Africa: its introduction, spread and impact. Nematology 8:801-817

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 EHA, 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 427:30-30

Qin L, Overmars H, Smant G, Popeijus H, Rouppe van der Voort J, Groenink W, van Koert P, Schots A, Bakker J, Helder J (2000) An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*. MPMI 13:830-836

Quénéhervé P, Valette C, Topart P, Tezenas du Montcel H, Salmon F (2009) Nematode resistance in bananas: screening results on some wild and cultivated accessions of Musa spp. Euphytica 165:136-

Raczynska KD, Le Ret M, Rurek M, Bonnard G, Augustyniak H, Gualberto J (2006) Plant mitochondrial genes can be expressed from mRNAs lacking stop codons. FEBS Letters 580:5641-5646

Ramakers C, Ruijter JM, Deprez RH, Moorman AF (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339:62–66

Ranganathan S, Nagaraj SH, Hu M, Strube C, Schnieder T, Gasser RB (2007) A transcriptomic analysis of the adult stage of the bovine lungworm, *Dictyocaulus viviparus*. BMC Genomics 8:311

Ravanat J-L, Di Mascio P, Martinez GR, Medeiros MHG, Cadet J (2000) Singlet oxygen induces oxidation of cellular DNA. J Biol Chem 275:40601-40604

Rehman S, Postma W, Tytgat T, Prins P, Qin L, Overmars H, Vossen J, Spiridon L-N, Petrescu A-J, Goverse A, Bakker J, Smant G (2009) A secreted SPRY domain-containing protein (SPRYSEC) from the plant-parasitic nematode *Globodera rostochiensis* interacts with a CC-NB-LRR protein from a susceptible tomato. MPMI 22:330-340

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16:276-277

Riepsamen AH, Blok V, Philips M, Gibson T, Dowton M (2008) Poly(T) variation within mitochondrial protein-coding genes in *Globodera* (Nematoda: Heteroderidae). J Mol Evol 66:197-209

Roberts F, Roberts CW, Johnson JJ, Kyle DE, Krell T, Coggins JR, Coombs GH, Milhous WK, Tzipori S, Ferguson DJP, Chakrabarti D, McLeod R (1998) Evidence for the shikimate pathway in apicomplexan parasites. Nature 393:801-805

Robertson L, Robertson WM, Sobczak M, Helder J, Tetaud E, Ariyanayagam MR, Ferguson MAJ, Fairlamb A, Jones JT (2000) Cloning, expression and functional characterisation of a peroxiredoxin from the potato cyst nematode *Globodera rostochiensis*. Mol Biochem Parasitol 111:41-49

Romero RM, Roberts MF, Phillipson JD (1995) Chorismate mutase in microorganisms and plants. Phytochemistry 40:1015-1025

Rosso MN, Favery B, Piotte C, Arthaud L, De Boer JM, Hussey RS, Bakker J, Baum TJ, Abad P (1999) Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. MPMI 12:585-591

Rost B, Fariselli P, Casadio R (1996) PROFhtm. Prot Sci 7:1704-1718

Rost B, Yachdav G, Liu J (2004) The PredictProtein server. Nucleic Acids Res 32:W321-W326

Routaboul JM, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, Einhorn J, Lepiniec Lc (2006) Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta 224:96-107

Roze E, Hanse B, Mitreva M, Vanholme B, Bakker J, Smant G (2008) Mining the secretome of the root-knot nematode *Meloidogyne chitwoodi* for candidate parasitism genes. Mol Plant Pathol 9:1-10

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell 127:1193-1207

Sakurai M, Watanabe Y, Watanabe K, Ohtsuki T (2006) A protein extension to shorten RNA: elongated elongation factor-Tu recognizes the D-arm of T-armless tRNAs in nematode mitochondria. Biochem J 399: 249–256

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning. A laboratory manual. In: Ford N, Nolan C, Ferguson M (Eds.), Cold Spring Harbor Laboratory Press, New York, USA.

Sandstrom DJ, Nash H (2004) Drug targets: turning the channel (on) for sedation. Current Biology 14:R185-R186

Sarah JL, Pinochet J, Stanton J (1996) The burrowing nematode of bananas, *Radopholus similis* Cobb,1913. Musa Pest Fact Sheet No. 1. INIBAP, Montpellier, France.

Sasso S, Ramakrishnan C, Gamper M, Hilvert D, Kast P (2005) Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*. FEBS Journal 272:375-389

Saverwyns H, Visser A, Van Durme J, Power D, Morgado I, Kennedy MW, Knox DP, Schymkowitz J, Rousseau F, Gevaert K, Vercruysse J, Claerebout E, Geldhof P (2008) Analysis of the transthyretin-like (TTL) gene family in *Ostertagia ostertagi* - Comparison with other strongylid nematodes and *Caenorhabditis elegans*. Int J Parasitol 38:1545-1556

Scarpulla RC (2008) Transcriptional paradigms in mammalian mitochondrial biogenesis and function. Physiol Rev 88:611-638

Schäfer B (2005) RNA maturation in mitochondria of *S. cerevisiae* and *S. pombe*. Gene 354:80-85

Schefe JH, lehmann KE, Buschmann IR, Unger T, Funke-Kaiser H (2006) Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's Ct difference" formula. J Mol Med 84:901-910

Schiex T, Gouzy J, Moisan A, de Oliveira Y (2003) FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. Nucleic Acids Res 31:3738-3741

