“Learn from yesterday, live for today, hope for tomorrow.
The important thing is not to stop questioning.”
Albert Einstein (1879-1955)

Promotor: **Prof. dr. Godelieve Gheysen**
Ghent University
Dpt. Molecular Biotechnology
Laboratory for Applied Molecular Genetics

Dean: **Prof. dr. ir. Guido Van Huylenbroeck**

Rector: **Prof. dr. Paul Van Cauwenberge**

MOLECULAR ANALYSIS OF MIGRATORY PLANT-PARASITIC NEMATODES WITH A FOCUS ON PLANT CELL WALL MODIFYING ENZYMES

ANNELIES HAEGEMAN

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

Moleculaire analyse van migratorische plantenparasitaire nematoden toegespitst op plantencelwandmodificerende enzymen

Cover illustration

Front: Fuchsin stained *Medicago truncatula* L. root infected by *Radopholus similis*
Back: 3D model of a putative endoxylanase protein from *Radopholus similis*

Printing

DCL Signs, Zelzate

Haegeman, A. (2009). Molecular analysis of migratory plant-parasitic nematodes with a focus on plant cell wall modifying enzymes. PhD thesis, Ghent University, Ghent, Belgium.

ISBN 978-90-5989-292-7

This research was supported by a postgraduate grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

Examination committee

Prof. dr. ir. Dirk Reheul (chairman)
Department of Plant Production
Plant Breeding and Sustainable Crop Production Unit
Faculty of Bioscience Engineering, Ghent University

Prof. dr. Els Van Damme (secretary)
Department of Molecular Biotechnology
Biochemistry and Glycobiology Unit
Faculty of Bioscience Engineering, Ghent University

Prof. dr. Godelieve Gheysen (promotor)
Department of Molecular Biotechnology
Applied Molecular Genetics Unit
Faculty of Bioscience Engineering, Ghent University

Prof. dr. Dirk De Waele
Department of Biosystems
Laboratory of Tropical Crop Improvement
Faculty of Bioscience Engineering, K.U. Leuven

dr. Marie-Noëlle Rosso
Biotic Interactions and Plant Health
Plant-Nematode Interactions Unit
French National Institute for Agricultural Research, Sophia-Antipolis, France

dr. Wim Bert
Biology Department
Nematology Unit
Faculty of Sciences, Ghent University

dr. ir. Bartel Vanholme
Department of Plant Systems Biology
Bio-energy Unit
Flanders Institute for Biotechnology
Faculty of Sciences, Ghent University

Table of contents

Abbreviations

AA	amino acid
AFLP	amplified fragment length polymorphism
ANOVA	analysis of variance
APN	animal-parasitic nematodes
BLAST	basic local alignment search tool
bp	base pair
CAI	codon adaptation index
CAZy	carbohydrate active enzymes
CBM	carbohydrate binding module
CDD	conserved domain database
cDNA	copy DNA
CMC	carboxymethylcellulose
DAPI	4’,6-diamidino-2-phenylindole
DC	dichroic
DIG	digoxigenin
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
DMT	5-methoxy-N,N-dimethyl tryptamine oxalate
dsRNA	double stranded RNA
DTT	dithiotreitol
EDTA	ethylenediaminetetraacetic acid
EM	emission
EPPO	European and Mediterranean plant protection organisation
EST	expressed sequence tag
EX	excitation
FITC	fluorescein isothiocyanate
FLN	free-living nematodes
F-primer	forward primer
GFP	green fluorescent protein
GHF	glycosyl hydrolase family
GO	gene ontology
h	hours
HAC	hospital antiseptic concentrate
HGT	horizontal gene transfer
HMM	hidden Markov model
ILD	incongruence length difference

kDa	kilo Dalton
LB	Luria Bertani
LEA	late embryogenesis abundant protein
min	minutes
ML	maximum likelihood
MLST	multilocus sequence typing
MP	maximum parsimony
mRNA	messenger RNA
MW	molecular weight
nt	nucleotides
ORF	open reading frame
PCR	polymerase chain reaction
PDB	protein databank
PMT	photomultiplier tube
PN	parasitic nematodes
PPN	plant-parasitic nematodes
RNA	ribonucleic acid
RNAi	RNA interference
R-primer	reverse primer
RT	reverse transcription
RT-PCR	reverse transcriptase polymerase chain reaction
s	seconds
S	Svedberg
SDS	sodium dodecyl sulphate
SL	spliced leader
SMART	simple modular architecture research tool
SP	signal peptide
SSC	saline sodium citrate
TAE	tris - acetic acid - EDTA
TBR	tree bissection-reconnection
TE	tris - EDTA
TGICL	TIGR gene indices clustering tool
U	unit
UTR	untranslated region
VAP	venom allergen protein
WAG	Whelan and Goldman
WEGO	web gene ontology annotation plot
WSP	*Wolbachia* surface protein

CHAPTER 1

INTRODUCTION

Nematodes: underestimated animals

Nematodes or roundworms are among the most ancient and successful creatures on earth. They have been around for an estimated one billion years (Wang et al., 1999), and have occupied almost every possible ecological niche since. Most of them have a free-living lifestyle, and can occur both in terrestrial environments as in freshwater or marine water. Others are parasites of plants, animals and humans. Nematodes are so ubiquitous that a famous nematologist once said “if all the matter in the universe except the nematodes were swept away, our world would still be dimly recognizable” (Cobb, 1915). They account for 80% of all individual animals on earth, and the number of different species is estimated to range from 100,000 to 1 million (Parkinson et al., 2004).

Although nematodes are relatively unknown to the public, they do have a quite large impact on human life. They cause severe yield reductions in food and fiber crops, affect livestock and companion animals, and can cause exotic human diseases such as elephantiasis or river blindness (Bird & Koltai, 2000; Bird & Kaloshian, 2003). Moreover, nematodes have become extremely important organisms for the scientific community. The free-living nematode *Caenorhabditis elegans* started to get attention in the seventies and has since become an important model organism in developmental, physiological and molecular biology. It was also the first multicellular organism from which the whole genetic code or genome was unraveled by the *C. elegans* sequencing consortium (1998). This opened tremendous opportunities for molecular biologists to investigate the function of different genes. To date, *C. elegans* remains one of the most important model organisms to study basic animal biology and diseases.

Some general features of nematodes

Despite their diversity in lifestyle, nematodes display a relatively conserved body plan. They are unsegmented, vermiform and bilaterally symmetric. The body consists of an external cylinder (the body wall including cuticle, epidermis and somatic musculature) and an internal cylinder (the digestive system) separated by a pseudocoelomic cavity filled with fluid under pressure and containing a number of other organs such as reproductive tracts and a nervous system (Decraemer & Hunt, 2006). Since nematodes do not have respiratory or circulatory systems, they depend on diffusion of water, gasses and metabolites in and out of their semi-permeable body walls. During their life cycle, all nematodes undergo four molts from the juvenile to the adult stage. They typically have an egg stage, four juvenile stages and the adult male and female. In some genera, the juvenile that hatches from the egg is the J1 stage, while in other species it is J2. Juveniles usually resemble the adult stage, differing in the absence of a mature reproductive system and in certain measurements and proportions. In some species,

the juvenile stages can be more resistant to environmental stress conditions (e.g. dauer stage) (Decraemer & Hunt, 2006).

The emergence of plant-parasitic nematodes

Parasitic animals benefit from a close symbiotic association with their host that is harmed by the parasite. Current molecular and paleontological evidence indicates that land plants originated between 425 and 490 million years ago, meaning that plants and terrestrial nematodes have coexisted in the earth's soil for an extensive period of time (Baldwin et al., 2004). Recently, nematodes were found in fossilised plant material of approximately 400 million years old, suggesting that nematodes invaded terrestrial plants quite early in plant evolution (Poinar et al., 2008). Plant parasitism of nematodes evolved at least three times independently: twice in the Enoplea class (orders Triplonchida and Dorylaimida) and once in the Chromadorea class (order Rhabditida) (Blaxter et al., 1998; De Ley & Blaxter, 2002). Similarly, animal parasitism among nematodes must have evolved at least four times independently (Blaxter et al., 1998). Although nematode parasites are generally assumed to have evolved from free-living ancestors, the precise origin and free-living sister taxa of each parasitic group are unknown (Blaxter et al., 1998). A phylogenetic tree illustrating the relationships between some selected nematode species is shown in Figure 1.1.

All plant-parasitic nematodes have evolved a stylet, a hollow needle-like structure to puncture the rigid plant cell wall. Through this stylet, they withdraw food from the plant cell, but also secrete proteins into the host tissue that facilitate parasitism. The three independently evolved nematode groups each have a different type of stylet (onchiostylet, odontostylet and stomatostylet). This repeated emergence of a stylet during evolution suggests functional and perhaps developmental constraints on alternative mechanisms to feed on plants (Baldwin et al., 2004). Other morphological adaptations to plant parasitism include the structure of the pharynx and two sets of pharyngeal glands, dorsal and subventral. These glands are important for the production of the proteins which are secreted through the stylet (Vanholme et al., 2004).

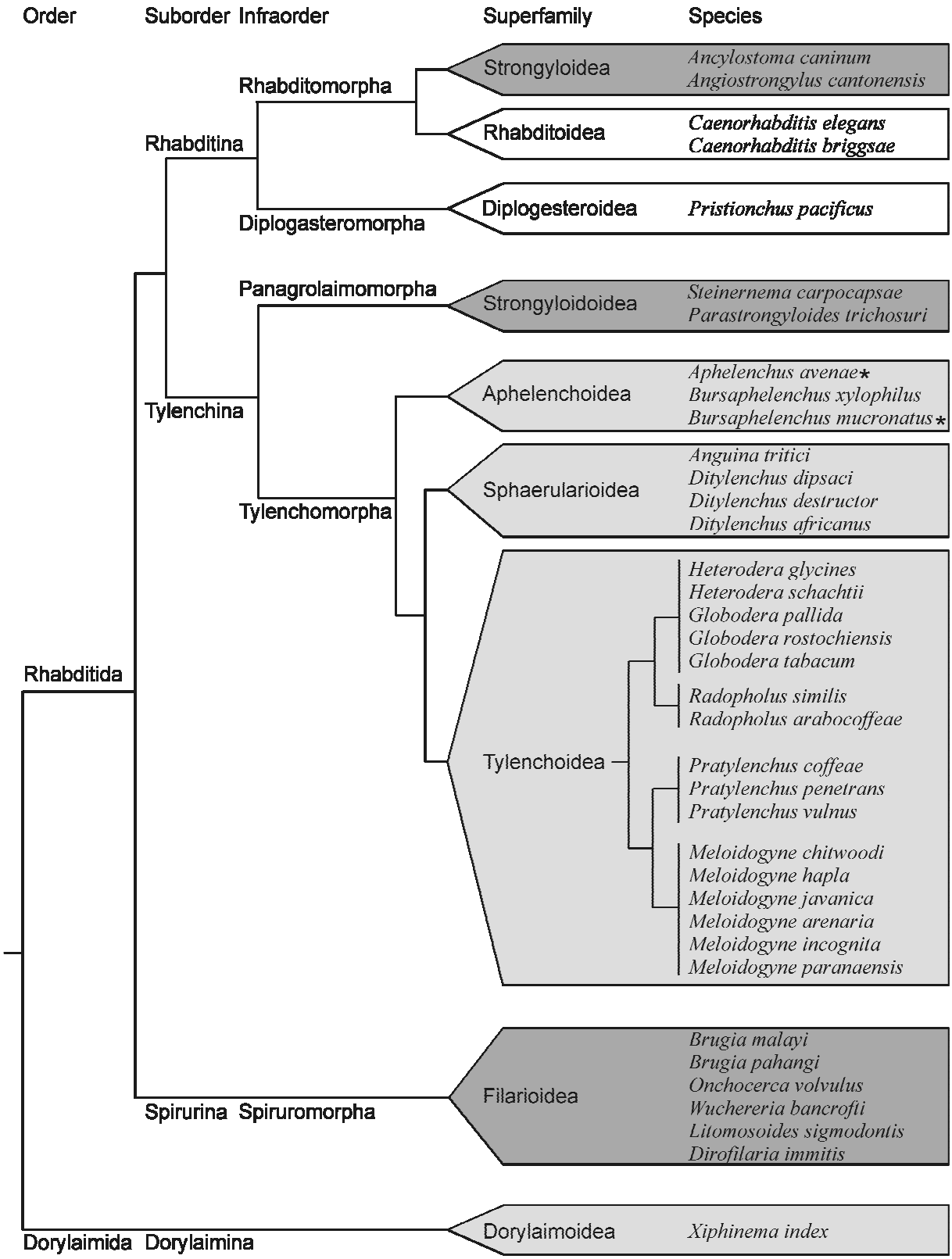

Figure 1.1: Phylogenetic relationships (not drawn to scale) between nematode species mentioned in this thesis as inferred from Bert et al. (2008), De Ley & Blaxter (2002), Holterman et al. (2009) & Mitreva et al. (2005). Light grey: plant-parasitic nematodes (stars indicate primarily fungal feeders); dark grey: animal-parasitic nematodes; white: free-living nematodes.

The different lifestyles of plant-parasitic nematodes

Even within the plant-parasitic nematodes, the mode of interaction with their host can be very variable. They can feed on all parts of the plant including roots, stems, leaves, flowers and seeds. Specialisations in plant-parasitic lifestyles include migratory ectoparasitism and burrowing endoparasitism as well as various types of independently evolved sedentary ecto- and endoparasitism and a full range of intermediates (Baldwin et al., 2004). The three main types of plant parasitism are (Decraemer & Hunt, 2006):

- Ectoparasitic: the nematode remains in the soil and feeds by using the stylet to puncture plant cells. The longer the stylet, the deeper it can feed within the plant tissues.
- Endoparasitic: the entire nematode penetrates the root tissue. Migratory endoparasites retain their mobility and have no fixed feeding site, whereas the more advanced sedentary endoparasites induce a specialised fixed feeding site such as a syncytium or a giant cell. This feeding site enhances the flow of nutrients from the host and therefore allows the females to become sedentary.
- Semi-endoparasitic: Only the anterior part of the nematode penetrates the root, with the posterior part remaining in the soil phase.

It must be noted that the above categories are not mutually exclusive as some genera may be for example semi-endoparasitic or migratory ecto-endoparasitic.

Parasitism genes

Nematode secretions

The most evolutionary advanced adaptations for plant parasitism by nematodes are the products of so-called parasitism genes (Gao et al., 2002). Parasitism proteins are secreted from the nematode and play a direct role in parasitism. These secretions mostly originate from the pharyngeal gland cells, but secretions from the chemosensory amphids might also be important (Davis et al., 2004). The arrangement of the pharyngeal glands is of taxonomic and phylogenetic importance (Decraemer & Hunt, 2006). Within the Triplonchida order five pharyngeal glands are present, while in the Dorylaimida, the number of gland cells can be four or five (Baldwin et al., 2004). Within the infraorder Tylenchomorpha (order Rhabditida), plant-parasitic nematodes have three gland cells, one dorsal and two subventral (Figure 1.2). Cytological observations in sedentary nematodes confirmed that the subventral glands are highly metabolically active during penetration and migration in the host tissue. Their secretory activity decreases with the establishment of the feeding site, at which time the dorsal gland activity increases (Vanholme et al., 2004). Although the phytonematode gland cells

function as true animal secretory cells with Golgi-derived membrane-bound secretory granules and subsequent exocytosis of granule contents through specialised valves into the pharyngeal lumen, parasitism protein secretion via the nematode stylet may be compared to type III secretion systems of plant-pathogenic bacteria with correspondingly similar effector molecule delivery to recipient host cells via a pilus (Davis et al., 2008). Interestingly, the adaptation of enlarged pharyngeal secretory cells is also present in nematode parasites of vertebrates, but absent in non-parasitic nematodes (Davis et al., 2008).

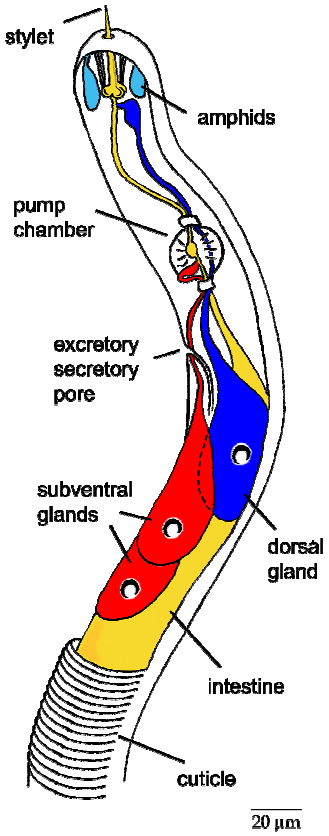

Figure 1.2: Body plan of a tylenchid nematode (Vanholme et al., 2004).

Identification of parasitism genes

Identification of parasitism genes has proven to be difficult due to the small size of plant-parasitic nematodes, which makes it hard to collect sufficient material for analysis (Vanholme et al., 2004). Immunoaffinity purification was used to enrich secreted proteins, resulting in the finding of a secreted protein (endo-1,4-β-glucanase) from the subventral glands of the potato cyst nematode *Globodera rostochiensis* (Smant et al., 1998). Another method is the analysis of collected nematode secretions by 2D gel electrophoresis and microsequencing. This has been successful for the beet cyst nematode *Heterodera schachtii* (De Meutter et al., 2001) and the root-knot nematode *Meloidogyne incognita* (Jaubert et al., 2002b). More recently, mass spectometry was used for direct identification of proteins secreted by *M. incognita*, revealing proteins with host cell reprogramming potential (Bellafiore et al., 2008). However, the transcriptomics approach became more popular than the proteomics approach. Expressed sequence tags (ESTs) can be used for homology searches to identify possible new parasitism genes. EST sequencing has resulted in several new interesting genes from different nematode species (Popeijus et al., 2000a; Dautova et al., 2001; Jaubert et al., 2002a). Other methods compare gene expression between preparasitic and parasitic stages of nematodes, such as differential display (Ding et al., 1998) or cDNA-AFLP (Qin et al., 2000). Elling et al. (2009) combined microarray expression data of all major life stages of *Heterodera glycines* with sequence analyses to identify parasitism associated proteins. Such integrated approaches will become increasingly important in the future as more sequence data is becoming available.

The plant cell wall

Plant cell walls are of fundamental importance for normal plant growth and development. The dynamic wall surrounding plant cells regulates cell expansion, controls tissue cohesion, is involved in the exchange of ions and forms a defense against external attacks from parasites or pathogens (Popper, 2008). Cell walls are composed of three types of layers: the middle lamella, the primary cell wall and in some cases the secondary cell wall (Figure 1.3).

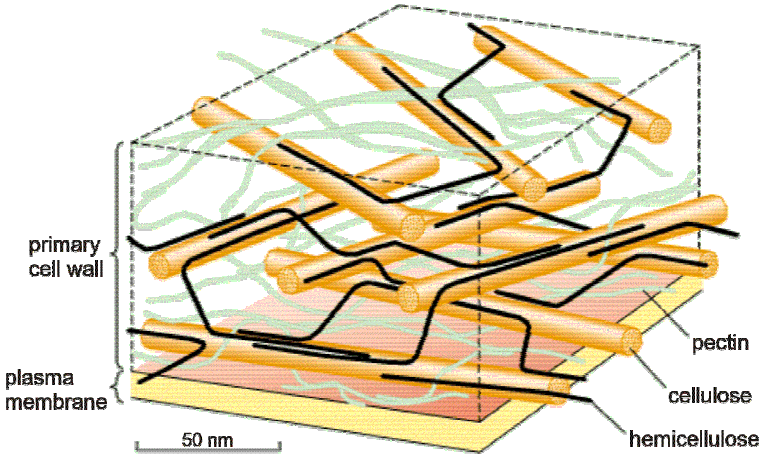

Figure 1.3: Structure of the plant cell wall (Lodish et al., 2000).

The middle lamella, which is rich in pectin, is deposited soon after mitosis and connects two adjacent plant cells (Cosgrove, 2005). The primary cell wall is formed throughout cell growth and expansion while the secondary cell wall, that is not always present, is deposited internally to the primary cell wall once cell growth has ceased.

The key structural component of the plant cell wall is cellulose, the most abundant biopolymer in the world. Cellulose is comprised of successive glucose residues which are inverted 180°, forming a flat ribbon with cellobiose as the repeating unit (Figure 1.4) (Taylor, 2008). These (1,4)-β-linked glucan chains are able to form extensive hydrogen bounds to adjacent glucan chains (Mutwil et al., 2008). Approximately 36 of these crystalline chains are arranged in parallel in 3 nm thick microfibrils forming insoluble cable-like structures. The cellulose microfibrils are among the longest molecules known in nature, since they are believed to consist of 8000 (primary cell wall) to 15,000 (secondary cell wall) glucose molecules (Somerville, 2006).

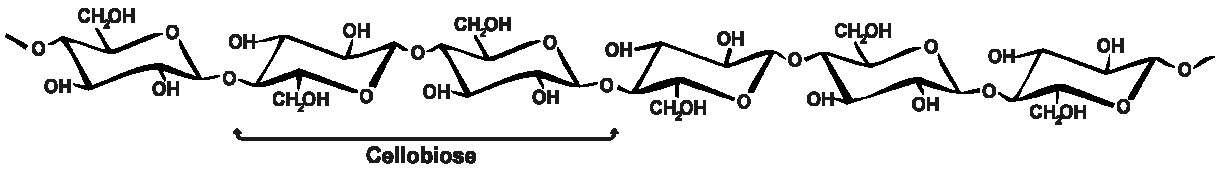

Figure 1.4: A fragment of a (1,4)-β-glucan chain showing inversion of adjacent sugar residues. The repeating unit cellobiose is indicated (Taylor, 2008).

Cellulose microfibrils are embedded in a matrix of complex polysaccharides, which are divided into two classes: pectins and hemicellulose. Pectins are a complex and heterogenous group of polysaccharides which are believed to be covalently linked together. Pectins are perhaps the most complex polysaccharides in the living world, including homogalacturonan, rhamnogalacturonans I and II, galactans, arabinans and other polysaccharides. They can be solubilised by aqueous buffers and dilute acidic solutions. Hemicelluloses on the other hand, require strong alkali for solubilisation. Hemicelluloses are cellulose-binding polysaccharides that form a strong but resilient network together with the cellulose. They include xyloglucans, xylans and mannans which have a (1,4)-β-linked backbone (D-glucan, D-xylan or D-mannan) similar to the structure of cellulose (Cosgrove, 2005).

Cell wall modifying enzymes in plant-parasitic nematodes

To invade plant tissue, plant-parasitic nematodes face the physical barrier formed by the plant cell wall. To overcome this barrier, plant-parasitic nematodes produce cell wall degrading and modifying proteins in the pharyngeal glands and secrete these enzymes through the stylet. The combinatorial effect of this enzymatic activity and physical damage caused by the protrusions of the stylet enable the nematode to break down the cell walls. Over the last decade, different cell wall degrading enzymes have been isolated from various plant-parasitic nematodes, mainly sedentary ones. These enzymes generally occur in extensive gene families, and recently the genome of *Meloidogyne incognita* was shown to possess more than 60 cell wall degrading enzymes (Abad et al., 2008).

Cellulases or endo-1,4-β-glucanases, that can degrade cellulose, have been found in the nematode genera *Heterodera*, *Globodera* and *Meloidogyne* (Smant et al., 1998; Rosso et al., 1999; Bera-Maillet et al., 2000; Goellner et al., 2000; Yan et al., 2001; Gao et al., 2002; Gao et al., 2004b; Ledger et al., 2006; Abad et al., 2008; Rehman et al., 2009a), which are all classified into the superfamily of the Tylenchoidea (order Rhabditida, suborder Tylenchina, infraorder Tylenchomorpha) (De Ley & Blaxter, 2002). Besides these sedentary nematodes, cellulases have also been found in the migratory nematode *Pratylenchus penetrans* (Uehara et al., 2001). Glycosyl hydrolases are classified into different families according to their sequence similarity (Henrissat & Bairoch, 1996). All endo-1,4-β-glucanases mentioned above belong to glycosyl hydrolase family 5 (GHF5). Interestingly, in *Bursaphelenchus xylophilus* which belongs to a different superfamily (Aphelenchoidea), endo-1,4-β-glucanases were cloned of GHF45 (Kikuchi et al., 2004). Moreover, in *Xiphinema index*, a nematode from a different order (Dorylaimida), GHF12 endoglucanases were found (Jones et al., 2005). The fact that different groups seem to have endoglucanases from a different glycosyl hydrolase family is probably a reflection of the separate evolutionary history of these gene families.

Endo-1,4-β-glucanases can degrade polysaccharides with a (1,4)-β-glucan backbone, such as cellulose, but in some cases also hemicelluloses like xylans. Another enzyme acting on

hemicellulose is endo-1,4-β-xylanase, identified in *Meloidogyne incognita* (Mitreva-Dautova et al., 2006; Abad et al., 2008).

Some nematode species migrate intercellularly through the middle lamina, rich in pectic polysaccharides. Pectate lyases can cleave internal (1,4)-α-linkages of pectate and have been identified in *Meloidogyne* spp., but also in intracellularly moving genera such as *Globodera*, *Heterodera* and *Bursaphelenchus* (Popeijus et al., 2000b; Doyle & Lambert, 2002; Huang et al., 2005; Kikuchi et al., 2006; Vanholme et al., 2007; Abad et al., 2008). Furthermore, a polygalacturonase that cleaves (1,4)-α-linkages in pectate by hydrolysis was identified in *Meloidogyne incognita* (Jaubert et al., 2002a; Abad et al., 2008). In *Heterodera* spp., a putative endo-1,4-β-galactosidase is thought to hydrolyse (1,4)-β-galactan in the hairy regions of pectin. The latter enzymes seem restricted to the genus *Heterodera*, since no homologs were found in the genome of *M. incognita* or *M. hapla* (Vanholme et al., 2009). Nevertheless, other putative cell wall degrading enzymes identified in the genome could have similar functions, such as arabinases (Abad et al., 2008).

Other cell wall modifying proteins do not have any hydrolytic activity. One of these proteins is expansin that weakens the non-covalent interactions between cellulose and hemicellulose, inducing plant cell wall extension. Expansin-like proteins were found in *Globodera rostochiensis* and *Bursaphelenchus* species (Qin et al., 2004; Kudla et al., 2005; Kikuchi et al., 2007). In the *M. incognita* genome, an extensive gene family of a total of 20 putative expansin-like proteins was identified (Abad et al., 2008). Another group of cell wall acting proteins without any enzymatic activity are cellulose-binding proteins, isolated from *H. glycines* and *M. incognita* (Ding et al., 1998; Gao et al., 2004a). These proteins may have an indirect role in plant cell wall degradation by activating a plant pectin methylesterase (Hewezi et al., 2008).

Other parasitism genes involved in plant-nematode interactions

Next to plant cell wall modifying enzymes, other proteins play a role in the infection process, especially in sedentary nematodes. Since sedentary nematodes induce a feeding site, several proteins are being secreted by the nematode that can for example alter the gene expression of the plant. *Heterodera glycines* secretes a protein similar to the CLAVATA/ESR (CLE) class of plant signal peptides, probably inducing feeding cell differentiation (Olsen & Skriver, 2003). Another example of a nematode protein interacting with plant pathways is chorismate mutase, found both in root-knot and cyst nematodes (Lambert et al., 1999; Jones et al., 2003; Bekal et al., 2003; Long et al., 2006). This enzyme can affect the synthesis of chorismate derived compounds among which are precursors of cell wall synthesis, plant hormone biosynthesis and plant defense compounds.

Plant-parasitic nematodes can also partly counter the defense mechanisms of the host plant. It was shown that nematodes secrete proteins involved in protein degradation, such as ubiquitin,

RING-H2 and SKP-1 (Gao et al., 2003). These proteasome complex proteins can possibly target specific host cell proteins for degradation and therefore contribute to the establishment of the feeding site. Several other proteins have been identified that could play a role in parasitism, although it remains unclear what their actual function is. Venom allergen proteins for example are being secreted from both cyst and root-knot nematodes into the plant tissue (Ding et al., 2000; Gao et al., 2001; Wang et al., 2007; Lu et al., 2007). These proteins are also secreted by animal-parasitic nematodes where they induce host immune responses, but their function remains unknown (Jasmer et al., 2003). Other putative parasitism proteins are SPRYSEC proteins, calreticulin, SXP-RAL2 and 14-3-3 family proteins which may play roles in alteration of host cell cycle, calcium binding, defense modulation and as cellular chaperones (Davis et al., 2008; Rehman et al., 2009b). Some putative parasitism proteins have nuclear localisation signals that are functional in plant cells (Elling et al., 2007), suggesting some host cell regulation by nematodes at the transcriptional level (Huang et al., 2006b).

Origin of parasitism genes

Some likely mechanisms for acquisition of parasitism genes include: 1) adaptation of pre-existing genes to encode new functions, 2) gene duplication and divergence of paralogs, and 3) horizontal gene transfer (HGT) (Scholl et al., 2003). One of the most remarkable findings concerning parasitism genes from plant-parasitic nematodes is that some of these genes are absent from all other nematodes and most other animals studied to date (Jones et al., 2005). Many of these genes resemble bacterial sequences, suggesting that these enzymes could have been acquired from bacterial plant pathogens through horizontal gene transfer. For example the nematode endo-1,4-β-glucanases from the Tylenchomorpha, which belong to GHF5, show very little similarity to plant endoglucanases but are rather homologous to bacterial sequences. This has led to the conclusion that these enzymes were at some point in evolution acquired from bacteria, and later extensive gene duplication resulted in gene families. Remarkably, the GHF45 endoglucanase from *Bursaphelenchus xylophilus* shows the highest homology to fungal sequences. Since *Bursaphelenchus xylophilus* is a facultative fungal feeder, it makes sense that this gene was acquired from fungi. Moreover, an endo-1,3-β-glucanase (GHF16) is present in *B. xylophilus* and the fungal feeder *B. mucronatus* with sequence characteristics that suggest it was acquired by HGT from bacteria (Kikuchi et al., 2005). Endo-1,3-β-glucans are important components of fungal cell walls and the protein is likely to play a role in fungal feeding. It is possible that these nematodes acquired endo-1,3-β-glucanase genes from bacteria to obtain a fungal feeding ability, and a subgroup subsequently acquired cellulase genes from fungi, which permitted them to parasitise plants (Jones et al., 2005).

Next to the parasitism genes, other genes in plant-parasitic nematodes have also been attributed to bacterial origins, for example a polyglutamate synthase gene or a pathway involved in the synthesis of vitamin B6 (Veronico et al., 2001; Craig et al., 2008). In animal-

parasitic nematodes and insects, it was proven that genes derived from their bacterial symbiont *Wolbachia* are present in the genome of the nematode or insect (Hotopp et al., 2007). This suggests that horizontal gene transfer may play a more important role in the evolution of nematodes than originally thought. However, the horizontal gene transfer hypothesis should be handled cautiously. The most commonly used - and in many cases the only - reason to claim that genes are of HGT origin is that no homologous genes can be found in other eukaryotes, only in bacteria. But one should be aware that the sequence data available nowadays is only from a very limited amount of species. An example of how this can lead to wrong assumptions is the presence of GHF9 endoglucanases in insects, which were thought to be acquired through HGT. However, as more sequence data became available, it seemed that it was rather a case of extensive gene loss in other lineages than HGT (Lo et al., 2003; Davison & Blaxter, 2005). Could this also be the case for the nematode GHF5 endoglucanases? It will probably become more clear in the future when more sequence data will become available, although endogenous GHF5 and GHF45 endoglucanases have also been found in insects and molluscs (Girard & Jouanin, 1999; Xu et al., 2001; Sugimura et al., 2003; Lee et al., 2004). The problem is that there are no strict objective rules about how to test whether a given gene was acquired from another non-related organism via HGT. Nevertheless, a combination of methods can be proposed that together may lift the weaknesses of the individual approaches. For example, a combination of phylogenetic methods, analysis of the distribution pattern, and habitat overlap between inferred donor and recipient could be applied (Mitreva et al., 2009).

Nematode under study: *Radopholus similis* (Cobb, 1893) Thorne, 1949

Introduction

Radopholus similis or the burrowing nematode was first discovered by the famous nematologist Nathan A. Cobb in 1891 when he studied banana roots from Fiji. In the same publication in 1893 he described the male and female separately as two different species, respectively *Tylenchus similis* and *Tylenchus granulosus*. Later (1906) he described a third species, *Tylenchus biformis*, when he found males and females together. It was only in 1968 that the nomenclatural problems were overcome and the name *Radopholus similis* was retained (Luc, 1987).

The burrowing nematode is a migratory endoparasitic nematode, widely known as a destructive pest of citrus, black pepper and banana. It is an active parasite on more than 250 different plant species throughout the tropical and semi-tropical world. The nematodes have completely destroyed black pepper plantations in Indonesia and have caused very large losses to citrus properties. They continue to be a major pest problem for the banana industry in many

parts of the world and have been introduced to European glasshouses by way of trade with ornamental plants, in particular *Anthurium* (EPPO diagnostics report, 2008).
Two physiological races of *R. similis* were recognised: one race parasitises banana and many other hosts but not citrus, while the other race parasitises both banana and citrus. Although no significant morphological differences were detected between the two races, the citrus race has been raised to sibling species rank and designated as *R. citrophilus* on the basis of putative biochemical, physiological and karyotypic differences (Huettel et al., 1984). However, more recent investigations suggest that citrus parasitism appears to be associated with only limited changes in the burrowing nematode genome and do not support assignment of sibling species status with respect to citrus parasitism (Kaplan & Oppermann, 1997). Moreover, citrus and non-citrus parasitic burrowing nematodes are not reproductively isolated and there is no possibility to distinguish races of *R. similis* attacking citrus and those not attacking citrus (EPPO diagnostics report, 2008). Therefore, *R. citrophilus* was later correctly considered as a junior synonym for *R. similis* (Valette et al., 1998; Kaplan et al., 2000).

Taxonomy and phylogeny

Radopholus similis is the only one of more than 30 species in the genus recognised as a pathogen of widespread economic importance (Duncan & Moens, 2006). It is a member of the family of the Pratylenchidae within the Tylenchoidea superfamily of the Rhabditid order. Molecular analyses have however proven that the Pratylenchidae family is polyphyletic and that *R. similis* is closely related to ectoparasitic and cyst-forming endoparasitic nematodes (Hoplolaimidae and Heteroderidae) (Subbotin et al., 2006; Bert et al., 2008; Holterman et al., 2009). The inclusion of the burrowing endoparastic nematodes in a single family Pratylenchidae is classically defined by similar morphological characteristics that are likely the result of convergent evolution related to similar feeding modes. With developing understanding of plant-nematode interactions in the economically very important genera *Radopholus* and *Pratylenchus*, wrong assumptions of monophyly of burrowing endoparasitic nematodes versus convergence have critical implications when extrapolating insights of one group (i.e. *Pratylenchus* spp.) to another (i.e. *Radopholus* spp.) (Bert et al., 2008).

Some morphological characteristics

The female burrowing nematode is 650-800 µm long by 20-24 µm diameter. The vulva is located at approximately 54% of the body length from the head. The head is rounded and slightly flattened, offset by a slight constriction, and is supported by a sclerotised framework. The stylet, 18 µm long, is plainly visible and has prominent knobs. The pharyngeal glands overlap the intestine dorsally. Males do not resemble females in gross appearance. They are 500-600 µm long and more slender than females. The head, set off by a conspicuous

constriction, is well rounded and non-sclerotised. The stylet is quite slender and indistinct, 12 µm long, and has small knobs, which are difficult to see (MacGowan, 1977) (Figure 1.5).

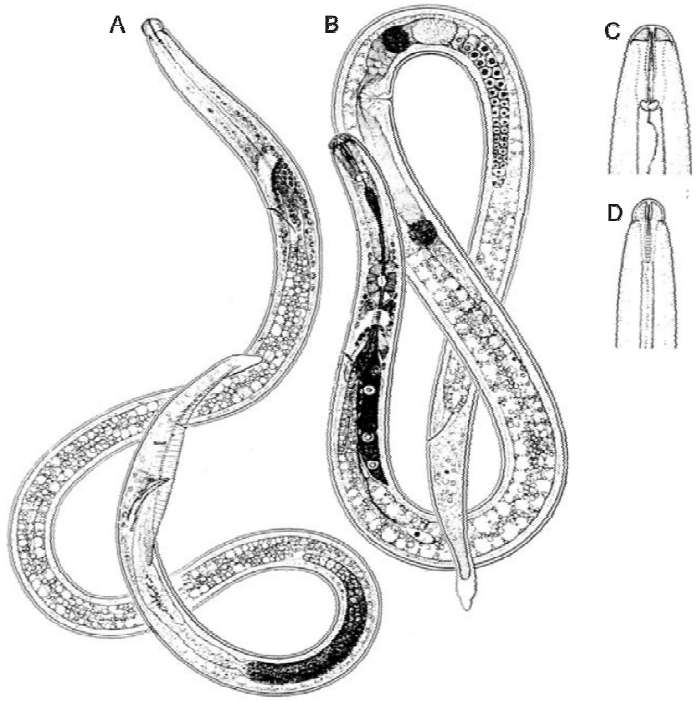

Figure 1.5: *Radopholus similis*. (A) Male; (B) Female; (C) Female head; (D) Male head (EPPO diagnostics report, 2008).

Biology

The egg to egg life cycle of the burrowing nematode is usually completed within 3 weeks. All stages are found inside the root. The nematode punctures the root epidermis cells with its stylet, digesting and sucking the cell contents as it burrows into the root. Entry is usually at the root tip or in the region of root hair production and takes less than 24 hours. The burrowing nematode attacks only tender young feeder roots and not hardened, suberised, senescing or decayed roots. Once inside the root, the worms feed and reproduce. The male, with its rudimentary stylet, is not known to penetrate roots nor to injure them to any extent. The female and juveniles feed on roots. Females may lay 1-6 eggs a day inside the roots. Lesions formed as a result of nematode activity enlarge and coalesce. Nematodes will at this point, or because of population pressure, leave the root and migrate into the soil. During this stage while the worm is seeking a new food source of healthy roots, the area of infestation spreads to adjacent plants (MacGowan, 1977).

The burrowing nematode's ability to survive adverse conditions is enhanced by 3 factors: 1) an extensive host range, 2) a short life cycle allowing rapid reproduction during favorable periods and 3) the ability of females to reproduce for one or two generations without males (MacGowan, 1977). The latter reproduction strategy was thought to be parthenogenesis, but it was shown that *R. similis* rather is a hermaphrodite. In females that have not mated with a

male 50 to 60 days after the fourth moult, self-fertilisation takes place with rod-shaped spermatids produced in the ovatestis and matured in the spermatheca (Kaplan & Opperman, 2000).

Most populations of *R. similis* reproduce best at intermediate (25°C) or high (30°C) rather than low (15-20°C) temperatures. Populations introduced into European parks and nurseries appear to have adapted to temperate conditions, reproducing at temperatures too low for most tropical populations. In Florida citrus plantations, the optimum temperatures occur longest each year in the deeper soil horizons where root infection is greatest. Temperature extremes in the surface soils are nearer the limits for nematode development, which may explain low population development in surface roots. The nematode does not have a known resting stage, so recurring moisture deficits typical of surface soils may also inhibit development near the soil surface (Duncan & Moens, 2006).

Symptoms and damage

Root damage caused by *R. similis* has typical symptoms caused by lesion nematodes: reddish, brownish to black lesions caused by cell wall collapse as nematodes move inter- and intracellularly (Duncan & Moens, 2006). The roots have extensive cavities and the phloem and cambium may be completely destroyed, leaving nematode-filled spaces separating the stele from the cortex. External cracks may appear over the lesion (EPPO diagnostics report, 2008). Tissue rot occurs following secondary infections with fungi and bacteria (Duncan & Moens, 2006). Infested plants also show poor growth, reduced leaf size and colour alterations, and are more susceptible to fungal infections. Infested citrus trees have fewer and smaller leaves and more dead twigs than healthy trees due to reduced uptake of water and nutrients. Because of the infection with *R. similis*, the root system of the plants is weakened, which leads to the toppling over of infected plants, especially those bearing fruits, after a strong gust of wind ("toppling over disease"). This is the most obvious symptom in the field of nematode infection (EPPO diagnostics report, 2008).

Figure 1.6: Toppling over of plants in a banana plantation in Martinique due to infection with *R. similis* (http://www.ird.fr/).

R. similis is considered to be among the ten most damaging plant-parasitic nematodes world-wide (EPPO diagnostics report, 2008). It is of great economic importance in the banana-growing areas of Australia, Central- and South America, Africa and the Pacific and Carribean Islands, causing what is variously called root rot, blackhead and toppling disease. In Central and South America for example, crop losses due to toppling over of banana plants fluctuate between 12 and 18% (Sarah et al., 1996). A study on plantains in Cameroon shows that the yield can be reduced by more than 50% over the first two crop cycles due to nematode infection (Fogain, 2000). *R. similis* was also responsible for the yellows disease of black pepper, which by 1953 had destroyed 90% of this crop in Indonesia. In Florida, the citrus pathotype causes so-called "spreading decline" in citrus fruits, reducing yields by 40-70% (oranges) and 50-80% (grapefruits) (EPPO quarantine pest data sheet on *R. citrophilus* and *R. similis*). As a result of these devastating effects of *R. similis*, the European and Mediterranean Plant Protection Organisation (EPPO) has classified the banana race as an A2 quarantine organism, and the citrus race as A1 since the latter is absent from Europe.

Control measures

Reducing nematode populations in the soil before planting and the use of nematode-free planting material are of primary importance in the control of *R. similis*. Nematode populations can be reduced to an undetectable level by fallow or crop rotation with non-host plants (e.g. sweet potato, pineapple), although six or seven weeks of flooding can be equally effective. However, the latter method is often difficult to apply as flooding requires a permanent water supply and the land to be leveled (Sarah et al., 1996). Additionally, a simple natural fallow period is often not sufficient to eliminate the burrowing nematodes and to clean the land properly. Several factors have to be considered, such as the efficiency of the destruction of the old banana plants and the evolution of the flora susceptible to *R. similis*. It was shown that the removal of old banana plants by a chemical (glyphosate) is much more efficient to reduce *R. similis* populations levels than mechanical destruction with a spading machine (Chabrier & Quénéhervé, 2003). Nematicides (generally organophosphates or carbamates) are traditionally the primary way to control nematodes. Nematicides can increase banana yields with 50% compared to untreated controls (Fogain, 2000). In large commercial plantations of banana, nematode control is based on the application of two to four nematicide treatments each year, having negative toxic side-effects on the environment (Chabrier & Quénéhervé, 2003). Nowadays, more attention is being paid to an integrated pest management approach. In Martinique for example, an alternative cropping system has been developed based on the cleanup of contaminated fields prior to planting. This cleanup is done through either a fallow period or an appropriate crop rotation, and then nematode-free *in vitro* banana plants are planted. As a consequence, growers are able to cultivate bananas for at least two to three years without nematicide application. Moreover, the introduction of ditches could efficiently isolate

field sectors and partially protect other banana plants from infestation (Chabrier & Quénéhervé, 2008).

Host resistance is an effective means of nematode control in many crops. In bananas and plantains, there has been little progress in breeding for resistance to nematodes due to the genetic complexity within the genus *Musa*, and to the easy choice of chemical nematicides in the past. The increasing awareness about the negative effects of pesticides has led to a restricted number of permitted nematicides, not only in the European Union, but also in the countries exporting to the EU. The genetic engineering of bananas may be an important strategy to reduce nematicides. It was shown that a transformed Cavendish banana expressing a rice cystatin, a nematode cysteine proteinase inhibitor, has a resistance level against *R. similis* of 70% (Atkinson et al., 2004). Alternatively, different banana hybrids are being screened for nematode resistance. Promising cultivars have been identified, even partially resistant to other nematodes as well, which will hopefully lead to an improved sustainability of banana production (Quénéhervé et al., 2009a; Quénéhervé et al., 2009b).

Nematode under study: *Ditylenchus africanus* Wendt, Swart, Vrain and Webster, 1995

Introduction

The peanut pod nematode, *Ditylenchus africanus*, was first found in hulls and seeds of peanut in South Africa in 1987. The species remained unnoticed for so long probably because the symptoms it causes are very similar to a fungal disease and because the nematodes themselves have a weak stylet and resemble harmless fungivorous nematodes. The nematodes isolated from the infected peanut pods were first identified as *Ditylenchus destructor*, the potato rot nematode (De Waele et al., 1997). *D. destructor* was known mainly as an important pest of potato tubers and bulbs of flowers in temperate regions. However, the South African isolate caused no damage to potato tubers and was therefore designated as a new race and ecotype (De Waele et al., 1991). Later, based on molecular analyses, the South African race of *D. destructor* was considered to be a new species and described as *D. africanus* (Wendt et al., 1995).

Taxonomy

D. africanus is a member of the Anguinidae family, superfamily Sphaerularioidea within the infra-order Tylenchomorpha of the order Rhabditida (De Ley & Blaxter, 2002). Of several economically important genera in the Anguinidae family, *Ditylenchus* spp. have the widest impact on agriculture (Duncan & Moens, 2006). The stem nematode *D. dipsaci*, a species complex with an extremely wide host range, is one of the most devastating plant-parasitic nematodes, especially in temperate regions (www.eppo.org).

Some morphological characteristics

The length of the male and female of *D. africanus* ranges from 550 to 1150 µm, while the width is between 20 and 30 µm. The nematodes have a flattened head with a delicate stylet of 8 to 10 µm long with distinct, separated, backwards sloping knobs. *D. africanus* differs from *D. destructor* mainly in the stylet length (8-10 µm vs. 10-14 µm) and spicule length (15.2-22 µm vs. 24-27 µm) (De Waele et al., 1997). More detailed studies should be done to confirm the morphological differences between the two species.

Biology

D. africanus is a migratory endoparasite which is mainly found in groundnut (*Arachis hypogea* L.), but can survive on weeds and on a variety of crops such as corn, soybean, tobacco and wheat. While *D. africanus* is present in the roots of peanut and the soil, about 90% of the total population at harvest is found in the pods. Soon after the peg of the groundnut plant has entered the soil and pod formation has been initiated, the nematodes enter the plant tissues at the base of the pod near the point of connection with the peg. They can subsequently invade the parenchymous regions of the pods, and eventually the seed testa. Mature pods however have a lignified hull which seems to act as a barrier to penetration of the inner pod tissues (Venter et al., 1995). Remarkably, *D. africanus* has also been observed feeding and reproducing on the hyphae of common plant-pathogenic fungi (De Waele et al., 1997). Since males and females occur in similar numbers, amiphimixis is thought to be the general mode of reproduction. The life cycle only comprises 6-7 days when the temperature is optimal (28°C). *D. africanus* can undergo dehydratation and enter a state of anhydrobiosis, which makes it possible for the nematodes to survive in the absence of host plants for at least 32 weeks (De Waele et al., 1997).