Schultz D, Yarus M (1996) On malleability in the genetic code. J Mol Evol 42:597-601

Schwarz EM, Antoshechkin I, Bastiani C, Bieri T, Blasiar D, Canaran P, Chan J, Chen N, Chen WJ, Davis P, Fiedler TJ, Girard L, Harris TW, Kenny EE, Kishore R, Lawson D, Lee R, Muller HM, Nakamura C, Ozersky P, Petcherski A,

Rogers A, Spooner W, Tuli MA, Van Auken K, Wang D, Durbin R, Spieth J, Stein LD, Sternberg PW (2006) WormBase: better software, richer content. Nucleic Acids Res 34:D475-D478

Sengupta S, Higgs PG (2005) A unified model of codon reassignment in alternative genetic codes. Genetics 170:831-840

Sengupta S, Yang X, Higgs PG (2007) The mechanisms of codon reassignments in mitochondrial genetic codes. J Mol Evol 64:662-688

Shimizu MM, Mazzafera P (2007) Polyphenoloxidase is induced by methyljasmonate and *Meloidogyne javanica* in soybean roots but is not involved in resistance. Nematology 9:625-634

Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM (1995) Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. Plant J 8:659-671

Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P (2002) PROSITE: A documented database using patterns and profiles as motif descriptors. Brief Bioinform 3:265-274

Sijmons PC, Grundler FMW, Mende N, Burrows PR, Wyss U (1991) *Arabidopsis thaliana* as a new model host for plant-parasitic nematodes. Plant J 1:245-254

Simon SA, Zhai J, Nandety RS, McCormick KP, Zeng J, Mejia D, Meyers BC (2009) Short-read sequencing technologies for transcriptional analyses. Annu Rev Plant Biol 60:305-333

Smant G, Stokkermans JPWG, Yan YT, De Boer JM, Baum TJ, Wang XH, Hussey RS, Gommers FJ, Henrissat 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 Nat Academy Sci of the U S A 95:4906-4911

Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, Rattray M, Hubbard SJ, Oliver SG, Talbot NJ (2008) Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLOS One 3:e2300

Sonnhammer ELL, Durbin R (1997) Analysis of Protein Domain Families in *Caenorhabditis elegans*. Genomics 46:200-216

Soriano IR, Asenstorfer RE, Riley IT (2004) Inducible flavone in oats (*Avena sativa*) is a novel defense against plant-parasitic nematodes. Phytopathology 94:1207-1214

Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, Coulson A, D'Eustachio P, Fitch DHA, Fulton LA, Fulton RE, Griffiths-Jones S, Harris TW, Hillier LW, Kamath R, Kuwabara PE, Mardis ER, Marra MA, Miner TL, Minx P, Mullikin JC, Plumb RW, Rogers J, Schein JE, Sohrmann M, Spieth J, Stajich JE, Wei C, Willey D, Wilson RK, Durbin R, Waterston RH (2003) The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. PLOS Biology 1:e45-e45

Stekel DJ, Git Y, Falciani F (2000) The comparison of gene expression from multiple cDNA libraries. Genome Res 10: 2055-2061

Stern MD, Anisimov SV, Boheler KR (2003) Can transcriptome size be estimated from SAGE catalogs? Bioinformatics 19:443-448

Stoffelen R, Verlinden R, Pinochet J, Swennen RL, De Waele D (2000) Host plant response of *Fusarium* wilt resistant *Musa* genotypes to *Radopholus similis* and *Pratylenchus coffeae*. International Journal of Pest Management 46:289-293

Strawn MA, Marr SK, Inoue K, Inada N, Zubieta C, Wildermuth MC (2007) Arabidopsis isochorismate synthase functional in pathogen-induced salicylate biosynthesis exhibits properties consistent with a role in diverse stress responses. J Biol Chem 282:5919-5933

Sunde M, Richardson SJ, Chang L, Pettersson TM, Schreiber G, Blake CC (1996) The crystal structure of transthyretin from chicken. Eur J Biochem 236:491-499

Swire J, Judson OP, Burt A (2005) Mitochondrial genetic codes evolve to match amino acid requirements of proteins. J Mol Evol 60:128-139

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony. 4th edition. Sinauer Associates, Sunderland, Massachusetts, USA.

Szakasits D, Heinen P, Wieczorek K, Hofmann J, Wagner F, Kreil DP, Sykacek P, Grundler FMW, Bohlmann H (2008) The transcriptome of syncytia induced by the cyst nematode *Heterodera schachtii* in *Arabidopsis* roots. Plant J 57:771-784

Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. BBA-Bioenergetics 1410:103-123

Taylor MJ, Bilo K, Cross HF, Archer JP, Underwood AP (1999) 16S rDNA phylogeny and ultrastructural characterization of *Wolbachia* intracellular bacteria of the filarial nematodes *Brugia malayi*, *B. pahangi*, and *Wuchereria bancrofti*. Exp Parasitol 91:356-361

Telford MJ, Herniou EA, Russell RB, Littlewood DT (2000) Changes in mitochondrial genetic codes as phylogenetic characters: Two examples from the flatworms. Proc Natl Acad Sci U S A 97:11359-11364

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680

Tielens AGM, Rotte C, van Hellemond JJ, Martin W (2002) Mitochondria as we don't know them. Trends Biochem Sci 27:564-572