Symptoms and damage

In contrast to most other *Ditylenchus* spp., *D. africanus* occurs only in underground parts of the plant, such as roots, pods and pegs. The first symptom is the appearance of dark brown tissues at the pod base where the peg joins the pod, which can cause the pod to break off during harvesting (De Waele et al., 1997). In more advanced infections, cells collapse and tunnels are formed that facilitate the migration of the nematodes (Venter et al., 1995). The most visible result of a severe infection is a black discoloration along the longitudinal veins of the hulls. Although *D. africanus* is present in the roots, there are no visible lesions. Infected seeds are usually shrunken and testae and embryos have a yellow to brown or black discoloration (Figure 1.7) (De Waele et al., 1989).

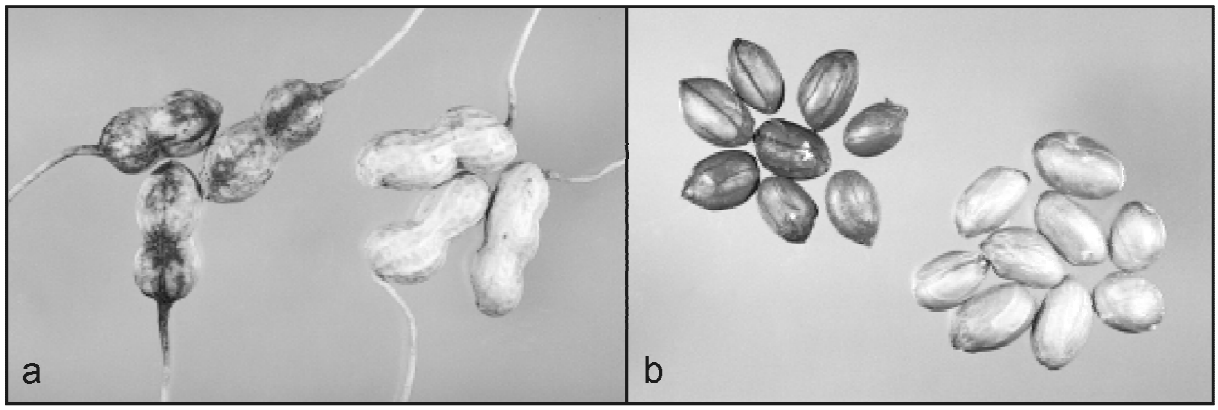

Figure 1.7: Symptoms on peanut caused by *Ditylenchus africanus*. (a) pods; (b) seeds. Left: infected; right: healthy (Kleynhans, 1999).

The peanut pod nematode has been found in all the major peanut production areas of South Africa where 73% of all seed samples examined were infected with *D. africanus* (De Waele et al., 1989). The widespread distribution in South Africa suggests that it may also be present in other southern African countries. In Mozambique, Malawi and Congo peanut pods showing symptoms typical of *D. africanus* infection have been reported (De Waele et al., 1997). The short life cycle of the nematode allows it to build up extremely high population densities during the growing season. When peanut seedlings are inoculated with 500 nematodes per plant, about 150,000 nematodes are present per 5g fresh hulls or seeds after 18 weeks. Moreover, as few as 50 *D. africanus* individuals per seedling can cause crop failure due to downgrading of the seed (Bolton et al., 1990). Annual yield losses of peanut due to nematodes are estimated at 12% worldwide (Venter et al., 1991), and in fields heavily infested with *D. africanus*, 40 to 60% of the pods and seeds can be destroyed (De Waele et al., 1989).

Control measures

Methods to contain the *D. africanus* infection and reduce the crop damage are scarce. Although populations of *D. africanus* can be suppressed by nematicides, the activity of nematicides usually lasts for about eight weeks, which means that nematodes surviving this period can still cause a considerable amount of damage (Basson et al., 1992). Different peanut cultivars have been tested on resistance, but all are susceptible to nematode damage (Venter et al., 1993). It was however shown that an early harvest could reduce the crop losses to only 12-13%, while a harvest of 15 days later results in losses of 45-49% (Venter et al., 1992). Furthermore, since eggs and nematodes that have undergone anhydrobiosis survive in the winter in decaying hulls, removal of the decaying pods after harvesting should reduce the infection level (Venter et al., 1991). *D. africanus* can also survive on several different weed species, which is important when peanut is not grown (De Waele et al., 1990).

Nematodes and molecular biology

During the last decade, a tremendous amount of molecular data has been generated for all sorts of organisms, particularly nematodes. Over 1,000,000 ESTs are available from 40 different species, both free-living and parasitic (www.nematode.net). Moreover, different genome projects have unraveled the complete genomes of free-living nematodes *C. elegans* (*C. elegans* sequencing consortium, 1998), *C. briggsae* (Gupta & Sternberg, 2003) and *Pristionchus pacificus* (Dieterich et al., 2008) as well as the animal-parasitic nematode *Brugia malayi* (Ghedin et al., 2007) and many other genome sequencing projects are in progress. Recently, the genome of two root-knot nematodes, *M. incognita* (Abad et al., 2008) and *M. hapla* (Opperman et al., 2008) became available, as well as a rough assembly of the cyst nematode *H. glycines* (Monsanto company).
Next to all this primary sequence information, nematology has also benefited from new molecular tools. The function of different genes has been extensively studied by single gene knockout with the RNAi technique. For *C. elegans*, every single gene has been targeted and all resulting phenotypes were documented, providing information about the possible function of all these genes (Rogers et al., 2008). Nevertheless, we are only beginning to understand the biology of nematodes, especially of parasitic ones. Genes that are specific for parasitic species are particularly interesting, but functional studies are very demanding and will take several years for nematologists and molecular biologists to elucidate.

Scope

Since sedentary nematodes are the most specialised and economically important plant-parasitic nematodes, nearly all molecular studies involve sedentary nematodes. Almost no effort has been done in the study of migratory nematodes, which can also cause a considerable amount of plant damage. The main goal of this thesis was to search for parasitism genes in migratory nematodes, especially cell wall modifying enzymes, and compare these to their homologs in sedentary nematodes.
The first part of this thesis describes the identification of endo-1,4-β-glucanases or cellulases in different migratory nematodes by homology based PCR cloning. GHF5 endoglucanases were found in *R. similis* (chapter 2), *D. africanus* (chapter 3) and *Pratylenchus coffeae* (chapter 4). Moreover, an evolutionary model concerning the gene structure of GHF5 endoglucanases within nematodes was constructed (chapter 4). Additionally, another cell wall modifying protein was identified by homology based cloning, namely an expansin-like protein in *D. africanus* (chapter 3).

Other techniques to search for novel parasitism genes involve a more high throughput EST analysis. This approach revealed the presence of an endo-1,4-β-xylanase in *R. similis* that is characterised in chapter 5, and was also applied to *D. africanus* (chapter 6). Finally, the *R. similis* ESTs revealed an unexpected amount of bacterial sequences with a high homology to a *Wolbachia* endosymbiont. The presence of this endosymbiont within the nematode was further investigated in chapter 7.

CHAPTER 2

A FAMILY OF GHF5 ENDO-1,4-BETA-GLUCANASES IN *RADOPHOLUS SIMILIS*

Adapted from:

Annelies Haegeman, Joachim Jacob, Bartel Vanholme, Tina Kyndt and Godelieve Gheysen. (2008). A family of GHF5 endo-1,4-beta-glucanases in the migratory plant-parasitic nematode *Radopholus similis*. Plant Pathology, 57, 581-590.

Abstract

Endo-1,4-β-glucanases, which can degrade cellulose, have been identified in a number of plant-parasitic nematodes, mainly sedentary endoparasites. We report the finding of four different endoglucanases of glycosyl hydrolase family 5 (GHF5) in the migratory endoparasitic nematode *Radopholus similis*. Spatial expression of these genes was analysed by *in situ* hybridisation, which showed the presence of transcripts in the pharyngeal gland cells. A semi-quantitative RT-PCR on different developmental stages was done to study the temporal expression pattern. Three of the endoglucanase genes have a reduced expression in adult males as opposed to females. This could be explained by the fact that males do not feed and are considered non-parasitic. Only one of the endoglucanase genes is expressed in juveniles. The four corresponding proteins have a putative signal peptide for secretion and a catalytic domain. Two of the proteins have an additional linker and carbohydrate binding module (CBM). Modelling of the catalytic domain resulted in the α/β-barrel typical for GHF5 endoglucanases. Mapping the conserved amino acids of the four endoglucanases onto the 3D structure revealed that most are positioned near the catalytic centre of the protein, whereas less conserved amino acids occur more often in the α-helices, pointing towards the outside of the protein. Analyses of the GC contents and codon adaptation indices indicate that the endoglucanase genes are well adapted to the codon usage of *Radopholus similis*. The GC and GC3 content of the endoglucanases is significantly higher compared to the average for all *Radopholus* ESTs.

Introduction

Plant-parasitic nematodes have different adaptations to penetrate the rigid plant cell wall. They damage the cells mechanically with their stylet, and enzymatically by secreting different cell wall modifying enzymes. The most extensively studied nematode cell wall degrading enzyme is endo-1,4-β-glucanase or cellulase, which can degrade cellulose. Cellulases were identified in the sedentary nematode genera *Heterodera, Globodera* and *Meloidogyne* (Smant et al., 1998; Bera-Maillet et al., 2000) These endo-1,4-β-glucanases (EC 3.2.1.4) belong to glycosyl hydrolase family 5 (GHF5) and are believed to be adopted from bacteria through horizontal gene transfer (HGT) (Jones et al., 2005). Interestingly, hypothetical proteins similar to GHF5 endoglucanases were recently found in the genome of the free-living nematode *Pristionchus pacificus*. It is not clear whether or not these enzymes are active or what their function is. Possibly, this is a kind of preadaptation towards a parasitic lifestyle (Dieterich et al., 2008).

Almost all GHF5 endoglucanases were identified in sedentary plant-parasitic nematodes, only two of the known GHF5 endoglucanases originate from a migratory nematode, namely *Pratylenchus penetrans* (Uehara et al., 2001). Another economically important migratory nematode in the Pratylenchidae family is the burrowing nematode *Radopholus similis*. All life stages of burrowing nematodes like *R. similis* penetrate into and migrate within the root cortex. This migratory lifestyle suggests that this nematode possesses different and multiple cell wall modifying enzymes. In this chapter we report on the cloning and characterisation of four endo-1,4-β-glucanases belonging to GHF5 in *R. similis*.

Materials and methods

Nematode culture

Radopholus similis was maintained on carrot disks in small petri dishes (∅ 35 mm) at a constant temperature of 25°C as described by Moody et al. (1973). Nematodes were collected by rinsing the petri dishes with sterile demineralised water 6-8 weeks after inoculation of the carrot disks with approximately 30 female adults. The collected nematodes were either used immediately for enzymatic assays or *in situ* hybridisation, or stored as a pellet at -80°C for DNA or RNA extraction. *Caenorhabditis elegans* Bristol N2 was obtained from the *Caenorhabditis elegans* Genetics Center (University of Minnesota, St. Paul). The strain was maintained under standard conditions as described by Brenner (1974).

Collection of secretions

Approximately 150,000 freshly harvested nematodes were collected in 2 ml sterile demineralised water with 400 µg/ml 5-methoxy-N,N-dimethyl tryptamine oxalate (DMT; Sigma-Aldrich, St. Louis, MO, USA) to stimulate the release of secretions (Goverse et al., 1994). Incubation was performed on a rotator for optimal aeration for 16h at room temperature. After incubation, the nematode suspension was centrifuged for 2 min at 250 g in a table-top centrifuge (Eppendorf, Hamburg, Germany). The supernatant, which contained the secretions, was collected and stored at 4°C up to a maximum of 6h. As a negative control, mixed stages of *R. similis* were collected and were allowed to settle in a tube. The supernatant was used as the negative control and was handled equally to the other samples.

Zymogram

Protein samples were dried in a SpeedVac vacuum centrifuge (Servant Instruments, Farmingdale, NY, USA) at room temperature and redissolved in 20 µl 100 mM Tris-HCl (pH 8.9), 10% glycerol and 0.0025% bromophenolblue (Sigma-Aldrich). Each sample was loaded on a native 10% polyacrylamide gel containing 0.1% carboxymethylcellulose (CMC; Acros Organics, Geel, Belgium), prepared in a Mini-Protean II Cell system (Bio-Rad Laboratories, Hemel Hempstead, UK). The SeeBlue Plus2 prestained protein marker (Invitrogen, Carlsbad, CA, USA) was used as a molecular weight standard. After running the gel (45 min at 45V followed by 1h at 150V), the gel was rinsed twice with demineralised water and incubated overnight at 37°C in 0.2 M Na_2HPO_4 and 0.2 M KH_2PO_4 (pH 6.0). The next day, the gel was rinsed twice with demineralised water and stained with 0.1% Congo Red solution for 30 min on a shaker. The gel was cleared by rinsing it twice with 1.0 M NaCl for 15 min.

Cellulase activity assays

Approximately 10,000 freshly harvested nematodes were crushed in 0.5 ml TE buffer (10 mM Tris-HCl, 1.0 mM EDTA, pH 8.0) to which 0.5 mM DTT was added. The sample was centrifuged for 5 min at 2000 g in a table-top centrifuge, and 2 µl of the supernatant was loaded on a 1.5% agar plate containing 0.5% CMC. Besides whole nematode extracts, 2 µl freshly purified secretions were spotted on the plate. After overnight incubation at 37°C, the plate was stained for 15 min with 0.1% Congo Red and subsequently washed for 15 min with 1.0 M NaCl. A protein extract from *C. elegans* as well as water rinsed from uninfected carrot disks were used as a negative control. As a positive control 1U cellulase of *Aspergillus niger* (Sigma-Aldrich) was used.

DNA and RNA isolation and cDNA synthesis

Genomic DNA was isolated from nematodes of mixed stages according to Bolla et al. (1988). RNA was extracted from mixed stages with TRIzol (Invitrogen) according to the

manufacturer's instructions. The RNA was treated with DNAse (Fermentas, St. Leon-Rot, Germany) to avoid DNA contamination. First strand cDNA was synthesised with SuperScript II Reverse Transcriptase (Invitrogen) in the presence of an oligodT primer. The reaction mixture contained 1 µg RNA as template, 4 mM dNTPs, 0.5 µM oligodT primer, 10 mM DTT, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$ and 200U SuperScript II Reverse Transcriptase. The mixture was incubated for 2h at 42°C. For the RT-PCR analysis, RNA was extracted from approximately 100 individuals and cDNA was prepared and amplified by the SMART PCR cDNA synthesis kit (Clontech, Palo Alto, CA, USA).

PCR amplification of gene fragments coding for endoglucanases

Genomic fragments were amplified by PCR using the degenerate primers ENG1 and ENG2 as described by Rosso et al. (1999) (Table 2.1). The 25 µl reaction mixture contained 150 ng DNA as template, 0.5 µM of each primer, 4 mM each of dNTPs, 1.5 mM $MgCl_2$, 20 mM Tris-HCl (pH 8.3), 50 mM KCl and 1U of *Taq* DNA polymerase (Invitrogen). The conditions for PCR were as follows: 2 min at 94°C followed by 35 cycles of 1 min at 94°C, 1 min at 52°C and 1 min at 72°C. The resulting fragments were separated on a 0.5x TAE 1.5% agarose gel, excised and purified by the QIAquick Gel Extraction kit (Qiagen, Hilden, Germany).

Table 2.1: Primers used for cloning and expression analyses of *Radopholus similis* endoglucanases

Primer	Primer sequence	Primer	Primer sequence
Rs-act-F	GAAAGAGGGCCGGAAGAG	Rs-act-R	AGATCGTCCGCGACATAAAG
ENG1	TAYGTIATHGTIGAYTGGCA	ENG2	GTICCRTAYTCIGTIACRAA
eng1Aup1	GGAACCGGCAACCGTTGTCTACC	eng1Aup2	TTGTCGTCGCTGCGGTACGTGTG
eng1Bup1	AGGTCTCGTACAGCACATGCGGGTAGG	eng1Bup2	GTGGGGATAAGCATCAGAAAAAATGGC
eng2up1	AACCCGATTCGGGACTGACCGTTC	eng2up2	TCGTCCTTGTTGTCGATGGCCCAG
eng3up1	GTTGTAGAGCTCGTAAATGATGTTTGG	eng3up2	TCCAGCGTTGCCATTGGGTATTTGTGG
eng1Adown1	GAACAGTCCGATCACGGGCCAAAAGAAC	eng1Adown2	TGCACTACTACGCCGCCACGCACAAAC
eng1Bdown1	ATTCCACTACTACGCAGCCACGCATGG	eng1Bdown2	GAAAAAGGACCGGAACAATGAATTGGG
eng3down1	TGCGTTTCGCTGCCCAAATGGCTTACC	eng3down2	GATAGGGTTGAACATGAACACGAATCC
eng1A-FL-F	GTAGTCAATCTCCCCTCAGCCCATC	eng1Astart	ATGAACTGCTTGTTCCTTTTGCCTC
eng1B-FL-F	ATCGGTTCATACGCGAGTTCGTTG	eng1Bstart	ATGTGCTGCGCCTCTGCACTGCT
eng2-FL-F	TTGGAGCGGCAACATGACATC	eng2start	ATGCTTAGCCTGTTCCTTTTGCC
eng3-FL-F	GAACCCAGCATATCCGAACTTG	eng3start	ATGGCTAAATTCCCTATTTTC
eng1A-IS-F	CGGTTGCCGGTTCCTGGAACATG	eng1A-IS-R	CCGGGCTGGTAAAGTGTTGTCAAT
eng1B-IS-F	ACATGCGAGTCGTCGGGCAGTG	eng1B-IS-R	CCAGACTGAAGTAGCTGGCCGTT
eng2-IS-F	ATGCTTAGCCTGTTCCTTTTGCC	eng2-IS-R	ATGATCACGTAGATGCCCTCCGCGATG
eng3-IS-F	TGCGTTTCGCTGCCCAAATGGCTTACC	eng3-IS-R	TCAAACTTTTCCGTCCTTTCCGC
eng1AspF	ATGAACTGCTTGTTCCTTTTGCCTC	eng1AspR	TTGTCGTCGCTGCGGTACGTGTG
eng1BspF	TCGCCTCAGCCACTGCATTGACT	eng1BspR	TGATTGGGTTCGCGGATGCGACTT
eng2spF	GGCCGAGTCCAAGACCTGGTGGAA	eng2spR	GCTCGTCCATCATCAGCAGCTCAC
eng3spF	GATTGTTGGGAGCAAGACGAATG	eng3spR	ATCGGTAAGCCATTTGGGCAGC
SP6	ATTTAGGTGACACTATAGAATACTCAAGC	T7	TAATACGACTCACTATAGGGCGAATTGG
oligodT	TTTTTTTTTTTTTTTTTTTTTTTTVN		

Purified fragments were ligated into pGEM-T (Promega, Madison, WI, USA) and transformed into heat shock competent *Escherichia coli* DH5α cells (Invitrogen). Transformed cells were selected on LB medium containing 100 µg/ml carbenicillin. Resistant colonies were checked for inserts by colony PCR with SP6 and T7 primers under PCR conditions as described above. Plasmids of the corresponding positive colonies were isolated using the Nucleobond AX kit (Macherey-Nagel, Düren, Germany) and the insert was sequenced at the VIB Genetic Service Facility (VIB-GSF, Antwerp, Belgium).

Cloning of 5' and 3' ends

Three genomic endoglucanase fragments were obtained. Full length gene sequences were obtained by genome walking using the Genome Walker Universal kit (Clontech). Two successive PCRs, each with a gene-specific primer and an adapter primer, were carried out on the different genomic libraries following the manufacturer's instructions. The longest obtained fragments were cloned and sequenced as described above. For the first fragment, the primers used in the first PCR were eng3up1 and eng3down1 to walk upstream and downstream respectively (Table 2.1). In the nested PCR the gene-specific primers eng3up2 and eng3down2 were used (Table 2.1). To clone the full length sequence of the second gene, a similar strategy was followed using the primers eng1Bup1, eng1Bup2, eng1Bdown1 and eng1Bdown2 (Table 2.1). For the third gene, downstream walking was done with primers eng1Adown1 and eng1Adown2 in two successive PCRs (Table 2.1). This downstream walking resulted in two different fragments corresponding to two different sequences. The primers used to obtain the 5' sequences of the third and fourth endoglucanase by upstream genome walking were eng1Aup1, eng1Aup2, eng2up1 and eng2up2 (Table 2.1). Based on the obtained genomic sequences, forward gene-specific primers, one in the 5' untranslated region and one at the start of the coding sequence were developed (eng1A-FL-F, eng1Astart, eng1B-FL-F, eng1Bstart, eng2-FL-F, eng2start, eng3-FL-F and eng3start; Table 2.1). These forward primers were used with an oligodT primer as reverse primer in a semi-nested PCR on a cDNA pool of mixed stages to amplify the corresponding cDNA fragments. All resulting fragments were cloned into pGEM-T and sequenced as described above.

Sequence analysis

Intron regions were identified by aligning the genomic sequences to the corresponding cDNA sequences using ClustalW with default parameters and manual adjustment (Thompson et al., 1994). Putative protein sequences were obtained by translating the cDNA sequences using the EMBOSS program "Transeq" (Rice et al., 2000). Signal peptides were predicted using SignalP 3.0 (Bendtsen et al., 2004), conserved domains were located by the NCBI Conserved Domain Database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=cdd) and N-glycosylation sites were predicted by NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/). Codon usage

of the retained sequences was calculated as described below. As reference data set, all available expressed sequence tags (ESTs; n=1154, may 2007) from *Radopholus similis* were clustered and assembled using TGICL (Pertea et al., 2003) and CAP3 (Huang & Madan, 1999) and were translated by prot4EST (Wasmuth & Blaxter, 2004), resulting in putative proteins. The corresponding open reading frames (ORFs) derived from the ESTs were used to calculate the overall codon usage by the EMBOSS program "cusp" (Rice *et al.*, 2000), which resulted in a codon usage table. Using this codon usage table, the codon adaptation index (CAI) of the endoglucanases was calculated by the EMBOSS program "cai" (Rice et al., 2000). In addition, the average CAI of *R. similis* transthyretin-like (ttl) ORFs *Rs-ttl1* to *Rs-ttl4* (GenBank accession numbers AM691120 to AM691124) (Jacob et al., 2007) was calculated for comparison. Overall GC content and GC contents of the nucleotides occurring at the first, second and third positions of codons were calculated by an in-house Perl program. The GC contents of the endoglucanases were compared to the average GC contents of all *R. similis* ORFs by a one sample T-test.

An alignment of the identified endoglucanases with the two endoglucanases from *Pratylenchus penetrans* (BAB68522 and BAB68523) (Uehara et al., 2001) was created with ClustalW and manually adjusted. Protein similarities and identities were calculated with MatGAT v2.02 (http://bitincka.com/ledion/matgat/). 3D structure of Rs-ENG2 (amino acid 20 to 302) was modelled on the crystal structure of a GHF5 cellulase of *Erwinia chrysanthemi* (1EGZC) using SWISS-MODEL and the result was visualised using DeepView spdbv 3.7 (Swiss-PdbViewer) (Guex & Peitsch, 1997) and rendered with POV-Ray version 3.6 (http://www.povray.org/).

Whole mount *in situ* hybridisation

An *in situ* hybridisation was carried out according to Vanholme et al. (2002) with minor modifications. Nematodes were fixed in 3% paraformaldehyde for 15h at 4°C, followed by an additional incubation for 4h at room temperature. Permeabilisation was performed by incubation in 0.5 mg/ml proteinase K (Merck, Whitehouse Station, NJ, USA) for 40 min at room temperature. Hybridisation was done overnight at 47°C. The templates for the production of DNA probes were generated by a first PCR under conditions as described above on the plasmid pGEM-T containing the cDNA clone of the corresponding endoglucanase. Subsequent linear PCRs using the first PCR products as templates with digoxigenin(DIG)-labeled oligonucleotides (Roche, Mannheim, Germany) and a single primer, generated single strand DNA probes (F-primers: sense probes; R-primers: antisense probes). The primer pairs used for the four endoglucanases were respectively eng1A-IS-F and eng1A-IS-R, eng1B-IS-F and eng1B-IS-R, eng2-IS-F and eng2-IS-R, eng3-IS-F and eng3-IS-R (Table 2.1).

Semi-quantitative reverse transcriptase PCR

Approximately 100 individuals of young embryos, juveniles, adult males and adult females were manually separated from a batch of freshly collected nematodes. RNA extraction and cDNA synthesis were done as described above. The endoglucanases were amplified in a semi-quantitative RT-PCR. The gene-specific primers for amplification of the endoglucanases were eng1AspF, eng1AspR, eng1BspF, eng1BspR, eng2spF, eng2spR, eng3spF and eng3spR (Table 2.1). The PCR was carried out in a total volume of 60 µl using conditions as described above. After a number of PCR cycles (21 for females, 24 for males, 27 for eggs and 30 for juveniles), a small aliquot (12 µl) was removed from each sample. This was repeated every three cycles to detect the exponential phase of the reaction. Actin (EU000540) was amplified as a positive control using the primers Rs-act-F and Rs-act-R (Table 2.1). The resulting fragments were separated on a 0.5x TAE 1.5% agarose gel.

Results

Radopholus similis secretes cellulose degrading enzymes

Total homogenates of *Radopholus similis* showed clear cellulase activity in the CMC plate assay (Figure 2.1). A similar activity was observed in purified secretions (results not shown). This proved that *R. similis* produces and secretes functional cellulases. No cellulase activity was detected in protein extracts of the free living nematode *C. elegans* confirming that the cellulolytic capacity of *R. similis* is a characteristic feature of plant-parasitic nematodes. A zymogram of the secretions corroborated the results of the plate assay. Three distinct bands were observed when proteins in the secretions were separated. The molecular weights of the bands were estimated to be approximately 55, 100 and 180 kDa (Figure 2.1). No cellulose degradation was observed in the negative control lane.

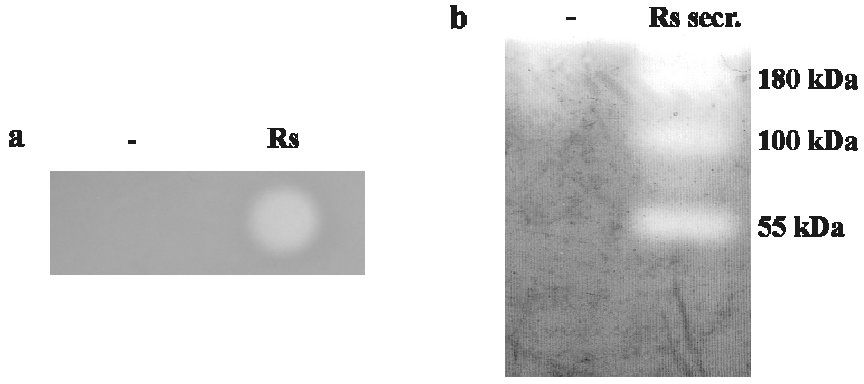

Figure 2.1: (a) CMC plate assay with negative control and *Radopholus similis* extract; (b) zymogram with negative control and secretions of *R. similis*. Estimated protein sizes are given.

Cloning of four endoglucanases genes

A PCR on *R. similis* genomic DNA using degenerate primers as described by Rosso et al. (1999), resulted in three fragments of 399 bp, 511 bp and 598 bp. A blastx search (http://www.ncbi.nlm.nih.gov/BLAST/) of the corresponding sequences confirmed that they were fragments of three different endoglucanases with endoglucanases from bacteria as well as other nematodes as top hits. Full length sequences of the genes were obtained by genome walking. During this cloning procedure, one additional endoglucanase was identified. Using a cDNA pool, corresponding coding sequences were amplified. All four cloned genes code for proteins with a catalytic domain and two of them have an additional carbohydrate binding module (CBM) (Boraston et al., 2004). The two endoglucanases with a CBM were named Rs-ENG1A (EF693940) and Rs-ENG1B (EF693941), whereas the other two proteins without a CBM were named Rs-ENG2 (EF693942) and Rs-ENG3 (EF693943). This is because most plant-parasitic nematode GHF5 endoglucanases indicated as 'ENG1' have a CBM (Gr-ENG1, AF056110; Gts-ENG1, AF182392; Hg-ENG1, AF006052; Hs-ENG1, AJ299386; Mi-ENG1, AF100549; Pp-ENG1, AB045780). Only two proteins with a CBM do not follow this nomenclature, being Mi-ENG3 (AY422836) of *M. incognita* and Mj-ENG3 (AM231138) of *M. javanica*. The respective lengths of the coding sequences of *Rs-eng1A, Rs-eng1B, Rs-eng2* and *Rs-eng3* are 1398, 1404, 960 and 1032 bp. Several introns were identified: one in *Rs-eng1A* (48 bp), six in *Rs-eng1B* (106, 46, 72, 44, 106 and 248 bp), one in *Rs-eng2* (56 bp) and five in *Rs-eng3* (54, 44, 45, 113 and 50 bp). The gene structure of the four genes is shown in Figure 2.2.

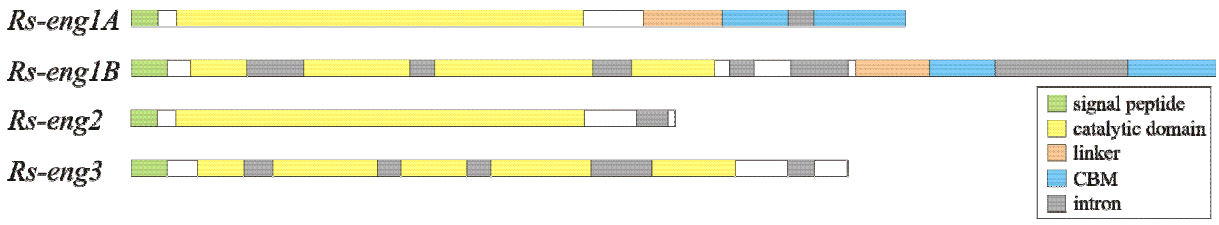

Figure 2.2: Genomic structure of the four cloned endoglucanases of *R. similis*. Different domains and introns are indicated by colours.

The GC contents of the first, second and third nucleotides (GC1, GC2, and GC3 respectively) of all codons of the endoglucanase ORFs were compared to the overall GC1, GC2 and GC3 contents derived from all *R. similis* ESTs (n=1154; Table 2.2). Overall GC and GC3 are significantly higher in the endoglucanases as opposed to the ESTs ($p<0.05$). In addition, the codon usage of the endoglucanase genes was compared to the average codon usage of *R. similis* (deduced from EST data) by means of the codon adaptation index (CAI). This index is a measure of the adaptation of a certain gene to the codon usage of a species (Sharp & Li, 1987). The CAI of *Rs-eng1A, Rs-eng1B, Rs-eng2* and *Rs-eng3* is respectively 0.86; 0.87; 0.86 and 0.83 whereas the average CAI of the *Rs-ttl* coding sequences is 0.82.

Table 2.2: GC, GC1, GC2 and GC3 content (%) of the four endoglucanase genes and the available *Radopholus similis* ESTs (n=1154)

	GC%	**GC1%**	**GC2%**	**GC3%**
Rs-eng1A	61.6	52.6	53.0	79.2
Rs-eng1B	59.9	42.5	59.0	78.2
Rs-eng2	62.1	52.8	41.6	91.9
Rs-eng3	57.3	56.1	40.1	75.6
Average	**60.2**	**51.0**	**48.4**	**81.2**
ESTs	53.7	56.6	42.6	61.9

Characteristics of the putative proteins

The putative endoglucanase proteins Rs-ENG1A, Rs-ENG1B, Rs-ENG2 and Rs-ENG3 have a molecular weight of respectively 49, 48, 39 and 35 kDa. The deduced protein sequences all have an N-terminal signal peptide for secretion. NetNGlyc predictions revealed two putative N-glycosylation sites in Rs-ENG1A: one in the catalytic domain (Asn87) and one in the CBM (Asn450). The site in the catalytic domain is also found in Rs-ENG2 (Asn88) (Figure 2.3). No glycosylation sites were predicted in Rs-ENG1B and Rs-ENG3. A conserved domain search confirmed that all sequences are cellulases belonging to GHF5 (pfam00150). Rs-ENG1A, Rs-ENG1B and Rs-ENG2 follow the GHF5 signature (Prosite PS00659). However, Rs-ENG3 has a glycine in the last position of the GHF5 signature. This amino acid was not yet described at this position of this motif. In addition to the catalytic domain, which is found in all four proteins, Rs-ENG1A and Rs-ENG1B contain a linker and a CBM of CBM family 2 (pfam00553). Both linkers are enriched in stretches of threonine (Figure 2.3). The CBMs of Rs-ENG1A and Rs-ENG1B are 76% similar. The similarity of the CBMs between Rs-ENG1A and Pp-ENG1 and Rs-ENG1B and Pp-ENG1 are respectively 61% and 63%. The catalytic domains were used to calculate the identity and similarity between the different proteins (Table 2.3).

Table 2.3: Protein similarities (below diagonal) and identities (above diagonal) (%) of the catalytic domains between the different endoglucanases of the migratory endoparasitic nematodes *Radopholus similis* (Rs) and *Pratylenchus penetrans* (Pp)

	Rs-ENG1A	**Rs-ENG1B**	**Rs-ENG2**	**Rs-ENG3**	**Pp-ENG1**	**Pp-ENG2**
Rs-ENG1A		61.7	80.0	51.6	64.8	66.1
Rs-ENG1B	75.5		62.0	44.6	56.3	51.4
Rs-ENG2	90.9	74.8		49.0	66.0	65.4
Rs-ENG3	66.0	59.4	66.0		46.2	48.3
Pp-ENG1	77.6	74.0	80.3	62.5		80.5
Pp-ENG2	78.5	69.5	81.3	62.5	89.5	

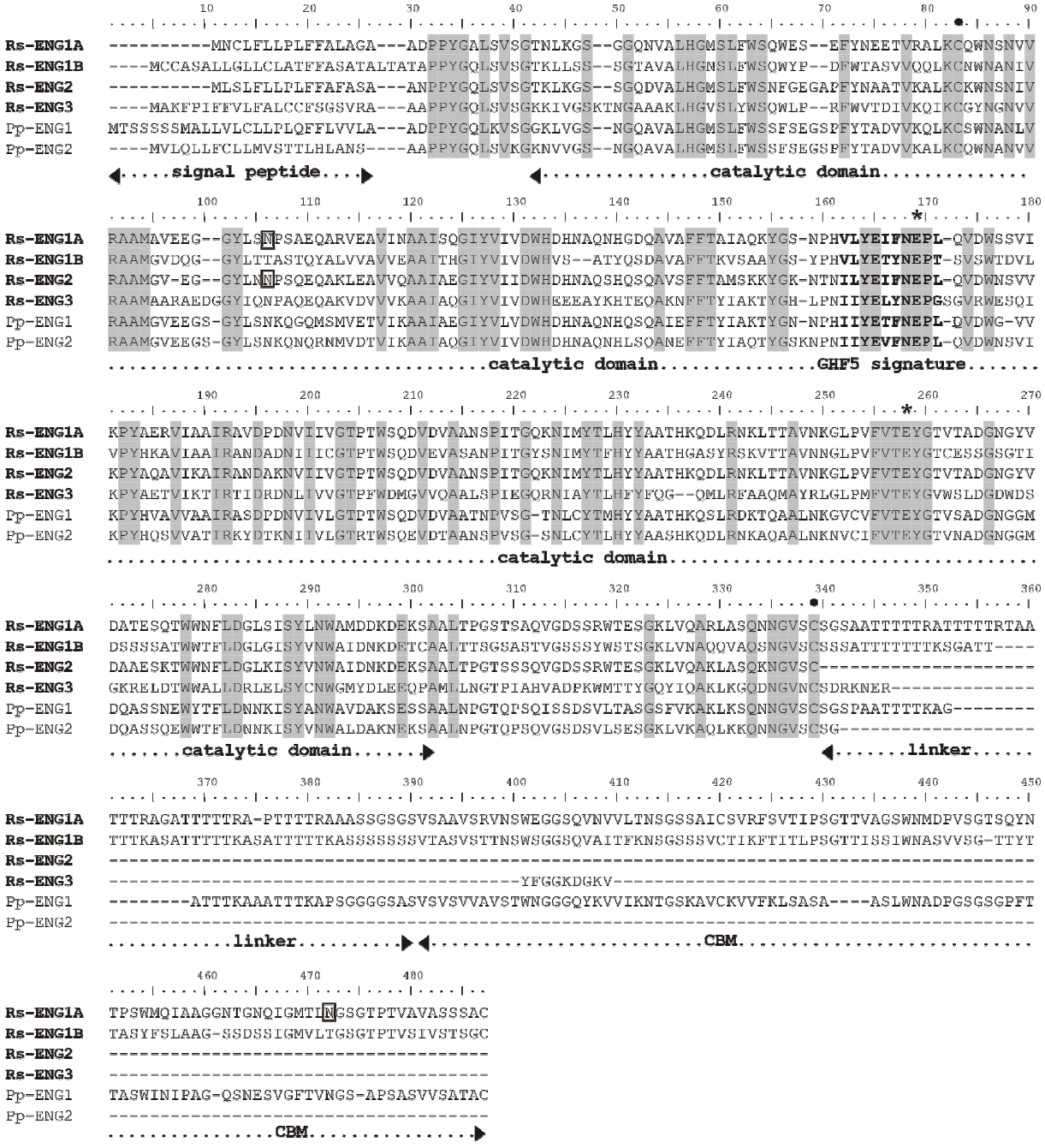

Figure 2.3: Protein alignment of *Radopholus similis* endoglucanases Rs-ENG1A, Rs-ENG1B, Rs-ENG2 and Rs-ENG3 and *Pratylenchus penetrans* endoglucanases Pp-ENG1 and Pp-ENG2. The location of the secretion signal peptide, catalytic domain, GHF5 signature, linker and CBM is indicated. Stars indicate the catalytic residues and dots the conserved cysteines. Asparagine residues in boxes are possible N-glycosylation sites.

Prediction of the secondary structure showed an alternation between β-strands and α-helices. Subsequent 3D modelling revealed an α/β-barrel, the typical GHF5 endoglucanase structure. Amino acids similar between the four endoglucanases were mapped onto the 3D model (Figure 2.4). This illustrated that most of the conservations in the sequences are located in the core of the protein, close to the active sites of the catalytic centre. Mutations are found in the loops or in the α-helices, pointing towards the outside of the protein. The catalytic domain

contains two conserved cysteine residues (Cys83 and Cys339; Figure 2.3), which are not present in bacterial endoglucanases. Unfortunately, Cys 339 was not retained in our model because of its C-terminal location in the protein. However, based on the available structure, it can be assumed that both residues are in close proximity, making disulfide bridge formation possible.

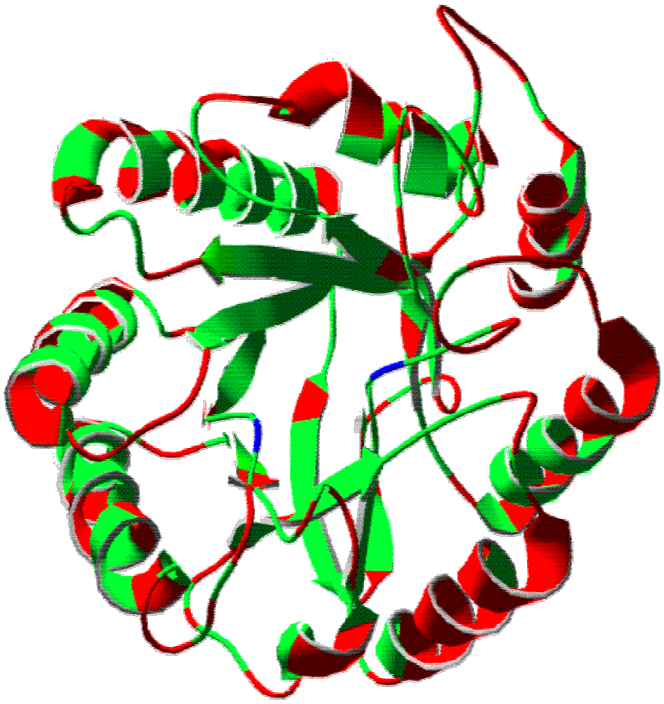

Figure 2.4: 3D model of Rs-ENG2. Amino acids that are identical in all *Radopholus similis* endoglucanases or amino acids that belong to the following strong groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY and FYW are indicated in green. The other non-conserved amino acids are indicated in red. The catalytic residues are indicated in blue.

Spatial and temporal expression of endoglucanases

To localise endoglucanase expression in *R. similis*, a whole-mount *in situ* hybridisation was performed. Gene-specific antisense probes targeting the different *Radopholus* endoglucanases stained the gland cell area, whereas the sense probes gave no signal (Figure 2.5). The gland region was strongly stained when probes against *Rs-eng2* and *Rs-eng1B* were used. Probes targeting *Rs-eng1A* and *Rs-eng3* resulted in weaker signals.

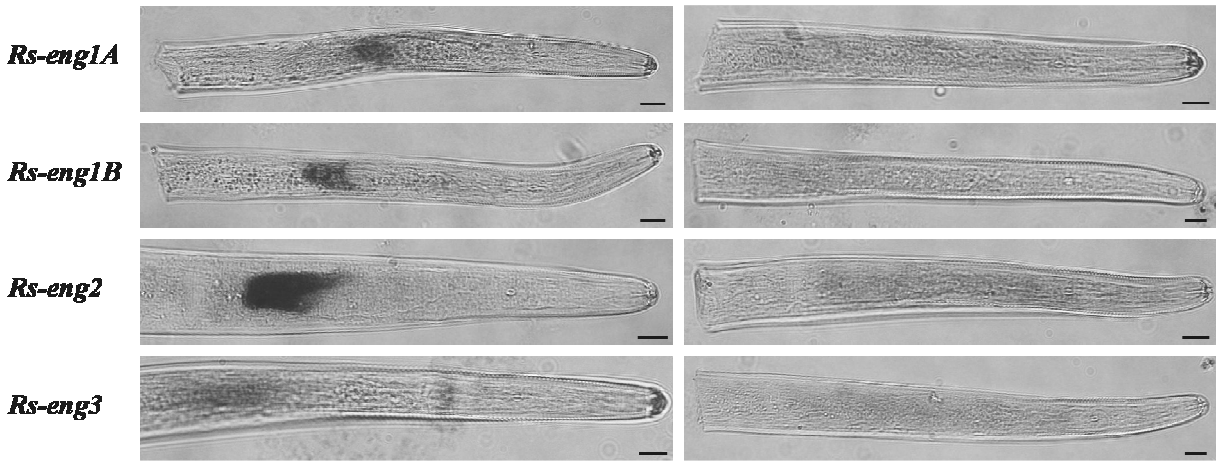

Figure 2.5 Whole mount *in situ* hybridisation on *Radopholus similis* females. Left panel: antisense *Rs-eng1A, Rs-eng1B, Rs-eng2* and *Rs-eng3* probe; Right panel: corresponding sense probes. Scale bar: 20 µm.

To reveal differences in temporal expression, a semi-quantitative RT-PCR was performed on different life stages. Gene-specific primers against the different endoglucanases were used in combination with actin primers as a reference. *Rs-eng1A* showed a slightly weaker expression level than the other endoglucanases. *Rs-eng1A*, *Rs-eng2* and *Rs-eng3* are expressed at a lower level in males than in females (Figure 2.6) as opposed to *Rs-eng1B*, which is equally expressed in males and females. Expression of *Rs-eng1A* is restricted to females and *Rs-eng3* could only be detected in juveniles. In eggs, expression of *Rs-eng3*, as well as a very low expression level of *Rs-eng1B* was observed.

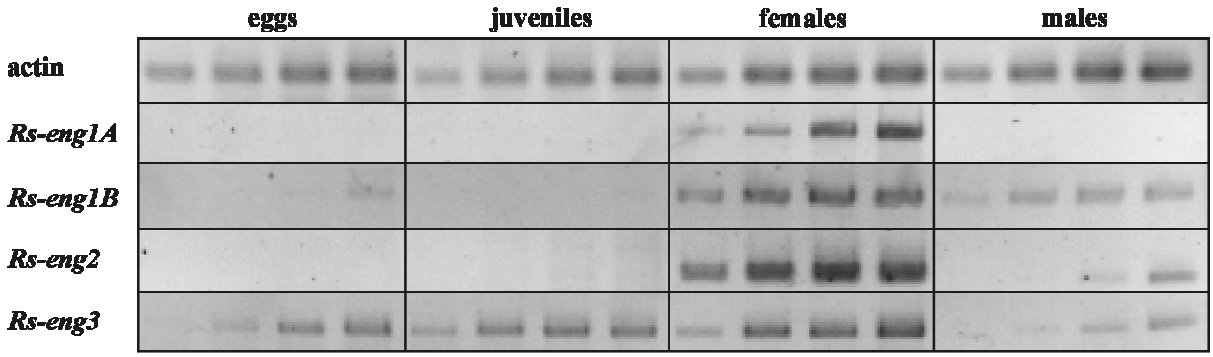

Figure 2.6 Expression of actin and the endoglucanases determined by semi-quantitative RT-PCR in developing embryos (27, 30, 33 and 36 cycles), juveniles (30, 33, 36 and 39 cycles), females (21, 24, 27 and 30 cycles) and males (24, 27, 30 and 33 cycles) of *Radopholus similis*.

Discussion

Four endoglucanases (Rs-ENG1A, Rs-ENG1B, Rs-ENG2, and Rs-ENG3) of GHF5 were identified from the migratory plant-parasitic nematode *Radopholus similis*. They all have a signal peptide for secretion and a catalytic domain. In addition, two of the proteins have a C-terminal CBM, which is connected by a linker to the catalytic domain. The catalytic domain is responsible for the cellulase activity, while the CBM is involved in the binding and recognition of the enzyme to its substrate (Boraston et al., 2004).

The current hypothesis is that the endoglucanases were acquired from bacteria by means of horizontal gene transfer (HGT) (Jones et al., 2005). This possible HGT event probably occurred relatively early in the evolution of plant-parasitic nematodes. Indeed, the codon adaptation indices for all *R. similis* endoglucanases are high and similar to the CAI of the *ttl*-genes, which means that the codon usage of the endoglucanase genes has been adapted well to the overall codon usage of the organism. Moreover, the genes and corresponding proteins depict features characteristic to eukaryotic sequences (e.g. introns, polyadenylation signal, eukaryotic signal peptide for secretion).