Torto-Alalibo T, Collmer CW, Lindeberg M, Bird D, Collmer A, Tyler BM (2009) Common and contrasting themes in host cell-targeted effectors from bacterial, fungal, oomycete and nematode plant symbionts described using the Gene Ontology. BMC Microbiology 9:S3

Trinh PQ, Nguyen CN, Waeyenberge L, Subbotin SA, Karssen G, Moens M (2004) *Radopholus arabocoffeae* sp. n. (Nematoda: Pratylenchidae), a nematode pathogenic to *Coffea arabica* in Vietnam, and additional data on *R. duriophilus*. Nematology 6:681-693

Tsang WY, Lemire BD (2003) The role of mitochondria in the life of the nematode, *Caenorhabditis elegans*. BBA-Mol Basis Dis 1638:91-105

Tucker ML, Xue P, Raina A, Ehrenfried ML, Asif M, Thai VK (2005) Characterization of several *Heterodera glycines* mRNA that encode small proteins with putative signal peptides. J Nematol 37:422-428

Tytgat T (2003) Identification of genes coding for pharyngeal gland secretions in cyst and root-knot nematodes. PhD-thesis, Ghent University, Ghent, Belgium, pp. 207

Tytgat T, Vanholme B, De Meutter J, Claeys M, Couvreur M, Vanhoutte I, Gheysen G, Van Criekinge W, Borgonie G, Coomans A, Gheysen G (2004) A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes. MPMI 17:846-852

Tytgat T, Vercauteren I, Vanholme B, De Meutter J, Vanhoutte I, Gheysen G, Borgonie G, Coomans A, Gheysen G (2005) An SXP/RAL-2 protein produced by the subventral pharyngeal glands in the plant parasitic root-knot nematode *Meloidogyne incognita*. Parasitol Res 95:50-54

Uehara T, Kushida A, Momota Y (2001) PCR-based cloning of two beta-1,4-endoglucanases from the root-lesion nematode *Pratylenchus penetrans*. Nematology 3:335-341

United Nations Environment Programme (1995) Report of the methyl bromide technical options committe. Montreal protocol on substances that deplete the ozone layer.

Van Leeuwen T, Vanholme B, Van Pottelberge S, Van Nieuwenhuyse P, Nauen R, Tirry L, Denholm I (2008) Mitochondrial heteroplasmy and the evolution of insecticed resistance: non-mendelian inheritance in action. Proc Natl Acad Sci U S A 105:5980-5985

van Overveld FWPC, Haenen GRMM, Rhemrev J, Vermeiden JPW, Bast A (2000) Tyrosine as important contributor to the antioxidant capacity of seminal plasma. Chem-Biol Interact 127:151-161

Vandekerckhove TTM, Coomans A, Cornelis K, Baert P, Gillis M (2002) Use of the *Verrucomicrobia*-specific probe EUB338-III and fluorescent *in situ* hybridization for detection of "*Candidatus Xiphinematobacter*" cells in nematode hosts. Appl Environ Microbiol 68:3121-3125

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:research0034.1-0034.11

Vanfleteren JR (1978) Axenic culture of free-living, plant-parasitic and insect-parasitic nematodes. Annu Rev Phytopathol 16:131-157

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 V88:731-733

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

Vanholme B, Gheysen G (2005) A method to prove secretion of proteins by the plant-parasitic nematode *Heterodera schachtii* using Dynabeads. Nematology 7:309-312

Vanholme B, Kast P, Haegeman A, Jacob J, Grunewald W, Gheysen G (2008) Structural and functional investigation of a secreted chorismate mutase from the plant-parasitic nematode *Heterodera schachtii* in the context of related enzymes from diverse origins. Mol Plant Pathol 10: 189 - 200

Vanholme B, Mitreva M, Van Criekinge W, Logghe M, Bird D, McCarter JP, Gheysen G (2006) Detection of putative secreted proteins in the plant-parasitic nematode *Heterodera schachtii*. Parasitol Res 98:414-424

Veech JA (1981) Phytoalexins and their role in the resistance of plants to nematodes. J Nematol 14:2-9

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484-487

Vercauteren I, Geldhof P, Peelaers I, Claerebout E, Berx G, Vercruysse J (2003) Identification of excretory-secretory products of larval and adult *Ostertagia ostertagi* by immunoscreening of cDNA libraries. Mol Biochem Parasitol 126:201-208

Vullo A, Frasconi P (2004) Disulfide connectivity prediction using recursive neural networks and evolutionary information. Bioinformatics 20:653-659

Wang J-PZ, Lindsay BG, Cui L, Wall PK, Marion J, Zhang J, dePamphilis CW (2006) Gene capture prediction and overlap estimation in EST sequencing from one or multiple libraries. BMC Bioinformatics 6:300

Wang XH, Allen R, Ding XF, Goellner M, Maier T, De Boer JM, Baum TJ, Hussey RS, Davis EL (2001) Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. MPMI 14:536-544

Wang XH, Mitchum MG, Gao BL, Li CY, Diab H, Baum TJ, Hussey RS, Davis EL (2005) A parasitism gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR (CLE) of *Arabidopsis thaliana*. Mol Plant Pathol 6:187-191

Washietl S, Hofacker IL, Stadler PF (2005) From the cover: fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci U S A 102:2454-2459

Wasmuth JD, Blaxter ML (2004) Prot4EST: Translating expressed sequence tags from neglected genomes. BMC Bioinformatics 5:187

Weaver LM, Herrmann KM (1997) Dynamics of the shikimate pathway in plants. Trends Plant Sci 2:346-351

Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogger JB (2007) Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol 144:32-42

Weischer B, Brown D (2000) An introduction to nematodes: General nematology. Pensoft Publishers, Sofia, Bulgaria, pp. 187.