The percentage of identity between the different domains suggests that the CBM evolves more rapidly than the catalytic domain, since the CBMs have a lower similarity (~62%) than the catalytic domains (~76%) when compared to the corresponding domains of a *Pratylenchus penetrans* endoglucanase (Pp-ENG1). The linkers between the catalytic domain and the CBM of Rs-ENG1A and Rs-ENG1B are both threonine rich. This is also the case for the linker of Pp-ENG1 of *P. penetrans*. In cyst nematode endoglucanases however, the linkers are either glycine-serine (Hg-ENG1 and Gr-ENG1) or alanine-lysine-proline rich (Gr-ENG2, an endoglucanase with linker but without CBM). In endoglucanases of root-knot nematodes, the linkers are glycine-serine and asparagine rich (Mi-ENG1 and Mi-ENG3). This illustrates that the amino acid composition of the linker is not conserved between taxa. Nevertheless there seems to be a bias towards the use of small amino acids. This was also observed in the linker sequences identified in bacteria, which show some identity within a species, but not between species (Gilkes et al., 1991). Most likely the evolutionary pressure on the linker is low as this part of the protein probably only has a structural function as an extended, flexible hinge between domains (Gilkes et al., 1991). This lack of specific functionality could have led to different linkers in different nematode taxa.

The average GC content of all *R. similis* ORFs obtained from the ESTs is 54%, the GC3 is 62%. This last value corresponds well to previously published data of Cutter et al. (2006), who found that *R. similis* has the highest GC3 value observed in nematodes so far (63.5%). The endoglucanase ORFs have an average GC and GC3 percentage of respectively 60% and 81%. Thus, the overall GC percentage of these endoglucanases is above average, which is mainly due to a much higher GC3 content. In the cyst nematode endoglucanases of

Heterodera glycines and *Globodera rostochiensis*, the mean GC is 50%, and the GC3 is 59%. Although the endoglucanases of *R. similis* and the cyst nematodes have similar structures, there is a rather big difference in GC content. There is no clear explanation for this observation, although it has been reported that there is a positive correlation between GC3 content and optimal temperature of the organism (Bernardi, 2000). Since *R. similis* lives in tropical regions and cyst nematodes mainly occur in temperate climates, this could fit in this so called thermal stability or thermal adaptation hypothesis. It should however be noted that this hypothesis has been undermined by different authors (Hurst & Merchant, 2001; Belle et al., 2002).

Whole-mount *in situ* hybridisation clearly showed expression of the genes in the gland cell area of *R. similis*. Cyst and root-knot nematodes have two subventral gland cells, which are mainly active during migration of the nematode through the plant tissue, and one dorsal gland cell, which is active in sedentary stages (Vanholme et al., 2004). In *Radopholus* species, the three pharyngeal gland cells are aligned and form a long dorsal overlap of the intestine (Luc, 1987), but no clear distinction can be made between the three cells (Trinh et al., 2004). Although the *in situ* hybridisation of the endoglucanases clearly showed staining in the gland cell area, we were not able to distinguish the different gland cells. To our knowledge, nothing is known about the activity of the different gland cells during the life cycle of migratory nematodes.

The zymogram revealed three different bands, of which only the lower one most likely corresponds to one of the characterised proteins. The difference between the observed band and the theoretical MW of this protein could be explained by the fact that Rs-ENG1A and Rs-ENG2 contain putative N-glycosylation sites. It is unclear what the higher molecular weight bands (100 and 180 kDa) in the zymogram represent. Bera-Maillet et al. (2000) also observed bands of high molecular mass in *M. incognita* homogenate, however these were not present in nematode secretions. It is possible that these are enzymatic complexes since it is known that cellulases can organise into high molecular weight cellulolytic complexes called cellulosomes (Bayer et al., 1998).

The spatial expression pattern, combined with the putative signal peptide for secretion, and the cellulase activity found in the secretions indicate that the characterised endoglucanases are secreted by the nematode. Once released in the plant tissue, the enzymes will soften the plant cell wall and facilitate the migration of the parasitic nematode in the plant tissue. Only one of the endoglucanases (*Rs-eng3*) was shown to be expressed in eggs and juveniles. Although unlikely, we cannot rule out the possibility that some of the developing embryos in the eggs were in a further stage of development, which could explain the presence of *Rs-eng3* transcripts. However, *Rs-eng1B* was also detected at a very low level in eggs, but not in juveniles. The other endoglucanases are only expressed at later life stages. The expression of few endoglucanases in juvenile stages is rather surprising, because juveniles are able to

migrate through the plant tissue and therefore one would expect a high expression level of various endoglucanases instead of only *Rs-eng3*. Nevertheless, plant-parasitic nematodes can have as much as 21 endoglucanase genes (Abad et al., 2008), and therefore other, as yet unidentified, endoglucanases might show complementary expression patterns compared to the four genes studied. Since *Rs-eng3* shows less homology to the other three cloned endoglucanases on protein level, it is possible that its different expression pattern is a result of different evolutionary paths.

In contrast to sedentary cyst nematodes, where no endoglucanase expression was detected in adult females (Smant et al., 1998), all endoglucanases identified in *R. similis* are highly expressed in adult females. This observation is obviously related to the different life style of the nematodes: *R. similis* adult females still migrate through the root tissue while cyst nematode adults are sedentary. In contrast to females, adult males of *R. similis* are believed to be non-parasitic, and probably only migrate in search of females in order to mate. However, since *R. similis* can also reproduce as a hermaphrodite or by parthenogenesis (Kaplan & Opperman, 2000), males are not essential for reproduction. It may be the case that during evolution, males are gradually losing their ability to migrate, and that this process is still ongoing. This is reflected in their degenerate stylet with reduced knobs and a reduced pharynx (Luc, 1987). In addition to the morphological differences, males are apparently accompanied by an altered expression of endoglucanase genes compared to females. Transcripts for one of the four described endoglucanases (*Rs-eng1A*) could not be detected in males and the other genes were expressed at a considerably lower level compared to females. This lower expression level of the endoglucanases in males could imply a reduced ability to migrate through the plant tissue.

Acknowledgements

We would like to thank Dr. Annemie Elsen (K.U. Leuven, Belgium) for kindly providing the initial *R. similis* culture.

CHAPTER 3

A GHF5 ENDOGLUCANASE (PSEUDO)GENE FAMILY AND AN EXPANSIN-LIKE GENE IN *DITYLENCHUS AFRICANUS*

Abstract

In this chapter, the finding of putative endoglucanases and a putative expansin-like protein in the migratory peanut pod nematode *Ditylenchus africanus* is described. Two cloned endoglucanase genes (*Da-eng1A*, *Da-eng1B*) include a signal peptide for secretion, a catalytic domain belonging to glycosyl hydrolase family 5 (GHF5), and a carbohydrate binding module (CBM). These are the first GHF5 endoglucanases found in the superfamily of the Sphaerularioidea (order Rhabditida). Moreover, four additional genes were cloned which have a very high similarity to the endoglucanases, but lack a part of the catalytic domain possibly due to homologous recombination. Two of these endoglucanase-like genes with deletions (*Da-engdel2*, *Da-engdel4*) code for erroneous proteins, while two others (*Da-engdel1*, *Da-engdel3*) code for hypothetical protein sequences with an intact CBM. The latter two genes could still exert a function similar to cellulose binding proteins.

An identified putative expansin-like protein (Da-EXP) consists of a signal peptide for secretion, an expansin-like domain and a CBM, a domain structure never found before in nematode expansin-like proteins. Interestingly, the CBM of the expansin-like protein is very similar to the endoglucanase CBMs, suggesting that *Da-exp* obtained its CBM through domain shuffling. The expression of both types of proteins appears to be equal among different life stages, although the endoglucanase expression is much stronger than the expansin-like gene expression. The expression of both *Da-eng1A* and *Da-exp* is located in the gland cell area of the nematode, pointing towards the secretion of these proteins in the plant tissue. They are believed to either degrade the cellulose in the plant cell wall, or loosen the cell wall to make it more accessible for cell wall degrading enzymes.

Introduction

The migratory nematode *Ditylenchus africanus* is a member of the Anguinidae family within the superfamily of Sphaerularioidea, order Rhabditida. It occurs on groundnut in South Africa, where it causes mainly qualitative damage, but it can also survive on the hyphae of fungi (De Waele et al., 1997). No molecular knowledge on parasitism genes is available on any nematode from this superfamily. To study the evolution and occurrence of cell wall modifying enzymes, it is interesting to look for certain genes in this nematode, for example GHF5 endoglucanase and expansin.

Expansins form a diverse plant protein family involved in various biological processes involving the re-arrangement of plant cell wall polysaccharides (Cosgrove, 2000a). Experimental evidence suggests that expansins loosen cell walls via a non-enzymatic mechanism that induces slippage of cellulose microfibrils in the plant cell wall (Sampedro & Cosgrove, 2005). The plant α- and β-expansins consist of a signal peptide and two domains: domain one related to the catalytic domain of GHF45 proteins, and domain two related to grass pollen allergens, which is possibly a carbohydrate binding module.

In animals, a functional expansin was first identified in the cyst nematode *Globodera rostochiensis* (Qin et al., 2004). It consists of two domains: the first domain is similar to the CBM of nematode endoglucanases while the second domain is similar to the first domain of plant β-expansins. Moreover, it was shown to be present in nematode secretions, and cell wall extension activity of the recombinant enzyme could be demonstrated. Qin et al. (2004) proposed that the synergistic action of expansin and other cell wall degrading enzymes could facilitate the invasion of the host by the nematode. Interestingly, the expansin domain is more similar to two hypothetical proteins from bacteria than to plant expansins. This could imply that the nematode expansins are acquired through horizontal gene transfer from bacteria, as suggested for other cell wall degrading enzymes (Kudla et al., 2005). Additional putative expansin-like proteins were found during an EST screen of *Bursaphelenchus xylophilus* and *B. mucronatus* (Kikuchi et al., 2007). In contrast to the endoglucanases, it seems that expansins present in cyst nematodes and pine wood nematodes are derived from the same source. Moreover, expansins are likely part of extensive gene families: in the genome of *Meloidogyne incognita*, 20 candidate expansins were identified (Abad et al., 2008).

In this chapter, the cloning of GHF5 endoglucanases and an expansin-like protein in *Ditylenchus africanus* is described. Surprisingly, the nematode also expresses different endoglucanase-like pseudogenes containing deletions.

Materials and Methods

Nematode culture, DNA and RNA extraction and first strand cDNA synthesis

Ditylenchus africanus was maintained on carrot disks as described in chapter 2 and collected approximately 6 weeks after inoculation of the carrot disks. The collected nematodes were either used immediately for the enzymatic assay or *in situ* hybridisation, or stored as a pellet at -80°C for DNA or RNA extraction. Genomic DNA was isolated from nematodes of mixed stages according to Bolla et al. (1988). RNA extraction and first strand cDNA synthesis were carried out as described in chapter 2. For the RT-PCR analysis, RNA was extracted from approximately 100 juveniles, adult males and adult females, and cDNA was prepared and amplified by the SMART PCR cDNA synthesis kit (Clontech).

Cellulase activity assay

A cellulase activity assay on 10,000 crushed nematodes was carried out as described in chapter 2. Water rinsed from uninfected carrot disks was used as a negative control while 1U cellulase of *Aspergillus niger* (Sigma-Aldrich) served as positive control.

Cloning endoglucanase and expansin-like genes

Genomic endoglucanase fragments were amplified by PCR using the degenerate primers ENG1 and ENG2 as described by Rosso et al. (1999) (Table 3.1). For expansin-like fragments, degenerate primers EXP1 and EXP2 (Table 3.1) were designed based on conserved regions in nematode expansin-like genes. Fragments were amplified by PCR, purified, cloned and sequenced as described in chapter 2.

To obtain the 5' and 3' genomic sequences of the endoglucanase fragments, the Genome Walker Universal kit (Clontech) was used. Two successive PCRs, each with a gene-specific primer and an adapter primer, were carried out on the different genomic libraries following the manufacturer's instructions. For the endoglucanase fragment, Da-eng1A-up1, Da-eng1A-up2, Da-eng1A-down1 and Da-eng1A-down2 (Table 3.1) were used to walk up- and downstream respectively. For the expansin-like fragment, the primers Da-exp-up1, Da-exp-up2 and Da-exp-down1 and Da-exp-down2 (Table 3.1) were used. The longest obtained fragments were cloned and sequenced as described in chapter 2. For both genes, another round of downstream walking was required using the primers Da-eng1A-down3 and Da-eng1A-down4 for the endoglucanase and Da-exp-down3 and Da-exp-down4 for the expansin-like fragment (Table 3.1).

cDNA sequences of the endoglucanase and expansin-like gene were amplified by PCR on the cDNA pool of *D. africanus* using gene-specific primers Da-eng1A-start and Da-eng1A-stop for the endoglucanase and Da-exp-start and Da-eng1A-stop (Table 3.1) for the expansin-like gene. These cDNA sequences were cloned and sequenced as described above.

Sequence analysis

Alignments, identification of introns, calculation of GC contents, translations and predictions of molecular weight, putative signal peptides and N-glycosylation sites were done as described in chapter 2. With the expansin-like domain of the putative expansin-like protein, a blastp search was done (E-value<1e-5) to identify homologous proteins. Similarly, a tblastn (E-value<1e-3) search was conducted against all clustered nematode ESTs on the nematode.net server (Wylie et al., 2004). An alignment of the putative protein sequences from the significant blast hits was created with ClustalW. An unrooted phylogenetic tree was constructed based on the expansin-like domain using Bayesian statistics with the software MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003). A mixed model for protein evolution was applied with 1,000,000 generations (sample frequency = 100), discarding the first 250,000 generations as burn-in.

The genomes of *Meloidogyne incognita* and *M. hapla* were searched for contigs containing expansin-like domains by tblastn (E-value<1e-3) on the respective websites of the genome projects (http://meloidogyne.toulouse.inra.fr/ and http://www.hapla.org/).

Whole mount *in situ* hybridisation

An *in situ* hybridisation was carried out according to Vanholme et al. (2002) with minor modifications. Nematodes were fixed in 4% paraformaldehyde for 4h at room temperature, 16h at 4°C, and again 4h at room temperature. Permeabilisation was performed by incubation in 0.5 mg/ml proteinase K for 1h 30 min at room temperature. Hybridisation was done overnight at 47°C. The templates for the production of DNA probes were generated as described in chapter 2. The F- and R-primers used for the endoglucanase and expansin-like gene were respectively Da-eng1A-start, Da-eng1A-R, Da-exp-F and Da-exp-R (Table 3.1).

Table 3.1 Primers used for cloning and expression analyses of the *D. africanus* endoglucanases and expansin-like gene

Primer	**Primer sequence**	**Primer**	**Primer sequence**
ENG1	TAYGTIATHGTIGAYTGGCA	ENG2	GTICCRTAYTCIGTIACRAA
EXP1	CYRGYGGMMGYGGMGCKTGYGG	EXP2	TCYGGGCATTTRTTRKTKATYGG
Da-eng1A-up1	CCAGCGTATTTTTGTGCCATTTGG	Da-eng1A-up2	CTCTCAACCAATATCTTTGACTTAACAGC
Da-eng1A-down1	AATAGTCGCGTTTAGGACGTGGACGTC	Da-eng1A-down2	TCCACTTCTACGCCGGCACTCAC
Da-exp-up1	TCGGGATAGTTGGATGGCACC	Da-exp-up2	TGGGCATTTGGATCAAACAGC
Da-exp-down1	GGGAATATGTCCGACTTTTTCAG	Da-exp-down2	CGGTTCGATCATAACTTTGC
Da-eng1A-down3	AATGGCGTGAGCTGCAGTGGTGGAAG	Da-eng1A-down4	GCGGTAACGGAAAATTGGTGGCTCAAG
Da-exp-down3	CGGTCTTCAATCTGGGGTAATAC	Da-exp-down4	CGCAACAGCTAATGCTAATCTCAG
Da-eng1A-start	ATGAAATTCTTCGCCAGCCTCG	Da-eng1A-stop	TCAGCAATATCCATAAGACACAACGC
Da-exp-start	ATGCAATTTGTCACACTTACCTTCAC	Da-eng1A-R	CGTCCACGTCCTGACTCCAAG
Da-exp-F	ATAAATGCCCGGAATGTGAA	Da-exp-R	ACCAGAGCCAACTGCATACC
Da-act-F	GGATACCGGCAGATTCCATA	Da-act-R	GAAATCGTCCGTGACATCAA

Semi-quantitative reverse transcriptase PCR

Approximately 100 individuals of juveniles, adult males and adult females were manually separated from a batch of freshly collected nematodes. RNA extraction and cDNA synthesis were done as described above. The endoglucanase and expansin-like gene were amplified in a semi-quantitative RT-PCR. The gene-specific primers for amplification of the endoglucanases were Da-eng1A-start, Da-eng1A-R, Da-exp-F and Da-exp-R (Table 3.1). As a control, actin was amplified with primers Da-act-F and Da-act-R (Table 3.1). The PCR was carried out in a total volume of 30 µl using conditions as described in chapter 2. The amount of cDNA added as template and the number of PCR cycles were optimised for each gene product to detect the exponential phase of the reaction. The resulting fragments were separated on a 0.5x TAE 1.5% agarose gel.

Results

Cellulase activity assay

A nematode extract was spotted on a CMC plate to determine whether *D. africanus* showed any cellulase activity. The positive control and the nematode extract both displayed a clear halo, while the negative control did not show any degradation of the substrate (Figure 3.1). This illustrates that *D. africanus* produces functional cellulase enzymes.

Figure 3.1 CMC plate assay with *Ditylenchus africanus* extract and negative control.

(Pseudo)endoglucanases

GHF5 endoglucanase fragments were amplified using degenerate primers, and by genome walking on one of the fragments, a genomic sequence of 2613 bp was obtained. Based on the obtained DNA sequence, primers were developed to clone the corresponding cDNA, which permitted the identification of introns. The gene, named *Da-eng1A* (ABY52965), has five introns of respectively 67, 110, 103, 183 and 149 bp. The introns are in respective phases 0, 2, 0, 0 and 0, and all introns have the common GT/AG splicing site (Blumenthal & Steward, 1997). The coding sequence is 1395 bp, hence the putative protein counts 464 amino acids. The putative protein has a molecular weight of 49 kDa and consists of a signal peptide for secretion and a GHF5 catalytic domain (pfam00150) linked to a CBM of family 2 (pfam00553). The GHF5 pattern of the catalytic domain (Prosite PS00659) is fully present (VLYETFNEPL). Two putative N-glycosylation sites are present, one in the catalytic domain

(Asn89) and one in the CBM (Asn447). The overall GC content of *Da-eng1A* is 49.7%; while GC1, GC2 and GC3 are respectively 47.7, 46.2 and 55.1%

A second genomic fragment of an endoglucanase was obtained (*Da-eng1B*). Genome walking of *Da-eng1B* resulted in a fragment of 1694 bp. *Da-eng1B* has four introns (with GT/AG splice sites) of respectively 85, 186, 100 and 178 bp. These four introns are located at the exact same positions as the first four introns of *Da-eng1A*. The putative protein consists of a signal peptide for secretion, a catalytic domain, a linker and a small part of the CBM. A full CBM is probably present in this gene, but the high homology of *Da-eng1B* to the other cloned genes made it impossible to isolate the remaining part of the fragment. The coding sequence of *Da-eng1B* has 69 point mutations compared to *Da-eng1A*, resulting in seven altered amino acids.

During the cloning process, four additional cDNAs emerged from the cDNA pool, each with a large deletion and therefore these probably are pseudogenes. The four cDNAs were named *Da-engdel1*, *Da-engdel2*, *Da-engdel3* and *Da-engdel4* and have deletions of 312, 637, 189 and 55 bp respectively. Interestingly, in three of the four genes, the sequence at the anterior part of the deletion is similar to the posterior part of the deleted sequence. An overview of the gene structure of all cloned endoglucanases is shown in Figure 3.2.

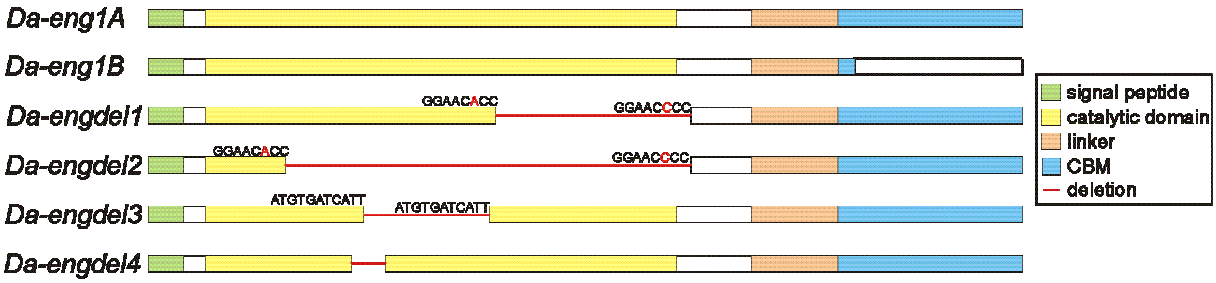

Figure 3.2 Gene structure of the (pseudo)endoglucanases of *Ditylenchus africanus*. Different domains and introns are indicated by colours. Similar DNA sequences before and in the deletion are indicated.

To check whether or not these cDNAs are artifacts from the cDNA synthesis or PCR, genomic fragments were cloned. For three of the four cDNAs, a corresponding DNA fragment with the deletion was amplified. Full length DNA sequences of *Da-engdel2* and *Da-engdel3* showed the presence of two and one intron respectively in the same positions as found for *Da-eng1A* and *Da-eng1B*. For *Da-engdel4*, only part of the 5' end of the genomic sequence was cloned, including one intron. Moreover, the first intron of both *Da-engdel2* and *Da-engdel3* has 100% identity to the corresponding intron of *Da-eng1A*, while the intron of *Da-engdel4* is more similar to the first intron of *Da-eng1B* (Figure 3.3). These DNA sequences including introns prove that the endoglucanase genes with the deletions are truly present in the genome, and could be pseudogenes.

Figure 3.3 Alignment of intron 1 of the different *D. africanus* endoglucanases. Shading threshold: 60% identity.

Putative protein sequences of these pseudogenes are very similar to the full endoglucanase genes. However, for *Da-engdel2* and *Da-engdel4*, the number of deleted bases is not a multiple of three, causing a frameshift leading to an erroneous translation. Hence, these two genes cannot be functional. For *Da-engdel1* and *Da-engdel3*, the deletion is in frame resulting in a truncated protein with disrupted catalytic domain but with an intact CBM. An alignment of the putative protein sequences illustrates the great similarity between them (Figure 3.4).

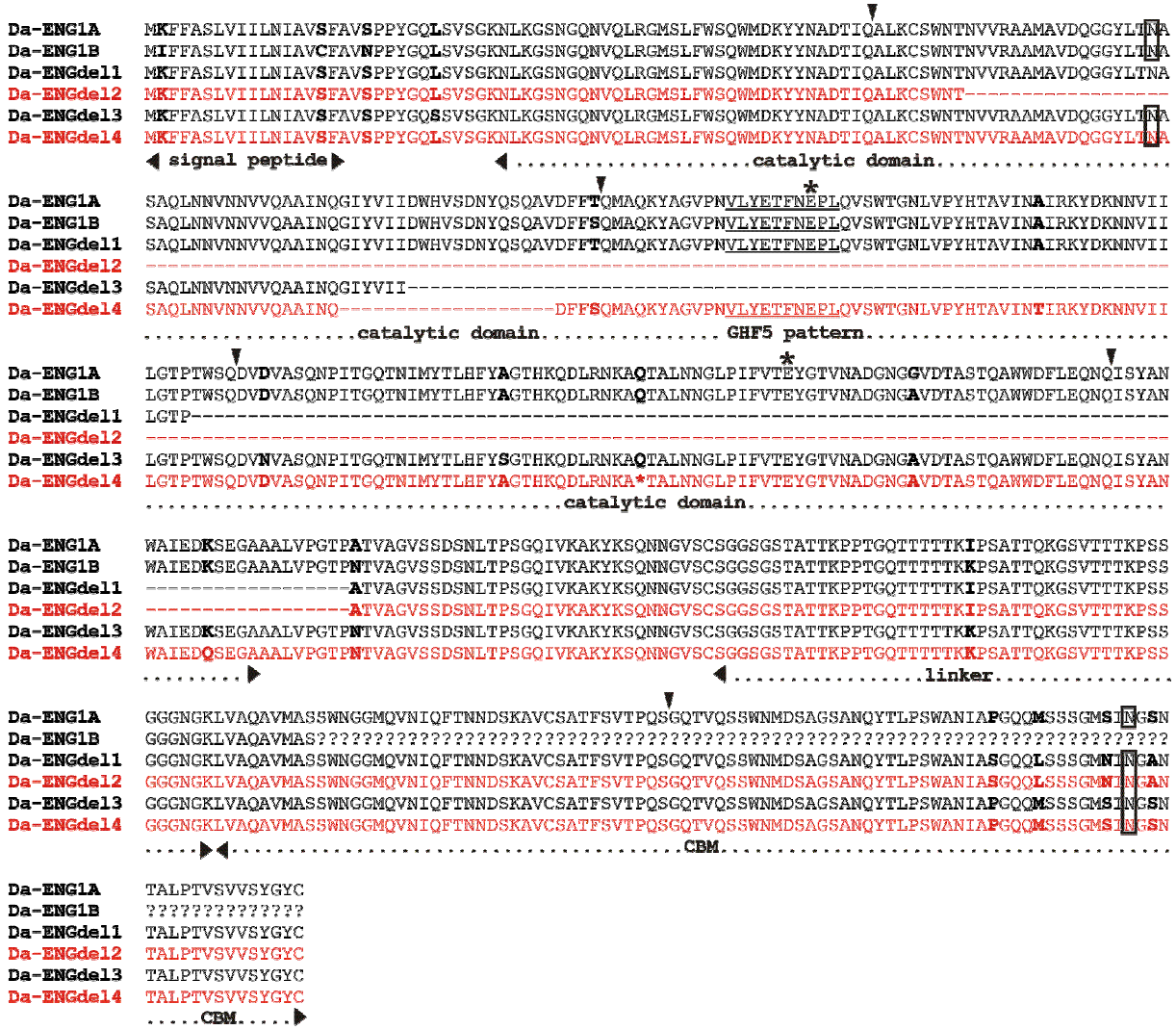

Figure 3.4 Alignment of the putative protein sequences of two *D. africanus* endoglucanases and four endoglucanase-like genes with deletions. The proteins marked in red have deletions that cause a frameshift, but for sequence comparison, this frameshift was corrected in the figure. Amino acids indicated in bold are not conserved among all proteins. Different domains are indicated, arrows show the positions of the introns, stars indicate the catalytic residues and putative N-glycosylation sites are boxed.

Expansin-like gene

Using degenerate primers, a putative expansin-like fragment was cloned, and by genome walking a DNA fragment of 1908 bp was obtained. The gene, called *Da-exp*, has two introns of 189 and 150 bp, both with a GT/AG splicing site. The coding sequence is 936 bp. The predicted protein (32 kDa) consists of a signal peptide for secretion, a putative expansin-like domain and a CBM of family 2. Two predicted N-glycosylation sites are present, one in the expansin-like domain (Asn146) and one in the CBM (Asn294). The overall GC content of *Da-exp* is 49.0%; while GC1, GC2 and GC3 are respectively 44.6, 54.8 and 47.8%.

A blastp search (E-value<1e-5) with the expansin-like domain against the Protein division of GenBank resulted in the following hits: six expansin-like proteins from *Bursaphelenchus* spp. (BAG16532, BAG16534, BAG16535, BAG16536, BAG16537, ABV60414), three expansin-like proteins from *Globodera rostochiensis* (CAC83611, CAC84564, CAD89105), two bacterial hypothetical proteins (AAD32751, CAC42207) and five putative pathogenicity factors or avirulence proteins from *Meloidogyne* spp (CAP59535, CAP59536, CAP59537, CAP59538, CAC27774). The average similarity of the expansin-like domain of Da-EXP to all *Bursaphelenchus* expansin-like proteins is 74.5%, while to *Globodera* expansin-like proteins, the similarity is only 58.4%. The average similarity of the expansin-like domain of Da-EXP to the C-terminal part of the avirulence proteins is 46.4%.

Searching clustered ESTs by tblastn (E-value<1e-3) resulted in 14 significant hits. These hits were derived from *Pratylenchus vulnus*, *Meloidogyne* spp., *Xiphinema index* and surprisingly from the animal-parasitic nematode *Parastrongyloides trichosuri*. Putative protein sequences were determined and an alignment was constructed (Figure 3.5). For none of these EST hits, other domains next to the expansin-like domain were detected. A closer inspection of the putative expansin-like sequence found in the EST library of the animal-parasitic nematode *Parastrongyloides trichosuri* revealed that it is identical to the expansin-like protein found in *Meloidogyne paranaensis*. This suggests that the EST from *P. trichosuri* is probably a contaminating EST from a *Meloidogyne* library.

A quick tblastn search against the nematode genomes of *M. incognita* and *M. hapla* revealed that these have respectively 12 and 7 contigs with significant similarity to the expansin-like domain.

The domain architecture of Da-EXP (SP / expansin-like domain / CBM) is particularly interesting. Almost all putative expansin-like genes found in the databases contain a signal peptide for secretion and an expansin-like domain only and in some cases a linker is present between those two domains. Only two additional genes have a CBM (Gr-EXP1 and Gr-EXPB1), but the domains are arranged in a different order (SP / CBM / expansin-like domain).

The nematode expansin-like domain does not have any significant similarity to plant expansins. It does have features of plant expansins, such as conserved cysteine residues and one of the conserved plant expansin motifs (HFD) (Cosgrove, 2000b) is partly conserved in nematode expansin-like proteins (H-D) (Figure 3.5).

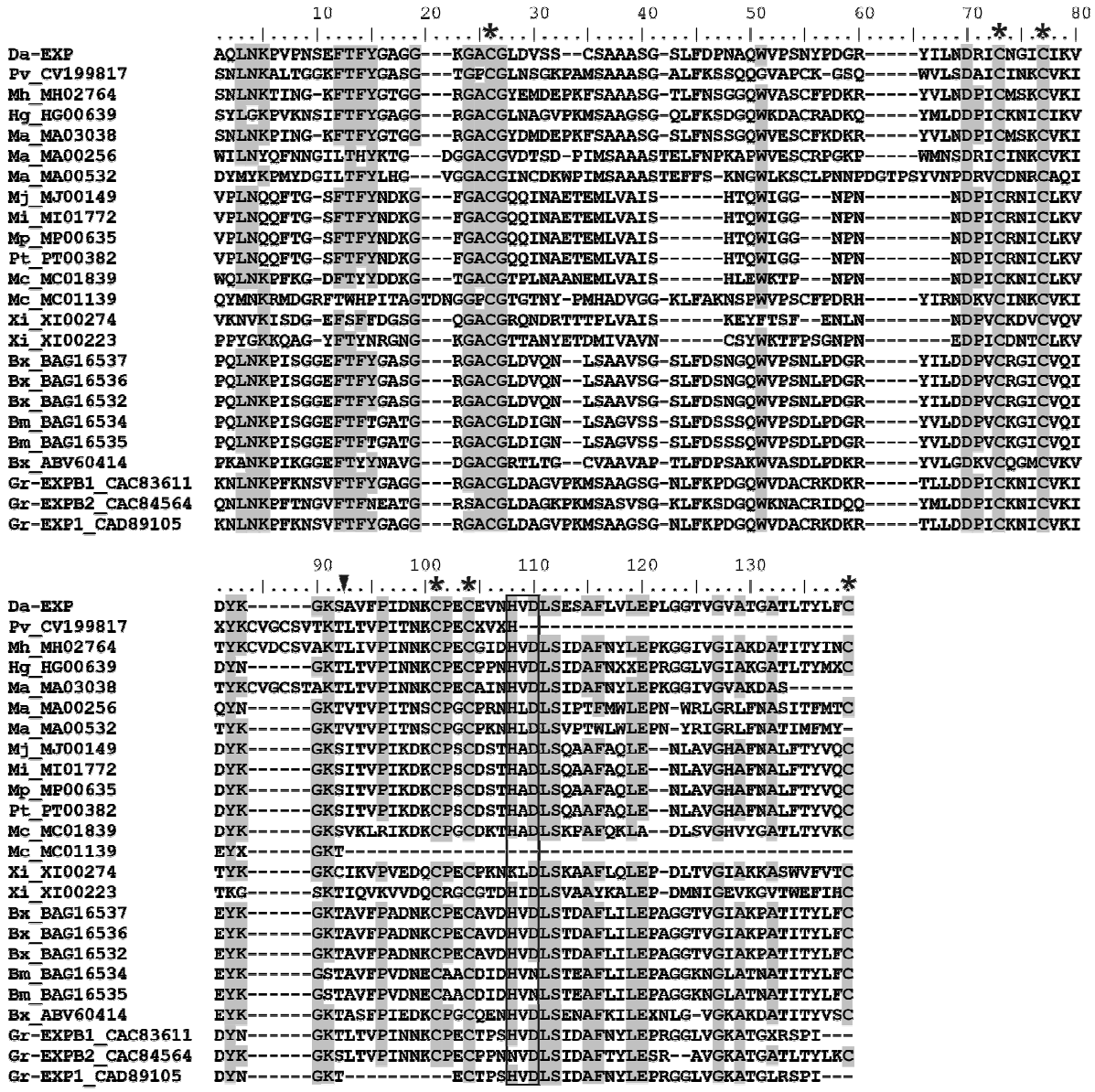

Figure 3.5 Protein alignment of the expansin-like domain of Da-EXP and its closest homologs. NemaGene cluster ids or accession numbers are given. The arrow indicates the position of the intron while stars indicate conserved cysteine residues. One of the conserved motifs in plant expansins (HFD) is boxed. Da: *Ditylenchus africanus*; Pv: *Pratylenchus vulnus*; Mh: *Meloidogyne hapla*; Hg: *Heterodera glycines*; Ma: *Meloidogyne arenaria*; Mj: *Meloidogyne javanica*; Mi: *Meloidogyne incognita*; Mp: *Meloidogyne paranaensis*; Pt: *Parastrongyloides trichosuri*; Mc: *Meloidogyne chitwoodi*; Xi: *Xiphinema index*; Bx: *Bursaphelenchus xylophilus*; Bm: *Bursaphelenchus mucronatus*; Gr: *Globodera rostochiensis*. Shading threshold: 75% identity.

The CBM of Da-EXP shows high similarity to the CBM of Da-ENG1A (81.8% similar), and the position of the intron is also conserved. The CBM of a *Globodera rostochiensis* endoglucanase (Gr-ENG1, AF004523) is less similar to the CBM of Gr-EXPB1 (53.7%) (Figure 3.6). The two tryptophans putatively involved in cellulose binding (Gilkes et al., 1991) are also conserved in the expansin-like proteins. The expansin-like domain and CBM are linked by 54 amino acids relatively rich in glycine (G), serine (S) and threonine (T).

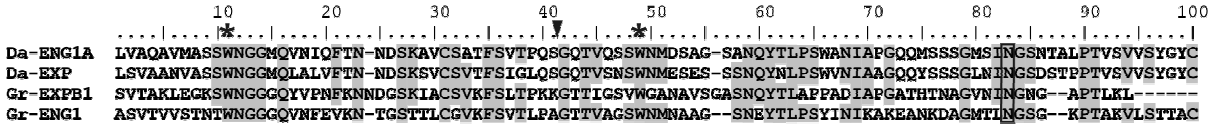

Figure 3.6 Protein alignment of the CBMs of expansin-like proteins of *D. africanus* and *G. rostochiensis* and the CBMs of endoglucanases Da-ENG1A and Gr-ENG1. The arrow indicates the position of the intron and the putative N-glycosylation site is boxed. The two conserved tryptophans (indicated by stars) are putatively involved in cellulose binding. Shading threshold: 75% identity.

A phylogenetic analysis of the expansin-like domain (Figure 3.7) illustrates that Da-EXP is most closely related to the expansin-like proteins of *Bursaphelenchus* spp., although this grouping is not strongly supported. Most expansin-like proteins of *Bursaphelenchus* spp. are grouped, however one of them is much less related to the others (ABV60414). Expansin-like proteins of cyst nematodes are placed together with high support. In contrast, the *Meloidogyne* expansin-like proteins are scattered over the tree. However, it does seem that expansin-like proteins with a similar linker length are more closely related to each other. For example, one strongly supported group of *Meloidogyne* proteins includes the expansin-like proteins where the signal peptide is immediately followed by the expansin-like domain. The expansin-like proteins from *Xiphinema index* are definitely less related to the other expansin-like proteins, and both proteins are also quite diverged from each other. The *Xiphinema* expansin-like proteins are most closely related to the bacterial expansin-like proteins. The latter proteins are not placed separately from the nematode expansin-like proteins, which is in favour of the horizontal gene transfer hypothesis. The presence of expansins in bacteria was indeed recently proven (Kerff et al., 2008).

The fact that expansin-like proteins of the same species are not necessarily grouped together implies that multiple gene copies of the protein exist, each with different evolutionary relationships. The putative pathogenicity and avirulence proteins form a separate group, although one of the *M. chitwoodi* expansin-like proteins (MC01139) seems to be equally related to the avirulence proteins as to other expansin-like proteins.

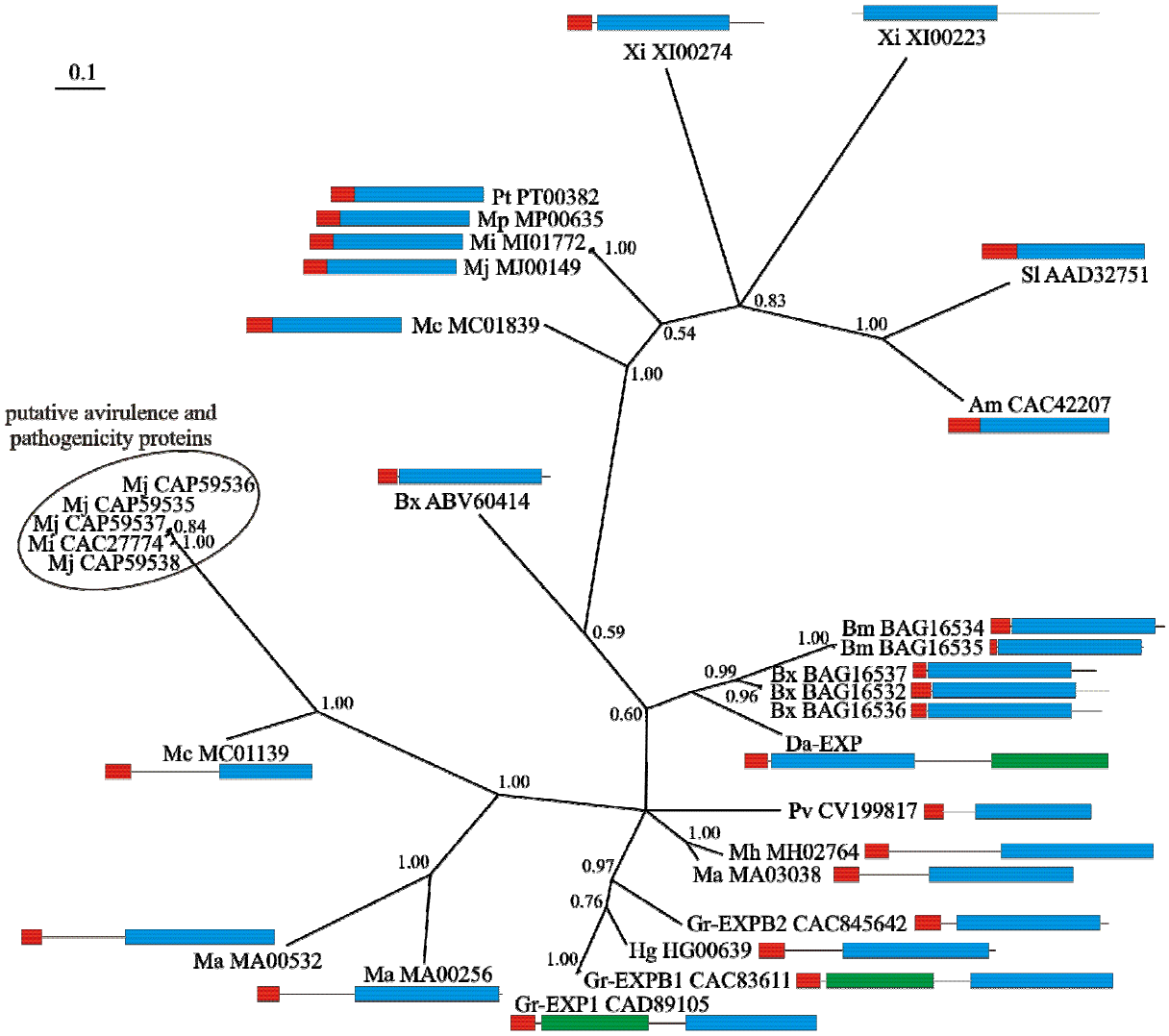

Figure 3.7 Bayesian unrooted tree inferred from the expansin-like domain of different proteins. For each protein the species is indicated by a two-letter code, followed by the corresponding NemaGene cluster id or accession number. The species two-letter codes are: Bm: *Bursaphelenchus mucronatus*; Bx: *Bursaphelenchus xylophilus*; Da: *Ditylenchus africanus*; Gr: *Globodera rostochiensis*; Hg: *Heterodera glycines*; Ma: *Meloidogyne arenaria*; Mc: *Meloidogyne chitwoodi*; Mh: *Meloidogyne hapla*; Mi: *Meloidogyne incognita*; Mj: *Meloidogyne javanica*; Mp: *Meloidogyne paranaensis*; Pt: *Parastrongyloides trichosuri*; Pv: *Pratylenchus vulnus*; Xi: *Xiphinema index*. The gene structure of the different expansin-like proteins is shown: red: signal peptide for secretion; blue: expansin-like domain; green: carbohydrate binding module. Putative avirulence and pathogenicity proteins from *Meloidogyne* spp. are encircled. Posterior probabilities are indicated on the branches.

Spatial and temporal expression pattern

To determine the spatial expression pattern, an *in situ* hybridisation was done with probes against *Da-eng1A* and *Da-exp* (Figure 3.8). Both probes stained the gland cell area of the nematode, the strongest being *Da-eng1A*. The negative control showed no staining. The temporal expression pattern was obtained by semi-quantitative RT-PCR on different stages of *D. africanus* (Figure 3.8). Both *Da-exp* and *Da-eng1A* have the same expression level in all stages, although the expression of *Da-eng1A* is much stronger.

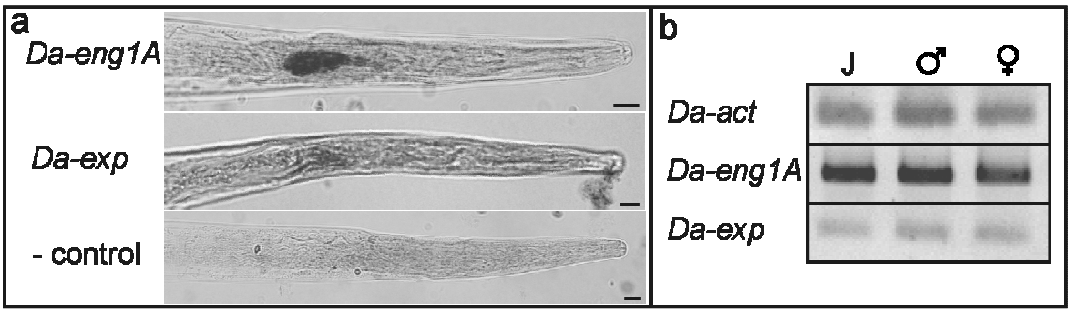

Figure 3.8 (a) *In situ* hybridisation on *Ditylenchus africanus* nematodes with antisense probes against *Da-eng1A* and *Da-exp*. As a negative control, a sense probe against *Da-eng1A* was used. Scale bars: 15 µm (b) Semi-quantitative RT-PCR (35 cycles) of actin (*Da-act*), endoglucanase (*Da-eng1A*) and expansin-like gene (*Da-exp*) on juveniles (J), male adults (♂) and female adults (♀) of *D. africanus.*

Discussion

The peanut pod nematode, *Ditylenchus africanus*, a member of the Sphaerularioideae superfamily (order Rhabidita, infraorder Tylenchomorpha), was demonstrated to express different cell wall modifying enzymes. Endo-1,4-β-glucanases that degrade cellulose probably form a gene family in this nematode. Two endoglucanases were cloned, *Da-eng1A* and *Da-eng1B*, which consist of a signal peptide for secretion, a GHF5 catalytic domain and a carbohydrate binding module (CBM) of family 2. These are the first plant-parasitic nematode GHF5 endoglucanases cloned outside the Tylenchoidea superfamily. Other endoglucanases have been cloned from *Bursaphelenchus* spp. in the Aphelenchoidea superfamily, but those belong to GHF45 and likely have a different evolutionary origin. The finding of a GHF5 endoglucanase family in *D. africanus* proves that the GHF5 endoglucanases are not restricted to the Tylenchoidea. A possible horizontal gene transfer event, as proposed to be the origin for the GHF5 endoglucanases (Jones et al., 2005), must have occurred in a common ancestor of both superfamilies. However there is some discussion about the placement of the Anguinidae family (containing *D. africanus*) within the Sphaerularioidea. It was proposed to consider a broader concept of the Tylenchidae family (superfamily Tylenchoidea) including the Anguinidae and at least part of the other Sphaerularioidea (Bert et al., 2008). Therefore, *D. africanus* could actually rather be a member of the Tylenchoidea, restricting the occurrence of GHF5 endoglucanases to this superfamily exclusively.

The *D. africanus* endoglucanase gene family unexpectedly holds endoglucanase-like genes with deletions. These genes can be described as pseudogenes, i.e. genes that are structurally similar to functional genes but that contain defects such as nonsense mutations, insertions, deletions or frameshifts which make them unable to produce functional proteins (D'Errico et al., 2004). Nevertheless, pseudogenes probably can acquire new functions during evolution. Since the presence of introns was shown in the endoglucanases with deletions, it concerns so-called non-processed pseudogenes that usually originate from gene duplication or rarely by

unequal crossing-over (D'Errico et al., 2004). Remarkably, the sequence just before the deleted part is the same as the end of the lost sequence in three of the four pseudo-endoglucanases. This suggests that homologous recombination could be the underlying mechanism for the origin of these deletions leading to pseudogenes. A comparison of the first intron shows that *Da-engdel1*, *Da-engdel2* and *Da-engdel3* probably arose from duplicates of *Da-eng1A*, while *Da-engdel4* is more likely to have originated from *Da-eng1B*. The deletions either give rise to truncated proteins (*Da-engdel1*, *Da-engdel3*) or erroneous proteins when a frameshift occurs (*Da-engdel2*, *Da-engdel4*). The latter genes, although no longer functional, only have a limited amount of mutations compared to the full endoglucanase genes. Therefore, these pseudogenes were probably formed very recently, otherwise it could be expected that more mutations would have accumulated already due to the lack of selective pressure. The two pseudogenes *Da-engdel1* and *Da-engdel3* code for a putative protein sequence with part of the catalytic domain deleted. The question is whether or not these proteins could still be functional. Cellulase activity is most probably lacking due to the loss of 3 α-helices and β-strands in the case of Da-ENGDEL1 and 2 α-helices and β-strands in the case of Da-ENGDEL3. Moreover, in both cases, one of the catalytic residues is missing, making it very unlikely that these proteins still have cellulose degrading activity. On the other hand, the CBM is intact in both cases and could still be functional. Both *Heterodera* spp. and *Meloidogyne* spp. secrete cellulose-binding proteins (CBP), that consist of a signal peptide for secretion, in some cases an unknown N-terminal sequence and a CBM (Ding et al., 1998; Gao et al., 2004a; Hewezi et al., 2008; Adam et al., 2008). Recently, a functional study suggested that CBP targets and/or activates a plant pectin methylesterase, to reduce the level of methylesterification of pectin in the cell wall. This allows improved access of other cell wall modifying enzymes to cell wall polymers, accelerating enzymatic activities (Hewezi et al., 2008). A similar activity could be assigned to the *D. africanus* endoglucanases with deletions, although experimental evidence is needed to elucidate the possible function of these truncated proteins. Nevertheless, these endoglucanases with deletions could either be an early stage of the evolution of new genes with possible new functions or just non-functional pseudogenes.

In this chapter, the isolation of another putative cell wall modifying enzyme from *D. africanus* is described, namely a putative expansin-like protein (Da-EXP). The expansin-like protein has a unique domain structure consisting of a signal peptide for secretion, an expansin-like domain and a CBM. The expansin-like domain shows significant similarity to expansin-like domains previously isolated from *Globodera rostochiensis* and *Bursaphelenchus* spp. and to two bacterial hypothetical proteins. No significant similarity to plant expansins was found, although some of the features of plant expansins are present, such as conserved cysteines and a conserved HFD motif, even though in nematodes only the histidine and aspartic acid are conserved. The question whether this protein truly is an expansin remains until functional studies prove its cell wall loosening activity, although this

was proven for the homologous expansin-like protein of *G. rostochiensis* (Qin et al., 2004). A search through all nematode ESTs revealed several hits, both in sedentary cyst and root-knot nematodes as well as in migratory nematodes such as *Pratylenchus vulnus*, *Bursaphelenchus* spp. (order Rhabditida, superfamily Aphelenchoidea) and *Xiphinema index* (order Dorylaimida). This illustrates the relatively widespread occurrence of the expansin-like proteins in plant-parasitic nematodes, in contrast to the GHF5 endoglucanases, which have so far not been found in *Bursaphelenchus* spp. or *Xiphinema index*. Moreover, it probably concerns a large gene family, since in both the *Meloidogyne incognita* and the *M. hapla* genome, several expansin-like contigs were identified. The finding of expansin-like proteins in several distantly related nematode taxa suggests that an early common ancestor of plant-parasitic nematodes probably already contained these proteins. A phylogenetic analysis showed that Da-EXP is more closely related to the *Bursaphelenchus* expansin-like proteins. This should however be handled cautiously, since the support for this group is low. More sequence information on expansin-like proteins in additional taxa should reveal more details about nematode expansin relationships and evolution in the future. Surprisingly, some putative avirulence proteins and pathogenicity factors of *Meloidogyne* spp. show significant similarity to the expansin-like domains. One of these putative avirulence proteins (*map-1*, CAC27774) was studied in more detail, and is believed to be involved in the early steps of recognition between resistant plants and avirulent nematodes (Semblat et al., 2001). At the time of discovery, *map-1* had no homologs in GenBank. It consists of a signal peptide and repetitive motifs linked to a C-terminal domain apparently resembling the expansin-like domains (46.4% similarity). The *map-1* gene and homologs may have a common origin with the expansin-like proteins and the evolutionary pressure of the constantly changing resistance gene caused an accelerated evolution leading to avirulence proteins only remotely related to expansin-like proteins. The recent finding that the perturbation of the cell wall (for example by the adhesion of CBMs) can elicit plant defense responses, supports this hypothesis (Dumas et al., 2008).

Perhaps the most interesting feature of Da-EXP is its CBM at the C-terminal end of the protein. Da-EXP has a similar domain architecture as the endoglucanases, and moreover, the CBMs of the two proteins have striking resemblance (81.8% protein similarity), and both have an intron at the exact same position. Even the linker sequence of both genes is enriched in the same amino acids, i.e. threonine, serine and glycine. This suggests that both CBM domains of Da-EXP and Da-ENGs have a common origin. Most nematode expansin-like proteins however do not have a CBM. A possible hypothesis could be that Da-EXP originated from Da-ENG, and that the pseudo-endoglucanases with deletions are an evolutionary intermediate between the two genes. This would imply that the expansin-like domain emerged from a GHF5 endoglucanase catalytic domain. It is true that nematode expansin-like proteins have features of plant expansins, but no significant similarity, pointing towards convergent

evolution. Nevertheless, the latter hypothesis is quite radical and probably inaccurate for different reasons. First of all, the expansin-like domain has been found in many nematode species, and has similarity to bacterial proteins, which makes an origin through horizontal gene transfer more plausible. Second, almost no other expansin-like proteins include a CBM, the link to the endoglucanases. Third, the CBM is very conserved between Da-EXP and the Da-ENGs, while the expansin-like domain does not resemble the endoglucanase catalytic domain. This would involve different rates of evolution for the different domains. Overall, a more reasonable hypothesis is domain shuffling, in which case an expansin-like domain was linked to a CBM from an endoglucanase. This hypothesis could also explain why in Gr-EXP1 and Gr-EXPB1, the CBM and expansin-like domain are in the opposite order compared to Da-EXP.

A whole-mount *in situ* hybridisation revealed that both *Da-eng1A* and *Da-exp* are expressed in the gland cell area of the nematode, although it could not be determined whether the expression was located in the dorsal or subventral gland cells. The presence of a signal peptide in both proteins and the expression in the gland cell area suggest that these proteins are secreted from the nematode. The expression of *Da-eng1A* in different life stages is considerably higher than the expression of *Da-exp*, which is consistent with the stronger signal of *Da-eng1A* obtained in the *in situ* hybridisation. The stronger expression level of *Da-eng1A* probably reflects its higher functional importance to invade the plant roots than *Da-exp*, although it is also possible that the primers were able to amplify additional gene duplicates (e.g. *Da-eng1B*). Both *Da-eng1A* and *Da-exp* show a similar expression level among different life stages (juveniles, male and female adults). This implies that both genes are equally important in juvenile and adult stages, which is in line with the constant migratory lifestyle of the nematode.

Acknowledgements

We would like to thank Prof. A. McDonald and Dr. H. Fourie (ARC-GCI, Potchefstroom, South Africa) for kindly providing the initial *D. africanus* culture.

CHAPTER 4

THE EVOLUTION OF THE GENE STRUCTURE OF NEMATODE GHF5 ENDOGLUCANSES

Adapted from:

Tina Kyndt, **Annelies Haegeman** and Godelieve Gheysen. (2008). Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: no evidence for an early domain shuffling event. BMC Evolutionary Biology, 8, 305.

Abstract

Endo-1,4-β-glucanases or cellulases from the glycosyl hydrolase family 5 (GHF5) have been found in numerous bacteria and fungi, and recently also in higher eukaryotes, particularly in plant-parasitic nematodes (PPN). Previous studies on the structure and evolution of these genes within the phylum Nematoda have focused primarily on data of sedentary nematodes, while in this chapter, extra data from migratory nematodes is included: i.e. four endoglucanases of *Radopholus similis* (chapter 2), one endoglucanase of *Ditylenchus africanus* (chapter 3) and a newly isolated endoglucanase of *Pratylenchus coffeae*.

Phylogenetic analyses show that the evolution of the catalytic domain and the carbohydrate binding module (CBM) follow a similar evolutionary tract. The exon/intron structure was compared among the PPN GHF5 endoglucanases and based on the location of the introns we inferred a model for the evolution of the GHF5 endoglucanase gene structure in plant-parasitic nematodes. Our model proposes the occurrence of an early duplication event of an endoglucanase gene including a CBM, and more recent gene duplications at genus or species level, in some cases accompanied by a loss of the CBM. The model implies the occurrence of some interesting evolutionary events, such as intron sliding or reverse transcription. Moreover, our data confirm a close relationship between *Pratylenchus* spp. and the root-knot nematodes, while some *Radopholus similis* endoglucanases are more similar to cyst nematode genes.

Introduction

Endo-1,4-β-glucanases (EC 3.2.1.4), or endoglucanases, can degrade the (1,4)-β-linkages of cellulose, the most abundant component of plant cell walls. The endoglucanases are classified in different glycosyl hydrolase families (GHF) on the basis of sequence similarity and hydrophobic cluster analysis (Henrissat & Bairoch, 1993). Animal endoglucanases belong to three structurally and presumably phylogenetically unrelated families: GHF5, GHF9 and GHF45 (Lo et al., 2003). GHF5 genes are found in various plant-parasitic bacteria and nematodes (Smant et al., 1998; Rosso et al., 1999; Uehara et al., 2001) from the order Rhabditida (infra-order Tylenchomorpha), which is part of one of the three evolutionary independent plant-parasitic nematode clades (Blaxter et al., 1998). In the other two plant-parasitic orders (Dorylaimida and Triplonchida), no GHF5 endoglucanases have been found so far. The nematode GHF5 endoglucanases consist of a catalytic domain, and in some cases an additional linker and carbohydrate binding module (CBM), which is thought to facilitate the interaction between the enzyme and its substrate (Boraston et al., 2004). Interestingly, hypothetical proteins similar to GHF5 endoglucanases were recently found in the genome of the free-living nematode *Pristionchus pacificus*, possibly a kind of preadaptation towards a parasitic lifestyle (Dieterich et al., 2008).

Horizontal gene transfer (HGT) from bacteria has been proposed to explain the origin of GHF5 endoglucanases in nematodes (Jones et al., 2005). At present, no hard evidence for the HGT hypothesis of GHF5 endoglucanases has been found. Two beetle species were also found to have endogenous GHF5 endoglucanases (Sugimura et al., 2003; Wei et al., 2006), and as more sequence data is becoming available, additional endoglucanases could be found in other animals as well. An evolutionary scheme proposed for GHF5 genes of sedentary nematodes by Ledger et al. (2006) hypothesised that after the possible HGT event, the endoglucanase genes must have been duplicated several times, and in some cases a sequential loss of the linker and CBM occurred. This model focused on genes from cyst and root-knot nematodes and one migratory nematode, all belonging to the Tylenchoidea superfamily. In order to have a broader overview of the endoglucanase evolution in the infra-order Tylenchomorpha, the gene structure of six additional genes was incorporated in our study. These include four *Radopholus similis* genes (chapter 2), one from *Ditylenchus africanus* (chapter 3) and a newly isolated one from *Pratylenchus coffeae*. In this chapter, we propose a more detailed evolutionary model of the gene structure of GHF5 endoglucanases within the Nematoda.

Materials and methods

Nematode culture

Pratylenchus coffeae was maintained on carrot disks in small petri dishes at a constant temperature of 25°C (Moody et al., 1973). Nematodes were collected by rinsing the petri dishes with sterile demineralised water approximately 8 weeks after inoculation of the carrot disks. Collected nematodes of mixed stages were stored as a pellet at -20°C for DNA or RNA extraction.

Cloning of endoglucanase gene from *P. coffeae*

Genomic DNA was isolated from the nematodes according to Bolla et al. (1988). RNA extraction and first strand cDNA synthesis were done as described in chapter 2.

A polymerase chain reaction (PCR) using degenerate primers ENG1 and ENG2 (Rosso et al., 1999) (Table 4.1) was performed on genomic DNA of *P. coffeae* and resulting fragments were cloned and sequenced as described in chapter 2.

To obtain the 5' and 3' genomic sequences of the endoglucanase fragments, the Genome Walker Universal kit (Clontech) was used according to the manufacturer's instructions. Primers used were Pc-eng1-up1 and Pc-eng1-up2 to walk upstream and Pc-eng1-down1 and Pc-eng1-down2 to walk downstream (Table 4.1). The longest obtained fragments were cloned and sequenced. The corresponding cDNA sequence of the endoglucanase gene was amplified by PCR using gene-specific primers Pc-eng1-start and Pc-eng1-stop on *P. coffeae* cDNA and cloned and sequenced as well (Table 4.1). The obtained sequence was submitted to the GenBank database (EU176871). The putative protein sequence was obtained by translating the cDNA sequence using the EMBOSS program "Transeq" (Rice et al., 2000). The molecular weight of the protein was estimated by the "pI/Mw" tool of the Expasy server (http://us.expasy.org). The presence of a signal peptide was predicted using SignalP 3.0 (Bendtsen et al., 2004).

Table 4.1: Primers used for cloning of *Pratylenchus coffeae* endoglucanase

Primer	Primer sequence	Primer	Primer sequence
ENG1	TAYGTIATHGTIGAYTGGCA	ENG2	GTICCRTAYTCIGTIACRAA
Pc-eng1-up1	TTGGAGTCAATGGCCCGGATGG	Pc-eng1-up2	GATGTTCGGATTGGAGCCATATTGC
Pc-eng1-down1	TCAAATCTCTGCTACACTCTCCAC	Pc-eng1-down2	CAAACAATCCCTCAGGGATAAGGC
Pc-eng1-start	ATGGCATTCACTTTGCTTTCC	Pc-eng1-stop	TGTGAGCGCCACTGCCTGCTAA

Sequence searches and alignments

All available nematode GHF5 endoglucanase sequences and some selected sequences from beetles, bacteria, fungi, protists and plants (protein, mRNA and DNA if available) were downloaded from GenBank and are shown in Table 4.2. Domains were identified using the

SMART tool (Schultz et al., 1998), which is based on the PFAM database. Sequences were aligned using ClustalW (Thompson et al., 1994) and then manually adjusted in BioEdit (Hall, 1999). The recently found GHF5 endoglucanase homologs in the free-living nematode *Pristionchus pacificus* (Dieterich et al., 2008) are not included in this study. Annotated versions of these genes are not available yet, and our own blast searches with different nematode GHF5 endoglucanase genes against the genome of *P. pacificus* (www.pristionchus.org) did not reveal any significant homologs.

Table 4.2: Sequence data used to construct alignments including gene name, accession number and available data

Gene	Acc	Data	Gene	Acc	Data
NEMATODA			***Pratylenchus penetrans***		
Meloidogyne incognita			*Pp-eng1*	AB045780	partial DNA, cDNA
Mi-eng1	AF323087	DNA, cDNA	*Pp-eng2*	AB045781	partial DNA, cDNA
Mi-eng2	AF323088	DNA, cDNA	***Pratylenchus coffeae***		
Mi-eng3	AY422836	cDNA	*Pc-eng1*	EU176871	DNA, cDNA
Mi-eng4	AY422837	cDNA	***Radopholus similis***		
Mi-eng5	AF323090	partial DNA	*Rs-eng1A*	EF693940	DNA, cDNA
Mi-eng6	AF323091	partial DNA	*Rs-eng1B*	EF693941	DNA, cDNA
Mi-eng7	AF323092	partial DNA	*Rs-eng2*	EF693942	DNA, cDNA
Mi-eng8	AF323093	partial DNA	*Rs-eng3*	EF693943	DNA, cDNA
Mi-eng9	AF323094	partial DNA	***Ditylenchus africanus***		
Mi-eng10	AF323095	partial DNA	*Da-eng1*	EU180235	DNA, cDNA
Meloidogyne javanica			**PLANTS**		
Mj-eng1	AF323099	partial DNA	*Arabidopsis thaliana eng1*	AC007357	DNA, cDNA
Mj-eng2	AF323100	partial DNA	*Arabidopsis thaliana eng2*	AL391147	DNA, cDNA
Mj-eng3	CAJ77137	cDNA	*Arabidopsis thaliana eng3*	AL391150	DNA, cDNA
Meloidogyne arenaria			**ANIMALS**		
Ma-eng1	AF323097	partial DNA	*Psacothea hilaris*	AB080266	cDNA, no introns
Heterodera schachtii			*Apriona germari*	AY771358	cDNA, no introns
Hs-eng1	CAC12958	cDNA	**BACTERIA**		
Hs-eng2	CAC12959	cDNA	*Cytophaga hutchinsonii*	YP_678708	DNA, cDNA
Heterodera glycines			*Bacillus licheniformes*	AAU23613	DNA, cDNA
Hg-eng1	AF052733	DNA, cDNA	*Bacillus subtilis*	AAK94871	DNA, cDNA
Hg-eng2	AF052734	DNA, cDNA	*Erwinia carotovora*	CAA53592	DNA, cDNA
Hg-eng3	AF056048	DNA, cDNA	*Erwinia chrysanthemi*	AAF18152	DNA, cDNA
Hg-eng4	AY325809	DNA, cDNA	*Saccharophagus degradans*	ABD82494	DNA, cDNA
Hg-eng5	AY336935	DNA, cDNA	*Cellvibrio japonicus*	CAA60493	DNA, cDNA
Hg-eng6	AY163572	DNA, cDNA	*Pseudoalteromonas atlantica*	ABG39929	DNA, cDNA
Globodera rostochiensis			*Thermobifida fusca*	AAC09379	DNA, cDNA
Gr-eng1	AF056110	DNA, cDNA	**FUNGI**		
Gr-eng2	AF056111	DNA, cDNA	*Trichoderma reesei eng2*	DQ178347	DNA, cDNA
Gr-eng3	AF408154	DNA, cDNA	*Trichoderma reesei eng3*	M19373	DNA, cDNA
Gr-eng4	AF408157	DNA, cDNA	**PROTISTS**		
Globodera tabacum			*Spirotrichonympha leidyi*	AB189037	DNA, cDNA
Gts-eng1	AF182392	cDNA	protist from *Cryptocercus*	AB274699	cDNA
Gts-eng2	AF182393	cDNA	protist from *Reticulitermes*	AB274531	cDNA

Exon-intron structure analyses

The presence, location and phase of introns were evaluated for each open reading frame (ORF) by comparing DNA and mRNA sequences. Intron positions were mapped onto the sequence alignments. Intron phase was named 0 when the intron was located in-between two consecutive codons; 1 if an intron was located between the first and second codon nucleotides and 2 if an intron was found between the second and third codon nucleotides. The intron

positions were mapped onto the secondary structure. Positions of α-helices and β-strands in the catalytic domain were deduced from an alignment of all nematode endoglucanases with an endoglucanase from *Erwinia chrysanthemi* (1EGZC) that has a known 3D structure. For the carbohydrate binding module, the location of the β-strands was inferred by aligning the sequences to a CBM from an endoglucanase from *Cellulomonas fimi* (1EXG). GC percentages of the introns were calculated with the internet tool GC calculator (http://www.genomicsplace.com/gc_calc.html). A consensus splice site of all introns was derived and represented using Weblogo 3 (Crooks et al., 2004).

Phylogenetic analyses

Maximum parsimony (MP) analyses were executed using PAUP* v4.0b10 (Swofford, 2003), on the informative characters of the mRNA and protein alignments using the heuristic search option with random sequence addition (100 random replications) and TBR branch-swapping. Support for the different clusters was evaluated by bootstrap analysis (100 replicates).
Hierarchical likelihood ratio tests were performed on all datasets using Modeltest 3.5 (Posada & Crandall, 1998) to determine the best-fitting evolutionary model. Estimated model parameters were applied in a maximum likelihood (ML) analysis in PAUP* v4.0b10 using the heuristic search option with random sequence addition, for DNA sequences. Tree-puzzle 5.2 (Schmidt et al., 2002) was used for ML analysis of protein sequences. Support for the different clusters was evaluated by bootstrap analysis (100 replicates).
Bayesian analysis was run using MrBayes version 3.1.2 (Ronquist & Huelsenbeck, 2003). For the DNA datasets the best-fitting evolutionary model, as identified using Modeltest 3.5, was applied. For the protein datasets a mixed model of amino acid evolution was applied to allow model-jumping for fixed-rate models. Bayesian inference was run for 1,000,000 generations, and the first 25,000 generations were discarded as burn-in.

Results

Endoglucanase gene of *Pratylenchus coffeae*

Endo-1,4-β-glucanase fragments were amplified by a degenerate PCR on genomic DNA of *P. coffeae*. This resulted in a fragment of 443 bp and the corresponding gene was named *Pc-eng1*. The full length sequence, obtained by genome walking, is 1527 bp from start till stop codon including 3 introns. The corresponding coding sequence amplified from a cDNA pool is 1377 bp. The putative endoglucanase protein Pc-ENG1 has an estimated molecular weight of 47 kDa. A conserved domain search confirmed that the protein is an endoglucanase belonging to glycosyl hydrolase family 5 (GHF5, pfam00150). The protein consists of a

predicted N-terminal signal peptide for secretion, a catalytic domain, a linker and a CBM of family 2 (pfam00553).

Phylogenetic analyses

One or more representative GHF5 endoglucanases were selected for different lineages for further analyses. These included GHF5 endoglucanase sequences from plants, animals, fungi, protists and bacteria (Table 4.2). Some of these genes consist of both a catalytic domain and a CBM, while others lack the CBM. Separate alignments were made for the catalytic domain and the CBM of the genes. Since not all GHF5 endoglucanase genes include a CBM, the alignment and hence also the phylogenetic tree for the CBM contained less taxa. For both domains, alignment of the sequences was performed at both DNA and protein sequence level.

Alignments showed that plant sequences are not sufficiently similar to PPN GHF5 endoglucanases to be useful for phylogenetic analyses. Preliminary unrooted phylogenetic tree construction with the other endoglucanases revealed the fungal sequences to be most divergent from PPN GHF5 endoglucanases. Therefore they were chosen as outgroup to construct rooted trees. All alignments were analysed using maximum parsimony (MP), maximum likelihood (ML) and Bayesian methods. All evolutionary trees, whether based on DNA or protein data, showed similar clustering, independent of the method used.

Figure 4.1 shows the Bayesian tree based on the protein alignment of the catalytic domain, using the fungal genes as outgroup.

Protist genes form a well-confirmed monophyletic group nested in the paraphyletic bacterial cluster. Animal genes also form a monophyletic group in which beetle genes cluster separately from the PPN genes. Within the PPN, Mi-ENG2, Da-ENG1 and Hg-ENG6 are at the base of the protein family but do not show a well-supported clustering. This grouping is correlated with their aberrant exon/intron structure (see further). The divergence of Da-ENG1 at sequence and structural level is not surprising since *Ditylenchus africanus* is the only species in our study belonging to the Sphaerularioidea and not to the Tylenchoidea.

A well-confirmed cluster groups the GHF5 endoglucanases of *Pratylenchus coffeae* and *P. penetrans* with those of the root-knot nematodes (genus *Meloidogyne*). Cyst nematode (*Heterodera* and *Globodera*) endoglucanases are grouped together with some endoglucanases from *Radopholus similis* (classified in the Pratylenchidae family), Rs-ENG1B and Rs-ENG3. Paralogy between groups of proteins is well-confirmed by high posterior probabilities. For instance, Rs-ENG3 and Hg-ENG5 are orthologous to each other and to two paralogs from *Globodera rostochiensis* (Gr-ENG3 and Gr-ENG4). Also, Rs-ENG1B is orthologous to some intra-species or intra-genus duplicates from *Heterodera* and *Globodera* spp.

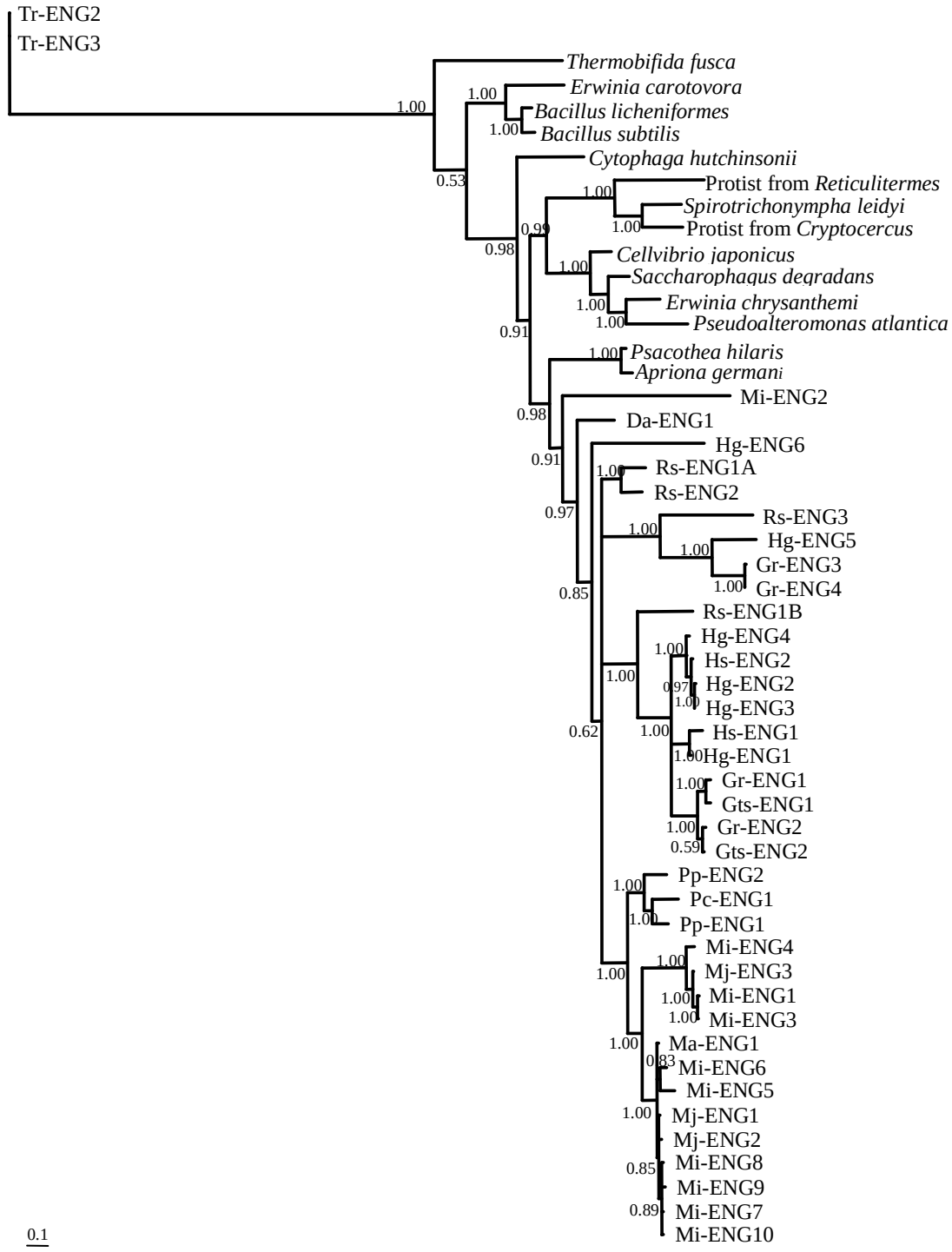

Figure 4.1: Bayesian phylogenetic tree of the GHF5 catalytic domain of endoglucanases from different prokaryotic and eukaryotic lineages, using the WAG model. The posterior probability is given on each node. The tree is rooted with the catalytic domain of two genes from the fungus *Trichoderma reesei*. The scale bar represents branch length (number of amino acid substitutions/site).

The Bayesian tree based on the CBM reveals a similar clustering, and grouping of most PPN paralogs is again supported by maximum posterior probability values (eg. Hs-ENG1 and Hg-ENG1; Pc-ENG1 and Pp-ENG1). However, posterior probabilities show low support for the relationships between genes from bacterial and PPN taxa. To compare this tree with the tree based on the catalytic domain, the Bayesian analysis for the catalytic domain was rerun including genes with a CBM only. Both resulting trees are compared in Figure 4.2. Although the CBM tree is not fully resolved, most clades are confirmed by both evolutionary trees with high posterior probabilities and/or bootstrap values. It must be noted that the protein similarities between CBMs of different genes are on average lower than the similarities of the corresponding catalytic domains. This was already noticed in chapter 2 within the Pratylenchidae, and can now be extended to all PPN GHF5 endoglucanases. This could suggest that the CBMs evolve faster than catalytic domains.

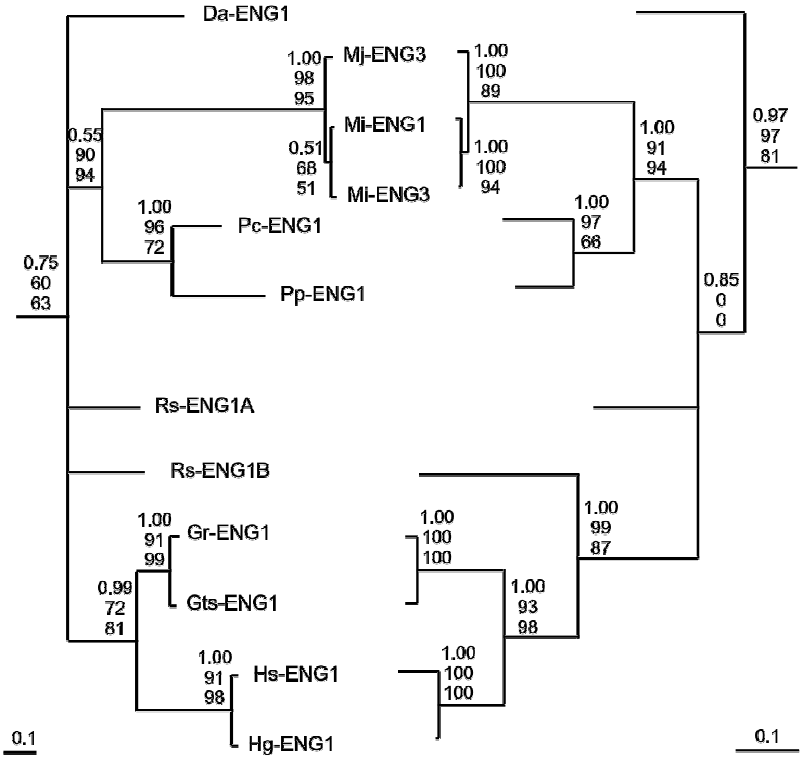

Figure 4.2: Comparison between the Bayesian phylogenetic tree of the carbohydrate binding module (left) and the tree of the catalytic domain (right) for all plant-parasitic nematode GHF5 endoglucanase genes with a CBM. Scales of the corresponding trees are given in the lower corners. On each node, three values are given. From top to bottom: the posterior probabilities of the Bayesian analyses; the bootstrap values of the maximum likelihood analyses and the bootstrap values of the maximum parsimony analyses.

No data from linker sequences have been included in this study (except for the presence of intron 22 in the linker of *Mi-eng1*, see further) because linkers show very little large-scale sequence homology. This is not surprising as the same observation has been made for linkers from bacterial endoglucanases (Gilkes et al., 1991). At DNA level, only linkers from very closely related genes show some degree of homology. In general, nematode linkers, as well as bacterial linkers, are short and mainly composed of small or tiny amino acids like glycine, serine, threonine, alanine, proline, and lysine.

Investigation of gene/domain structure and intron properties

The exon/intron gene structure of the PPN GHF5 endoglucanases was compared to the multiple protein sequence alignment. It was proposed before that introns appear preferably at the borders of domains or in between secondary structures in proteins to facilitate domain shuffling (Gilbert, 1978). This theory apparently does not hold for the nematode GHF5 endoglucanases: there is no tendency for introns to occur at the borders of domains or α-helices or β-strands, as illustrated in Figure 4.3. Only intron 1 is located in-between the signal peptide and the catalytic domain, and the well-conserved introns 20 and 21 are in the transition zone between the catalytic domain and the linker.

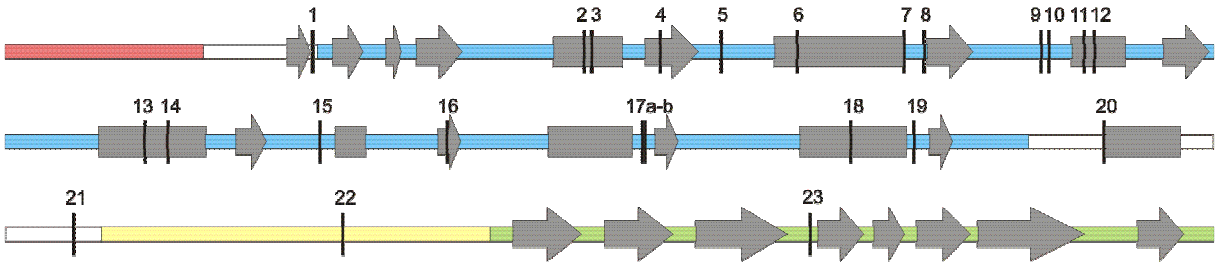

Figure 4.3: Intron positions of a typical PPN GHF5 endoglucanase in relation to the secondary gene structure and different domains. The signal peptide, catalytic domain, linker and CBM, are indicated in red, blue, yellow and green respectively. Transition zones are white. Arrows represent β-strands; boxes α-helices. Introns are represented by a vertical black line and numbered in order of their occurrence, as further specified in Figure 4.4.

All introns and their properties are represented in Figure 4.4. The ORF of PPN GHF5 endoglucanase genes is interrupted by maximum 8 introns (*Hg-eng1*) with an average of 5.28 introns. The average length of each intron is 132 bp, making a total of 697 bp of intron per gene, which is smaller than the (already small) average intron size observed in the nematode lineage (467 bp per intron) (Deutsch & Long, 1999).

The GC percentage per intron varies from 5% (intron 10 of *Mi-eng5*) to 58% (intron 2 of *Gr-eng1*), with an average for all introns (n=110) of 31%. Apparently there are differences in GC percentage according to the species. For instance, the average GC% of all introns of *M. incognita* endoglucanase genes (n=23) is 22%, whereas for *R. similis* introns (n=13) it is 42%. This is probably related to the overall GC% of the coding sequences of a certain species. *M. incognita* coding sequences generally have a relatively low GC content (37%) (Mitreva et al.,

2006), whereas *R. similis* coding sequences have a high GC content (54%) (Jacob et al., 2008).

67% of the analysed introns are in phase 0, while only 17% correspond to phase 1 or phase 2. 49% (54/110) of the exons are symmetrical. As a consequence of this high over-representation of phase 0 introns, symmetrical exons of type 0-0 (44%, 48/110) are largely in excess compared to symmetrical exons of type 1-1 (4%, 4/110) or type 2-2 (2%, 2/110). Phase 0 is highly over-represented in PPN endoglucanases, as observed by Ledger et al. (2006) as well, with twice the expected value for an addition of intron with equal probability for the three codon sites. This is also considerably higher than the observed prevalence in *C. elegans* (47%) (Deutsch & Long, 1999). The observed excess of phase 0 introns can be due to the preferential retention of these introns by natural selection (Sverdlov et al., 2003) or as a consequence of the burst of domain shuffling that has been shown to coincide with the big bang of metazoan radiation in the Cambrian period (Patthy, 1999; Kaessmann et al., 2002).

Exon-intron boundaries were compared for all introns observed in the available PPN GHF5 endoglucanase gene sequences. The great majority of the detected introns are of the GU-AG type. Only three intron positions (intron 13, 17a and 17b) show the alternative, but rare (0.6% in *C. elegans*) GC-AG splice site in a minority of the homologs. This non-classical splice site was already reported by Yan et al. (1998). A consensus 5’ and 3’ splice site was deduced from the exon-intron boundaries and is shown in Figure 4.5. Comparing this consensus with the previously described *C. elegans* splice site consensus (Blumenthal & Steward, 1997) shows a low degree of conservation both at the 5’ and 3’ side. Low 5’ splice site conservation has been shown to be correlated with a large intron number and the ability to generate alternatively spliced mRNAs (Irimia et al., 2007). Protosplice site tendency in PPN GHF5 endoglucanases is moderately high as 46 % (11/24) of the introns show a consensus splice site that is completely consistent with the protosplice site (MAG/R), pointing to a probable recent origin of these introns.

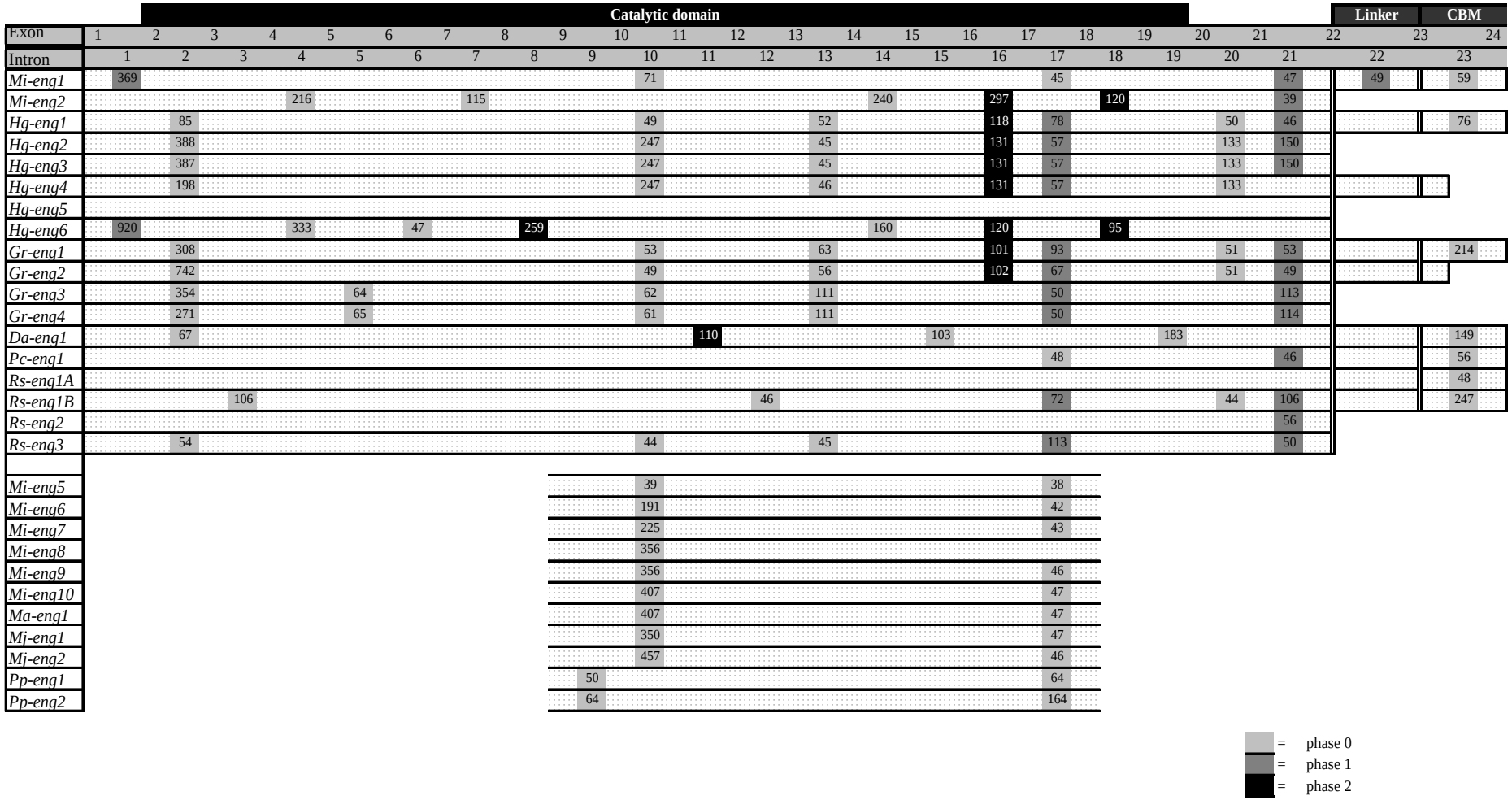

Figure 4.4: Schematic overview of the exon/intron structure of all available genomic sequences from PPN GHF5 endoglucanase genes. Introns are considered to be homologous when they appear at exactly the same location in the amino acid alignment. Phase of the introns is shown by gray-scale boxes. Length of the introns (in bp) is given inside each box. Domains, as identified by the SMART tool (PFAM database), are represented above the genes. Exon/intron and domain length are not drawn to scale.

	5' splice site											
	-3	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
A	58	59	5	0	0	65	83	19	35	45	43	42
C	19	39	2	0	12	4	7	8	11	17	26	18
G	34	11	85	116	0	29	7	68	7	21	14	18
U	5	7	24	0	104	18	19	21	63	33	33	38
Consensus												
***C. elegans* consensus**	A/C	A	G	G	U	A/G	A	G	U	U	U	
All organism consensus	A/C	A	G	G	U	A/G	A	G	U	U	U	

	3' splice site										
	-9	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**
A	24	28	27	20	11	28	42	116	0	29	19
C	31	38	28	24	13	8	41	0	0	9	21
G	19	11	5	9	3	15	5	0	116	78	14
U	42	39	56	63	89	65	28	0	0	0	62
Consensus											
***C. elegans* consensus**			U	U	U	U	C	A	G	A/G	
All organism consensus							C/T	A	G	G	

Figure 4.5: Nucleotides at the 5' and 3' splice sites of all studied introns of PPN GHF5 endoglucanases. Splicing occurs at the vertical lines. Positions are numbered with respect to the splice sites and the consensus splice site is visually represented. For comparison, the consensus splice site of *C. elegans* and all organisms (Blumenthal & Steward, 1997) is included below the deduced PPN consensus sequence.

Even though many clusters of closely spaced introns were detected in the aligned sequence, no evidence was found in our study, or in other studies, that proves their possible homology. Although intron sliding cannot be completely ruled out, the influence of this process is considered to be negligible (except for sliding by a few bases), and intron position diversity is indicated to be primarily arising by the gain or loss of introns during eukaryotic evolution (Stoltzfus et al., 1997). Therefore, a pair of introns was required to occur in the exact same position in the aligned sequence of the orthologous/paralogous genes to be considered homologous. Noteworthy, there is a strong conservation of the location and phase of introns within the PPN GHF5 gene family (Ledger et al., 2006). This conservation is evident among orthologous genes, but also among paralogs. Some intron positions are very well conserved (intron 2, 10, 13, 16, 17b, 21 and 23) while others only occur in one gene (intron 3, 6, 7, 8, 11, 12, 15, 19, 22).

Since spliceosomal introns diverge so rapidly that sequence similarity indicative of homology quickly vanishes, significant intron sequence conservation can only be expected in genes resulting from very recent duplications. Sequence comparisons between introns that are conserved in the majority of the dataset indeed showed similarities between paralogous genes.

For instance, introns of *Gr-eng1* and *Gr-eng2* as well as *Gr-eng3* and *Gr-eng4* reveal remarkable similarity. This is also the case for introns in *Hg-eng2*, *Hg-eng3* and *Hg-eng4*. The conserved introns of the *Meloidogyne* endoglucanase genes are also very similar, confirming the homology of these conserved introns and the paralogy of the genes.

A model for the evolution of introns in the PPN GHF5 endoglucanases

A model was constructed for nematode endoglucanase genes by combining the sequence-based evolutionary tracts with the intron data. Figure 4.6 shows a Bayesian tree obtained from the catalytic domain for all genes for which the exon/intron structure is available.

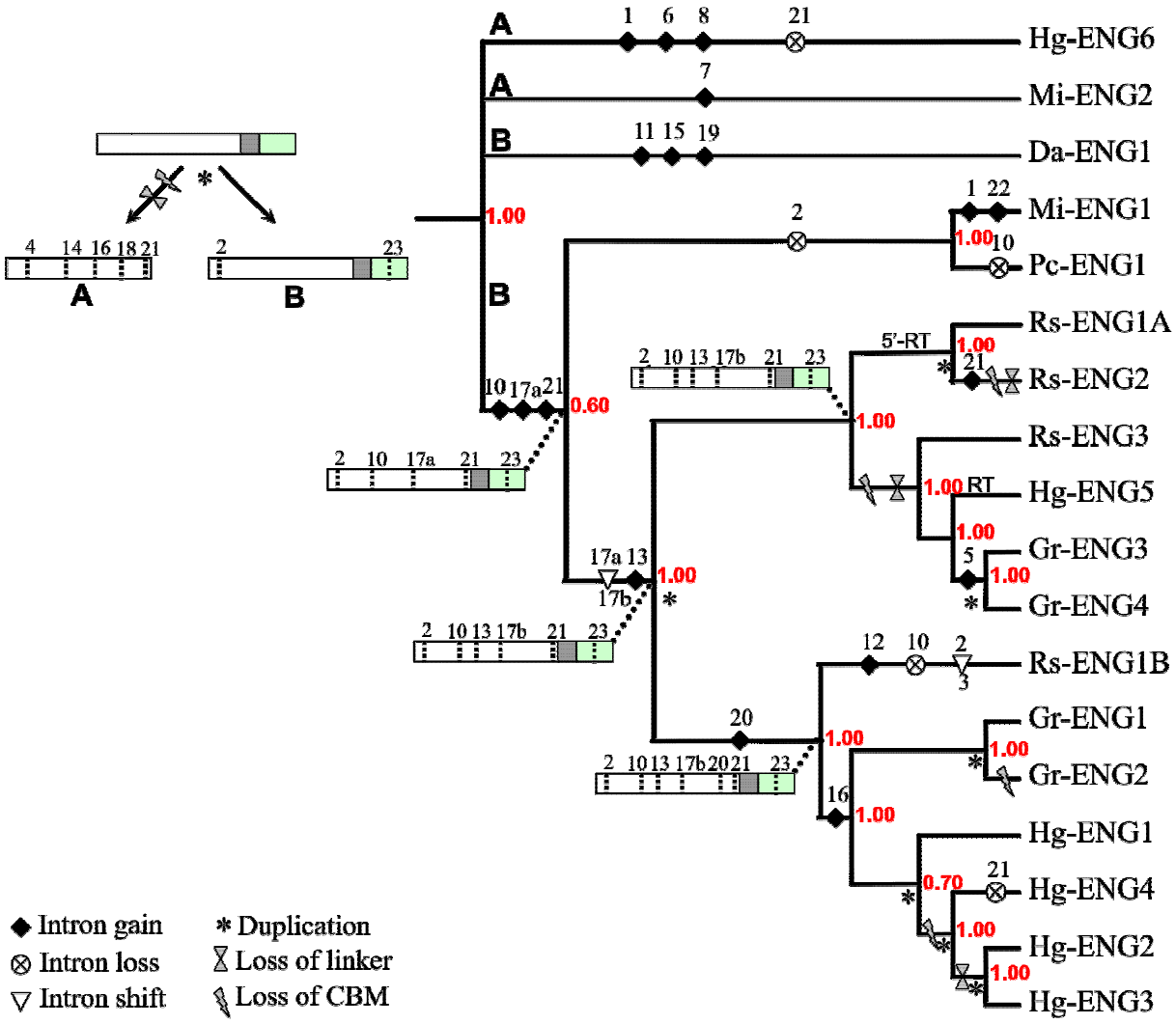

Figure 4.6: Proposed model for GHF5 endoglucanase gene structure evolution in plant-parasitic nematodes based on a Bayesian tree obtained from genes for which gene structure information was available. Posterior probabilities are indicated in red for each node. Introns are numbered as specified in Figure 4.4. Assumed gene structures are schematically drawn for certain nodes: the catalytic domain is represented by a white box, the linker is shown as a grey box and the CBM as a green box; Intron presence is shown by dotted lines. RT: Reverse Transcription.