Wildermuth MC, Dewdney J, Wu G, Ausubel FM (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence. Nature 414:562-565

Williams SA, Lizotte-Waniewski MR, Foster J, Guiliano D, Daub J, Scott AL, Slatko B, Blaxter ML (2000) The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of Brugia malayi. Int J Parasitol 30:411-419

Wojtczak A (1997) Crystal structure of rat transthyretin at 2.5 Å resolution: first report on a unique tetrameric structure. Acta Biochim Pol 44:505-517

Wright AJ, Hunter CP (2003) Mutations in a β-tubulin disrupt spindle orientation and microtubule dynamics in the early *Caenorhabditis elegans* embryo. Mol Biol Cell 14:4512-4525

Wubben MJE, Jin J, Baum TJ (2008) Cyst nematode parasitism of *Arabidopsis thaliana* is in hibited by salicylic acid and elicits uncouple SA-independent pathogenesis-related gene expression in roots. MPMI 21:424-432

Wuyts N (2006) Infection and reproduction of nematode in transgenic and mutant Arabidopsis and tobacco with an altered phenylpropanoid metabolism. In: Interactions between plant parasitic nematodes and plant secondary metabolism with emphasis on phenylpropanoids in roots. PhD-thesis, Katholieke Universiteit Leuven, Leuven, Belgium, pp. 59-91.

Wuyts N, Lognay G, Verscheure M, Marlier M, De Waele D, Swennen R (2007) Potential physical and chemical barriers to infection by the burrowing nematode *Radopholus similis* in roots of susceptible and resistant banana (*Musa* spp.). Plant Pathol 56:878-890

Wuyts N, Swennen R, De Waele D (2006) Effects of plant phenylpropanoid pathway products and selected terpenoids and alkaloids on the behaviour of the plant-parasitic nematodes *Radopholus similis*, *Pratylenchus penetrans* and *Meloidogyne incognita*. Nematology 8:89-101

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

Wyss U, Robertson WM, Trudgill DL (1988) Oesophageal bulb function of *Xiphinema index* and associated root cell responses, assessed by video-enhanced contrast light microscopy. Revue Nématologie 11:253-261

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434:338-345

Yan J, Marr TG (2005) Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. Genome Res 15:369-375

Yan YT, Smant G, Davis E (2001) Functional screening yields a new beta-1,4-endoglucanase gene from *Heterodera glycines* that may be the product of recent gene duplication. MPMI 14:63-71

Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34:W293-W297

Yokobori Si, Suzuki T, Watanabe K (2001) Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. J Mol Evol 53:314-326

Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol 25:99-120

Zhu J, He F, Wang J, Yu J (2008) Modeling transcriptome based on transcript-sampling data. PLoS ONE 3:e1659-e1659

Zhu XQ, Chilton NB, Gasser RB (2005) Detection of sequence variation in parasite ribosomal DNA by electrophoresis in agarose gels supplemented with a DNA-intercalating agent. Electrophoresis 19:671-674

Zhu-Salzman K, BI JL, LIU TX (2005) Molecular strategies of plant defense and insect counter-defense. Insect Sci 12:3-15

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406-3415

Zunke U (1990) Ectoparasitic fedding behaviour of the root lesion nematode, *Pratylenchus penetrans*, on root hairs of different host plants. Revue Nématologie 13:331-337

De laatste loodjes…

Ook al is de twijfel een eerbetoon aan de waarheid, ooit moeten er knopen doorgehakt worden. En de knoop die rechtstreeks tot dit boekje geleid heeft, deed zich bij dit schrijven vier jaar geleden aan mij voor. De beste methode om moeilijke beslissingen te nemen is het buikgevoel lichtjes bij te spekken met rationele overwegingen. Maar omdat ik toen nog student was, kan het ook zijn dat er wel een pita te veel op mijn maag gelegen had.

Dus toen de nieuwe lichting doctoraatsstudenten in onze vakgroep van wal stak, werd ook ik direct opgenomen in de vakgroep van Lieve, waarvoor dank. Bij het nabeschouwen van dat begin, komen we echter allemaal tot de conclusie dat de 20 jaar schoolse opvoeding ons toch niet volledig voorbereid heeft op wat komen ging, en ongetwijfeld ook niet helemaal op wat nog volgen zal. Met andere woorden, in mijn geval was ik een naïeve enthousiaste hemelbestormer, vol wilde plannen en ideeën. Ik had een duidelijk beeld van wat wetenschap is en moest zijn — tot er discussies ontstonden over schijnbaar de onnozelste zaken, zoals 'wat zijn parasieten nu eigenlijk?', en minder plezante vragen als 'welke reviewers zetten we op het artikel??', en frustrerende vragen als 'waar in de vriezer heb ik dat tubetje nu weer gelaten???' Aan dat laatste hebben we direct wat aan gedaan, initieel onstuimig aan de efficiëntie werkend. Maar over andere vragen kan ik nog altijd gaten in mijn broek zitten denken…

"Woordjes met Joachim" - l: lou·te·ren, het moreel beter maken.

b.v. Ik ben gelouterd, mijn moreel is weer te pruimen, ☺

De wetenschap zoals ik ze de laatste jaren ervaren heb, is een intrigerend beestje. De overtuiging dat 'droge' feiten en data de basis vormde van wetenschappelijk onderzoek, ruimde beetje bij beetje — en toegegeven, ook een beetje laat — plaats voor de menselijke kantjes van de wetenschap, die hun licht (al dan schaduw?) werpen op het feitelijke. Eén van de meest gehoorde complimenten betreffende nieuwe wetenschappelijke resultaten is daar zeer kenmerkend voor: "Hey, proficiat met je nieuwe artikel hé. Ik vind het een *schoon verhaal*." En om het met de woorden van de *prof* uit te drukken, "we zitten daar met PR-achtige systemen", inderdaad. En uiteindelijk blijkt het misschien ook zo het beste te lopen: in wetenschap moet je proberen de *emotie* proberen over te brengen, de emotie gekenmerkt door een verhoogde hartslag, rechtstaande haartjes, verwijde pupillen, te veel aan energie, die je overvalt als een bliksem bij klaarlichte dag. Inderdaad, die dagelijkse vijf kopjes koffie kunnen misschien hetzelfde veroorzaken, maar vaak was het een gevolg van een nieuw inzicht, of een nieuwe bevinding. En zo probeer je die energie vast te houden, om je aan het loodzware werk te zetten je resultaten neer te pennen doorspekt met een rationele dosis enthousiasme.

*"In fact, in our field you don't need so many facts,
unfortunately."*

Het is duidelijk dat de wetenschapper in deze tijd meer en meer een pluralistisch beestje moet zijn - zeker als hij zijn eerste stappen in een wetenschappelijke carrière zet, in het overgrote deel van de gevallen als doctoraatsstudent. Opmerkelijk genoeg wordt deze openingszin vaak bevestigd door gevestigde wetenschappelijke waarden. Bovenal moet de wetenschapper, als hij wil *overleven*, schrijver, of beter, (co)auteur zijn — maar wetenschap en taalgevoel combineren in één persoon… een zeldzaamheid me dunkt. Dus de wetenschapper ploetert verder - dan nog in de Engelse taal - want publiceren is gebod nummer één. Het tweede belangrijk werkpunt voor de succesvolle wetenschapper, is de zoektocht naar gelijkgezinden, naar een 'netwerk'. En het moet gezegd, weinig zaken gaven me zo een *boost* en vernieuwde werklust als het maken van nieuwe contacten, vaak na een buitenlands congres. Ik ben Lieve en de organisatoren van de COST actie dan ook zeer dankbaar voor de geboden mogelijkheden tot internationale interactie.

"Woordjes met Joachim" - 2: de lak·moes·proef, figuurlijk test die definitief over iets uitsluitsel geeft. Dit is een test die aangeeft of iets écht deugt. Het wordt gebruikt in contexten waarin er (nog) getwijfeld wordt aan bijvoorbeeld de kwaliteit, het bestaansrecht, de invloed of het belang van iets; de lakmoesproef is 'de ultieme test'.

Ik voelde mij zoals een schrijnwerker bij het schrijven van artikels — urenlang schaven, schuren en vijlen tot een aanvaardbaar opstel, mogelijke ongevallen zoals overtollige woorden en herhalingen vermijdend, om de lezer — en belangrijk, de reviewers! — een aangename leeservaring te bezorgen. Vaak voelde ik mij zelfs beetje verkoper bij voordrachten. Ik voelde mij ambtenaar bij het invullen van documenten ter regulatie van het wetenschappelijke werk — want het zijn soms *snoodaards*, die wetenschappers! En raadgever bij het motiveren en begeleiden van nieuwe studenten. Portier bij het openen van de deur van *het tweede verdiep*. En een gevoel dat misschien het best te omschrijven valt als nutteloos bij vele vergaderingen…

"Woordjes met Joachim" - 3: snood·aard, formele aanduiding voor een slecht mens.

Maar af en toe, tussen al die andere zaken door, was het ook de wetenschapper in mij die de kans kreeg om te groeien. In het begin had die het reilen en zeilen daar op de wetenschappelijke zee nog niet onder de knie. Vooral de eerste fases, te vergelijken een zig-zag koers van een wanordelijk vlotje, met een stranding her en der, in de gaten gehouden vanop de verre kaai door een geldschieter, terwijl ondertussen mijn kaart van stress en onkunde ondersteboven bestudeerd werd. Maar gelukkig werkte mijn kompas en wees me beetje bij

beetje een juiste richting uit (1). Toegegeven, mijn drang om álles te willen zien heeft me zeer lang, tot op de dag van vandaag eigenlijk, parten gespeeld — de zee is zooo groot! Maar de eerste ontdekkingetjes tekenden zich aan de horizon af… Euforie alom! Een eerste klein wetenschappelijk resultaatje, en de toon was gezet. De koers werd steeds scherper gesteld, naar de overkant van die grote plas. Wie weet welke peilloze diepten we onwetend overgevaren hebben? En hoewel we misschien vaak anders dachten, we voeren nooit alleen over deze zee (2). Het vlotje waarmee we van wal staken breidde langzamerhand uit tot volwaardige zeilscheepjes. Met stormen en verraderlijke ijsschotsen werd vlotjes omgegaan. De kaarten werden gedetailleerder, zandbanken vermeden en uiteindelijk kregen geldschieters kregen wat ze verwachten (3). Nieuwe resultaten hebben me vaak in de richting geleid van nieuwe collega's, die onderweg toevallige dezelfde richting uitvoeren en aan wie ik tijdens de reis veel heb aan gehad (4).