The gene evolution model plotted onto this tree is based on the parsimony principle, meaning that the occurrence of intron gain and loss events during evolution is minimised. The evolutionary tree based on the catalytic domains shows the genes with the same structure grouped in the same terminal branches, although the length and sequence of introns have diverged.

Our working hypothesis proposes that the divergence of the gene structure of the PPN GHF5 gene family is associated with the expansion of the number of members from an ancient or early eukaryotic ancestral gene. The possible congruent evolutionary tract of modules, as observed in Figure 4.2, suggests that the ancestral PPN GHF5 gene probably contained a CBM. Bacterial endoglucanases show significant sequence homology in both domains and if the HGT hypothesis is true then the analysed GHF5 endoglucanases likely evolved from a bacterial ortholog containing both domains. Based on the recent indications that introns emerged during the earliest phases of eukaryotic evolution (Sverdlov et al., 2007) our model starts from an intronless ancestral PPN GHF5 endoglucanase gene.

The exon/intron structure of *Hg-eng6* and *Mi-eng2* and their sequence divergence from the other genes indicate that they have probably originated from a different type of GHF5 gene (type A). This implies that the ancestral PPN GHF5 endoglucanase must have been duplicated already early in the Tylenchomorpha lineage. This early duplication event was also suggested by Ledger et al. (2006). During or shortly after this duplication one of the copies (type A) must have lost its linker and CBM. All other genes are proposed to have evolved from the second copy (type B) containing linker and CBM. Whether intron 23 has been gained before, during or shortly after the duplication cannot be evaluated based on the available sequence data. In the ancestral nematode of the lineage leading to the cyst nematodes and *Radopholus similis*, a second gene duplication took place.

At intra-species level, the conserved gene structure of certain paralogs and the observation of some sequence conservation in their introns indicate a history of recent duplication events in the present-day genomes of plant-parasitic nematodes. In some cases, intra-species duplicates lost their CBM and occasionally also their linker region. Generally, the observed variation in exon/intron structures between homologous and paralogous genes implies that introns have been gained and lost along all major lineages of the tree and at different time points during evolution. Concordant with the fact that the conservation of protosplice sites strongly supports the recent origin of an intron (Yoshihama et al., 2006), recently gained introns in our model (intron 5, 6, 12, 15, 19 and 22) show a stronger conservation of the protosplice site (MAG|R) than older introns.

Our model implies some special evolutionary gene structure processes for which little information is available and that are therefore difficult to prove.

The phylogenetic hallmark of intron sliding is a distribution in which one intron is nested within the distribution of another (Stoltzfus et al., 1997). We propose the occurrence of intron

sliding for two introns in our model, namely sliding of intron 17a to intron 17b and intron 2 to intron 3, and this hypothesis is based on their close vicinity (1 nucleotide between 17a and 17b and 3 nucleotides between 2 and 3) and the apparent nested distribution. While the nested clustering theory of Stoltzfus et al. (1997) pertains for these introns and they are located in very close vicinity, the fast evolution of intron sequences precludes the ability of finding significant sequence similarity at intron level. In literature, the few established reports of intron sliding all involve very recent events (Kloek et al., 1993; Schafer et al., 1999; Sato et al., 1999; Rogozin et al., 2000). Alternative mechanisms like separate gain and loss events might equally likely have been involved in the observed pattern of intron distribution.

The genes with no or only one intron probably have lost their introns by reverse transcription and subsequent homologous reinsertion of the synthesised cDNA, affecting either the whole gene or only a part of it. *Hg-eng5* is an example of a gene where this process might have happened. We propose a similar mechanism for *Rs-eng1A* and *Rs-eng2*, involving a reverse-transcribed mRNA containing only a catalytic domain, but in this case the recombination must have been followed by a duplication event. The loss of introns through reverse transcription of mRNAs followed by the recombination of the synthesised cDNA in the genome (Liaud et al., 1992; Charlesworth et al., 1998; Sverdlov et al., 2004) is hard to prove and an alternative gene duplication event cannot be ruled out because the genomes of plant-parasitic nematodes contain more GHF5 endoglucanase genes than currently available in databases, for example in the *Meloidogyne incognita* genome, 21 endoglucanases have been found (Abad et al., 2008). If a gene duplication would have given rise to these genes, then a closely related paralog, generated in the same duplication event, should be found in the genome (Boudet et al., 2001). The impossibility to establish a clear relation between the intronless gene *Hg-eng5* and any other gene from the same species suggests that reverse transcription followed by a homologous reinsertion is most probable. For *Rs-eng1A* and *Rs-eng2*, the reverse transcription event is supposed to have taken place earlier in evolution, involving only the 5' end, and should have been followed by a duplication event, the loss of linker and CBM and a gain of intron 21 in one of the paralogs.

Discussion

Glycosyl hydrolase family 5 (GHF5) endo-1,4-β-glucanase genes have been identified in various plant-parasitic nematodes. These genes have different structures: all have a signal peptide for secretion and a catalytic domain, some have an additional linker and carbohydrate binding module (CBM) and others only have an additional linker but no CBM. This modular gene structure could have arisen by exon or domain shuffling, a process by which a chimeric protein is created by domains or segments of two different genes.

Present-day PPN GHF5 endoglucanase genes exhibit a large diversity of exon/intron structures, but protein sequences in the catalytic domain are relatively well-conserved. Generally, PPN GHF5 endoglucanases contain a rather high number of introns. It is known that paralogs undergo functional diversification, which tends to be accompanied by a weakened selection and the acceleration of sequence evolution (Lynch & Conery, 2000). Babenko et al. (2004) observed two clear trends in the evolution of paralogous gene families: the evolution of their gene structure is notably dynamic, and this evolution involves more intron gains than losses. In addition, Logsdon et al. (1995) noticed that nematodes have a particularly high rate of intron turnover compared to other animals. These tendencies are also recognizable in the PPN GHF5 gene family.

While still a lot of uncertainty exists about the ancient origin of the PPN GHF5 endoglucanases, in this chapter we assume that the full gene cassette (i.e. catalytic domain and CBM) was the ancestral endoglucanase gene, since both domains have similar evolutionary tracts. We discard the possibility of independently acquired domains that occasionally are assembled as a result from domain shuffling, as described for an endoglucanase in the bacterium *Myxococcus xanthus* (Quillet et al., 1995). Some additional trends in the gene structure data also support the absence of early domain shuffling. Firstly, exon or domain shuffling is more likely to involve large introns as recombination frequency through cross-overs is known to be proportional to DNA length. Although we cannot exclude the possibility that larger introns were present in an ancestral GHF5 PPN endoglucanase, intron sizes in the present-day nematode genes are extremely small (also for introns 20 and 21) and are therefore less likely to be involved in recombination. Secondly, we could not find a significant correlation between specific locations of introns and secondary structure elements, which would imply a role for introns in the evolution of proteins as proposed by Gilbert (1978). Moreover, successful domain shuffling requires that the domains are bordered by introns that are of the same phase, that is, that the domain is symmetrical in accordance with the phase-compatibility rules of exon shuffling (Patthy, 1999), because shuffling of asymmetrical exons/domains will result in a frame-shift. The only possible symmetric, and hence movable, domain would be a 1-1 domain bordered by intron 1 and intron 21. Although 1-1 symmetrical domains are suggested to be frequently associated with domain shuffling events (Kaessmann et al., 2002), the low conservation of intron 1 among the investigated gene structures (it is only present in 2 genes) as well as its well-conserved protosplice site point to a quite recent origin, and therefore make its involvement in a possible domain shuffling event very unlikely. While we assume that exon shuffling did not lie at the origin of the emergence of the PPN GHF5 endoglucanase gene structure, more recent exon/domain shuffling events at lower taxonomic levels cannot be ruled out.

A well-confirmed cluster groups the GHF5 endoglucanases of *Pratylenchus coffeae* and *P. penetrans* with those of the root-knot nematodes (genus *Meloidogyne*). Cyst nematode

(*Heterodera* and *Globodera*) genes are grouped together with some genes from *Radopholus similis*. These observations corroborate recent phylogenetic trees derived from rRNA data (Subbotin et al., 2006; Bert et al., 2008; Holterman et al., 2009) revealing *Radopholus similis* to be more closely related to Heteroderidae than to Pratylenchidae. This distinction was also observed on a morphological level by Luc (1987), who found that the genus *Radopholus* can be distinguished from other genera in Pratylenchidae by its strong secondary sexual dimorphism and the distinctive lip pattern.

Ledger et al. (2006) proposed an evolutionary scheme for the evolution of GHF5 endoglucanases in plant-parasitic nematodes. Their model proposed two ancestral copies of the GHF5 endoglucanase gene: one with and one without CBM, which is in agreement with our data. However, the model of Ledger et al. (2006) does not take duplication events in ancestral taxa (e.g. as we have detected in the ancestor of *Radopholus similis* and the cyst nematodes) into consideration and is therefore confusing in regard of gene and species evolution. Our evolutionary model proposes a first duplication event already early during the evolution of the ancestral GHF5 endoglucanase, leading to a copy without and a copy with the linker and CBM domain, similar to the model of Ledger et al. (2006). Of the first copy, only two descendants with a remarkably similar exon/intron structure have been found up till now (*Hg-eng6* and *Mi-eng2*). The second copy is much more represented among the known GHF5 endoglucanases. Selection might have played a role in the preferential conservation of this type, possibly due to a positive effect of the CBM. After this early duplication, a second duplication event must have taken place in the common ancestor of the cyst nematodes and *R. similis*. These nematodes all contain two divergent types of GHF5 endoglucanases within their genomes. At intra-species level, the presence of some completely and largely conserved structures between paralogs and the possibility of finding some sequence conservation in their introns indicate the presence of relatively recent events of duplication in the present-day genomes of plant-parasitic nematodes.

The lack of support for evolutionary relationships between nematode genes and genes from other major lineages precludes drawing conclusions about possible HGT events between bacterial and eukaryotic genomes which is why in this study the HGT hypothesis has neither been accepted nor rejected. Additional sequence data from other major eukaryotic lineages are awaited before a thorough and sound investigation of this problem can be attained.

Acknowledgments

We would like to thank Dr. Annemie Elsen (K.U. Leuven, Belgium) for kindly providing the initial *Pratylenchus coffeae* culture.

CHAPTER 5

CHARACTERISATION OF A PUTATIVE ENDOXYLANASE IN *RADOPHOLUS SIMILIS*

Adapted from:

Annelies Haegeman, Bartel Vanholme and Godelieve Gheysen. (2009). Characterization of a putative endoxylanase in the migratory plant-parasitic nematode *Radopholus similis*. Molecular Plant Pathology, 10, 389-401.

Abstract

Plant-parasitic nematodes have developed an arsenal of enzymes to degrade the rigid plant cell wall. In this chapter we report the presence of a putative endoxylanase in the migratory endoparasitic nematode *Radopholus similis*. This enzyme is thought to facilitate the migration of the nematode through the plant tissue, as it breaks down xylan, the major component of hemicellulose. The corresponding gene (*Rs-xyl1*) was cloned and the sequence revealed three small introns. Interestingly the position of all three introns was conserved in a putative endoxylanase from *Meloidogyne hapla* and one of them in two endoxylanases of *M. incognita*, which suggests a common ancestral gene. Spatial and temporal expression of the *Rs-xyl1* gene was examined by *in situ* hybridisation and semi-quantitative RT-PCR. The putative protein consists of a signal peptide for secretion, a catalytic domain and a carbohydrate binding module (CBM). The catalytic domain showed similarity to both glycosyl hydrolase family 5 (GHF5) and GHF30 enzymes. Using Hidden Markov Model profiles and phylogenetic analysis, we were able to show that Rs-XYL1 and its closest homologs are not members of GHF5 as previously suggested, but rather form a subclass within GHF30. Silencing the putative endoxylanase by dsRNA targeting the CBM region resulted in an average decrease in infection of 60%, indicating that the gene is important for the nematode to complete its lifecycle.

Introduction

The major component of hemicellulose, a complex of polymeric carbohydrates in plant cell walls, is xylan. Next to cellulose, xylan is the second most abundant polysaccharide in nature. It is composed of (1,4)-β-linked xylopyranose units and can have various substituents and variable structures according to the plant species (Collins et al., 2005). Xylan is found at the interface between lignin and cellulose, where it plays a role in fiber cohesion and plant cell wall integrity. In monocotyledons, it is located in the primary cell wall whereas in dicotyledons it is the major constituent of the secondary cell wall. Endo-1,4-β-xylanases (EC 3.2.1.8) depolymerise the nonhydrolysed xylan polymer by the random cleavage of the xylan backbone (Subramaniyan & Prema, 2002). These enzymes are mainly produced by plant pathogenic bacteria and fungi as well as by endosymbionts located in the digestive tracts of various wood-boring insects (Brennan et al., 2004). Glycosyl hydrolases are classified into different families according to their sequence similarity (Henrissat & Bairoch, 1996). Endo-1,4-β-xylanases are mostly classified in glycosyl hydrolase families 10 and 11, but some have also been designated to GHF 5, 7, 8, 16, 26, 43, 51 and 62 (Collins et al., 2005) (Carbohydrate Active Enzymes website http://www.cazy.org) (Coutinho & Henrissat, 1999). To our knowledge only three endoxylanases have been found in animals: one in Gastropoda (*Ampullaria crossean* AAY46801, GHF10), one in Insecta (*Phaedon cochleariae* CAA76932, GHF11) and one in Nematoda (*Meloidogyne incognita* AAF37276, GHF5). The nematode endoxylanase, Mi-XYL1, has similarity to a range of bacterial endoxylanases (Mitreva-Dautova et al., 2006). Most of these have been designated to GHF5, although there is some confusion about this classification since the proteins have similarity to GHF30 enzymes as well. Originally, the first enzyme of this group, was classified into a new family located between families 5 and 30 (Keen et al., 1996). Other authors also suggested to classify them into a new family (Suzuki et al., 1997; Hurlbert & Preston, 2001). However, since the first crystallographic structure of one of these xylanases was unraveled, they have been inconsistently classified as GHF5 (Larson et al., 2003; Mitreva-Dautova et al., 2006; John et al., 2006).

A recent EST study on the migratory nematode *Radopholus similis* revealed some interesting genes, including an EST with homology to an endo-1,4-β-xylanase (Jacob et al., 2008). In this chapter we report the further characterisation of the corresponding gene. It is the first putative endoxylanase in a migratory plant-parasitic nematode, and moreover it is the first characterised animal endoxylanase that includes a putative carbohydrate binding module (CBM).

Materials and methods

Nematode cultures, DNA extraction, RNA extraction and cDNA synthesis

Radopholus similis was cultured on carrot discs as described in chapter 2. Mixed stages were collected by rinsing the petri dishes with distilled water 6-8 weeks after inoculation. *Meloidogyne incognita* was cultured on the roots of pea (*Pisum sativa*). The seeds were sterilised by soaking them for 30 min in sterile distilled water, 5 min in 100% ethanol, 15 min in 5% NaOCl and 1% Tween-20 and afterwards washed 5 times in sterile water. The sterile seeds were germinated on KNOP medium (Sijmons et al., 1991) and grown at 21°C. After 2 to 3 weeks, the root tips were infected with nematodes of the J2 stage and the plants were incubated at 28°C. About 7 weeks after inoculation, galls with egg masses were dissected and allowed to hatch in sieves in sterile water at 28°C.

Genomic DNA was isolated from approximately 5000 individuals as described by Bolla et al. (1988). RNA extraction and first strand cDNA synthesis were done as described in chapter 2.

Activity assay and cloning of *Rs-xyl1*

Approximately 10,000 freshly harvested nematodes were crushed in 0.5 ml TE buffer (10 mM Tris-HCl, 1.0 mM EDTA, pH 8.0) to which 0.5 mM DTT was added. The sample was centrifuged for 5 min at 2000 g in a table-top centrifuge, and 2 µl of the supernatant was loaded on a 1.5% agar plate containing 0.5% xylan from beechwood (Sigma). After overnight incubation at 37°C, the plate was stained for 15 min with 0.1% Congo Red and subsequently washed for 15 min with 1.0 M NaCl. Water rinsed from uninfected carrot discs was used as a negative control.

Primers Rs-xyl-F and Rs-xyl-R (Table 5.1) constructed on EST EY194441 were used in a PCR with conditions as described in chapter 2. Resulting fragments were purified, cloned and sequenced as described in chapter 2. Based on the obtained sequence, primers were developed for upstream and downstream walking using the Genome Walker Universal Kit (Clontech) according to the manufacturer's instructions in two successive PCRs. Primers used for upstream walking were Rs-xyl-up1 and Rs-xyl-up2, those for downstream walking Rs-xyl-down1 and Rs-xyl-down2 (Table 5.1). The fragments of interest were cloned and sequenced. An additional primer set (Rs-xyl-start and Rs-xyl-stop, Table 5.1) was developed to amplify the full length coding sequence of the putative endoxylanase from a cDNA pool. The resulting cDNA fragment was cloned in pGEM-T and sequenced.

Primers Mi-xyl-F and Mi-xyl-R (Table 5.1) were developed based on the endoxylanase coding sequence of *Meloidogyne incognita* (AF224342). These primers were used in a standard PCR reaction on DNA of *Meloidogyne incognita* and resulting fragments were cloned and sequenced as described in chapter 2.

Sequence analysis

The identification of introns, calculation of GC contents, translations and predictions of molecular weight, putative signal peptides and N-glycosylation sites were done as described in chapter 2. Tertiary structure predictions were done in SWISS-MODEL (Guex & Peitsch, 1997) using the first approach mode and the crystal structure of a xylanase of *Erwinia chrysanthemi* (PDB entry 1NOF) (Larson et al., 2003) as template (E-value 2.0e-81). The obtained PDB coordinates were visualised using Deepview spdbv 3.7 (Swiss-PdbViewer) (Guex & Peitsch, 1997) and rendered with POV-Ray version 3.6 (http://www.povray.org/). Due to lack of similarity to 1NOF, the first 12 amino acids of the mature protein were not modelled by the first approach mode, and these were manually fitted into the model. To identify possible domains in the putative protein sequence of the endoxylanase, a domain search was done with InterProScan (Zdobnov & Apweiler, 2001). A blastp search with default settings was performed with the catalytic domain and the putative carbohydrate binding module (CBM) (September 2008). Sequences of the closest homologs were retrieved from GenBank and a protein alignment was constructed with the ClustalW algorithm (Thompson et al., 1994) in BioEdit 7.0.5.3 (Hall, 1999). A local Hidden Markov Model (HMM) search was done with HMMER 2.3.2 (http://hmmer.janelia.org) against the closest homologs of the Rs-XYL1 catalytic domain with profiles from GHF5 and GHF30 downloaded from the PFAM website (Finn et al., 2006). The CAZy website (http://www.cazy.org) was searched to find which carbohydrate binding modules the closest homologs of the Rs-XYL1 CBM were assigned to. A local HMMer search was done against the closest homologs of the CBM with the following CBM profiles downloaded from the PFAM website: CBM4_9 (PF02018), CBM6 (PF03422), CBM13 (PF00652). The HMM profile of CBM35 was constructed locally from raw sequences. To identify homologous sequences in other nematode species, tblastn searches with the catalytic domain and the CBM were performed against all nematode ESTs. The nematode genome of *M. hapla* was searched for putative endoxylanases by tblastn with Rs-XYL1 as query on the website of the genome project (http://www.hapla.org) (Opperman et al., 2008).

A phylogenetic tree was constructed including the 100 best hits from a blastp search with the catalytic domain. Ten protein sequences which belong to the PFAM seed alignment from GHF5 were selected, as diverse as possible (both mannanases and endoglucanases from bacteria and fungi). Three protein sequences belonging to GHF10 were chosen as outgroup. The phylogenetic analysis was done by Bayesian statistics using the WAG model in MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003), with 1,000,000 generations and a sample frequency of 100. The first 250,000 generations were considered as burn-in. The resulting tree was visualised in Treeview 1.6.6 (Page, 1996).

Spatial and temporal expression of *Rs-xyl1*

An *in situ* hybridisation was performed as described in chapter 2. Probes were generated by a linear PCR with a single primer (Rs-xyl-F: sense probe; Rs-xyl-R: antisense probe) in the presence of DIG-labeled oligonucleotides (Roche, Mannheim, Germany). As template for the probe generating PCR, the product of a first PCR reaction was used. This PCR was done on a plasmid containing the corresponding xylanase cDNA fragment with primers Rs-xyl-F and Rs-xyl-R. All PCR reactions were carried out as described in chapter 2.

A semi-quantitative RT-PCR on different life stages was done as described in chapter 2. The primers used were Rs-xyl-F and Rs-xyl-R and as a positive control, actin (EU000540) was amplified using the primers Rs-act-F and Rs-act-R (Table 5.1).

Generation of dsRNA, soaking, infection tests and semi-quantitative RT-PCR

Two different regions of *Rs-xyl1* were selected as target for silencing by RNAi, one in the catalytic domain (286 bp) and one in the putative CBM (242 bp) of *Rs-xyl1*. These regions were amplified by PCR under standard conditions with the T7 promoter sequence incorporated at the 5' end of either the sense or the antisense strand. Primers used for the first region were Rs-xyl-catdomF and Rs-xyl-catdomR and the primers for the second region were Rs-xyl-CBMF and Rs-xyl-CBMR. Double-stranded RNA against green fluorescent protein (*gfp*) was prepared from a *gfp* containing construct using primers GFP-F and GFP-R. PCR products were used as templates for *in vitro* transcription reactions using a Megascript RNAi kit (Ambion, Huntingdon, UK). The dsRNA was quantified spectrophotometrically. A few thousand nematodes of mixed stages were soaked in 50 mM octopamine, 3 mM spermidine, 0.05% gelatin and 0.5 mg/ml dsRNA for 24h at room temperature on a rotator in the dark. As an extra negative control, nematodes were incubated in the same solution but without dsRNA. To check whether *R. similis* is stimulated to ingest molecules in the presence of octopamine, nematodes were soaked under the same conditions with the fluorescent dye fluorescein isothiocyanate (FITC) added. The uptake of FITC was visualised 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 done with filter cubes of the following composition (EX: excitation, DC: dichroic, EM: emission): EX: 465-495; DC: 505; EM: 515-555. Images were acquired with NIS-Elements software version 2.10 (http://www.nis-elements.com/) and analysed with ImageJ (Abramoff et al., 2004). After soaking, nematodes were sterilised for 1h in 0.33% hospital antiseptic concentrate (HAC) on a rotator at room temperature and subsequently washed three times with sterile water. Seedlings from *Medicago truncatula* L. var. Jemalong were grown in six well plates on modified Strullu Romand medium (Elsen et al., 2000). After three weeks, the seedlings were infected with 50 sterilised soaked nematodes per plant, 11 plants for each condition. The plants were grown at 22°C and light/dark cycles of 16/8h. Ten days later, the roots were stained with acid fuchsin

(Byrd et al., 1983), destained using acidified glycerol and the nematodes inside the roots were counted using a dissection microscope. Results were statistically analysed in S-PLUS 7.0 (http://www.insightful.com/). Normality was checked with the Kolmogorov-Smirnov test, homoscedasticity with a modified Levene test and ANOVA analysis was done with the Tukey method. The soaking and infection tests were repeated three times. For semi-quantitative RT-PCR, approximately 5000 soaked nematodes were used. RNA extraction and cDNA synthesis were done as described above. The primers Rs-act-F and Rs-act-R were used to amplify actin, whereas Rs-xyl-F2 and Rs-xyl-R2 were used to amplify *Rs-xyl1*. The amount of cDNA added as template and the number of PCR cycles were optimised for each gene product to detect the exponential phase of the reaction. Resulting products were separated on a 0.5x TAE gel.

Table 5.1: Primers used for cloning, *in situ* hybridisation, RT-PCR and RNAi of *Radopholus similis* and *Meloidogyne incognita* endoxylanases

Primer	**Primer sequence**	**Primer**	**Primer sequence**
Rs-xyl-F	AAGGGCAAGCAGGTGTGGATG	Rs-xyl-R	TGATGCTCAGTTGGAGACGAAG
Rs-xyl-up1	CATCATACTGGTCCAGTCGAAGC	Rs-xyl-up2	TAGTGCTCCGTCATCCACACCTGC
Rs-xyl-down1	AAGAGCGGCGCATTCACCTACTC	Rs-xyl-down2	CTTCGTCTCCAACTGAGCATCA
Rs-xyl-start	ATGTTCGCCCTTCTCGTTCCTG	Rs-xyl-stop	TCAGTTGGAGACGAAGGTGGTGACG
Rs-act-F	GAAAGAGGGCCGGAAGAG	Rs-act-R	AGATCGTCCGCGACATAAAG
Mi-xyl-F	TTTGGTGGTTCTAGTGCTTGG	Mi-xyl-R	TGTTGTTGGATTTTCAGTAGCA
Rs-xyl-catdomF	CCGCCTCGATGAAGTCCAACAAC	Rs-xyl-catdomR	GGGTCAGTGAGTGCCTTGTTG
Rs-xyl-CBMF	CCAACGTGTTTCTGAGCGCCTAC	Rs-xyl-CBMR	TCAGTTGGAGACGAAGGTGGTGAC
GFP-F	ATCCGCCACAACATCGAGG	GFP-R	TTGTACAGCTCGTCCATGC
Rs-xyl-F2	GGTATGCAGTTGGCTCTGGT	Rs-xyl-R2	GACACAACGCTGACAGTTGG

Results

Putative endoxylanases in *Radopholus similis* and *Meloidogyne* species

The xylan activity assay on nematode homogenate from *R. similis* showed a clear halo which was not present in the negative control, indicating that *R. similis* contains active enzymes that are able to break down xylan (Figure 5.1).

Figure 5.1: Xylan activity assay with *Radopholus similis* extract and negative control.

Screening over 7000 ESTs of *R. similis* revealed a tag of 521 bp with homology to an endoxylanase of *Aeromonas punctata* (blastx E-value = 2e-34) (Jacob et al., 2008). Based on this sequence primers were constructed to clone both the genomic (1337 bp) as well as the coding sequence (1203 bp) of the corresponding gene. The gene was named *Rs-xyl1* and was submitted to the GenBank database (EU190885). Comparing the genomic DNA to the cDNA revealed three short introns (43, 44 and 47 bp). The first two introns, respectively in phase 2 and phase 0, have the consensus GT/AG splice site (Blumenthal & Steward, 1997) whereas the third intron (phase 0) has the alternative splice site GC/AG. The coding sequence has an overall GC percentage of 61%, a GC1 and GC2 of each 49% and a GC3 of 85%. The introns have a GC percentage of 60%.

Nematode endoxylanases were previously only described for the sedentary nematode *Meloidogyne incognita* (Mitreva-Dautova et al., 2006) and in the recently published genome of *M. incognita* six endoxylanases were identified (Abad et al., 2008). Since to date only coding sequences are publicly available for these genes and we were interested in the number and specific position of introns, primers were constructed to isolate the corresponding genomic clones. Two genomic fragments were amplified from *M. incognita* of respectively 1503 and 1357 bp and were named *Mi-xyl2* (EU475875) and *Mi-xyl3* (EU475876). Their deduced coding sequences showed minor differences to the original cDNA sequence of *Mi-xyl1* as presented by Mitreva-Dautova et al. (2006) (8 bp in case of *Mi-xyl2* and 15 bp in case of *Mi-xyl3).* These isolated genomic sequences are probably other copies of the gene in the genome, or it also could be allelic variants. Both genomic sequences contain a single intron in phase 0, holding the canonical GT/AG donor/acceptor motif. The intron of *Mi-xyl2* is 642 bp, whereas the intron of *Mi-xyl3* is shorter (496 bp). The introns only show similarity at the extensions: approximately 80 nucleotides at the 5' end of the introns and 120 nucleotides at the 3' end. Interestingly, both introns were located at the same position as the second intron of *Rs-xyl1*.

A tblastn with Rs-XYL1 against the genome of *Meloidogyne hapla* revealed one contig (contig 2188) with significant similarity (E-value = 4e-24). This putative endoxylanase of *M. hapla* includes the putative CBM and contains three predicted introns of respectively 148, 214 and 275 bp, all with a GT/AG splice site. The introns are located at the exact same position as the three introns from *Rs-xyl1*.

Protein properties and structure

The putative protein sequence of 400 amino acids (AA), called Rs-XYL1, contains a putative 21 AA signal peptide for secretion. The predicted molecular mass of the mature endoxylanase is 40 kDa. NetNGlyc predictions revealed one putative N-glycosylation site at position 270 of the mature protein.

A search for protein domains resulted in hits for a glucosylceramidase of glycoside hydrolase family 30 (GHF30; EC 3.2.1.45) (E-value=3.3e-07) and, although with low support, a carbohydrate binding module of family II (E-value=3.7e+02). Alternatively, the CATH database (Pearl et al., 2003) was searched and high support was found for a glycoside hydrolase domain (E-value=1.4e-68). Thus, the protein consists of an N-terminal signal peptide for secretion of 21 AA, and two putative modules: a glycoside hydrolase catalytic domain of 274 AA and a C-terminal carbohydrate binding module (CBM) of 90 AA. This putative endoxylanase is the first animal xylanase found with a putative CBM.

The putative endoxylanase found in the *M. hapla* genome has a predicted signal peptide of 24 AA, a catalytic domain of 272 AA and also a putative CBM of 91 AA. One possible N-glycosylation site is present. The protein sequence deduced from *Mi-xyl2* of *M. incognita* is identical to Mi-XYL1, whereas Mi-XYL3 differs in three amino acids from Mi-XYL1. An alignment of Rs-XYL1 with the endoxylanase sequences of *M. incognita* and *M. hapla* is shown in Figure 5.2.

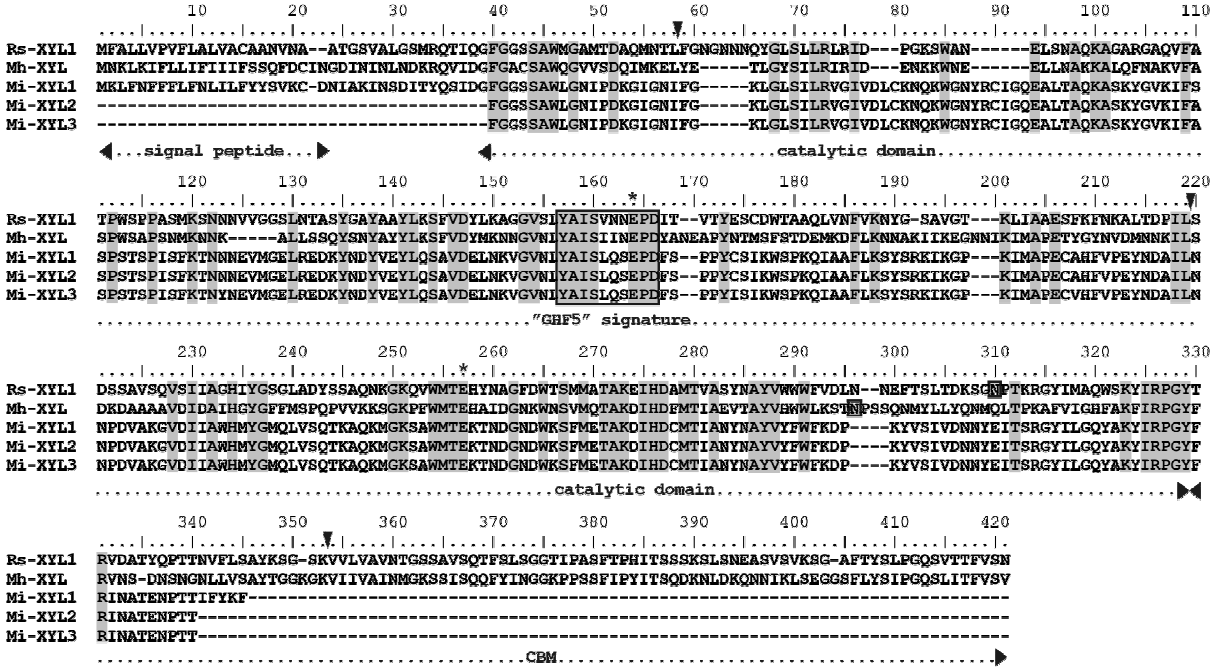

Figure 5.2: Alignment of Rs-XYL1 with Mi-XYL1, the isolated sequences Mi-XYL2 and Mi-XYL3 and the putative Mh-XYL. The location of the signal peptide, catalytic domain, the putative GHF5 signature and the putative carbohydrate binding module (CBM) are indicated. Stars indicate the two catalytic residues. The three arrows indicate the positions of the introns in Rs-XYL1 and Mh-XYL. Possible N-glycosylation sites are boxed. The threshold for shading is 100%.

A 3D model was created using a xylanase from *Erwinia chrysanthemi* as template (PDB identifier 1NOF) (Figure 5.3). Since the proteins are 45% identical and 75% similar, the simplest homology modelling scenario could be used.

The α- and β-chains of the catalytic domain fold into an $(\alpha/\beta)_8$ barrel structure (indicated in green in Figure 5.3) whereas eight consecutive β-strands of the CBM form a β-sandwich (indicated in yellow in Figure 5.3). One β-strand at the N-terminal end of the mature protein (13 amino acids) contributes to the C-terminal β-sandwich (indicated in purple in Figure 5.3).

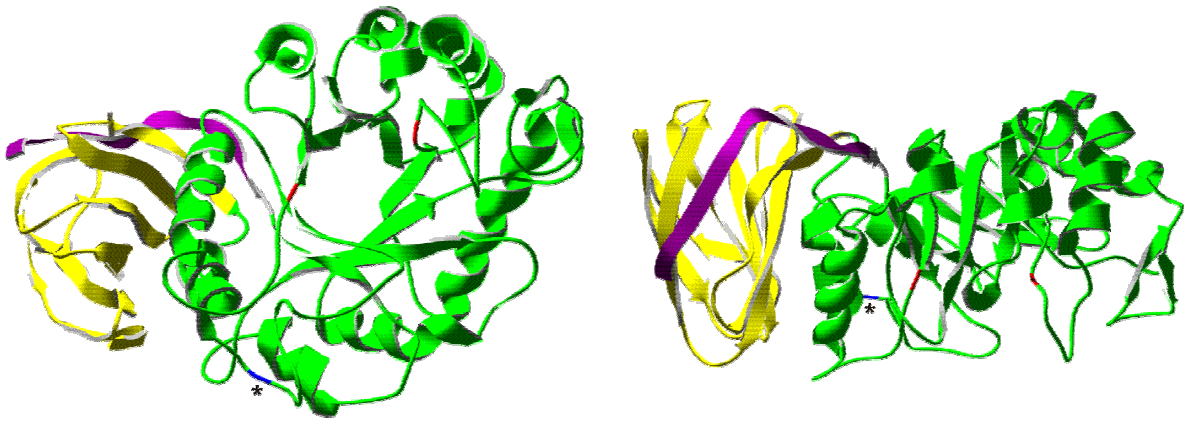

Figure 5.3: 3D model of the putative endoxylanase Rs-XYL1 from *Radopholus similis*, based on the 3D model of the *Erwinia chrysanthemi* endoxylanase 1NOF (left: front view: right: side view). The catalytic domain is indicated in green and the carbohydrate binding module (CBM) in yellow. The β-strand at the N-terminal end of the protein which contributes to the CBM is in purple. The catalytic residues (Glu153 and Glu240) are in red and the possible N-glycosylation site (Asn291) is indicated in blue and with a star.

Database searches and phylogenetic analysis

A blastp search was performed using the catalytic domain as query. There were two groups of blast results. The first group included the top hits (E-value between 1e-73 and 1e-22) and consists of 24 bacterial sequences and the nematode endoxylanases. Both the literature as well as the CAZy website (http://www.cazy.org) classifies most of these proteins as GHF5 enzymes although they all show higher similarity to GHF30 enzymes according to the conserved domain database (E-value<1e-3). Similarly, HMMer searches using the GHF30 profile classified all proteins (including Rs-XYL1) as GHF30, except for Mi-XYL1. No significant matches were found when using the GHF5 profile. These observations question the classification of the proteins to the GHF5 subclass. The second group of blast hits, which show less homology to Rs-XYL1 (E-value between 1e-14 and 1e-4) appeared to be GHF30 enzymes from both bacteria and fungi, mostly named glycosyl hydrolase or glucosylceramidase. It should however be noted that only two of these hits are classified as GHF30 according to the CAZy database. The others have not been classified at all, although they all have a significant match to the GHF30 profile.

A phylogenetic tree including the two groups of blast hits and some GHF5 enzymes revealed a clear separation of two clusters, highly supported with a posterior probability value of 1.00 (Figure 5.4). The first cluster includes the “true” GHF5 enzymes, whereas the second cluster includes all proteins which showed homology to the catalytic domain of Rs-XYL1. From now on, we will call the first cluster “GHF5 cluster” and the second cluster “GHF30 cluster”. Within the GHF30 cluster, a first group containing fungal sequences and one bacterial sequence is separated from the other proteins with high support (posterior probability of 1.00). The rest of the GHF30 cluster is not well resolved (posterior probability values less than 0.60). However, Rs-XYL1 and all the blast top hits, with very low E-values, form a monophyletic group. This group is not well resolved from the other blast hits with higher E-values, as the posterior probability is only 0.56, which is mainly due to the protein ZP_01576533. Discarding this protein from the dataset results in a monophyletic cluster of all other top blast hits with high support (posterior probability of 1.00). Despite the fact that more basal branches within the GHF30 clusters are not well resolved, the GHF30 cluster itself does form a monophyletic group with high support.

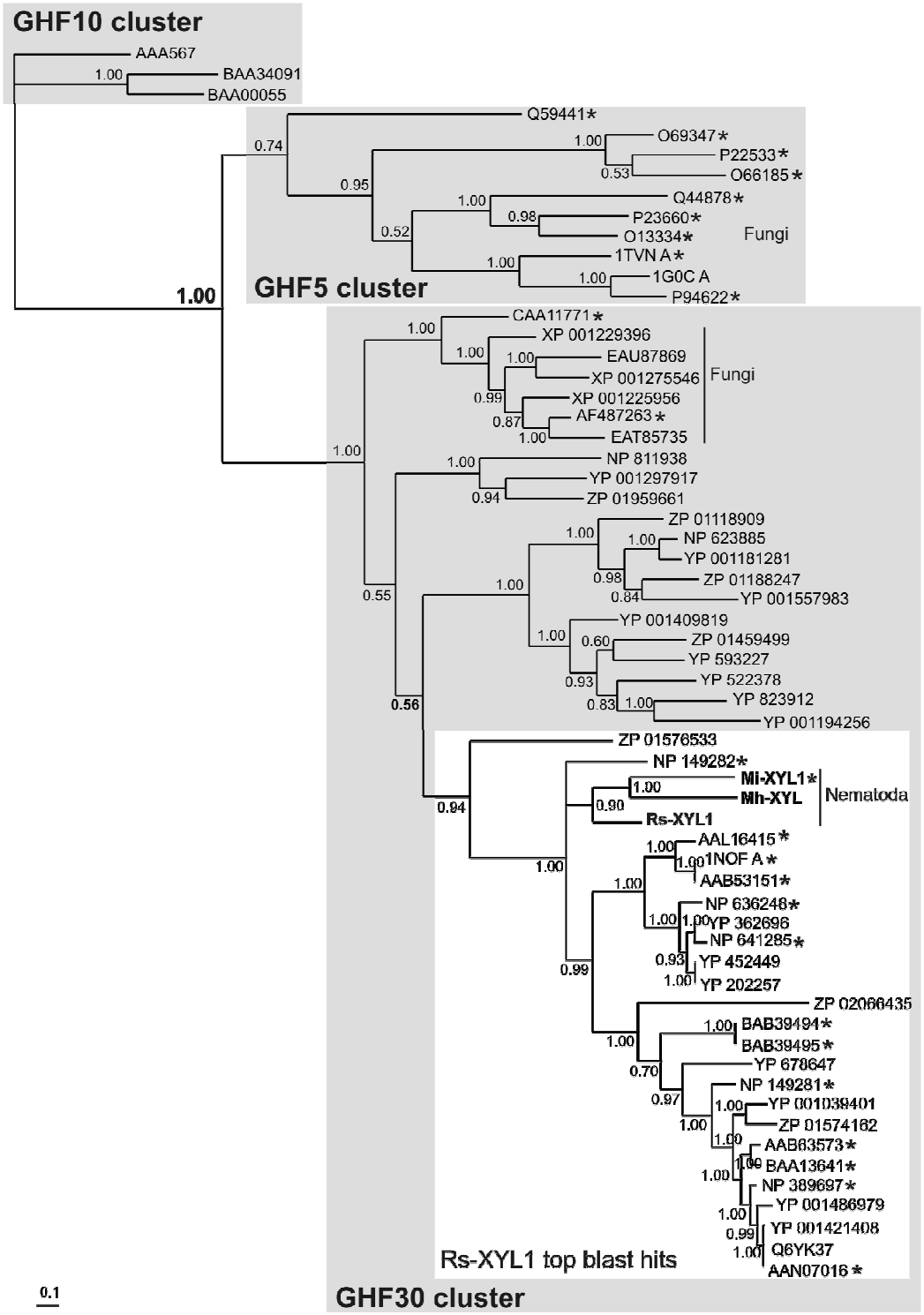

Figure 5.4: Phylogenetic tree including the closest homologs of the catalytic domain of Rs-XYL1. Ten GHF5 enzymes were included and three GHF10 enzymes were chosen as outgroup. The tree was deduced by Bayesian analysis with 1,000,000 generations with a sample frequency of 100. Posterior probabilities are shown on the branches. The GHF5 cluster includes true GHF5 enzymes selected from the PFAM seed alignment for GHF5. The lower cluster, indicated as Rs-XYL1 top blast hits, includes all blast hits from a blastp search with the catalytic domain of Rs-XYL1 as query with E-values between 1e-73 and 1e-22. The other sequences within the GHF30 cluster include blastp hits with E-values between 1e-14 and 1e-4. Fungal and nematode sequences are indicated, all other sequences are bacterial. Stars indicate enzymes that are classified as GHF5 according to the CAZy classification.

A blastp search with the putative carbohydrate binding module resulted in fewer hits, all from bacteria (Table 5.2). An alignment is shown in Figure 5.5. Only the first two hits (both xylanases from *Aeromonas punctata)* have a significant match in the conserved domain database (CDD) and are classified as CBM6 (E-values=1e-15 and 7e-10). The other blast hits have no matching domains in CDD, although some are classified in different CBM groups by the CAZy classification (Table 5.2). A HMMer search was performed with profiles for different CBMs against the different proteins retained by the blast search. HMM profiles for which a significant hit was found (positive bit score) are indicated in Table 5.2. The closest homologs, from *A. punctata*, have been assigned to CBM35 according to CAZy. Nevertheless, some homologs belong to CBM13 and CBM22 whereas others have not been assigned to a CBM family at all. A HMMer search with different CBM profiles resulted in even more confusion, as some CBMs showed positive bit scores to different CBM families, for example CBM35 members were also recognised as CBM6 (Table 5.2). None of the HMM profiles matched the putative CBM of Rs-XYL1.

Table 5.2: Closest homologs to the Rs-XYL1 carbohydrate binding module (CBM) as identified by blastp (E-value<1e-4). The species name, accession number and description of the protein are given. The CAZy classification (http://www.cazy.org) and HMM profiles that resulted in a significant hit (positive bit score) are indicated.

Species	ACC	Description	blastp E-value	CAZy	HMM search
Radopholus similis	EU190885	endo-1,4-β-xylanase	-	-	-
Aeromonas punctata	AAB63573	xylanase D	2.00E-14	CBM35	CBM6, CBM35
Aeromonas punctata	BAA13641	endo-xylanase	9.00E-14	CBM35	CBM6, CBM35
Clostridium cellulolyticum	ZP_01574162	Carbohydrate binding family 6	2.00E-12	-	-
Bacillus subtilis	NP_389697	hypothetical protein	4.00E-11	-	-
Bacillus subtilis	Q6YK37	Glucuronoxylanase xynC preprecursor	5.00E-11	-	-
Bacillus amyloliquefaciens	YP_001421408	YnfF	7.00E-11	-	-
Clostridium acetobutylicum	NP_149281	Possible xylan degradation enzyme	3.00E-09	CBM13	CBM13
Clostridium acetobutylicum	NP_149282	Possible xylan degradation enzyme	3.00E-08	CBM13	CBM13
Clostridium cellulolyticum	ZP_01576533	cellulosome enzyme, dockerin type I	8.00E-08	-	CBM6, CBM35
Ruminococcus albus	BAB39494	xynC	8.00E-08	CBM22	CBM4(9)
Ruminococcus albus	BAB39495	xylanase C	3.00E-07	CBM22	CBM4(9)
Clostridium thermocellum	YP_001039401	Carbohydrate binding family 6	5.00E-07	-	CBM6
Cytophaga hutchinsonii	YP_678647	CHU large protein, candidate xylanase	7.00E-07	-	-
Bacillus subtilis	AAN07016	YnfF	2.00E-05	-	-

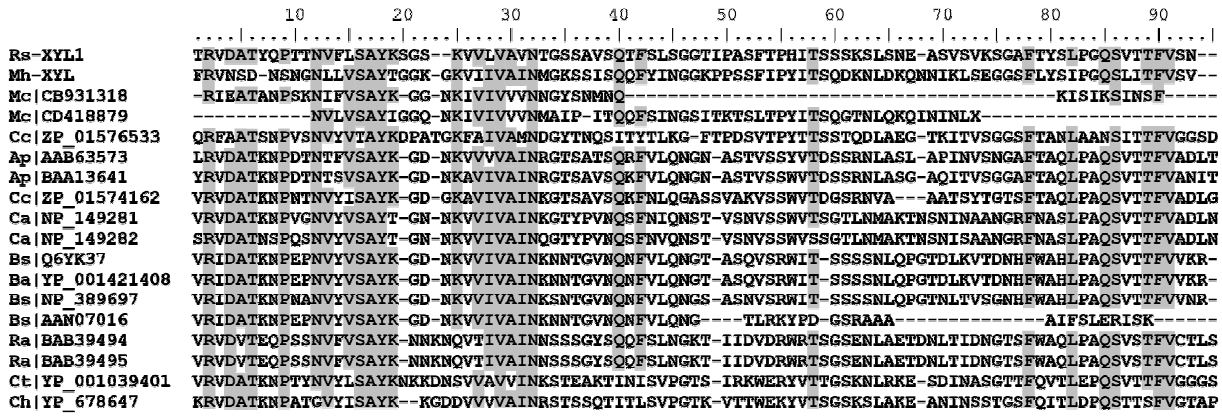

Figure 5.5: Alignment of the carbohydrate binding module of Rs-XYL1 with its closest homologs (blastp hits with E-value<1e-4), Mh-XYL and two translated ESTs from *Meloidogyne chitwoodi*. Cc: *Clostridium cellulolyticum*; Ap: *Aeromonas punctata*; Ca: *Clostridium acetobutylicum*; Bs: *Bacillus subtilis*; Ba: *Bacillus amyloliquefaciens*; Ra*: Ruminococcus albus*; Ct: *Clostridium thermocellum*; Ch: *Cytophaga hutchinsonii*; Mc: *Meloidogyne chitwoodi*; Mh: *Meloidogyne hapla*. The threshold for shading is 75%.

All nematode ESTs were searched for homologs to the endoxylanase domains. The top hits (E-value<1e-20) of the tblastn search with the catalytic domain were all ESTs derived from the genus *Meloidogyne*; being *M. chitwoodi* (CB931318, CD418879), *M. arenaria* (CF357210, CF357155) and *M. javanica* (CF350477, CF350376). Other hits (E-value<1e-4) were from *Globodera pallida* (BM415401, CV578872) and from *Caenorhabditis elegans* (AU208008, BJ762920, CB388002). The latter show less similarity and are most likely not endoxylanases but other glycosyl hydrolases and glucosylceramidases, probably involved in glycan degradation and sphingolipid metabolism (EC3.2.1.45, BRENDA database) (Barthelmes et al., 2007). A tblastn search with the CBM resulted in only two hits (E-value<1e-3), both from *M. chitwoodi* (CB931318, CD418879) (Figure 5.5).

Temporal and spatial expression of *Rs-xyl1*

The expression level of *Rs-xyl1* was compared for different life stages by semi-quantitative RT-PCR (Figure 5.6). In developing embryos and juveniles, no expression of *Rs-xyl1* was detected whereas males and females revealed approximately similar expression levels. To localise the transcripts, a whole mount *in situ* hybridisation was carried out (Figure 5.7). The antisense probe showed clear staining in the gland cell area. The sense probe used as negative control showed no staining.

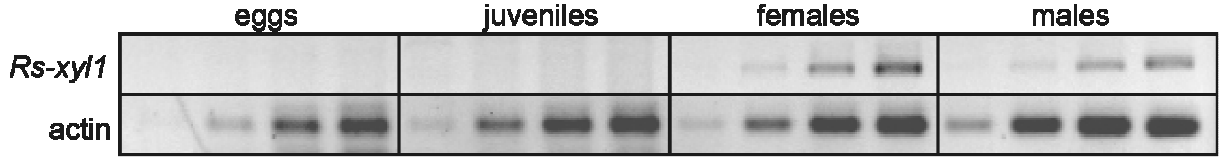

Figure 5.6: Expression of *Rs-xyl1* and actin in developing embryos (27, 30, 33 and 36 cycles), juveniles (30, 33, 36 and 39 cycles), females (21, 24, 27 and 30 cycles) and males (24, 27, 30 and 33 cycles) of *R. similis*.

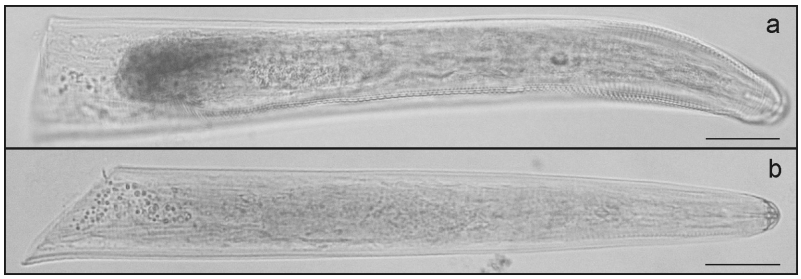

Figure 5.7: Whole mount *in situ* hybridisation on *Radopholus similis*. (a) *Rs-xyl1* antisense probe; (b) *Rs-xyl1* sense probe. Scale bar: 20 µm.

Silencing of *Rs-xyl1* by RNAi and infection tests

Batches of freshly harvested nematodes were soaked in dsRNA against green fluorescent protein (*gfp*), water or dsRNA against *Rs-xyl1* (either targeting the catalytic domain or the putative CBM). The uptake of FITC in the presence of octopamin was successful, suggesting a similar dsRNA uptake (Figure 5.8). After soaking, the expression level of *Rs-xyl1* was checked by a semi-quantitative RT-PCR. Actin was amplified to approximately the same level in all samples. In the infection experiments, nematodes in *Medicago* roots were stained with fuchsin and counted (Figure 5.8).

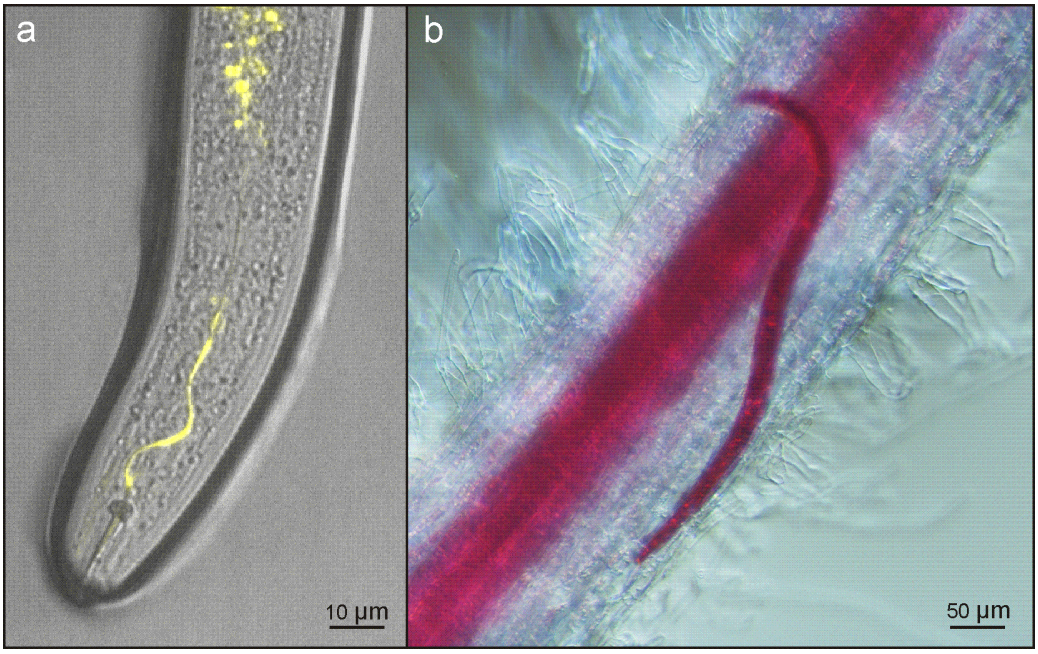

Figure 5.8: (a) Uptake of FITC by *Radopholus similis*: overlay of transmitted light image and fluorescence image; (b) Fuchsin stained *Medicago truncatula* root infected by *R. similis*. After destaining, the root's vascular system and the nematode stay pink. Scales are indicated.

In one of the three experiments, no differences in expression level of *Rs-xyl1* could be detected between the different treatments. The corresponding infection experiment showed no significant differences in infection of the soaked nematodes compared to the controls.

In the other two experiments, *Rs-xyl1* had a lower expression level in the nematodes soaked in dsRNA targeted to the CBM as compared to the controls (dsRNA against *gfp* and water). No silencing of *Rs-xyl1* could be detected in nematodes soaked in dsRNA targeted to the catalytic domain. Both corresponding infection experiments showed a significant ($p<0.01$) decrease in infection (53% and 66%) of nematodes soaked in dsRNA targeted against the CBM of *Rs-xyl1* compared to nematodes soaked in dsRNA against *gfp* or water. Interestingly, nematodes soaked in dsRNA targeting the catalytic domain of *Rs-xyl1*, appeared to show a slight decrease in infection, but this was not significantly different from the controls. The data of one of these experiments is graphically represented in Figure 5.9.

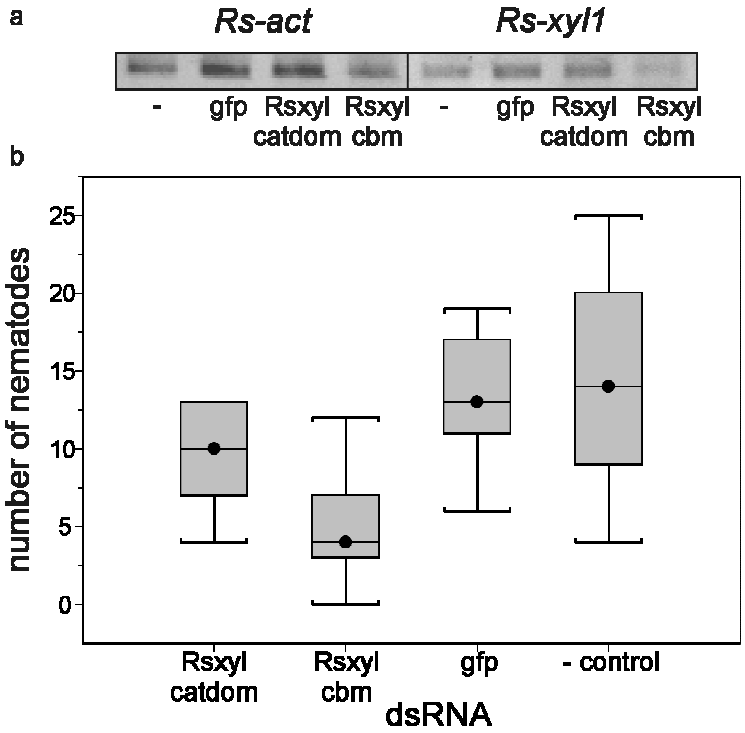

Figure 5.9: Silencing of *Rs-xyl1* (a) Semi-quantitative RT-PCR with primers against actin (30 cycles) and *Rs-xyl1* (36 cycles) on RNA extracted from nematodes soaked in dsRNA against *Rs-xyl1* located in the catalytic domain (Rsxyl-catdom), dsRNA against *Rs-xyl1* located in the putative CBM (Rsxyl-cbm), dsRNA against *gfp* and water (- control) (b) Box-plot of the number of *Radopholus similis* nematodes in the root tissue per plant ten days after inoculation with soaked nematodes. Eleven plants were used for each treatment. The length of the vertical bar denotes the interquartile range, the middle bar represents the median and the upper and lower horizontal bars denote the upper and lower quartiles respectively.

Discussion

During the last decade, different endogenous cell wall degrading enzymes have been identified in plant-parasitic nematodes. These enzymes facilitate the penetration and migration of the nematode through the rigid plant tissue by softening the plant cell walls. Although xylan is an important component of this cell wall, endoxylanases were only described for the sedentary nematode *Meloidogyne* species (Mitreva-Dautova et al., 2006).

In this study we report the finding and characterisation of the first putative endoxylanase (*Rs-xyl1*) of a migratory nematode, *Radopholus similis*. The overall GC content of the coding sequence (61%) is higher than the average value (54%) (Jacob et al., 2008) and is due to the high GC3 (85% instead of 65%). The gene contains three short introns (less than 50 bp). Interestingly, a putative endoxylanase found in the genome sequence of *Meloidogyne hapla* (contig 2188, http://www.hapla.org/), revealed three predicted introns at the same position of the introns of *Rs-xyl1*, and two isolated genomic endoxylanase sequences of the root-knot nematode *M. incognita* had a single intron at the same position of the second intron of *Rs-xyl1*. This indicates that these introns were probably already present in the common ancestor endoxylanase gene of these species. The phylogenetic analysis confirmed the close phylogenetic relationship of the nematode endoxylanases. The isolated genomic sequences of *M. incognita* endoxylanases (*Mi-xyl2*, *Mi-xyl3*) are very similar to the previously described endoxylanase *Mi-xyl1* (Mitreva-Dautova et al., 2006). If these genes are not allelic variants, they probably originated from the same ancestral gene through duplications.

The main difference between the putative endoxylanase of *R. similis* and the previously identified endoxylanase of *M. incognita* is the presence of a putative carbohydrate binding module (CBM) in Rs-XYL1 (Figure 5.2). This putative CBM was also present in the genomic sequence found in the *M. hapla* genome, and blast searches against the *M. incognita* genome revealed that it is probably also associated with the six identified endoxylanases in the genome (Abad et al., 2008). 3D modelling of the protein revealed that the CBM consists of eight β-strands and that the first β-strand at the N-terminal end of the mature protein contributes to the β-sandwich, as is the case for the *Erwinia chrysanthemi* xylanase 1NOF. The presence of a putative CBM in Rs-XYL1 is not surprising, because most bacterial endoxylanases are also associated with a CBM. Although difficult to prove, it is a possibility that nematode endoxylanases have originated through the horizontal gene transfer of a bacterial endoxylanase as was also suggested for nematode endoglucanases (Keen & Roberts, 1998). In this view, it is more likely to find an endoxylanase with a CBM in nematodes. The putative endoxylanase found in the genome of *M. hapla* also includes a putative CBM. A search through the nematode ESTs with the putative CBM of Rs-XYL1 resulted in two hits from *M. chitwoodi*. These tags were also retained by a blastx search using the catalytic domain as query. Therefore it seems that other nematode endoxylanases possibly also have a

CBM. Endoxylanases without a CBM, such as Mi-XYL1, could have lost this domain during evolution.

The amino acid sequence of the catalytic domain showed significant similarity to the *M. incognita* endoxylanase Mi-XYL1 as well as to different bacterial endoxylanases of GHF5 and of GHF30. Since the closest homologs of the protein have been assigned to GHF5, it is tempting to classify Rs-XYL1 in the same subclass. However, it should be noted that the GHF5 signature, which consists of a consensus sequence motif of 10 amino acids, shows three mismatches both in Rs-XYL1 and in Mi-XYL1, and two mismatches in Mh-XYL and most bacterial endoxylanases that appeared as best hit in our blast search, questioning their classification as GHF5 proteins. Moreover none of these proteins show significant similarity to the HMM GHF5 profile whereas all proteins (except for Mi-XYL1) have a significant match to the GHF30 profile. The fact that Mi-XYL1 is not recognised by the HMM profile of GHF30 is probably due to the limited seed alignment from only 3 sequences on which the profile has been based. Keen et al. (1996) suggested classifying this group of endoxylanases in an intermediate class between GHF5 and GHF30. Unfortunately this idea never caught on, and the enzymes have later been assigned to GHF5, resulting in an inconsistent classification. In the phylogenetic analysis of this study, we were able to show that Rs-XYL1 and related endoxylanases form a monophyletic cluster with other GHF30 enzymes. We can conclude that the endoxylanase homologs are a subclass of GHF30 rather than GHF5 and that previous classification of some of these enzymes into GHF5 is incorrect.

In analogy to the difficulties to classify the catalytic domain, it is not clear which family the CBM belongs to. Similarity was found to proteins belonging to different CBM families, and a lot of homologs are not properly classified themselves. This implies that the CBM classification needs to be revised for some families.

Expression of *Rs-xyl1* could only be demonstrated in the adult stages of *R. similis* and not in eggs or juveniles. This suggests that the putative endoxylanase only plays a role in later life stages, which is not what we would expect, since both juvenile and adult stages migrate and feed within the root tissue. Although adult males of *R. similis* do not feed, the expression level of *Rs-xyl1* in males is the same as in females. The expression of *Rs-xyl1* is located in the gland cell area of *R. similis*, as shown by *in situ* hybridisation (Figure 5.7). The fact that the putative protein sequence contains a signal peptide for secretion and that the gene is expressed in the gland cell area, suggests that the protein is secreted through the stylet of the nematode. Once released in the plant tissue, the putative endoxylanase can break down xylan present in the plant cell wall. The functional importance of the putative endoxylanase was proven by RNAi and infection experiments. When dsRNA targeted to the putative CBM was applied to nematodes, partial silencing of *Rs-xyl1* occurred, while no silencing could be observed when the dsRNA was based on the catalytic domain. This difference in silencing effect caused by different target regions in the same gene has also been observed for a pectate lyase of

Heterodera glycines (Sukno et al., 2007). Moreover, in one of the experiments no silencing could be detected at all. Therefore it seems that the silencing effect can vary by the position of the target sequence of the gene, but also by the replica of the experiment itself. When *Rs-xyl1* had a lower expression level after treatment with dsRNA targeted to the CBM, a significant decrease in infection of on average 60% was detected. In the one experiment where there was no silencing effect, no significant decrease in infection was observed. This suggests that *Rs-xyl1* indeed plays an important role in the infection process. We did observe a slight but statistically unsignificant decrease in infection with the nematodes soaked in dsRNA targeted to the catalytic domain, but since we did not see a clear reduction in expression, we can not draw any conclusions for this construct. Nevertheless, this experiment proves that the RNAi technique, so far almost exclusively studied in sedentary nematodes (Lilley et al., 2007), can be successful to reduce infection of migratory nematodes as well. In one other migratory nematode, *Bursaphelenchus xylophilus*, it was recently shown that RNAi could effectively reduce gene expression and cause lethality, but effects on plant parasitism were not studied (Park et al., 2008).

A search for other xylanase homologs in nematode ESTs did not reveal any new candidate xylanase genes. Homologs were found in *M. chitwoodi*, *M. arenaria* and *M. javanica*, as also described by Mitreva-Dautova et al. (2006). Despite the available EST sequence information, endo-1,4-β-xylanases have not been found extensively in contrast to for example endoglucanases. Nevertheless, in this chapter we have shown that a putative endoxylanase of a migratory nematode, *R. similis*, is important in the infection process.

CHAPTER 6

EXPRESSED SEQUENCE TAGS OF *DITYLENCHUS AFRICANUS*: THE FIRST TRANSCRIPTOME ANALYSIS OF AN ANGUINID NEMATODE

Adapted from:

Annelies Haegeman; Joachim Jacob; Bartel Vanholme; Tina Kyndt; Makedonka Mitreva & Godelieve Gheysen. (2009). Expressed sequence tags of the peanut pod nematode *Ditylenchus africanus*: the first transcriptome analysis of an Anguinid nematode. Molecular and Biochemical Parasitology, doi: 10.1016/j.molbiopara.2009.04.004.

Abstract

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

Introduction

Expressed sequence tag (EST) analysis is a relatively cheap and rapid method to obtain a first molecular impression of a species. The technique consists of a random selection of clones from a cDNA library and sequencing of their inserts. Although it is mainly used for gene discovery (McCarter et al., 2003), it can also be used for other goals, e.g. estimation of gene expression level (Munoz et al., 2004), detection of single nucleotide polymorphisms (Picoult-Newberg et al., 1999), or improving genome annotation (Bailey et al., 1998). In nematology, the technique is widely used and to date over one million ESTs from over 60 species are available (dbEST, NCBI GenBank). Some of these ESTs are derived from cDNA libraries generated from specific life stages or from specific tissues of the nematode. In the case of plant-parasitic nematodes, the transcriptional activity of the pharyngeal glands is of particular interest since gland proteins are injected in the plant tissue during the nematode-host interaction. EST analyses have led to the discovery of many of these parasitism genes that code for hydrolytic enzymes, such as pectate lyase (de Boer et al., 2002) and xylanase (Mitreva-Dautova et al., 2006). They are involved in maceration of the plant cell walls during migration of the nematode in the plant tissue.

Ditylenchus africanus is an interesting nematode to subject to EST analysis for several reasons. First, almost no molecular knowledge is available for this species or any other member of the Anguinidae. Second, it has a different taxonomic classification in contrast to other EST studies examining plant-parasitic nematodes, which focused mainly on species from the superfamily Tylenchoidea (grouping cyst and root-knot nematodes as well as migratory nematodes such as *Radopholus* and *Pratylenchus*). Third, it is a plant-parasitic nematode facultatively feeding on fungi. Only one nematode with a similar feeding habit has been investigated using a comparable approach (*Bursaphelenchus xylophilus*; superfamily Aphelenchoidea) (Kikuchi et al., 2007). Finally, it adds to the EST dataset of migratory nematodes, therefore empowering comparative studies. The focus for small-scale EST projects is gradually shifting from sedentary nematodes towards migratory nematodes. In the last few years, EST data have become available for several migratory nematodes such as *Pratylenchus penetrans* (Mitreva et al., 2004), *Radopholus similis* (Jacob et al., 2008), *Bursaphelenchus xylophilus* (Kikuchi et al., 2007) and *Xiphinema index* (Furlanetto et al., 2005).

Materials and methods

Nematode culture, cDNA library construction and EST generation

The culture and RNA extraction of *D. africanus* was done as described in chapter 3. A cDNA library was constructed using the SMART cDNA Library Construction Kit (Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions, starting from 1 µg RNA. The resulting *D. africanus* mixed stage library contained over 800,000 primary transformants. Random colonies were sequenced using the M13 forward or reverse primer at the Genome Center (Washington University, St.-Louis, MO, USA). Resulting sequences were submitted to the EST division of GenBank (dbEST).

Cleaning and clustering of the EST sequences

The sequences were cleaned using Seqclean (www.tigr.org) with a locally downloaded vector database and default parameter settings. The cleaned dataset was clustered using the TIGR Gene Indices Clustering Tool (TGICL) (Pertea et al., 2003) and sequences were assembled by CAP3 (Huang & Madan, 1999) using default settings. The fragmentation (i.e. the percentage of unigenes which are redundant) was estimated with ESTstat (Wang et al., 2005). The obtained unigenes served as a basis for following analyses.

Homology searches

Basic Local Alignment Search Tool (blast) analyses (Altschul et al., 1990) were done both locally and by using Netblast. Blastx and blastn searches were conducted with all unigenes against the NCBI Protein and Nucleotide database. Additionally, blastn and tblastx searches were done against the genomes of *M. incognita* (http://meloidogyne.toulouse.inra.fr/) (Abad et al., 2008) and *M. hapla* (http://www.hapla.org/) (Opperman et al., 2008). Since most data available for nematodes are in the EST database, a tblastx search against all nematode ESTs was done. In-house perl scripts parsed the resulting hits for species names and unigenes were subsequently classified into different categories (nematodes, invertebrates, plants, animals, fungi, prokaryotes) according to the species names derived from the blast hits. In parallel, a blastx search was conducted against proteins of the model organism *Caenorhabditis elegans*. Resulting top hits were searched for RNAi phenotypes using Wormbase (Rogers et al., 2008). For all *C. elegans* homologs with an RNAi lethal phenotype, Gene Ontology (GO) terms were retrieved and visualised with WEGO (Ye et al., 2006). Blastn against all mitochondrial nematode genes revealed putative mitochondrial unigenes.

To annotate gene ontology (GO) terms, BLAST2GO (Conesa et al., 2005) was used on all unigenes using default parameters. Blastx value cut-off was chosen at E-value<1e-4. GOSlim view was used and GO graphs were generated with a node scoring filter of 25 for "molecular function", 50 for "biological process" and 12 for "cellular component".

Translation into putative proteins

All unigenes were translated with OrfPredictor (Min et al., 2005). The blastx output was used to select the correct reading frame for translation. The minimum amino acid number for predicted protein sequences was set to 40. Nucleotide sequences were trimmed to their coding parts and the overall GC content as well as the average GC content of the first (GC1), second (GC2) and third nucleotide (GC3) of the codons was calculated. Unigenes without predicted open reading frame were considered to be non-coding, and for these sequences only the GC content was calculated. Signal peptides for secretion in the predicted proteins were predicted by SignalP 3.0 (Bendtsen et al., 2004). A signal peptide was only assigned to a sequence when both the Hidden Markov Model (HMM) and the neural network predicted its presence. Moreover, a transmembrane domain search (http://protfun.net/services/TMHMM/) on the mature putative proteins revealed whether the protein was retained in the membrane.

Dot blot analysis

Forward (F) and reverse (R) primers were developed to amplify fragments of CL1, CL7, CL270, CL371, CL406, FE921742 and FE922861 (Table 6.1) by polymerase chain reaction (PCR). The reaction mixture contained 300 ng purified plasmids of the cDNA library as template, and reaction conditions were as described in chapter 2 and resulting fragments were cloned and sequenced. Probes were generated by a PCR with purified plasmid with the correct insert as template. Resulting PCR products were radioactively labeled with [α-^{32}P]-dCTP with the DecaLabel DNA Labeling Kit (Fermentas) according to the manufacturer's instructions. As a positive control, 100 ng of unlabeled PCR product was spotted on a Hybond N+ membrane (GE Healthcare, Uppsala, Sweden), together with 1.5 µg carrot RNA as a negative control and 1.5 µg *D. africanus* mixed population RNA, both extracted with TRIzol (Invitrogen) according to the manufacturer's instructions. The membrane was baked at 80°C for 2h and prehybridised for 1h at 65°C in hybridisation solution (5x SSC, 5x Denhardt's solution, 0.5% SDS, 100 µg/ml salmon sperm DNA). After prehybridisation, the hybridisation buffer was replaced and the corresponding probe was added for hybridisation overnight at 65°C. Washing was done at hybridisation temperature for 5 min in 2x SSC with 0.1% SDS and 15 min in 2x SSC. The blots were exposed to a Fujifilm Imaging Plate for 6h and scanned with a FLA-5100 imaging system (Fujifilm, Düsseldorf, Germany).

Table 6.1: Primers used for dot blot analysis and *in situ* hybridisation

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

In situ hybridisation

A whole mount *in situ* hybridisation was carried out as described in chapter 3. The templates for the production of DNA probes against unigenes with homology to expansin-like proteins (CL496) and venom allergen proteins (FE920532) were generated by a first PCR as described in chapter 2 on the plasmid pGEM-T containing the cDNA clone of the corresponding unigene. Subsequent linear PCRs using the first PCR products as templates with digoxigenin(DIG)-labeled oligonucleotides and a single primer, generated single strand DNA probes (Table 6.1).

Results

Cleaning and cluster analysis

A total of 4847 ESTs were obtained from a cDNA library prepared from mixed stages of the plant-parasitic nematode *Ditylenchus africanus*. Cleaning the EST sequences resulted in the removal of 602 ESTs, due to vector contamination, poly(A) tails and sequences shorter than 100 nt. The average length of the 4245 retained ESTs was 286 bp. Clustering to reduce data redundancy resulted in 2456 ESTs forming 807 contigs (merged overlapping sequences) classified into 778 clusters (group of contigs with minor sequence variation) and 1789 singletons, resulting in a total of 2596 unigenes. After clustering, the average length of a unigene increased to 335 bp and the average GC content is 38.9%. A graphical representation of the cluster size distribution is shown in Figure 6.1. The largest single cluster represents 84 ESTs, which is 2.0% of the total number of ESTs. Most of the clusters (88.0%) consist of four or less ESTs. The fragmentation or underclustering of the dataset was estimated at 3.5%, which means that the dataset represents approximately 2505 genes.

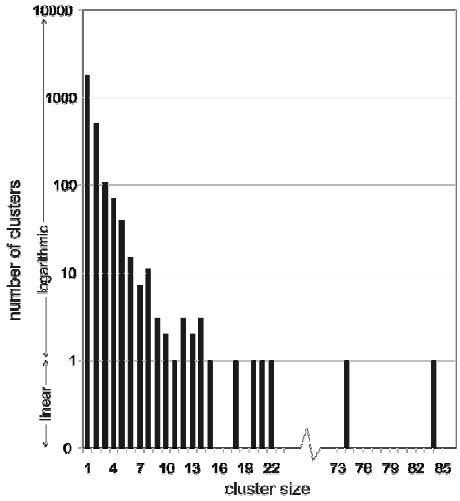

Figure 6.1: Cluster size distribution. The chart shows the number of clusters with a particular cluster size (the number of ESTs present in this cluster) in the EST dataset of *Ditylenchus africanus*.

Transcript abundance and homology searches

Blast searches against the Protein and Nucleotide division of GenBank were performed, as well as against all nematode ESTs. The 20 most abundant unigenes in the EST dataset represent 9.6% of the total number of ESTs (Table 6.2). Eight of these unigenes, including cluster 1, showed no significant hit in any of the blast searches conducted (E-value<1e-3). In Table 6.3 unigenes are shown which have homologs involved in anhydrobiosis, oxidative stress and parasitism.

Table 6.2: The 20 most prevalent unigenes with the number of ESTs they contain and with their best blast hit description (either derived from blastx against the Protein division of GenBank with cut-off 1e-3, blastn against the Nucleotide division of GenBank with cut-off 1e-5 or tblastx against all nematode ESTs with cut-off 1e-3), accession number, blast E-value, database of the top blast hit (Protein, Nucleotide or EST), percentage protein identity and GC content.

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

Table 6.3: Some selected unigenes based on homology results. The best blast hit description is given (either derived from blastx against the Protein division of GenBank with cut-off 1e-3, blastn against the Nucleotide division of GenBank with cut-off 1e-5 or tblastx against all nematode ESTs with cut-off 1e-3), accession number, blast E-value, database of the top blast hit (Protein, Nucleotide or EST) and percentage protein identity.

Unigene	Top hit species and descriptor	Accession	E-value	db	%id
Anhydrobiosis					
CL23	*Aphelenchus avenae* LEA1	Q95V77	6.00E-23	Prot	48.4
CL11	*Pratylenchus penetrans* LEA5 homolog	BQ627245	1.00E-21	EST	55.4
CL79	*Pratylenchus vulnus* fatty acid desaturase (fat-6) homolog	EL890688	1.00E-83	EST	77.4
FE922042	*Haemonchus contortus* C-type lectin homolog	CB333334	6.00E-10	EST	35.3
FE921359	*Caenorhabditis elegans* C-type lectin (clec-53)	NP_491247	4.00E-04	Prot	28.2
CL475	*Globodera rostochiensis* stomatin (sto-5) homolog	EE268121	4.00E-51	EST	74.8
FE922269	*Aphelenchus avenae* trehalose 6-phosphate synthase homolog	AJ811572	1.00E-11	Nuc	75.6
Oxidative stress					
FE922062	*Wuchereria bancrofti* Cu/Zn-superoxide dismutase homolog	CK726411	7.00E-07	EST	54.8
FE925119	*Bursaphelenchus xylophilus* glutathione reductase homolog	CJ987986	3.00E-45	EST	75.9
CL450	*Globodera rostochiensis* glutathione S-transferase homolog	BM344760	1.00E-18	EST	36.8
FE924154	*Meloidogyne incognita* glutathione S-transferase homolog	CK983784	4.00E-23	EST	42.4
CL500	*Dictyocaulus viviparus* peroxiredoxin homolog	EV850422	5.00E-28	EST	87.0
CL130	*Heterodera glycines* peroxiredoxin homolog	CA940959	4.00E-98	EST	75.4
CL452	*Ancylostoma caninum* superoxide dismutase	EX542471	6.00E-13	EST	61.4
CL226	*Chlamydomonas reinhardtii* superoxide dismutase	XM_001699025	1.00E-21	Nuc	83.3
CL21	*Haemonchus contortus* superoxide dismutase	CB018828	8.00E-41	EST	74.7
Allergens					
CL752	*Toxoascaris leonina* ABA-1 allergen homolog	ES880608	6.00E-10	EST	54.0
FE923927	*Meloidogyne paranaensis* polyprotein allergen homolog	CN477518	5.00E-11	EST	42.9
CL43	*Globodera pallida* major allergen homolog	BM415278	3.00E-33	EST	52.6
FE923357	*Heterodera glycines* vap-1	AAK60209	5.00E-23	Prot	61.7
Putative parasitism genes					
CL10	*Zeldia punctata* SEC-2 protein homolog	AW783768	4.00E-72	EST	76.3
FE922606	*Pratylenchus vulnus* transthyretin-like protein homolog	EL889277	2.00E-20	EST	64.9
CL678	*Globodera rostochiensis* transthyretin-like protein homolog	BM355233	5.00E-10	EST	80.0
CL536	*Meloidogyne chitwoodi* transthyretin-like protein homolog	CD682816	5.00E-52	EST	72.6
CL112	*Meloidogyne chitwoodi* transthyretin-like protein homolog	CF801754	6.00E-10	EST	48.3
CL665	*Haemonchus contortus* calreticulin homolog	AI723603	2.00E-69	EST	86.4
CL573	*Strongyloides ratti* 14-3-3b protein homolog	FC819840	1.00E-70	EST	57.5
CL301	*Globodera rostochiensis* 14-3-3b protein homolog	EE267463	1.00E-103	EST	95.8
CL8	*Meloidogyne chitwoodi* putative expansin-like protein homolog	CB831016	6.00E-16	EST	46.4
CL496	*Bursaphelenchus xylophilus* expansin-like protein homolog	CJ981766	4.00E-26	EST	57.5

Classifying the unigenes based on species names of blastx homologous sequences revealed 298 animal specific sequences of which 160 unigenes were nematode specific. These nematode specific unigenes were classified according to their homologs in the blastx search and in the nematode EST database (Figure 6.2). Most of these sequences have homologs in both free-living, animal-parasitic and plant-parasitic nematodes, suggesting a role in general nematode development and metabolism. Only one unigene (CL496) has homologous protein sequences exclusively from plant-parasitic nematodes, i.e. an expansin (top blastx hit: expansin-like protein BAG16537 from *Bursaphelenchus xylophilus*, E=1e-20). Unigenes with homologs to animal-parasitic nematodes exclusively (4) are all homologs to proteins with unknown function. Twelve unigenes are found in both animal- and plant-parasitic nematodes exclusively, and include several allergen homologous sequences (Table 6.3) as well as sequences with unknown function.

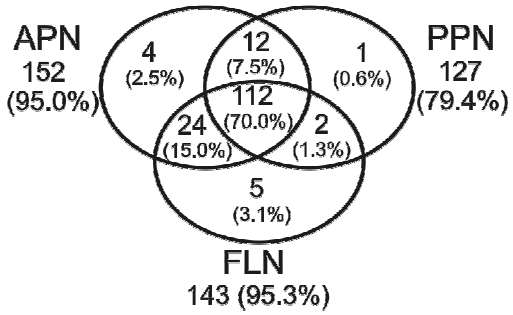

Figure 6.2: Classification of the 160 nematode specific unigenes with blastx homology (E-value<1e-3) into animal-parasitic (APN), plant-parasitic (PPN) and free-living (FLN) nematodes based on nematode EST homology. The number of unigenes in each class is given, including the percentage of the total nematode specific genes.

Remarkably, some unigenes had unexpected blastx homologs originating from prokaryotic sequences exclusively (15). Similarly, two unigenes matched fungal hypothetical proteins only: one (FE921742) with hits from *Coprinopsis cinerea* (E=3e-04) and two ESTs from *Globodera rostochiensis*, and the other (FE922861) with hits from fungal plant pathogens (*Gibberella* sp., *Sclerotinia* sp. and *Botryotinia* sp.; E=4e-14). Searching the genomes of *Meloidogyne incognita* and *M. hapla* revealed a tblastx homolog for respectively 30.6% and 35.4% of the unigenes (E-value<1e-3). Thirty unigenes of putative mitochondrial origin were retrieved corresponding to 40 ESTs or 0.94% of the total dataset. The GC content of the putative mitochondrial unigenes is 34.5%, which is lower than the overall GC content of 38.9% (Figure 6.3). For 1113 unigenes (42.9%), no homologous sequences could be found in either protein, nucleotide or nematode EST databases or in the genomes of *M. incognita* and *M. hapla* (E-value<1e-3). These sequences had a shorter average sequence length (260 bp) compared to unigenes with homologs (391 bp) (Figure 6.3), which is the most obvious reason why no homology was found. These unigenes of unknown origin could correspond either to non-coding sequences (e.g. UTRs), regulatory and structural RNA or to novel protein coding genes.

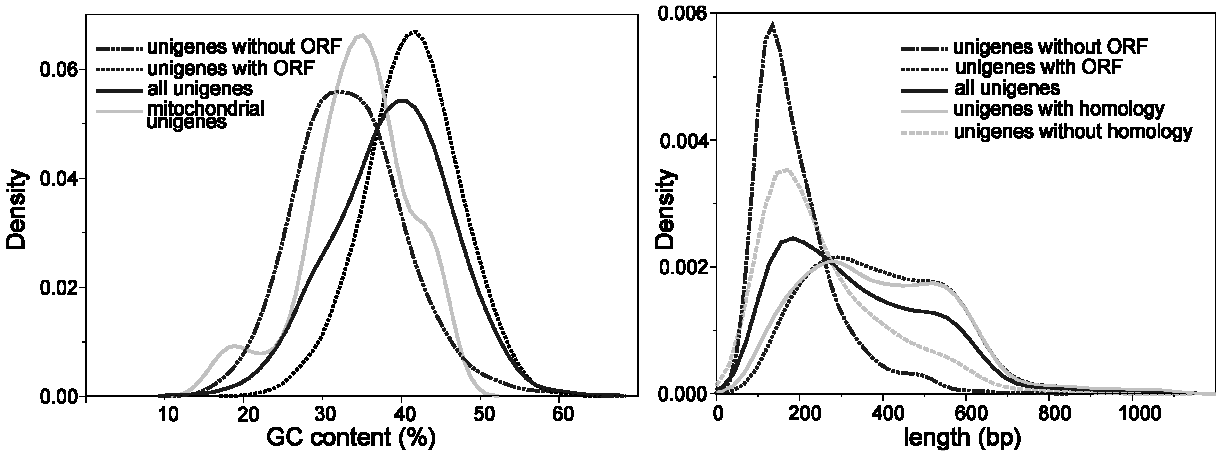

Figure 6.3: Density plots of the GC content and length of different subsets of unigenes.

Gene Ontology

To have a general overview of the functions of the unigenes, Gene Ontology (GO) terms were assigned to the unigenes with BLAST2GO. Of the 931 unigenes for which BLAST2GO could identify at least one blast hit, 801 were annotated with at least one GO term. The number of GO terms per unigene varied from 1 to 53. In total, 2648 different GO terms were retrieved: 1658 of biological process, 653 of molecular function and 337 of cellular component (Figure 6.4). In the molecular function category the majority of the putative proteins are involved in protein binding (43.7%) whereas other GO terms occur between 5.8% (hydrolase activity) and 10.9% (structural constituent of ribosome). For the biological process category, 23.0% of the terms are involved in embryonic development and 17.9% in larval development. The most prevalent cellular component terms are protein complex (25.1%), integral to membrane (16.7%) and mitochondrion (13.6%).

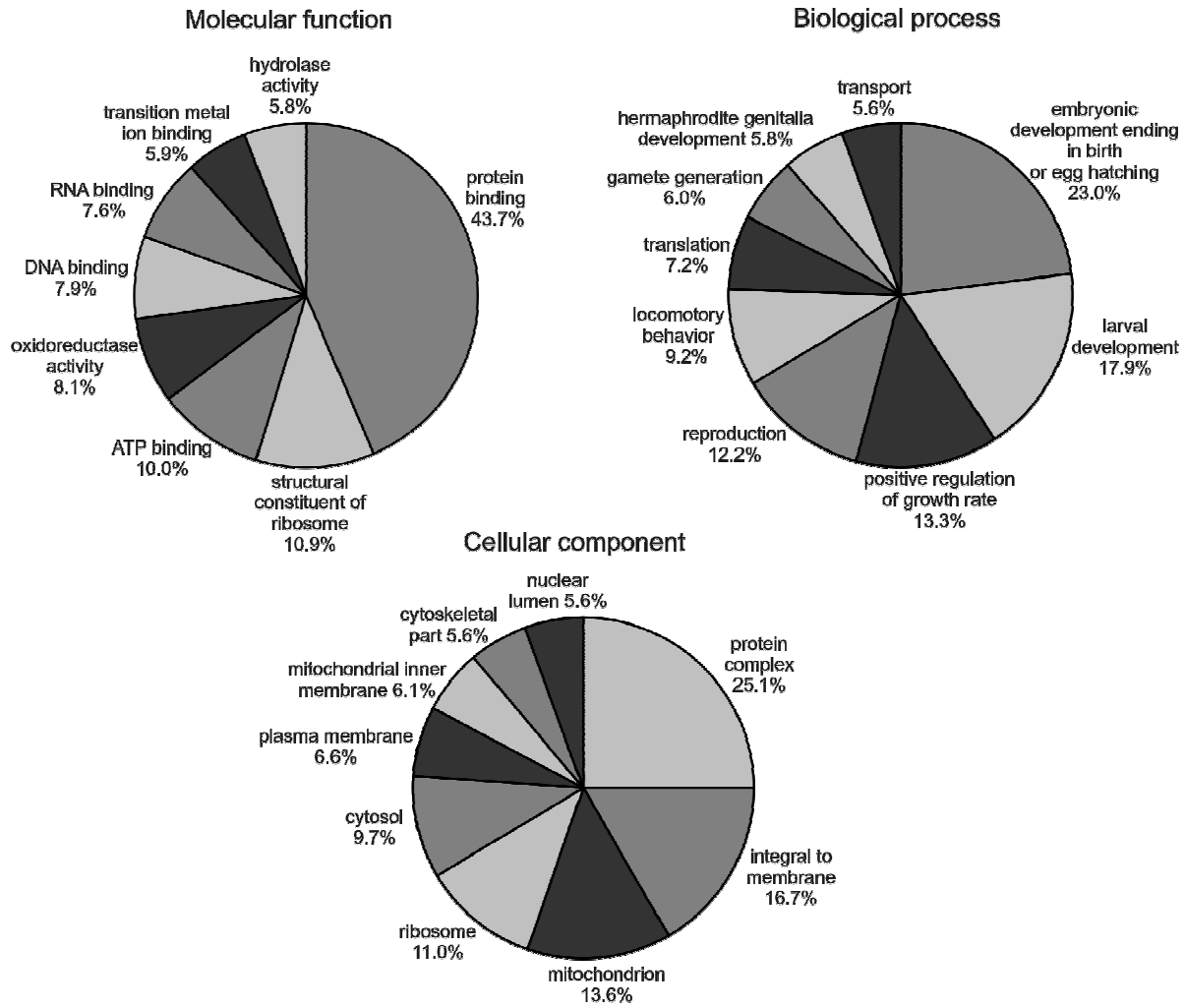

Figure 6.4: Representation of gene ontology (GO) mappings for all unigenes as calculated by BLAST2GO. Different pie charts are given for the terms molecular function, biological process and cellular component with a node score cut-off of 25, 50 and 12 respectively.

Translation into putative proteins

Unigenes were translated into putative protein sequences on the basis of the blastx output or based on the longest ORF in case of no blastx homology. For 1769 unigenes a putative ORF of minimum 40 AA was found, while 827 were predicted to be non-coding by OrfPredictor. Overall GC contents of the coding sequences is 41.4%, while GC1, GC2 and GC3 are 50.8; 38.5 and 48.5%, respectively. The overall GC content of putative non-coding sequences is 33.6% (Figure 6.3). A signal peptide for secretion was found in 101 of the 1769 putative proteins (5.7%). Of these, 90 had no transmembrane domain (5.1%). These are possibly being secreted by the nematode and hence could play a role in parasitism. Only 13 showed similarity to known, mostly hypothetical, proteins (blastp search; E-value<1e-3). The majority (72) of these 90 putative secreted proteins did not show any similarity to other known nematode ESTs (tblastn search; E-value<1e-3). Only 7 candidate parasitism genes were retained which had homology restricted to parasitic nematode ESTs (Table 6.4).

Table 6.4: Novel candidate parasitism genes. Putative proteins of these unigenes have a predicted signal peptide and tblastn searches against nematode ESTs (E-value<1e-4) reveal hits with parasitic nematodes only (PPN: EST hits in plant-parasitic nematodes only, APN: EST hits in animal-parasitic nematodes only, PN: EST hits in both animal- and plant-parasitic nematodes). The top hit species, accession number, blast E-value, percentage protein identity, the number of amino acids, the length of the putative signal peptide (SP) and the presence of a start methionine are given.

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

RNAi phenotypes

RNAi phenotypes were assigned to the unigenes by homology with *C. elegans* proteins. 965 unigenes had a *C. elegans* homolog (blastx search; E-value<1e-3). After removing redundant protein hits, 815 different *C. elegans* proteins were retained. For 473 or 58.0% of the total number of unigenes with a *C. elegans* homolog, an RNAi phenotype was retrieved from the Wormmart database. The majority (333) of these proteins have a lethal phenotype (WBPhenotype0000050, WBPhenotype0000054, or WBPhenotype0000062). GO terms for these *C. elegans* proteins were retrieved and the terms of the second level of the “biological process” category are shown in Figure 6.5. All of them have the GO terms “developmental process” and “multicellular organismal process”, suggesting that the corresponding genes are involved in the general development of the organism. Other GO terms that occur in the

majority of these genes are “reproduction”, “metabolic process”, “cellular process”, “growth” and “biological regulation”. 366 unigenes (14.1% of the total number of unigenes) correspond to *C. elegans* genes with lethal phenotype. A subset of these unigenes (34) are promising future candidates for parasitic control due to their nematode specific nature.

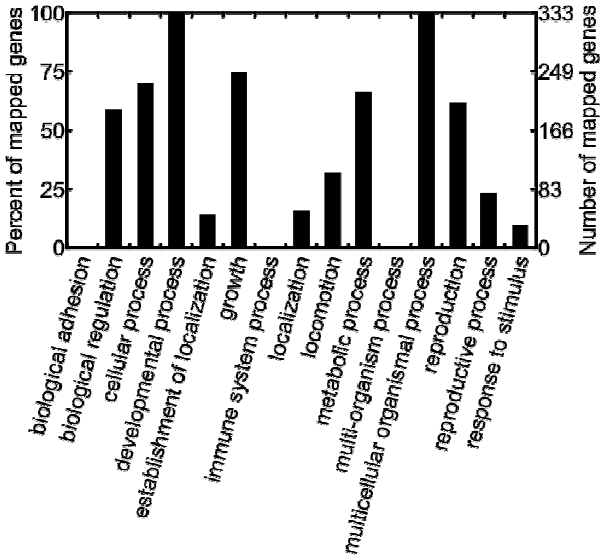

Figure 6.5: Graphical representation of the occurrence of GO terms of biological process (level 2) among the *C. elegans* homologs of *Ditylenchus africanus* unigenes with a lethal RNAi phenotype.

Dot blot analysis

To find out whether the largest clusters without any homology were really present in the mRNA pool, a dot blot analysis was carried out. Correct fragments were obtained by PCR on the cDNA library for both CL1 and CL7. The dot blot analysis showed a clear signal for both fragments against RNA extracted from *D. africanus* nematodes, while no signal was obtained against carrot RNA (Figure 6.6). This shows that these clusters are not artifacts or contaminants from the cDNA library since they are also present in independently isolated nematode RNA.