> Birger: *"Het is vandaag vers-vlees-dag, want de groentjes zijn toegekomen."* (de aankomst van de lichting eerstejaars bedoelend).

Toch. Het *ambetante* is dat er een enorme discrepantie bestaat tussen het *individuele en tijdelijke* karakter van het onderzoek van de doctoraatsstudent en de nood aan *samenwerking* innen de universitaire omgeving om vooruit te geraken. De te beperkte budgetten is de directe oorzaak van de navenante competitie voor werkingsmiddelen. Ik denk dat deze strijd zich vaak veruitwendigd door niet-aflatende publicatiedrang en gebrek aan samenwerking tussen labo's. Doordat het waarde-oordeel van een wetenschapper bepaald wordt door een verre anonieme instantie op basis van het aantal publicaties, gecombineerd met de tijdelijkheid van de positie van de doctoraatsstudent, leidt vaak tot schijnbare willekeur en een serieuze '*brain drain*' op het niveau van de labo's. Mede daarom is het juist als gesteld wordt dat momenteel vaak niet een wetenschappelijke 'kritische massa' gehaald wordt aan de vakgroepen. Misschien is het inderdaad hoog tijd voor een grondige herziening van de financiering van het wetenschappelijk onderzoek aan de academische wereld.

> *"I will help you carry this burden, as long as it's yours to carry."*
> (Lord of The Rings)

Toch kan veel goedgemaakt worden door *communicatie*, het sleutelwoord tot een soepel draaiende onderzoeksgroep. Naar mijn gevoel is er nog te weinig steun om ideeën, voorstellen en resultaten uit te wisselen — ook op wetenschappelijk vlak. Initieel bestond bij mij de drang om enkel met positieve resultaten naar buiten te komen. Maar in zeer veel gevallen verdienen negatieve resultaten meer aandacht, omdat deze momenten unieke mogelijkheden bieden tot groei en vergaren van nieuwe kennis. Vaak vergt dit inderdaad enige moed — maar het loonde de moeite. Verschillende keren kwam iemand te horen over het gene en gindse negatieve resultaat van mij, waarop die persoon de kennis bleek te hebben mij uit mijn negatief dal te

halen en het werk terug op het goede spoor te zetten. Voor deze bijdragen en de leuke tijd wil ik dan ook van ganser harte iedereen van de vakgroep Moleculaire Biotechnologie (5) bedanken.

"He who wonders discovers that this in itself is wonder."
M.C. Escher

En natuurlijk wil ik ook mijn ouders vermelden, voor de kansen die ze mij geboden hebben, voor hun onvoorwaardelijke steun, en voor de prettige verbouwingswerken die we samen de laatste jaren in de Kiekenstraat verwezenlijkt hebben. Voor het onvoorwaardelijk plezier kon ik ook altijd rekenen op vele anderen, die me regelmatig hielpen herinneren aan dat spreekwoord met die boog en dat gespannen staan (6).

"The whole problem with the world is that fools and fanatics
are always so certain of themselves, but wiser people so full
of doubts."
Bertrand Russell

Laat die zetel in de buro maar komen, voorlopig lijkt mijn schip even aangemeerd te zijn. Op het moment van schrijven weet ik nog niet zeker waar het mij de komende maanden brengen zal, maar ik weet dat ik altijd zal kunnen terugvallen op iemand die ik altijd zal kunnen liefhebben, *again and again*, mijn liefste Sofie.

(1) Veel dank aan Bartel, wiens favoriete woordje veranderde van 'nematode' naar '*koedie-koediekoedie*' toen hij papa werd, voor de inleiding in de wondere wereld van wetenschap.

(2) Dank aan de medezeilers, die vanaf dezelfde kaai, in hetzelfde jaar de tocht aangevat hebben en steeds voor een plezierige sfeer gezorgd hebben: Annelies, Birger, Delphine, Caroline, Gianni, Maté, Leander.

(3) Dank aan het Bijzonder Onderzoeksfonds (BOF) om mij eraan te herinneren niet te vergeten te publiceren.

(4) Veel dank aan de mensen van het labo gewasbescherming, voor de prettige gesprekken, samenwerking en uitwisseling van resultaten: Thomas, Tijs, Pieter ("'t is inderdaad soms *dzjolen* hé") en Wannes.

(5) Bedankt, Wim, Winnok, Tina, Bernard, Isabel, Elke, Fien, Joseph, Soumi, Njira, Lilli, Kamrun en alle anderen van de vakgroep!

(6) Merci Kristof, Sebastien, Barbara, Stijn, Gert, Liesbeth, Manu, Dillis, Alexander, Frederik, iedereen van het LEEJO,... kortom iedereen die bijgedragen heeft aan talrijke momenten van onbekommerd plezier!