Some sequences showed similarity to prokaryotic or fungal sequences exclusively in a blastx search, for example CL270, CL371, CL406, FE921742 and FE922861. To check if these sequences are derived from contamination, a similar dot blot analysis was done. For three of the five selected sequences (CL406, FE921742 and FE922861), the corresponding fragment could not be amplified from the cDNA library. This most likely means that these sequences are not present in the cDNA library, and are probably the result of contamination. For CL270 and CL406, the correct fragment could be cloned, however in a dot blot, no signal was obtained neither for nematode nor carrot RNA (Figure 6.6). This suggests that CL270 and CL406 are from contaminating origin, although it cannot be ruled out that the detection limit of the dot blot is too low to detect the corresponding mRNA.

In situ hybridisation

An *in situ* hybridisation with probes against CL496 (putative expansin-like protein) and FE920532 (putative venom allergen protein) was carried out. Both probes showed staining in the gland cell area of the nematode, while the corresponding sense probes showed no staining (Figure 6.6). Whether the expression was located in the subventral or dorsal gland cells could not be determined. These results suggest that both proteins are possibly being secreted by the nematode and hence could play a role in parasitism.

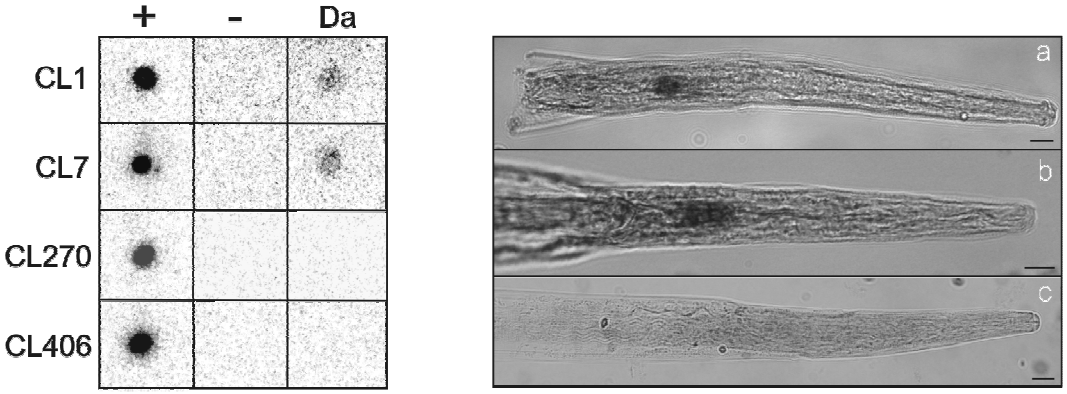

Figure 6.6: Left: Dot blot analysis with probes against the selected unigenes CL1, CL7, CL270 and CL406. (+) PCR fragment of the corresponding probe; (-) carrot RNA; (Da) RNA isolated from mixed stages of *Ditylenchus africanus*; Right: *In situ* hybridisation on *Ditylenchus africanus*. (a) CL496 (expansin-like protein) antisense probe; (b) FE920532 (venom allergen protein) antisense probe; (c) CL496 sense probe (negative control). Scale bars: 15 µm.

Discussion

An increasing amount of sequence information derived from plant-parasitic nematodes is becoming available through EST projects, especially from sedentary nematodes from the superfamily Tylenchoidea within the order of the Rhabditida. In contrast, *Ditylenchus africanus* is a migratory endoparasite belonging to the family of the Anguinidae, superfamily Sphaerularioidea. This family comprises important seed and stem nematodes such as *Anguina tritici* and *Ditylenchus dipsaci*. This study is the first molecular survey of a nematode in this family, *D. africanus*, by means of a small scale EST analysis. Only 57% of the clustered unigenes showed homology to known proteins, nematode ESTs and the genomes of the plant-parasitic nematodes *Meloidogyne incognita* and *M. hapla*. This means that at least 43% of all unigenes did not have any homology to either all known proteins, nucleotide sequences, nematode ESTs or the *Meloidogyne* genomes. The amount of unigenes without homology is unusually high, as compared to for example *Bursaphelenchus xylophilus* (27% unigenes without homology) (Kikuchi et al., 2007), a nematode which is phylogenetically further removed from the heavily sampled cyst and root-knot nematodes than *D. africanus*. The most

obvious reason for this high amount of orphan sequences is the relatively short average length of the inserts of the cDNA library. The shorter the insert, the less likely it is to find a significant homolog. Moreover, the length distribution of unigenes with homology is similar to unigenes which are predicted to be coding. Other sequences without homology are shorter in length and could be either non-coding regions, regulatory or structural RNAs or pioneer sequences. The sequences predicted to be non-coding have a lower average GC content (34%) compared to coding sequences (41%). In *C. elegans*, the GC content of coding sequences is around 45%, while it is only 34% in intergenic regions (Shin et al., 2008). These similar numbers suggest that a large portion of the *D. africanus* non-coding sequences are 5' or 3' UTRs. On the other hand, since we are dealing with the first sequences from a representative of the Anguinidae family, it is to be expected that there are a number of species or family specific unigenes without any homology in currently available sequence databases. Interestingly, a considerable number of these unknown unigenes seems to be highly expressed. Eight of the 20 most abundant clusters including the most abundant cluster (CL1) have no homologs at all. A dot blot for two of these unknown, highly expressed unigenes confirmed that these are truly present in the nematode mRNA pool. A high fraction of the largest clusters lacking homology has also been observed in other nematode EST datasets, and these ESTs are thought to be of mitochondrial origin (Dubreuil et al., 2007; Jacob et al., 2008). Other highly expressed genes represented by a large number of ESTs are for example major sperm protein (important in the motility of sperm), ferritin (iron-storage protein), tropomyosin (an actin-filament regulator) and *sec-2* or *far* (fatty acid and retinol binding protein, previously hypothesised to be involved in defense) (Prior et al., 2001). The latter two have been reported earlier as highly abundant transcripts in ESTs (McCarter et al., 2003; Vanholme et al., 2006; Kikuchi et al., 2007; Jacob et al., 2008). Gene ontology mapping revealed that most unigenes are involved in the biological processes of reproduction and development. This is not surprising, since *D. africanus* has a short life cycle of only 6-7 days and has a high reproductive rate.

Interestingly, nematode specific unigenes predominantly had EST homologues from animal-parasitic species (95%), while 79% also had homologs in plant-parasitic nematodes. This apparent closer relationship to animal-parasitic nematodes can be due to the fact that there is more sequence information available for animal-parasitic compared to plant-parasitic nematodes (24,575 vs. 1227 proteins and 294,843 vs. 143,883 ESTs). Only one unigene (CL496) had homologs within plant-parasitic nematodes exclusively (an expansin identified in *Bursaphelenchus xylophilus*). In addition, CL8 shows the highest similarity to an EST from *Meloidogyne chitwoodi* with homology to expansins. Although plant-parasitic nematodes are known to secrete an arsenal of different cell wall modifying enzymes (Vanholme et al., 2004), unigenes similar to expansin genes were the only ones found in the EST dataset. The

expression of CL496 was demonstrated to be in the gland cell region of the nematode, pointing towards the secretion of the protein.

Some unigenes showed similarity to putative parasitism genes such as *sec-2* (Prior et al., 2001), calreticulin (Jaubert et al., 2005) and 14-3-3b protein (Jaubert et al., 2004). However all of these enzymes should be more thoroughly investigated in order to find out their true role in parasitism. For example, calreticulin and 14-3-3b protein have been linked to the formation of a feeding site by sedentary nematodes (Jaubert et al., 2005). The fact that these secreted proteins are also present in migratory nematodes could suggest a more general role in parasitism. The unigenes were searched for new putative parasitism genes which should code for proteins equipped with a signal peptide. In addition, the best candidates should only have homologs in parasitic nematode ESTs. Although this approach rules out the finding of species or family specific parasitism genes, since it is dependent on available nematode EST data, this condition was included to reduce the number of false positives. Seven promising novel candidate parasitism genes were found (Table 6.4). These genes are interesting for future functional studies and could reveal new insights about plant parasitism.

Several nematode parasitism genes have been proposed to originate through horizontal gene transfer (HGT) (Jones et al., 2005). New HGT candidates should have homology with prokaryotes or fungi exclusively, although prokaryotic sequences could also point to contamination or the presence of an endosymbiont (Jacob et al., 2008). Several unigenes identified in our EST dataset showed similarity to prokaryotes exclusively, but most of these unigenes show homology to widely occurring non endosymbiotic species, suggesting some bacterial contamination. For other unigenes resembling prokaryotic or fungal sequences, it could not be demonstrated that these were present in independently isolated RNA of the nematode, hence these are probably of contaminating origin as well.

Our EST dataset contains several genes that are predicted to be involved in anhydrobiosis, a dehydrated state of the nematode. One of these genes is a LEA (late embryogenesis abundant protein) homolog, proteins found to be highly expressed in the anhydrobiotic nematodes *Aphelenchus avenae* and *Steinernema carpocapsae*, as well as in plants where they prevent protein aggregation during seed desiccation (Browne et al., 2002; Tyson et al., 2007). In addition, fatty acid desaturase, stomatin and trehalose-6-phosphate synthase were found, mainly involved in maintaining the integrity of the cell membrane (Tyson et al., 2007). Other unigenes commonly upregulated in anhydrobiotic stages are involved in oxidative stress, such as superoxide dismutase, glutathione-S-transferase and peroxiredoxin. These enzymes and also LEA proteins are represented by a quite large number of ESTs, which indicates that oxidative stress as well as drought stress are important factors affecting the nematode. In agreement with this finding, as much as 8% of the GO mappings in the “molecular function” category (node score>25), are associated with oxidoreductase activity.

Another group of unigenes showed remarkable similarities to polyprotein allergens from animal-parasitic nematodes and the venom allergen protein (VAP) of *Heterodera glycines*. Gao et al. (2001) found that the latter is expressed in the gland cells, which was confirmed for a *vap* homolog in *D. africanus* by *in situ* hybridisation. In *Meloidogyne incognita* it was shown that a venom allergen AG5-like protein is expressed during preparasitic and parasitic J2 stages exclusively (Ding et al., 2000). Animal-parasitic allergen genes have been associated with the immune response of the host animal (Murray et al., 2001). However, their presence in plant-parasitic nematodes suggests a more general role for these proteins in parasitism.

The RNAi gene knockout technique has shown promising results in sedentary plant-parasitic nematodes (Lilley et al., 2007). Transcript levels of a certain gene can be reduced when dsRNA homologous to that gene is fed to the nematode. Interesting RNAi targets for experiments with *D. africanus* were identified by looking for *C. elegans* homologs with lethal RNAi phenotypes. Retained unigenes were all involved in major biological processes (development, metabolic processes, reproduction) as revealed by the Gene Ontology terms of their *C. elegans* homologs. Most of these unigenes have homologs both in free-living and animal-parasitic nematodes, due to the limited genomic data available for plant-parasitic nematodes. Unigenes showing homology to nematodes exclusively (34) are interesting targets for control measures since these are very unlikely to have off-target effects on other organisms. Future experiments will have to prove the efficacy of these candidate RNAi targets and whether or not RNAi has similar effects in *D. africanus* as in sedentary nematodes, although the technique was shown to be effective in the migratory nematode *Radopholus similis* (chapter 5).

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

Acknowledgments

We would like to thank the Washington University School of Medicine for generating the ESTs, supported by NIH research grant AI 46593 to Makedonka Mitreva.

CHAPTER 7

AN ENDOSYMBIOTIC BACTERIUM IN *RADOPHOLUS SIMILIS*: MEMBER OF A NEW *WOLBACHIA* SUPERGROUP

Adapted from:

Annelies Haegeman; Bartel Vanholme; Joachim Jacob; Tom Vandekerckhove; Myriam Claeys; Gaetan Borgonie & Godelieve Gheysen. (2009). An endosymbiotic bacterium in a plant-parasitic nematode: member of a new *Wolbachia* supergroup. International Journal for Parasitology, doi: 10.1016/j.ijpara.2009.01.006.

Abstract

Wolbachia is an endosymbiotic bacterium widely present in arthropods and animal-parasitic nematodes. Despite previous efforts, it has never been identified in plant-parasitic nematodes. Random sequencing of genes expressed by the burrowing nematode *Radopholus similis* resulted in several sequences with similarity to *Wolbachia* genes. The presence of a *Wolbachia*-like endosymbiont in this plant-parasitic nematode was investigated using both morphological and molecular approaches.

Transmission electron microscopy, fluorescent immunolocalisation and DAPI staining confirmed the presence of the endosymbiont within the reproductive tract of female adults. *16S rDNA*, *ftsZ* and *groEL* gene sequences showed that the endosymbiont of *Radopholus similis* is distantly related to the known *Wolbachia* supergroups. Finally, based on our initial success to find sequences of this endosymbiont by screening an expressed sequence tag (EST) dataset, all nematode ESTs were mined for *Wolbachia*-like sequences. Although the retained sequences belonged to six different nematode species, *Radopholus similis* was the only plant-parasitic nematode with traces of *Wolbachia*.

Based on our phylogenetic study and the current literature we designate the endosymbiont of *R. similis* to a new supergroup (supergroup I) rather than considering it as a new species. Although its role remains unknown, the endosymbiont was found in all individuals tested, pointing towards an essential function of the bacteria.

Introduction

In 1924, Hertig and Wolbach discovered an endosymbiotic bacterium in the ovaries of the mosquito *Culex pipiens* (Hertig & Wolbach, 1924). The bacterial species was named *Wolbachia pipientis* and catalogued as a *Rickettsiaceae* member belonging to the *Alphaproteobacteria*. Since this first report, *Wolbachia* species have been found in numerous arthropods including insects, crustaceans, spiders, and mites (Lo et al., 2007). Although widespread in arthropods, the distribution in nematodes is more restricted. Until a few years ago they were only described in filarial nematodes of the family Onchocercidae which includes the important animal parasites *Brugia malayi*, *Onchocerca volvulus* and *Wuchereria bancrofti* (Fenn & Blaxter, 2006). Genetic evidence for the presence of *Wolbachia* in the rat lungworm *Angiostrongylus cantonensis*, a non-filarial nematode, suggested a more widespread occurrence of *Wolbachia* among animal-parasitic nematodes (Tsai et al., 2007), but this finding was afterwards claimed to be the result of contamination with DNA from arthropods and filarial nematodes (Foster et al., 2008). Bordenstein et al. (2003) investigated 21 non-filarial nematode species, but none were found to harbour *Wolbachia*.

In contrast to their symbiotic lifestyle in most arthropods, *Wolbachia* of nematodes are considered to be mutualists since all adult nematodes in species infected with *Wolbachia* have the endobacteria (Werren et al., 2008). Moreover, treatments with antibiotics that would eradicate *Wolbachia* have unsuspected and dramatic effects on the host, causing delayed moulting, reduced growth rates, aberrant embryogenesis and eventual death (Casiraghi et al., 2002; Fenn & Blaxter, 2006). Remarkably, similar treatments of insects have only minor or no effects. A possible explanation for their beneficial role in relation to nematodes was given upon sequencing of the *Wolbachia* genome of *B. malayi* (Foster et al., 2005). Careful annotation of the sequences revealed that *Wolbachia* could provide nematode-essential metabolites, such as riboflavin, heme, glutathione, glycolytic enzymes, and compounds necessary for the biosynthesis of purines and pyrimidines (Foster et al., 2005). Within filarial nematodes, *Wolbachia* occurs in the lateral chords of the hypodermis in both males and females, as well as throughout the whole reproductive tract of females (Kramer et al., 2003).

The establishment of a conserved taxonomy is still a subject of controversy within the *Wolbachia* community. Lo et al. (2007) have proposed that all *Wolbachia* endobacteria be declared as one species, being *Wolbachia pipientis*. Pfarr et al. (2007) disagree with this idea mainly based on some remarkable differences in the genome sequences of *Wolbachia* of *B. malayi* and *Drosophila melanogaster* (wBm and wMel respectively). Whereas wBm lacks genes necessary for recombination resulting in little or no genomic recombination, wMel has an active recombination system (Wu et al., 2004). In combination with the biological differences between arthropod and nematode *Wolbachia*, Pfarr et al. (2007) state that there is enough evidence to consider the *Wolbachia* of filarial nematodes as a different species.

However, it was recently shown that nematode *Wolbachia* are not monophyletic (Bordenstein et al., 2009). Evolutionary relationships between *Wolbachia* strains have been inferred mainly from *16S rRNA, wsp* (*Wolbachia* surface protein), *ftsZ* (cell division protein) and *groEL* (heat shock protein 60) sequences. Based on these evolutionary trees, *Wolbachia* is divided into different clades, referred to as "supergroups" (Casiraghi et al., 2005). Supergroups A and B include most of the *Wolbachia* found in arthropods (Werren et al., 1995), whereas C and D harbour the majority of the *Wolbachia* found in filarial nematodes (Bandi et al., 1998). Supergroup E consists of *Wolbachia* from springtails (Collembola) (Vandekerckhove et al., 1999; Czarnetzki & Tebbe, 2004) and supergroup F comprises *Wolbachia* from termites, weevils, true bugs, scorpions (Lo et al., 2002; Rasgon & Scott, 2004; Baldo et al., 2007), and the filarial nematodes *Mansonella* spp. (Casiraghi et al., 2001; Keiser et al., 2008). Two relatively new supergroups are supergroup G with Australian spiders as hosts (Rowley et al., 2004), and H with Isopteran species (Bordenstein & Rosengaus, 2005). Finally, some *Wolbachia* strains such as those identified in *Dipetalonema gracile* (filarial nematode), *Ctenocephalides* spp. (flea) and *Cordylochernes scorpioides* (pseudoscorpion) could not be classified in any of these supergroups (Casiraghi et al., 2005; Zeh et al., 2005; Bordenstein et al., 2009).

During a recent screen of expressed sequence tags (ESTs) derived from the burrowing nematode *Radopholus similis* approximately 1% of the unigenes showed similarity to *Wolbachia* sequences (Jacob et al., 2008). However, it was not clear if these could be nematode genes acquired by lateral gene transfer from an ancient and long-lost endosymbiont (Dunning-Hotopp et al., 2007) or whether there was still a true endosymbiont present in *R. similis*. In this chapter, we confirm the presence of a *Wolbachia*-like endosymbiont in this plant-parasitic nematode, using molecular and microscopic evidence. Other symbionts have already been identified in plant-parasitic nematodes: *Verrucomicrobia* species in *Xiphinema index* (Vandekerckhove et al., 2000), *Bacteroidetes* species in *Heterodera glycines* and *Globodera rostochiensis* (Noel & Atibalentja, 2006). This is, however, to our knowledge the first report on a *Wolbachia*-like bacterium in a plant-parasitic nematode.

Materials and methods

Searching for indications of *Wolbachia* in nematode ESTs

To look for *Wolbachia* ESTs in nematode EST databases, we used a reciprocal blast search approach utilising the blast algorithm. All 805 *Wolbachia* proteins from the *Brugia malayi Wolbachia* genome (WolBm, NC_006833) were downloaded from the Genome division of GenBank (http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome). To identify nematode ESTs with significant homology to WolBm proteins, a tblastn search with the WolBm proteins as

query was done locally by StandAloneBlast (ftp://ftp.ncbi.nih.gov/blast/) against all available nematode ESTs downloaded from GenBank (December 2007; *n* = 903,682; E-value<1e-2). All significant hits were retained, and sequences with the term "*Caenorhabditis*" in the description line were removed to reduce the size of the dataset. To verify whether these selected nematode ESTs could be derived from *Wolbachia*, a blastx search with the identified ESTs was conducted against all known proteins using NetBlast (blastcl3 network client, ftp://ftp.ncbi.nih.gov/blast/) (February 2008). Only when the top hit was a *Wolbachia* protein, was the nematode EST considered as a putative candidate sequence originating from *Wolbachia*.

Culture of *Radopholus similis* and sterile DNA extraction

Radopholus similis was cultured as described in chapter 2. All following manipulations for DNA extraction were performed under sterile conditions. For surface sterilisation, nematodes were soaked for 30s in 0.1% benzalkonium chloride (Sigma-Aldrich). Subsequently, nematodes were spun in a microcentrifuge (Eppendorf) at 1000 g for 3 min and washed three times with sterile demineralised water. The nematodes were pelleted and used for genomic DNA extraction (Bolla et al., 1988) under sterile conditions. The DNA pellet was dissolved in 50 µl of sterile demineralised water.

Cloning of *Wolbachia* genes

To amplify the *16S rRNA* gene of *Wolbachia*, a PCR was performed on 150 ng of isolated DNA. The 25 µl reaction mixture was prepared under sterile conditions and contained 0.5 µM of each universal bacterial *16S rRNA* primers (16S-F and 16S-R) (Edwards et al., 1989) (Table 7.1), 4 mM of each dNTP, 1.5 mM $MgCl_2$, 20 mM Tris-HCl (pH 8.3), 50 mM KCl and 1U of *Taq* DNA polymerase (Invitrogen).

Table 7.1: Primer sequences used for cloning of the *16S rRNA*, *ftsZ* and *groEL* genes

Primer	**Primer sequence**	**Primer**	**Primer sequence**
16S-F	AGAGTTTGATCCTGGCTCAG	16S-R	AAGGAGGTGATCCAGCCGCA
wsp-F	TGGTCCAATAAGTGATGAAGAAACTAGCTA	wsp-R	AAAAATTAAACGCTACTCCAGCTTCTGCAC
ftsZ-up1	ATTCCTCTGCCGCACCTCTAC	ftsZ-up2	CATCAGGTAAAGCACCAGCA
ftsZ-down1	CCGGATTTACCTGTACTACACC	ftsZ-down2	CTGTGGTGGGAGTAGGTGGT
ftsZ-down3	GCTGATAATGTGCTGCATATAGG	ftsZ-down4	CTGATTTAATGGTTATGCCAGG
groEL-up1	CCAGCAGAAATGGGCTTTAAAGC	groEL-up2	ACACTGAGAACAACTTTGAGC
groEL-down1	GCTTGAGCGTGGTTATGCTTC	groEL-down2	GAGTGCCATACATCCTTTAATACC
AP1	GTAATACGACTCACTATAGGGC	AP2	ACTATAGGGCACGCGTGGT
Rs-act-F	TGTCCGTCTTTGTCATTTGC	Rs-act-R	TGATCACCGTCGGAAACG
ftsZ-F	CTGTGGTGGGAGTAGGTGGT	ftsZ-R	AGCAGCCTTTGCTTCTCTTG
groEL-F	AAAGCCCATTTCTGCTGGTA	groEL-R	GGAAGCATAACCACGCTCAA
ftsZ-F2	CCGGATTTACCTGTACTACACC	ftsZ-R2	AGCAGCCTTTGCTTCTCTTG

The PCR consisted of 2 min at 94°C followed by 35 cycles of 1 min at 94°C, 1 min at 54°C and 1 min 30s at 72°C. The PCR product was separated on a 0.5× TAE 1.5% agarose gel, purified, cloned and sequenced as described in chapter 2.

An attempt was made to amplify the *wsp* gene using primers wsp-F and wsp-R (Braig et al., 1998). For *ftsZ* and *groEL*, part of the sequence was available as an EST (EY193345 and EY195553 respectively). The full length coding sequences were obtained by genome walking on *R. similis* DNA libraries using the Genome Walker Universal kit (Clontech) according to the manufacturer's instructions. For both genes, two successive primers were developed based on the EST sequences to walk both up- and downstream (ftsZ-up1, ftsZ-up2, ftsZ-down1, ftsZ-down2, groEL-up1, groEL-up2, groEL-down1, groEL-down2), in combination with the adapter primers from the Genome Walker Universal kit (AP1 and AP2) (Table 7.1). For the *ftsZ* gene, another round of downstream genome walking was required (ftsZ-down3, ftsZ-down4) (Table 7.1). The longest resulting fragments were gel excised, purified, cloned and sequenced as described in chapter 2. From the resulting coding sequences, GC percentages were calculated with an in-house perl script.

Phylogenetic analyses

The NCBI CoreNucleotide database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore) was searched for *Wolbachia* sequences from different host species. In our analysis we included only *Wolbachia* strains for which sequences of all three genes (*16S rRNA*, *ftsZ*, and *groEL*) were available. This approach ruled out *Wolbachia* strains belonging to supergroups E and G, but did include species of the different supergroups A, B, C, D, F and H. Accession numbers of the sequences used in this study are shown in Table 7.2.

A multiple alignment for the three genes was constructed with ClustalW (Thompson et al., 1994) and manually edited. Two species of the genera most closely related to *Wolbachia* were added to the alignment (*Anaplasma marginale* and *Ehrlichia canis*). The chosen outgroup, *Rickettsia felis*, is like *Wolbachia*, a member of the *Rickettsiaceae*. Congruence between the datasets obtained from the different genes was tested using the partition homogeneity test (Johnson & Soltis, 1998), which is analogous to the incongruence length difference (ILD) test of Farris et al. (1994), as implemented in PAUP* v4.0b10 (Swofford, 2003) using the heuristic search option with random sequence addition (100 random replications) and TBR branch-swapping. Bayesian analyses were conducted with MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003) for each gene separately and for a straightforward concatenated dataset, constructed by combining the alignments of the three genes. The HKY model (Hasegawa et al., 1985) was used with 500,000 generations, and resulting trees were visualised in TreeView 1.6.6 (Page, 1996).

Table 7.2: (Host) species, taxonomic classification, *Wolbachia* supergroup (SG) and GenBank accession numbers of sequences used for phylogenetic analyses of *16S rRNA*, *ftsZ* and *groEL* genes

(Host) species	Phylum	class	order	*16S rRNA*	*ftsZ*	*groEL*	SG
Brugia malayi	Nematoda	Chromadorea	Rhabditida	AF051145	AJ010269	NC006833	D
Brugia pahangi	Nematoda	Chromadorea	Rhabditida	AJ012646	AJ010270	AJ609654	D
Dirofilaria repens	Nematoda	Chromadorea	Rhabditida	AJ276500	AJ010273	AJ609653	C
Litomosoides sigmodontis	Nematoda	Chromadorea	Rhabditida	AF069068	AJ010271	AF409113	D
Mansonella sp.	Nematoda	Chromadorea	Rhabditida	AJ279034	AJ628414	AJ628412	F
Onchocerca gibsoni	Nematoda	Chromadorea	Rhabditida	AJ276499	AJ010267	AJ609652	C
Radopholus similis	Nematoda	Chromadorea	Rhabditida	EU833482	EU833483	EU833484	?
Tribolium confusum	Arthropoda	Insecta	Coleoptera	X65674	U28194	AY714798	B
Aedes albopictus	Arthropoda	Insecta	Diptera	AF397408	DQ842305	DQ243927	A
Drosophila melanogaster	Arthropoda	Insecta	Diptera	DQ235275	DQ235339	DQ235379	A
Drosophila simulans wAu	Arthropoda	Insecta	Diptera	AF390865	DQ235342	DQ235382	A
Drosophila simulans wHa	Arthropoda	Insecta	Diptera	DQ235279	DQ235341	DQ235381	A
Drosophila simulans wMa	Arthropoda	Insecta	Diptera	AF390864	AY508999	AY714799	B
Drosophila simulans wNo	Arthropoda	Insecta	Diptera	DQ235288	DQ266426	AY714800	B
Drosophila simulans wRi	Arthropoda	Insecta	Diptera	DQ235278	U28178	AY714806	A
Laodelphax striatellus	Arthropoda	Insecta	Hemiptera	AB039036	AB039038	DQ356890	B
Sogatella furcifera	Arthropoda	Insecta	Hemiptera	AB039037	AB039039	DQ243918	B
Trichogramma cordubensis	Arthropoda	Insecta	Hymenoptera	X65675	U95749	AY714803	B
Coptotermes acinaciformis	Arthropoda	Insecta	Isoptera	DQ837197	DQ837186	AJ627384	F
Kalotermes flavicollis	Arthropoda	Insecta	Isoptera	Y11377	AJ292345	AJ609660	F
Microcerotermes sp.	Arthropoda	Insecta	Isoptera	AJ292347	AJ292346	AJ628411	F
Zootermopsis angusticollis	Arthropoda	Insecta	Isoptera	AY764279	AY764283	AY764278	H
Zootermopsis nevadensis	Arthropoda	Insecta	Isoptera	AY764280	AY764284	AY764277	H
Ephestia kuehniella	Arthropoda	Insecta	Lepidoptera	X65671	U62126	AB081644	A
Hypolimnas bolina bolina	Arthropoda	Insecta	Lepidoptera	AB052745	AB167352	AB167350	B
Hypolimnas bolina jacintha	Arthropoda	Insecta	Lepidoptera	AB085178	AB167399	AB167398	B
Armadillidium vulgare	Arthropoda	Malacostraca	Isopoda	X65669	U28208	AM087231	B
Chaetophiloscia elongata	Arthropoda	Malacostraca	Isopoda	AJ223241	AJ223246	AM087239	B
Anaplasma marginale	*Proteobacteria*	*Alphaproteobacteria*	Rickettsiales	CP000030	CP000030	CP000030	-
Ehrlichia canis	*Proteobacteria*	*Alphaproteobacteria*	Rickettsiales	CP000107	CP000107	CP000107	-
Rickettsia felis	*Proteobacteria*	*Alphaproteobacteria*	Rickettsiales	CP000053	CP000053	CP000053	-

Detection of *Wolbachia* in individual nematodes and additional populations

Forty single adult female and 40 male nematodes were manually collected in 40 µl buffer (10 mM Tris-HCl pH 8.0, 1 mM DTT, 0.45% Tween-20) and centrifuged for 5 min at 8000 g. For DNA extraction, single nematodes were sonicated on ice twice for 5s (Branson sonifier S-250). Proteinase K was added to a final concentration of 60 ng/µl, whereupon the nematode suspension was incubated for 30 min at 37°C. Proteinase K was then inactivated by a 5 min incubation in a boiling water bath. After centrifugation for 5 min at 8000 g, 5 µl of the supernatant was used for PCR. PCR conditions were as described above, with the exception of a higher number of cycles (n=40) and different primer pairs (Rs-act-F and Rs-act-R to amplify actin as a positive control, ftsZ-F and ftsZ-R to amplify the *ftsZ* gene and groEL-F and groEL-R for the *groEL* gene; Table 7.1). PCR mix without template served as a negative control. An additional population (*R. similis* from Indonesia) and related species (*R.*

arabocoffeae from Vietnam) were checked for the presence of *Wolbachia* by cloning the *ftsZ* gene with primers ftsZ-F2 and ftsZ-R2 (Table 7.1). DNA extraction and PCR conditions were as described above. PCR mix without template served as a negative control.

DAPI staining and immunolocalisation

To visualise genetic material from both the nematode and the endosymbiont, a staining with 4',6-diamidino-2-phenylindole (DAPI) (Invitrogen) was performed. 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 visualised 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 420LP. Images were acquired using NIS-Elements software version 2.10 (http://www.nis-elements.com/).

Immunolocalisation was carried out with rabbit antiserum against purified *Wolbachia* bacteria from *Aedes albopticus* (kindly provided by Claude Nappez, Université de la Méditerranée, Marseille, France). Nematodes were fixed overnight at 4°C in 4% paraformaldehyde in M9 buffer (42 mM Na_2HPO_4, 22 mM KH_2PO_4, 86 mM NaCl, 1 mM $MgSO_4.7H_2O$, pH 7.0) followed by additional fixation for 4h at room temperature. The fixative was removed and replaced by M9 buffer. Nematodes were cut and subsequently washed three times with M9 buffer by centrifuging for 1 min at 3000 g. Incubation in 2 mg/ml proteinase K for 30 min at room temperature permeabilised the nematode sections. A washing step with M9 buffer removed the proteinase K. The nematode pellet was chilled on ice for 15 min, and incubated for 30s on ice in 1 ml of cold methanol. Upon removing the methanol, one ml cold acetone was added and the pellet was incubated for 1 min on ice, then washed with distilled water and M9 buffer. Blocking was done overnight at 4°C with 1% blocking reagent (Roche) dissolved in M9 buffer. After removal of the blocking buffer, nematode fragments were incubated for another 3h at 4°C in blocking buffer to which primary antibodies were added (1:750). The nematodes were then washed with M9 buffer three times and incubated for 10 min in 0.5% blocking buffer at room temperature. The supernatant was removed after centrifuging for 4 min at 500 g, whereupon 0.5% blocking buffer with 1:20 diluted secondary antibody (fluorescein-labeled goat anti-rabbit IgG (H+L), KPL, Guildford, UK) was added. After 3h incubation at 4°C, secondary antibodies were removed by three cycles of washing with M9 buffer and centrifuging for 4 min at 500 g. As a negative control, the same procedure was executed without adding primary antibodies. Nematodes were mounted on a glass slide, embedded in Vectashield (Invitrogen), and viewed with a Nikon Eclipse TE300 epifluorescence microscope equipped with a Biorad Radiance 2000 confocal system.

Fluorescein was excited with a 488 nm Argon ion laser and detected with a photomultiplier tube (PMT) through a 528/50 nm HQ BP filter. To reduce aspecific signal, Kalman averaging (n=3) was applied during acquisition. The Argon laser was simultaneously used to acquire transmission images. The objective lens was as described above. Digital images were obtained with Lasersharp 2000 software and analysed using ImageJ (Abramoff et al., 2004).

Transmission electron microscopy

Female adult nematodes were fixed in Karnovsky solution (2% paraformaldehyde, 2.5% glutaraldehyde in 0.134 M sodium cacodylate buffer, pH 7.2) at 60°C. After 30 min the heads and tails of the nematodes were removed. The fixed nematode pieces were then incubated in Karnovsky solution at 4°C overnight on a rotator. After this primary fixation, nematodes were washed for 8h at room temperature in 0.134 M sodium cacodylate buffer (pH 7.2). The washing solution was renewed three times during this period. Postfixation was done in reduced osmium, a mixture of 1 ml OsO_4 (4%), 3 ml sodium cacodylate (0.134 M) and 66 mg $K_3Fe(CN)_6$ overnight at 4°C on a rotator. The nematode pieces were then washed with distilled water and subsequently dehydrated in a 50%, 70%, 90% and 100% ethanol series at room temperature for three times at 10 min each. The samples were then transferred to absolute alcohol to which $CuSO_4$ bars were added to remove any remaining water. The specimens were then infiltrated with a low-viscosity embedding medium (Spurr, 1969) and polymerised at 70°C for 8h. Ultrathin (70 nm) longitudinal sections were cut on a Reichert Ultracut S Ultramicrotome (Leica, Vienna, Austria) with a diamond knife (Diatome Ltd., Biel, Switzerland). Formvar-coated single slot copper grids were used (Agar Scientific, Stansted, United Kingdom). The sections were poststained with uranyl acetate and lead citrate stain (EM stain, Leica) and visualised with a Jeol JEM 1010 (Jeol Ltd., Tokyo, Japan) transmission electron microscope operating at 60kV. Digital recordings were made with a DITABIS system (Pforzheim, Germany).

Results

Wolbachia sequences in nematode ESTs

A proportion of nematode EST sequences (30,909 out of 903,682) showed significant homology to one of the *Wolbachia* proteins of *Brugia malayi Wolbachia*. After removal of ESTs with the term "*Caenorhabditis*" in the description line (n=17,606) to reduce the number of sequences, the remaining 13,303 ESTs were searched for homology with all known proteins in GenBank (February 2008). The resulting ESTs with a *Wolbachia* sequence as the top hit (n=110) belonged to seven different nematode species as summarised in Table 7.3. Three species with *Wolbachia* ESTs have never been experimentally confirmed to contain

Wolbachia: *Radopholus similis*, *Ancylostoma caninum* and *Pristionchus pacificus*. However the latter two could be false positives as only one putative *Wolbachia* EST was found despite the large amount of available ESTs for these nematodes. The retained EST from *P. pacificus* (AW097859) also showed high homology to a *Caenorhabditis elegans* protein in a blastx search and can therefore be considered as a false positive. However, the EST retained from *A. caninum* (EX545240) only showed significant homology to *Wolbachia* proteins and may indicate the presence of *Wolbachia* in this nematode. Regardless of the low frequencies of *A. caninum* and *P. pacificus* putative *Wolbachia* ESTs, the percentage of *Wolbachia* sequences in the EST libraries of nematode species varies between 0.021% (*W. bancrofti*) and 0.623% (*R. similis*) (Table 7.4).

Cloning of *16S rRNA*, *ftsZ* and *groEL* sequences

A 1494 bp fragment of *16S rRNA* was amplified from the sterile extracted DNA from *R. similis* (population Uganda), using universal bacterial primers. A blastn search revealed a *16S rRNA* sequence of *Wolbachia* isolated from *Drosophila simulans* wRi as the top hit (supergroup A; 95% identity). Using primers based on EST fragments of the *ftsZ* and *groEL* genes, full length coding sequences for both genes could be cloned. The full length coding sequence of the *ftsZ* gene is 1161 bp. Genome walking revealed a 299 bp fragment of another gene, 270 bp upstream of the *ftsZ* gene. Blastx revealed that the most homologous protein is a cytidine triphosphate (CTP) synthetase of *Wolbachia* from *Drosophila melanogaster* (65% identity). Furthermore, the full length coding sequence of the *groEL* gene is 1638 bp. No introns are present in full length open reading frames of either *ftsZ* or *groEL*. GC percentages of the coding sequences of the *16S rRNA*, *ftsZ* and *groEL* genes were respectively 48%, 38% and 37%. Despite efforts using different PCR conditions and nested PCR, amplification of *wsp* did not succeed. Additional *ftsZ* fragments were cloned from a different *R. similis* population (Indonesia), and from the closely related species *R. arabocoffeae* (Vietnam). In both cases a fragment of 369 bp was obtained, which differed in only two nucleotides from the *Wolbachia ftsZ* sequence amplified from the *R. similis* population Uganda. All obtained sequences were submitted to GenBank (accession numbers EU833482, EU833483, EU833484, FJ168559, FJ168560).

Table 7.3: All *Wolbachia*-like ESTs found among nematode ESTs by a local tblastn search with as query all proteins from *Wolbachia* from *Brugia malayi*. From left to right in the table: EST accession number, nematode species name, tblastn E-value, WolBm protein accession number and WolBm protein description

EST nr	species	E-value	WolBm hit	WolBm hit description
AA471650	*Brugia malayi*	1.00E-49	YP_197839	Glycine/serine hydroxymethyltransferase
AA509094	*Brugia malayi*	1.00E-74	YP_197845	Permease of the major facilitator superfamily
AI087737	*Brugia malayi*	2.00E-78	YP_197883	Signal recognition particle GTPase
AW682783	*Brugia malayi*	1.00E-107	YP_197893	30S ribosomal protein S1
BM889394	*Brugia malayi*	6.00E-40	YP_197914	SsrA-binding protein
AA842543	*Brugia malayi*	4.00E-44	YP_197979	Kef-type K+ transport system, membrane component
AW225352	*Brugia malayi*	8.00E-45	YP_197986	Zn-dependent carboxypeptidase
AW257559	*Brugia malayi*	1.00E-83	YP_198002	CTP synthetase
AA114455	*Brugia malayi*	2.00E-13	YP_198088	DNA mismatch repair protein
AA842397	*Brugia malayi*	1.00E-71	YP_198088	DNA mismatch repair protein
AA842686	*Brugia malayi*	4.00E-25	YP_198127	Ankyrin repeat-containing protein
BE758394	*Brugia malayi*	1.00E-63	YP_198156	50S ribosomal protein L18
AA228179	*Brugia malayi*	5.00E-30	YP_198169	50S ribosomal protein L2
AA257567	*Brugia malayi*	3.00E-57	YP_198169	50S ribosomal protein L2
AW041947	*Brugia malayi*	4.00E-32	YP_198170	50S ribosomal protein L23
CB338440	*Brugia malayi*	5.00E-11	YP_198179	Predicted aminomethyltransferase related to GcvT
AA406977	*Brugia malayi*	2.00E-34	YP_198181	chaperonin GroEL
AA842541	*Brugia malayi*	3.00E-72	YP_198217	DNA-directed RNA polymerase subunit omega, RpoZ
AW888279	*Brugia malayi*	1.00E-15	YP_198236	Rieske Fe-S protein
AA088159	*Brugia malayi*	1.00E-78	YP_198256	phosphatidylserine decarboxylase
AA123612	*Brugia malayi*	8.00E-54	YP_198264	DNA polymerase III, gamma/tau subunit
AA180693	*Brugia malayi*	5.00E-06	YP_198268	Protein-disulfide isomerase
AA842713	*Brugia malayi*	2.00E-81	YP_198269	Oligoketide cyclase/lipid transport protein
AA842209	*Brugia malayi*	1.00E-13	YP_198302	Threonyl-tRNA synthetase
AA842102	*Brugia malayi*	4.00E-52	YP_198304	NADH:ubiquinone oxidoreductase, NADH-binding chain F
AW179640	*Brugia malayi*	2.00E-70	YP_198305	Short-chain alcohol dehydrogenase family enzyme
AA406876	*Brugia malayi*	6.00E-51	YP_198357	IMP dehydrogenase, GuaB
BE758486	*Brugia malayi*	3.00E-66	YP_198371	preprotein translocase subunit SecB
AA509095	*Brugia malayi*	2.00E-46	YP_198417	hypothetical protein Wbm0587
AA842374	*Brugia malayi*	2.00E-96	YP_198477	DNA-directed RNA polymerase, RpoB/RpoC
AW225431	*Brugia malayi*	5.00E-30	YP_198477	DNA-directed RNA polymerase, RpoB/RpoC
AA509209	*Brugia malayi*	2.00E-09	YP_198482	Transcription antiterminator
AW043475	*Brugia malayi*	2.00E-37	YP_198523	hypothetical protein Wbm0693
H35903	*Brugia malayi*	9.00E-09	YP_198531	50S ribosomal protein L21
AW179871	*Brugia malayi*	3.00E-10	YP_198534	aspartyl/glutamyl-tRNA amidotransferase subunit B
AA471531	*Brugia malayi*	9.00E-30	YP_198581	hypothetical protein Wbm0751
AA257602	*Brugia malayi*	7.00E-34	YP_198594	DNA gyrase, topoisomerase II, B subunit, GyrB
AA471424	*Brugia malayi*	2.00E-18	YP_198601	hypothetical protein Wbm0771
BE758437	*Brugia malayi*	7.00E-50	YP_198605	Cytochrome b subunit of the bc complex
AI815321	*Onchocerca volvulus*	3.00E-52	YP_197836	Aspartate aminotransferase family enzyme
BE636257	*Onchocerca volvulus*	3.00E-04	YP_197839	Glycine/serine hydroxymethyltransferase
AI066948	*Onchocerca volvulus*	2.00E-35	YP_197842	Predicted 3'-5' exonuclease
AI438198	*Onchocerca volvulus*	2.00E-03	YP_197913	prolipoprotein diacylglyceryl transferase
AW739479	*Onchocerca volvulus*	5.00E-22	YP_198013	hypothetical protein Wbm0180
AW610590	*Onchocerca volvulus*	2.00E-29	YP_198014	ATP-binding subunit of Clp protease
AW626403	*Onchocerca volvulus*	8.00E-03	YP_198014	ATP-binding subunit of Clp protease
BE230823	*Onchocerca volvulus*	9.00E-29	YP_198014	ATP-binding subunit of Clp protease
AW330455	*Onchocerca volvulus*	2.00E-88	YP_198181	chaperonin GroEL
AW288295	*Onchocerca volvulus*	5.00E-05	YP_198194	2-methylthioadenine synthetase
AW330443	*Onchocerca volvulus*	6.00E-04	YP_198241	AICAR transformylase/IMP cyclohydrolase PurH
AA618711	*Onchocerca volvulus*	7.00E-35	YP_198257	recombinase A
AW330450	*Onchocerca volvulus*	8.00E-53	YP_198278	succinate dehydrogenase flavoprotein subunit
AW499039	*Onchocerca volvulus*	6.00E-04	YP_198329	DNA polymerase III alpha subunit
AI313874	*Onchocerca volvulus*	2.00E-14	YP_198358	transcription termination factor Rho
AI205420	*Onchocerca volvulus*	4.00E-21	YP_198367	translation initiation factor IF-2
BE230839	*Onchocerca volvulus*	2.00E-53	YP_198507	spermidine/putrescine transport system, ATPase comp
BQ455744	*Dirofilaria immitis*	7.00E-22	YP_197889	Predicted membrane-associated Zn-dependent protease
BQ456662	*Dirofilaria immitis*	3.00E-09	YP_198389	Dihydrolipoamide dehydrogenase E3 component
BQ455897	*Dirofilaria immitis*	9.00E-80	YP_198427	Glycerol-3-phosphate dehydrogenase
BQ455358	*Dirofilaria immitis*	4.00E-42	YP_198496	pyruvate dehydrogenase subunit beta
BQ457020	*Dirofilaria immitis*	3.00E-43	YP_198496	pyruvate dehydrogenase subunit beta
CD374832	*Wuchereria bancrofti*	1.00E-125	YP_198169	50S ribosomal protein L2
EX545240	*Ancylostoma caninum*	1.00E-09	YP_198620	orotate phosphoribosyltransferase
AW097859	*Pristionchus pacificus*	1.00E-11	YP_197866	tRNA modification GTPase TrmE
EY191273	*Radopholus similis*	1.00E-43	YP_197853	polynucleotide phosphorylase/polyadenylase
EY193019	*Radopholus similis*	3.00E-34	YP_197896	hypothetical protein Wbm0062
EY193107	*Radopholus similis*	4.00E-11	YP_198014	ATP-binding subunit of Clp protease
EY193983	*Radopholus similis*	6.00E-15	YP_198026	hypothetical protein Wbm0193
EY193015	*Radopholus similis*	6.00E-51	YP_198075	Malate dehydrogenase
EY195691	*Radopholus similis*	9.00E-32	YP_198111	Type IV secretory pathway, component VirB9
EY192932	*Radopholus similis*	7.00E-40	YP_198129	Glyceraldehyde-3-phosphate dehydrogenase, GapA
EY194312	*Radopholus similis*	2.00E-27	YP_198163	30S ribosomal protein S17
EY192361	*Radopholus similis*	3.00E-38	YP_198165	50S ribosomal protein L16
EY190439	*Radopholus similis*	2.00E-77	YP_198166	30S ribosomal protein S3
EY190640	*Radopholus similis*	3.00E-38	YP_198166	30S ribosomal protein S3
EY193646	*Radopholus similis*	2.00E-42	YP_198172	50S ribosomal protein L3
EY192020	*Radopholus similis*	2.00E-18	YP_198177	30S ribosomal protein S12
EY191545	*Radopholus similis*	2.00E-46	YP_198181	chaperonin GroEL
EY195553	*Radopholus similis*	1.00E-80	YP_198181	chaperonin GroEL
EY189814	*Radopholus similis*	3.00E-32	YP_198223	DNA-directed RNA polymerase sigma 70 subunit, RpoD
EY190120	*Radopholus similis*	1.00E-23	YP_198223	DNA-directed RNA polymerase sigma 70 subunit, RpoD
EY195600	*Radopholus similis*	1.00E-03	YP_198269	Oligoketide cyclase/lipid transport protein
EY192071	*Radopholus similis*	7.00E-66	YP_198273	bifunctional GMP synthase/glutamine amidotransferase
EY193596	*Radopholus similis*	3.00E-14	YP_198287	F0F1-type ATP synthase, subunit b
EY190008	*Radopholus similis*	1.00E-03	YP_198314	hypothetical protein Wbm0484
EY190340	*Radopholus similis*	8.00E-04	YP_198314	hypothetical protein Wbm0484
EY194910	*Radopholus similis*	8.00E-04	YP_198314	hypothetical protein Wbm0484
EY192068	*Radopholus similis*	1.00E-20	YP_198320	Integral membrane protein, interacts with FtsH
EY194703	*Radopholus similis*	7.00E-14	YP_198332	50S ribosomal protein L9
EY192659	*Radopholus similis*	6.00E-27	YP_198342	carbamoyl phosphate synthase large subunit
EY190670	*Radopholus similis*	8.00E-24	YP_198363	Molecular chaperone GrpE (heat shock protein)
EY189806	*Radopholus similis*	1.00E-48	YP_198368	transcription elongation factor NusA
EY190112	*Radopholus similis*	1.00E-48	YP_198368	transcription elongation factor NusA
EY193335	*Radopholus similis*	1.00E-48	YP_198368	transcription elongation factor NusA
EY194109	*Radopholus similis*	1.00E-38	YP_198381	ATP-dependent Lon protease
EY191616	*Radopholus similis*	6.00E-56	YP_198383	Protease subunit of ATP-dependent Clp protease
EY191902	*Radopholus similis*	5.00E-55	YP_198383	Protease subunit of ATP-dependent Clp protease
EY191828	*Radopholus similis*	6.00E-22	YP_198432	cell division protein FtsZ
EY192572	*Radopholus similis*	2.00E-25	YP_198432	cell division protein FtsZ
EY192973	*Radopholus similis*	5.00E-03	YP_198432	cell division protein FtsZ
EY193345	*Radopholus similis*	1.00E-31	YP_198432	cell division protein FtsZ
EY193624	*Radopholus similis*	3.00E-53	YP_198477	DNA-directed RNA polymerase, RpoB/RpoC
EY195901	*Radopholus similis*	2.00E-60	YP_198477	DNA-directed RNA polymerase, RpoB/RpoC
EY195545	*Radopholus similis*	3.00E-44	YP_198487	50S ribosomal protein L20
EY193639	*Radopholus similis*	2.00E-03	YP_198509	ABC-type transport system, permease component
EY195762	*Radopholus similis*	5.00E-70	YP_198569	3-oxoacyl-(acyl-carrier-protein) synthase
EY191572	*Radopholus similis*	5.00E-13	YP_198581	hypothetical protein Wbm0751
EY189968	*Radopholus similis*	9.00E-10	YP_198623	Type IV secretory pathway, VirB6 components
EY190296	*Radopholus similis*	4.00E-12	YP_198623	Type IV secretory pathway, VirB6 components
EY193937	*Radopholus similis*	4.00E-12	YP_198623	Type IV secretory pathway, VirB6 components

Table 7.4: Overview of the number of *Wolbachia* ESTs found among all nematode ESTs. For each species is given its lifestyle, its total number of ESTs in the GenBank database (december 2007), whether or not *Wolbachia* has been described in literature, the number of *Wolbachia* ESTs found, and the percentage of *Wolbachia* ESTs in relation to the total amount of ESTs (december 2007). APN: animal parasitic nematode; PPN: plant-parasitic nematode; FLN: free-living nematode

Species	**lifestyle**	**total ESTs**	***Wolbachia*?**	***Wolbachia* ESTs**	**% *Wolbachia* ESTs**
Ancylostoma caninum	APN	46,965		1	0.0021
Brugia malayi	APN	26,215	yes	39	0.1488
Onchocerca volvulus	APN	14,974	yes	17	0.1135
Wuchereria bancrofti	APN	4847	yes	1	0.0206
Dirofilaria immitis	APN	4005	yes	5	0.1248
Litomosoides sigmodontis	APN	2699	yes	-	
Angiostrongylus cantonensis	APN	1279		-	
Brugia pahangi	APN	28	yes	-	
Pristionchus pacificus	FLN	14,663		1?	0.0068
Radopholus similis	PPN	7380		46	0.6233
all other PPN species	PPN	157,814		-	

Phylogenetic analyses

16S rRNA, *ftsZ* and *groEL* genes from *Wolbachia* belonging to different supergroups (Table 7.2) were aligned with the corresponding sequences isolated from *R. similis Wolbachia* and used for Bayesian analyses. Congruency between the three *Wolbachia* gene datasets was statistically confirmed by a partition homogeneity test ($p>0.01$). All constructed phylogenetic trees had indeed the same topology and therefore only the tree of the combined dataset is shown (Figure 7.1). The supergroups A, B, C, D, F and H are clustered with high posterior probability values. The *Wolbachia*-like symbiont from *R. similis* does not appear to belong to any of the known supergroups. Extra analyses with the limited data available for supergroups E (springtails), G (spiders) and the unclassified species *Ctenocephalides felis* and *Dipetalonema gracile* also did not show any relatedness to these supergroups or species (data not shown). However it does form a monophyletic group with all other *Wolbachia* sequences with a high posterior probability value (1.00). Nevertheless, it remains difficult to say to which supergroup the *Wolbachia* of *R. similis* is most closely related.