"Laat daarom uw ziel uw rede exalteren tot de hoogte van een hartstocht, en laat
hem uw hartstocht leiden met rede, zodat uw hartstocht zijn eigen dagelijkse
opstandingen mag overleven en als een feniks uit zijn as zal herrijzen."
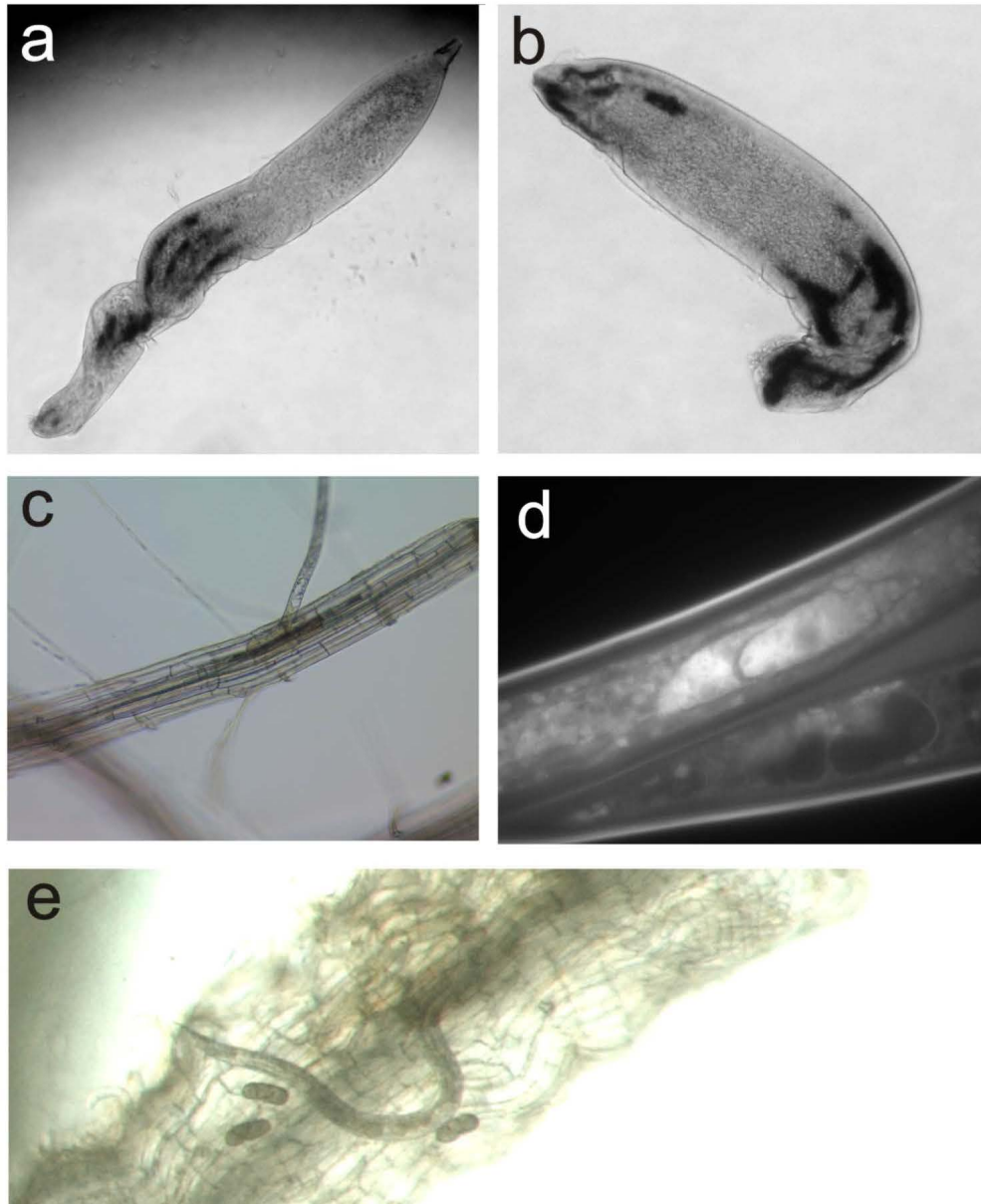

Figure 69: Als afsluiter enkele foto's die niet in de hoofdstukken van dit proefschrift hun plaats gevonden hebben. **a** en **b**, *in situ* hybridisatie op parasitaire stadia van *Heterodera schachtii* van het *sec-2* (aka *far*) gen, dat tot expressie komt in de hypodermis. **c**, *H. schachtii* in een *Arabidopsis* wortel, vijf dagen na infectie. **d**, een kleuring van de faryngeale klieren van *Radopholus similis* door middel van DAPI. **e**, *R. similis* vrouwtje met kroost in wording, in een wortel van de klaver *Medicago truncatula*.

Curriculum Vitae

Joachim Jacob

joachim.jacob@gmail.com   ☎ 0486060938

Address   Kiekenstraat 22, 9000 Gent
Birth   27/08/1982, Lokeren

Education

PhD training, Ghent University, 2005-2009.

Bio-engineer, Faculty of Bioscience Engineering, Ghent University, 2000-2005, great distinction.
    Thesis: "Investigation in the expression variability of two divergent positioned transgenes"(in cooperation with Bayer Cropscience, Zwijnaarde, Belgium).

High School, Greek-Sciences, Sint-Lodewijkscollege, Lokeren, 1994-2000.

Professional carreer

BOF fellowship ("Bijzonder Onderzoeksfond", Ghent University), 2005-2009.
    Project: "Plant-parasitic nematodes: from genomics to functional analysis of parasitism genes."
    Promotor: Prof. dr. G. Gheysen, Dept. Molecular Biotechnology,
    Faculty of Bioscience Engineering, Ghent University.

Publications

An endosymbiotic bacterium in a plant-parasitic nematode: Member of a new *Wolbachia* supergroup. Annelies Haegeman, Bartel Vanholme, <u>Joachim Jacob</u>, Tom T.M. Vandekerckhove, Myriam Claeys, Gaetan Borgonie, Godelieve Gheysen. International Journal for Parasitology (2009) 39(9):1045-1054.

Expressed sequence tags of the peanut pod nematode *Ditylenchus africanus*: The first transcriptome analysis of an Anguinid nematode. Annelies Haegeman, <u>Joachim Jacob</u>, Bartel Vanholme, Tina Kyndt, Makedonka Mitreva, Godelieve Gheysen. Molecular and Biochemical Parasitology (2009) 167(1):32-40.

Arabinogalactan endo-1,4-β-galactosidase: a putative plant cell wall-degrading enzyme of plant-parasitic nematodes. Bartel Vanholme, Annelies Haegeman, Joachim Jacob, Bernard Cannoot, Godelieve Gheysen (2009) In press.

Exploring the transcriptome of the burrowing nematode *Radopholus similis*. <u>Joachim Jacob</u>, Makedonka Mitreva, Bartel Vanholme, Godelieve Gheysen. Molecular genetics and genomics, (2008) 280(1):1-17.

A family of endo-1,4-beta-glucanases in the migratory plant-parasitic nematode *Radopholus similis*. Annelies Haegeman, <u>Joachim Jacob</u>, Bartel Vanholme, Tina Kyndt, Godelieve Gheysen. Plant Pathology (2008) 57(3):581-590.

Structural and functional investigation of a secreted chorismate mutase from the plant-parasitic nematode *Heterodera schachtii* in the context of related enzymes from diverse origins. Bartel Vanholme, Peter Kast, Annelies Haegeman, <u>Joachim Jacob</u>, Wim Grunewald, Godelieve Gheysen. Molecular Plant Pathology (2008) 10(2):189-200.

Four transthyretin-like genes of the migratory plant-parasitic nematode *Radopholus similis*. Members of an extensive nematode-specific family. <u>Joachim Jacob</u>, Bartel Vanholme, Annelies Haegeman, Godeliev Gheysen. Gene (2007) 402(1):9-19.

International congresses and Meetings

Oral presentation

Current status of the molecular characterization of migratory plant-parasitic nematode *Radopholus similis*. Jacob, Joachim; Haegeman, Annelies; Kyndt, Tina; Vanholme, Bartel; Gheysen, Godelieve. 59th International Symposium on Crop Protection, Gent, Belgium, 22 May, 2007.

Exploring the transcriptome of the burrowing nematode *Radopholus similis*. Jacob Joachim, Mitreva Makedonka, Vanholme Bartel, Gheysen Godelieve. Proceedings of the second COST 872 Annual Meeting, Postojna, Slovenia, 26-29th May, 2008.

Flavonoids and early defence response in cyst nematode infection of *Arabidopsis thaliana*. Jacob Joachim, Melis Peter, Vanholme Bartel, Cannoot Bernard, Gheysen Godelieve. 61st International Symposium on Crop Protection, Gent, Belgium, 20 May, 2009.

The mitochondrial genome of the burrowing nematode *Radopholus similis*. Jacob Joachim, Vanholme Bartel, Van Leeuwen Thomas, Gheysen Godelieve. Proceedings of the third COST 872 Annual Meeting, Toledo, Spain, 25-28th May, 2009.

Poster presentation

In search of parasitism genes in the burrowing nematode *Radopholus similis*. Jacob, Joachim; Haegeman, Annelies; Kyndt, Tina; Vanholme, Bartel; Gheysen, Godelieve. Proceedings of the first COST 872 Annual Meeting, La Colle-sur-Loup, France, 9-11th May, 2007.

NEXT, the Web-based Nematode EST Exploration Tool. Jacob Joachim. Proceedings of the third COST 872 Annual Meeting, Toledo, Spain, 25-28th May, 2009.

Participation

28th International Symposium of the European Society of Nematologists, Blagoevgrad, 5-9th June, 2006.

Courses PhD training

- General Morphology: Nematoda, Prof. W. Decraemer, UGent
- Bio-informatica 2, Prof. W. Vancrieckinge, Ugent
- Onderzoeksmethodiek, Prof. S. Matthys, IVPV
- Fysisch milieu en ontwikkeling, Prof. M. De Dapper, Ugent
- Management of Research & Development at industry and university, Prof. W. Soetaert, Ugent

Other projects

Guidance of master thesis "De rol van flavonoieden bij infectie door de nematode *Heterodera schachtii* in *Arabidopsis thaliana*", Peter Melis, 2008.

Selection of *Radopholus similis* candidate target genes for *in planta* generation of double-stranded RNA in banana (*Musa* sp.), Project by Soumi Joseph, Ph.D. student.

Supervision of students

PR Activitities

- "Opendeurdag" March 18th, 2006.
- "Vlaamse Wetenschapsweek", October 19-29th, 2006.
- "Alumnidag 20 jaar Alumni", October 21st, 2006.
- "Vlaamse Wetenschapsweek", October 17-26th, 2008.