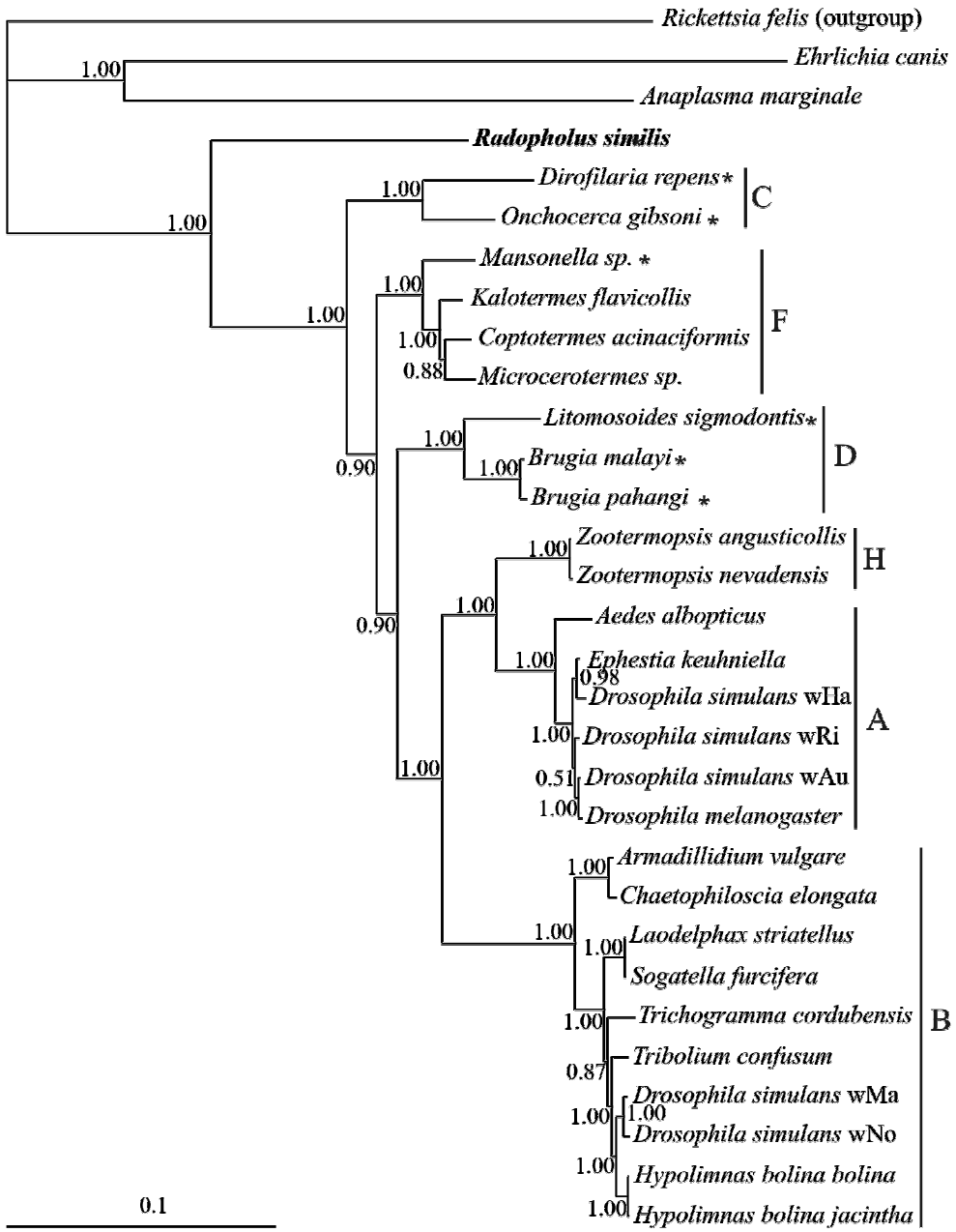

Figure 7.1: Phylogenetic tree inferred from *16S rRNA*, *ftsZ* and *groEL* gene data with Bayesian statistics (500,000 generations). All *Wolbachia* taxa are represented by their host species names. Nematode hosts are marked with asterisks. Posterior probabilities are shown at branching points. *Wolbachia* supergroups A, B, C, D, F and H are indicated.

The presence of *Wolbachia* in single nematodes

Forty adult females and 40 adult males were tested for the presence of the *Wolbachia*-like symbiont. A positive control PCR was performed with actin primers and a correct fragment was amplified from 28 females and 29 males. No amplification was obtained in the negative control. The actin positive samples were subsequently tested for the presence of the symbiont, using both *ftsZ* and *groEL* primers. The PCRs resulted in amplification products of the expected lengths for all 28 females and 29 males, whereas no amplification product was

detected in the negative controls. From these results we conclude that likely all *R. similis* individuals are infected, as is the case in animal-parasitic species.

Visualisation of the endosymbiont

DAPI staining showed strong signals for the different cell nuclei of the nematode. Around the oocyte nuclei in the ovaries, tiny fluorescent dots were observed, which may be of endosymbiotic origin (Figure 7.2).

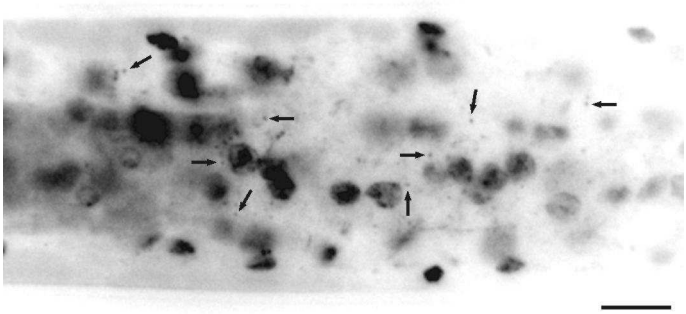

Figure 7.2: Female adult nematode stained with DAPI, at the level of an ovary. Nuclei of the ovary cells show intense staining. Small dots (arrows) indicate genetic material of endosymbionts. Scale bar: 5 µm.

Likewise, immunolocalisation using polyclonal antibodies against *Wolbachia* produced strongly stained particles in the female ovaries, both inside and surrounding the oocytes (Figure 7.3). In some cases a signal was observed on the cuticle. Since the antibodies used are polyclonal and were raised against whole *Wolbachia* cells, these antibodies are not fully *Wolbachia* specific and may also recognise other bacteria. The signal on the cuticle is most probably derived from external bacteria that are attached to the surface of the nematode.

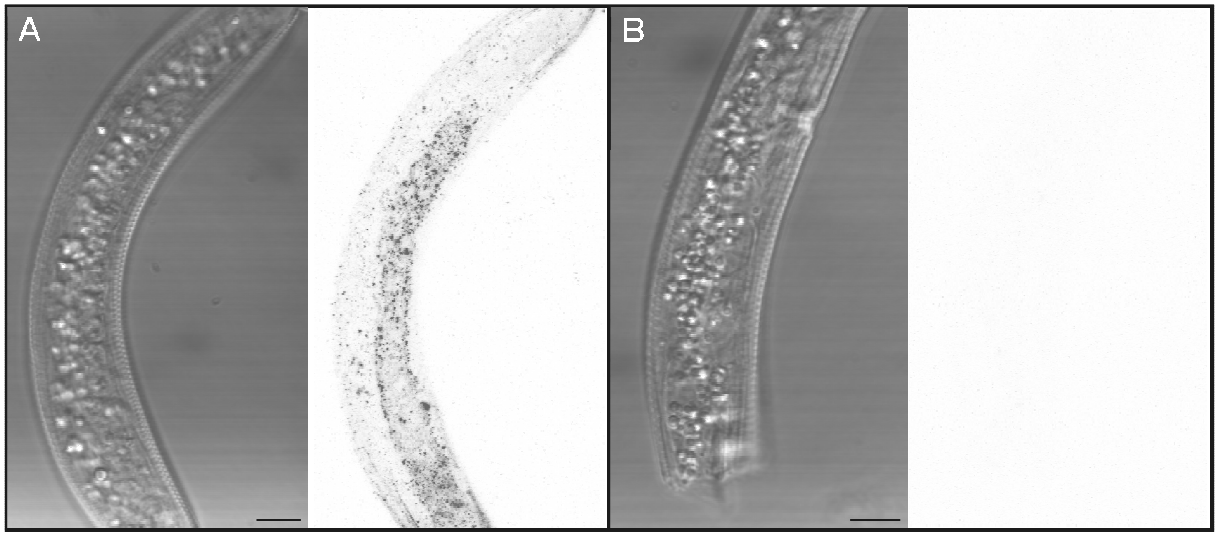

Figure 7.3: Female adult nematodes subjected to immunolocalisation with polyclonal antibodies against *Wolbachia*. (A) nematodes treated with both primary and secondary antibody; (B) negative control (only secondary antibody). Left panel: transmitted light; right panel: epifluorescence. Scale bars: 10 µm.

Transmission electron microscopy of female adults confirmed the presence of endosymbiotic bacteria in the ovary (Figure 7.4). These bacteria had a large variety of shapes, ranging from rod-shaped to round, U-shaped, and even club-shaped. The cross-sections of the observed endosymbionts ranged from approximately 350 to 900 nm. In many cases, the three membranes could be distinguished, two of bacterial origin and one of host origin. Inside the bacterial cell, ribosomes could be seen which typically appear smaller than ribosomes in the surrounding host tissue. Next to the ovary, bacterial cells inhabit the uterus, albeit in lower numbers.

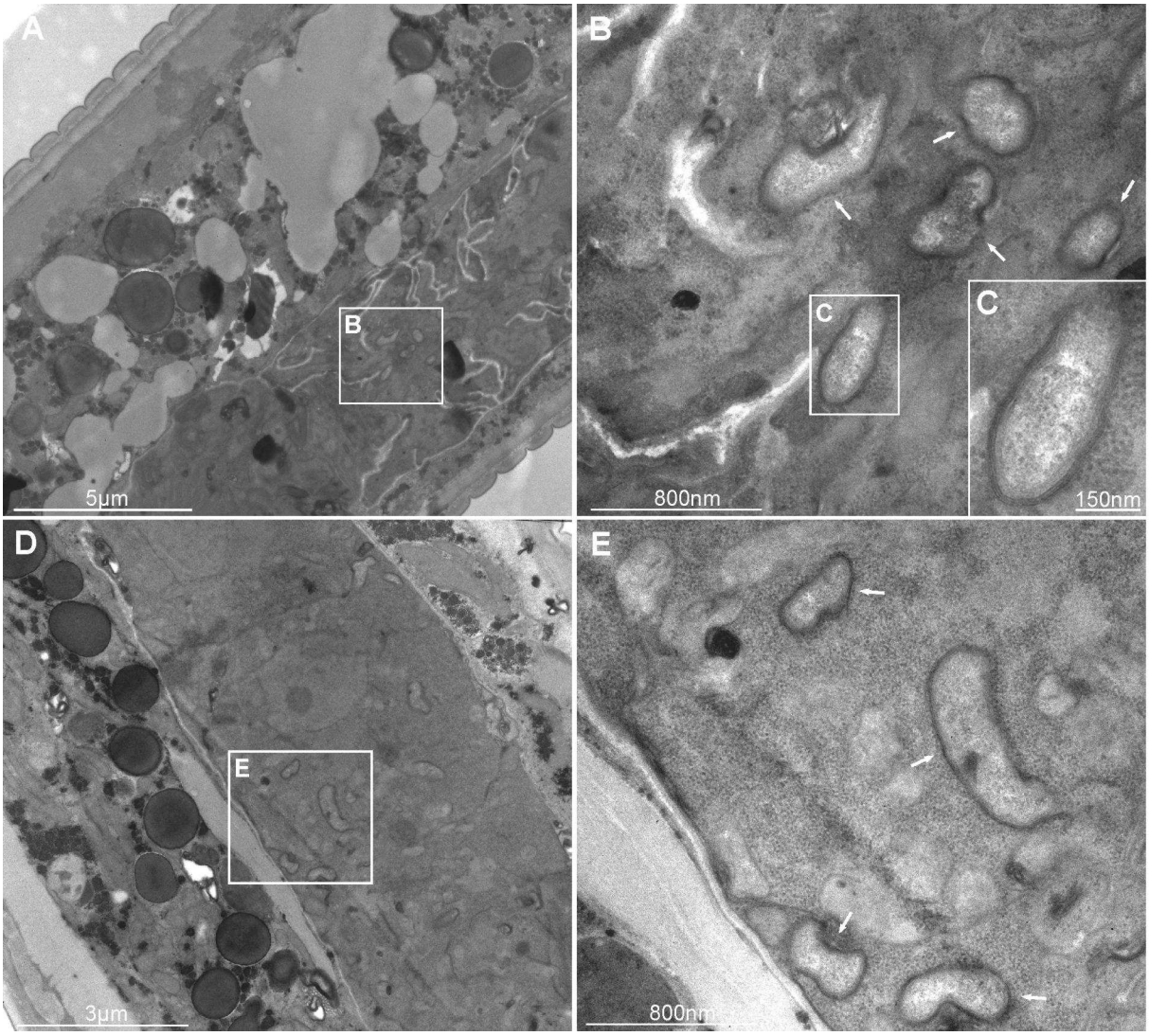

Figure 7.4: Transmission electron microscopy sections of female adult nematodes. B and C are detailed pictures of A in the uterus (organ characterised by numerous invaginations), while E is a detail of D in the ovary. Endosymbionts are indicated with white arrows. cu: cuticle; hd: hypodermis; sm: somatic musculature; int: intestine; ut: uterus; ov: ovary; oo: oocyte; nu: nucleus; mt: mitochondria.

Discussion

We investigated the presence of a *Wolbachia*-like bacterium in the plant-parasitic nematode *Radopholus similis*. Despite past efforts (Bordenstein et al., 2003) we believe this is the first discovery of *Wolbachia* in a plant-parasitic nematode. A first indication came from a comprehensive analysis of an EST dataset derived from a mixed-stage population of *Radopholus similis* (Jacob et al., 2008), revealing a subset of sequences (30 unigenes composed of 38 ESTs) with high similarity to *Wolbachia* proteins. Extending this *in silico* approach by exploring all nematode ESTs for putative *Wolbachia* material revealed seven nematode species with EST sequences reminiscent of *Wolbachia*. For four of those, the presence of *Wolbachia* was previously described in the literature, which indicates that our approach was adequate in retaining sequences derived from *Wolbachia*-infected nematode species. Although the presence of *Wolbachia* was described in other nematode species as well (*Brugia pahangi* and *Litomosoides sigmodontis*) this could not be confirmed by our EST screen. It should however be noted that for these species the number of ESTs in the database is relatively low (less than 3000). For three nematode species, candidate *Wolbachia* tags were found whereas the endosymbiont itself has not yet been described. For two of these species only one tag was found among 14,663 (*Pristionchus pacificus*) or 46,965 (*Ancylostoma caninum*) ESTs. The EST from *P. pacificus* is probably a false positive since it also showed high homology to a *C. elegans* gene. Nevertheless, the EST from the *A. caninum* library showed high homology to *Wolbachia* proteins exclusively. In contrast to these few hits, for *R. similis* 46 putative *Wolbachia* sequences among 7380 ESTs were retained, adding eight ESTs to the sequences identified by Jacob et al. (2008) due to a higher E-value cut-off. The number of putative *Wolbachia* sequences found is too high to be considered as false positives. Moreover, contamination of the cDNA library is highly unlikely since our laboratory has never worked with insects or filarial nematodes, so contaminating *Wolbachia*-containing organisms have never been present. Since the construction of most cDNA libraries is based on polyA tails, which are absent in bacterial transcripts, one would not expect bacterial sequences in EST libraries. This implies that the identified ESTs are not necessarily of bacterial origin, but could be derived from the nematode genome itself. It was shown that large pieces of the symbiont genome can be integrated into the host genome without much adaptation, and that the transferred genes can be transcriptionally active (Dunning-Hotopp et al., 2007). *Wolbachia*-like ESTs in the cDNA libraries are therefore not conclusive evidence of the presence of the symbiont itself. However, we were able to confirm the presence of an endosymbiont in *R. similis* by using different visualisation methods. DAPI staining and immunolocalisation showed signs of endosymbionts in the ovaries of adult females. Transmission electron microscopy confirmed the presence of endosymbionts surrounded by three membranes in the ovaries and in lower numbers in the uterus. This occurrence pattern is

similar to that in filarial nematodes (Kramer et al., 2003), yet in our case no bacteria were found in the hypodermis. The endosymbionts could be clearly distinguished from other cell organelles such as mitochondria.

For different plant-parasitic nematodes, such as *Heterodera glycines*, *Globodera rostochiensis*, *Bursaphelenchus xylophilus*, and several *Meloidogyne* spp., a substantial number of ESTs are available (over 10,000 per species). The fact that our approach could not identify *Wolbachia*-like sequences here most likely suggests that there is no *Wolbachia* present in these species. However it cannot be ruled out that other strains of these species do contain *Wolbachia*. Nonetheless, the occurrence of *Wolbachia* in plant-parasitic nematodes seems less widespread than in filarial nematodes.

Sequence analysis of the *ftsZ*, *groEL* and *16S rRNA* genes of the *R. similis* endosymbiont confirmed their similarity to *Wolbachia* genes from other hosts. However, the assumption that these genes are of bacterial origin should be made with caution because it cannot be ruled out that *Wolbachia*-like sequences have been integrated into the nematode host genome. An indication that these genes are indeed of bacterial origin is the average GC content (41%) which is lower than the average GC content of *R. similis* coding sequences (54%) (Jacob et al., 2008). Moreover, the transcriptionally active genes (because those are present in ESTs) lack introns and have only a short untranslated region (UTR). The relatively high AT content could be one reason why bacterial transcripts emerged from the nematode cDNA library, since the oligo(dT) primer can bind to regions rich in adenine in addition to polyA tails of eukaryotic transcripts. It should however be noted that a recent horizontal transfer to the nematode genome would not yet reveal many eukaryotic features or changes in GC content. This was shown for *Drosophila ananassae*, where more than 90% nucleotide identity was found between the transferred fragment and the corresponding endosymbiont genes (Dunning-Hotopp et al., 2007). Nonetheless, with compelling microscopic evidence in support, it is clear that we are dealing with a full bacterial genome indeed. After all, if the observed endosymbiont was not the origin of the sequences and was therefore not *Wolbachia*, we would expect other *16S rRNA* sequences to be amplified with the universal primers.

Phylogenetic analyses of the cloned sequences showed that the *Wolbachia* from *R. similis* does not cluster with one of the known supergroups. Independent analysis of the three genes confirmed its distant relationship to the other supergroups. Yet it still forms a monophyletic cluster with the other *Wolbachia* sequences, and shows less similarity to the closest known *Wolbachia* relatives *Ehrlicha* and *Anaplasma*. The relationships among the different supergroups in the phylogenetic tree should be handled with caution. It was recently shown that phylogenetic artifacts such as long branch attraction and the limitation of sequence models can lead to erroneous but highly supported tree reconstructions (Bordenstein et al., 2009). Nevertheless, it remains clear that the sequences from the *R. similis Wolbachia* are distantly related to sequences from the other supergroups. Based on the phylogenetic analysis,

and the fact that a new host type (a plant-parasitic nematode) is involved, we propose the assignment of the endosymbiont of *R. similis* to a new *Wolbachia* supergroup, namely supergroup I. Baldo et al. (2006) suggest that enough sequence information must be available to assign a *Wolbachia* strain to a new supergroup. Based on this constriction, Baldo et al. (2007) plead for the removal of supergroup G since it has only been based on the highly recombinant *wsp* gene and they propose a multilocus sequence typing (MLST) system to properly characterise *Wolbachia* strains based on five conserved genes. Despite the fact that only one of these (*ftsZ*) was found among the *Wolbachia* ESTs, we feel confident that in our specific case enough evidence is supplied to justify the establishment of a new supergroup, mainly because our phylogeny is based on three genes: *16S rRNA*, *groEL* and *ftsZ*. Nevertheless, it would be interesting to sequence the other genes of the MLST system for future comparisons. One might still argue that the identified endosymbiont is a different bacterial species closely related to *Wolbachia*. The phylogenetic analysis shows that the cloned sequences are quite distantly related to the other *Wolbachia* sequences, with *16S rRNA* similarities ranging from 93 to 97%. As a general rule in microbiology, a strain is considered as a new species when *16S rRNA* sequence similarity to its nearest neighbour is less than 97% (Janda & Abbott, 2007). Therefore the *Wolbachia* from *R. similis* could be considered as a different species. However, this 97% rule is not followed by the *Wolbachia* community due to unresolved relationships among the different supergroups. Recently, it has been suggested that the endosymbionts of filarial nematodes should be considered as a separate species (Pfarr et al., 2007). Nevertheless, even under the 97% criterion the question remains whether the *Wolbachia* strain of *R. similis* could be a member of the filarial nematode *Wolbachia* species group; strictly speaking, it should be considered as an independent *Wolbachia* species. Attempts to clone the *Wolbachia* surface protein (*wsp*) failed, probably due to the high sequence divergence as opposed to other *Wolbachia* strains. This strengthens the idea that the use of the *wsp* gene is inappropriate for detection and phylogeny of *Wolbachia* as it is highly recombinant (Baldo & Werren, 2007).

Despite the apparent sequence divergence between *Wolbachia* strains, it is plausible that the strains have a common ancestor. Within the phylum Nematoda, animal parasitism and plant parasitism have originated independently multiple times (Blaxter et al., 1998). Both the filarial nematodes and some plant-parasitic nematodes including *R. similis* form a monophyletic clade, suggesting both groups had a common ancestor at some point in history (Holterman et al., 2006). If this ancestral nematode was infected with a *Wolbachia*-like endosymbiont, it is possible that this strain managed to survive in some species during evolution whereas it was lost in other nematodes. Perhaps the presence of *Wolbachia* is just an evolutionary relic from the common ancestor of plant-parasitic and animal-parasitic nematodes. In this respect, *Wolbachia* would have started to play an important role in animal-parasitic nematodes during evolution, whereas in plant-parasitic nematodes it became less

important, accounting for its merely sporadic existence in only a few current species. An alternative hypothesis states that *Wolbachia* strains of filarial and plant-parasitic nematodes were acquired independently.

The role of *Wolbachia* in *R. similis* remains unknown. In *Drosophila melanogaster*, it was recently shown that *Wolbachia* makes its host resistant to RNA viruses (Teixeira et al., 2008; Hedges et al., 2008). A similar effect on *R. similis* would result in a clear evolutionary benefit for the nematodes harbouring *Wolbachia*. The high infection rate indicates that *R. similis* cannot survive without its endosymbiont, which is the case for all infected filarial nematodes. Additionally, it is present in another population and the closely related nematode species *R. arabocoffeae*. It is possible that *Wolbachia* provides its host with essential metabolites, as described for *Brugia malayi* (Foster et al., 2005). Whether it has any direct effect at all on reproduction could not be demonstrated; such effects have been reported for arthropods only. Although parthenogenesis and self-fertilisation have been described for *R. similis* (Brooks & Perry, 1962; Kaplan & Opperman, 2000), it is difficult to investigate whether *Wolbachia* plays a role in these processes. Nevertheless, the presence of *Wolbachia* in a plant-parasitic nematode remains intriguing and further research is needed to determine its role and its impact on the nematode host. Moreover, if the function of the endosymbiont could be elucidated, it might open new perspectives in the control of *R. similis*.

Acknowledgements

We thank Claude Nappez and Prof. Didier Raoult (Unité des Rickettsies, Faculté de Médecine, Université de la Méditerranée, Marseille, France) for kindly providing the polyclonal antibodies against *Wolbachia*. We would also like to thank Dr. Annemie Elsen (K.U. Leuven, Belgium) and Prof. Maurice Moens (ILVO, Merelbeke, Belgium) for providing *R. similis* from Indonesia and *R. arabocoffeae* from Vietnam respectively.

CHAPTER 8

GENERAL DISCUSSION AND PERSPECTIVES

Introduction

The understanding of the genetically defined mechanisms by which plant pathogens can invade their hosts is crucial to modern plant pathology. Several genomes of plant pathogenic bacteria are available, and recently, the molecular knowledge of plant-parasitic nematodes has exploded as well. In the genome of the root-knot nematode *Meloidogyne incognita*, 61 plant cell wall degrading enzymes have been found (gene families of cellulase, xylanase, polygalacturonase, pectate lyase, arabinase and invertase), and an additional 20 candidate expansins (Abad et al., 2008). Remarkably, it was recently proposed that endoparasitism among plant-parasitic nematodes within the Tylenchomorpha evolved at least ten times out of ectoparasitism: six times resulting in migratory endoparasitism and four times in sedentary endoparasitism. Only in one case, (*Meloidogyne* spp.), sedentary endoparasitism appears to have evolved directly from migratory endoparasitism (Bert et al., 2008; Holterman et al., 2009). These apparent separate evolutionary paths suggest that differences in the arsenal of parasitism genes probably exist between different nematodes with comparable lifestyles. For example, the presence of an arabinogalactan endo-1,4-β-galactosidase was demonstrated in cyst nematodes of the genus *Heterodera*, while the *Meloidogyne* genomes lack this enzyme (Vanholme et al., 2009). Within the same genus, the number of parasitism genes can differ. In the genome of *M. hapla* for example, 22 pectate lyase genes were found (Opperman et al., 2008) while there are 30 in *M. incognita* (Abad et al., 2008). A similar trend was observed in the genomes of plant pathogenic bacteria, where strikingly different numbers and combinations of genes encoding cell wall degrading enzymes occur (Van Sluys et al., 2002). These genus- or species-specific adaptations towards plant parasitism are essential to understand the evolution of plant parasitism. Since most data on parasitism genes in plant-parasitic nematodes was available for sedentary nematodes, in this thesis migratory nematodes were investigated for the presence of parasitism genes.

Identification of cell wall modifying enzymes in plant-parasitic nematodes

Different approaches exist to identify homologous or novel cell wall modifying enzymes in additional nematode species. A first method to identify specific genes is by PCR using degenerate primers based on conserved regions within a gene. This technique has some shortcomings. First of all, a sufficient amount of sequence information has to be available for a certain gene in multiple nematodes, preferably in species closely related to the investigated species. Secondly, conserved regions of adequate size to design primers must be present. Thirdly, the discovery of novel genes is precluded since the technique is PCR-based. Despite its limitations, the use of degenerate primers has proven to be successful to clone cell wall

modifying protein homologs in migratory nematodes. Using degenerate primers designed for GHF5 endoglucanases by Rosso et al. (1999), multiple endoglucanases could be cloned from *Radopholus similis*, *Ditylenchus africanus* and *Pratylenchus coffeae* (chapter 2, 3 and 4). A similar approach identified an expansin-like protein in *D. africanus*, although this failed to detect any homologs in *R. similis*. Other attempts included the design of degenerate primers against GHF16 endo-1,3-beta-glucanases, first identified in the facultative fungal feeder *Bursaphelenchus xylophilus* and thought to originate from horizontal gene transfer from fungi (Kikuchi et al., 2005). Since *D. africanus* is also able to feed on fungi, this gene could be present in this nematode as well. However, no evidence was found for endo-1,3-beta-glucanases in *D. africanus*. This can either be due to the limitations of the primers, or to the genus- or family-specific nature of the gene. The latter explanation would fit in the finding that fungal feeding evolved at least twice within nematodes: in the Aphelenchoididae (comprising *B. xylophilus*) and in the other Tylenchomorpha (comprising *D. africanus*) (Bert et al., 2008).

A more extensive approach to identify parasitism genes is an expressed sequence tag (EST) analysis. During this thesis, EST analyses were conducted for *R. similis* (Jacob et al., 2008) and *D. africanus* (chapter 6). Surprisingly, only a very limited number of cell wall modifying enzymes were identified in these EST datasets. In the *R. similis* ESTs, one EST similar to endoglucanase and one similar to endoxylanase (chapter 5) were identified in 7007 ESTs (Jacob et al., 2008). In the case of *D. africanus*, two unigenes (consisting of 16 ESTs) similar to expansin-like proteins were found among a total of 4847 ESTs (chapter 6). A similar trend was observed in a *Pratylenchus penetrans* (1928 ESTs) (Mitreva et al., 2004) and a *Xiphinema index* (1400 ESTs) dataset (Furlanetto et al., 2005), both containing no ESTs derived from cell wall modifying enzymes. In the sedentary nematode *M. incognita* however, over 80 ESTs similar to endoglucanase and pectate lyase were found among a total of 5700 (McCarter et al., 2003). This could suggest that migratory plant-parasitic nematodes have a lower expression of genes encoding cell wall modifying enzymes than sedentary nematodes or that they possess fewer copies of these genes. One might think that cell wall modifying enzymes are even more important in migratory nematodes than in sedentary ones due to their constantly migrating lifestyle, but this assumption is certainly not reflected in current EST datasets. It should however be noted that for all these datasets derived from migratory nematodes, the number of ESTs generated is limited, and more tags resembling cell wall modifying enzymes could be expected when more ESTs become available. This is illustrated in a larger study including 13,000 ESTs of *Bursaphelenchus xylophilus*, of which 26 are similar to cellulase and 9 to pectate lyase (Kikuchi et al., 2007). Nevertheless, it remains striking that only a very small part of the transcriptome seems to be involved in the invasion of the host plant. One potential explanation is that the corresponding proteins may possibly be quite stable, which would not require large amounts of transcripts. Analyses on the

biochemical properties of some nematode endoglucanases have shown that these are effective in a large spectrum of temperature and pH conditions (Bera-Maillet et al., 2000; Gao et al., 2004b). The apparent low amount of cell wall modifying enzymes secreted by the nematode could also be attributed to unknown additional factors that facilitate the migration through the plant root. In sedentary nematodes for example, it was demonstrated that these could increase the expression of different host plant cell wall modifying enzymes during feeding site development (Gheysen & Fenoll, 2002). It is unknown whether migratory nematodes exert a similar action, but it could explain why they secrete only a limited amount of cell wall modifying enzymes.

A personalised enzyme cocktail

Interestingly, it appears that every nematode species or genus has evolved its own arsenal of cell wall modifying enzymes adapted to the needs of the organism. Most cell wall modifying enzymes exist in multigene families, as was demonstrated for the endoglucanases of *R. similis* and *D. africanus* in chapter 2 and 3. Moreover, the endoglucanases of *R. similis* revealed a diverse expression pattern among life stages, suggesting functional specialisation. For example, males are known to be non-parasitic and consequently have a lower endoglucanase expression level. In *D. africanus*, there is no known difference in parasitic ability of males and females, hence the endoglucanase expression level is the same in males and females. Furthermore, in *R. similis*, three out of four endoglucanases as well as the endoxylanase show very little expression in juveniles compared to adults while the endoglucanase and expansin-like gene of *D. africanus* are equally expressed in juveniles and adults. There is no clear explanation for this observation, although it is likely that these expression patterns correlate with a slightly different lifestyle of juveniles of both nematode species. It definitely illustrates the plasticity of the genes to adapt to the nematode's needs, and reflects the complexity of the evolution of this gene family. Additionally, it seems that genes can acquire domains from other genes through domain shuffling. In the identified expansin-like protein of *D. africanus*, a carbohydrate binding module (CBM) is present with a high similarity to the CBM of an endoglucanase. Moreover, this domain order is so far unique among nematode ESTs, suggesting that this domain shuffling event happened quite recently, possibly at family or genus level. These kind of domain shuffling events lead to the formation of new genes that can extend the function of existing proteins.

Carbohydrate binding modules (CBMs)

A constant finding during this thesis is that most cell wall modifying enzymes include a CBM. In the putative endoxylanase of *R. similis*, a possible new type of CBM not identified in eukaryotes yet is present. A closer look into the *Meloidogyne* genome sequences showed that putative endoxylanases are also associated with the same type of CBM. Most endoglucanases include a CBM as well, as is the case for the expansin-like protein of *D. africanus*. This high occurrence of CBMs suggests an important and perhaps underestimated function for these domains. The expression by *D. africanus* of endoglucanase pseudogenes with an interrupted catalytic domain but an intact CBM confirms the possible functional importance of the CBM. In bacteria, cell wall degrading enzymes act in multi-enzyme high molecular weight complexes, called cellulosomes. These contain different subunits including CBMs to facilitate the anchoring to the plant cell (Bayer et al., 2004). Whether or not different proteins of plant-parasitic nematode are organised in a similar complex is unknown, but if this would be the case, proteins consisting of a CBM only would definitely be useful in such a kind of complex. Another possible function of CBMs can be to improve the access of other cell wall modifying enzymes to cell wall polymers. Cellulose binding proteins, that have a similar domain architecture, activate plant pectin methylesterases (Hewezi et al., 2008). These reduce the level of methylesterification of pectin. CBMs could exert a similar action, improving the efficiency of other cell wall modifying enzymes. On the other hand, Dumas et al. (2008) suggest that CBMs of the oomycete *Phytophthora* may also act as an elicitor of defense responses in plants. It is not yet clear if this is a general phenomenon of CBMs released by plant pathogens and to what extent nematode CBMs induce, either directly or indirectly through cell wall modifications, defense responses. More research is required to fully understand the importance of CBMs in cell wall modifying enzymes.

Other putative functions of cell wall modifying enzymes

Although cell wall modifying enzymes are of vital importance for the successful penetration and migration of nematodes through the plant cell wall, only a tiny part of the transcriptome consists of cell wall modifying enzymes. This could be the result of a long co-evolution between parasite and host. Cell wall modifying proteins have probably gained additional roles in the plant-nematode interaction during evolution. As described in the previous paragraph, CBMs could be elicitors of plant defense responses (Dumas et al., 2008). Bacterial endoxylanases can likewise trigger plant defense systems (Belien et al., 2006). Moreover, the *D. africanus* expansin-like protein showed significant similarity to putative avirulence proteins and pathogenicity factors, suggesting that expansin-like proteins may also be

recognised by the plant. The resulting defense responses can include the production of plant inhibitors of fungal and bacterial cell wall modifying enzymes, such as a polygalacturonase-inhibiting protein or a xylanase inhibiting protein (Juge, 2006). Whether or not these or similar inhibitors are active against nematode cell wall modifying enzymes as well remains to be elucidated, but the high similarity of nematode enzymes to bacterial enzymes is an important indication that the nematode enzymes may also be targeted by the plant inhibitors. The delicate balance between nematode secreted proteins and plant defense responses has during evolution probably led to a careful and economic selection of secreted cell wall modifying enzymes by the nematode.

Evolution of nematode GHF5 endoglucanases

Due to the relatively high amount of nematode sequence data for GHF5 endoglucanases, several hypotheses about the evolution of these genes have been put forward. A model proposed by Ledger et al. (2006) was adapted and extended in this thesis. The evolution of the endoglucanases also reflects the species evolution: similar relationships between species have been found comparing rRNA genes. Remarkably, the endoglucanase genes of *R. similis* are most similar to the genes from cyst nematodes, confirming earlier reports that question the placement of *R. similis* within the Pratylenchidae (Subbotin et al., 2004; Bert et al., 2008; Holterman et al., 2009). This apparent parallel gene and species evolution suggests that the ancestral endoglucanase emerged early in the evolution of Tylenchomorpha. The gene family extended already early during evolution, since some endoglucanases have an aberrant genomic structure as a result from an early duplication event. The evolution is still ongoing as demonstrated by recent duplication events and also the discovery of pseudo-endoglucanases in *D. africanus*. The latter genes could be evolving towards a new function. In other genes, such as the endoxylanase, a similar evolutionary pattern can be expected. Although this gene family is less extensive (6 copies in *M. incognita* genome, while 21 copies for endoglucanase), duplication events also occurred to expand the gene family. Since the introns of *M. incognita* and *R. similis* endoxylanases are in the same position, both most probably originated from a common ancestral endoxylanase.

Horizontal gene transfer as the origin of cell wall modifying enzymes?

The current hypothesis is that cell wall modifying enzymes originate from multiple independent horizontal gene transfer (HGT) events from bacteria (Jones et al., 2005). For most of these genes, homologs are found exclusively in bacteria or fungi, hence HGT is the

most parsimonious explanation for their presence in plant-parasitic nematodes. As more sequence data is becoming available, it should become more clear whether or not the HGT hypothesis stays valid (Mitreva et al., 2009). The endoglucanase, endoxylanase and expansin-like genes cloned in this thesis all have a typical eukaryotic gene structure, including introns and a polyA tail. This suggests that possible HGT events occurred early in the evolution of plant-parasitic nematodes, as sequence features of eukaryotic genes are very common in these genes. In the *R. similis* endoglucanases however, an aberrant GC content was found (very high GC3) when compared to other *R. similis* genes. The same trend was observed in the *R. similis* endoxylanase and the *D. africanus* endoglucanase, although not in the *D. africanus* expansin-like gene. Could this aberrant nucleotide composition be an ancient relic of a bacterial ancestor? Probably not, as all other characteristics of bacterial genes are missing, but still the reason for this high GC3 content remains unclear.
The finding of a *Wolbachia* endosymbiont in *R. similis* is interesting in the light of the HGT hypothesis. The endobacteria in plant-parasitic nematodes could be considered as possible donors of horizontally transferred genes. Indeed, incorporation of pieces of the *Wolbachia* genome into the host genome has been proven for at least eight different host species (four insects and four nematodes), and these genes are transcriptionally active (Dunning-Hotopp et al., 2007). Although it seems very unlikely that cell wall modifying enzymes find their origin in *Wolbachia*, these genes could be traces of another, long lost endosymbiont. As such it would be interesting to investigate whether pieces of the *Wolbachia* genome have been inserted into the *R. similis* genome. Should this be the case, then it can serve as a model for how plant-parasitic nematodes may have obtained genes from prokaryotes through HGT.

It is not all cell wall modifying enzymes

The formation of a specialised feeding site by sedentary nematodes is generally assigned to an extra set of secreted parasitism and regulatory proteins absent in migratory nematodes. As such, it is tempting to assume that cell wall modifying proteins are sufficient for migratory nematodes to colonise the plant roots. Nevertheless, other proteins are likely to play a role in the plant-nematode interaction. Genes important to counteract oxidative stress and drought were very abundant in the *D. africanus* EST dataset. This suggests that the nematode is constantly struggling against environmental stress factors, and these genes form a much larger part of the transcriptome than cell wall modifying enzymes. As described earlier, the plant reacts to the nematode infection by various defense responses, which the nematode tries to counter. One of the major responses of plants to nematode infection is the production of damaging free radicals including hydrogen peroxide (Gheysen & Jones, 2006). Nematodes can secrete enzymes in their surface coat to neutralise the oxidative stress, such as

peroxiredoxin, superoxide dismutase or glutathione-S-transferase, all present in the *D. africanus* ESTs. Moreover, the nematode genes that are recognised by the plant and trigger immune responses are constantly evolving in an attempt to circumvent the recognition of the plant. Examples of such proteins are allergen proteins and galectins, also present in the *D. africanus* ESTs, that have an unknown function but are known to elicit immune responses in animal-parasitic nematodes (Murray et al., 2001; Young & Meeusen, 2002) and could have comparable host responses in plants. These genes are probably rapidly evolving since the selection pressure is expected to be high. Both the nematode and the host plant are in a kind of evolutionary conflict, where new plant recognition capacities are matched by efforts to escape resistance by the nematode. Plant pathogenic bacteria and fungi secrete several effector proteins important for the virulence of the pathogen as well as to evade plant defense responses. Effector proteins can interfere with each step of plant defense including pre-existing physical and chemical barriers as well as inducible responses ranging from recognition of the pathogen by the plant, defense signaling, transcriptional changes and even hormone signaling (Chisholm et al., 2006; Göhre & Robatzek, 2008). Several genes with unknown function are upregulated in the early parasitic stages of *M. incognita* and could possibly be a kind of effector proteins (Dubreuil et al., 2007). Moreover, additional secreted proteins with homology to plant proteins were identified with domains that resemble known effector functions or that have the ability to reprogram plant cells (Bellafiore et al., 2008). Whether or not these genes can be considered as true parasitism genes remains a matter of debate until their function has been established. Other genes with similarity to genes putatively involved in plant parasitism were identified in the *D. africanus* ESTs, such as genes encoding SEC-2, transthyretin-like proteins, calreticulin-like proteins, and a 14-3-3b protein. The latter two proteins are thought to be involved in the formation of the feeding site of sedentary nematodes (Jaubert et al., 2004; Jaubert et al., 2005), but their presence in migratory nematodes could suggest a more general role in parasitism. However, the function of these genes in the parasitic process is far from clear. Moreover, these genes are involved in general nematode metabolism as well, as it is assumed that duplicates have evolved to gain specific roles in parasitism. The identified homologues in *D. africanus* could therefore equally well be genes needed for the general metabolism of the nematode instead of parasitism genes. This illustrates that assumptions based on homology searches should be handled cautiously and that functional studies remain necessary to draw accurate conclusions.

Perspectives

In this thesis, several new genes were cloned from different migratory nematode species for which no molecular knowledge was available. However, this is only the beginning of the molecular understanding of these animals. As demonstrated in chapter 6, EST analyses have limitations in detecting novel parasitism genes. This is also illustrated by *M. incognita*, for which over 20,000 ESTs are available, and yet when its genome was sequenced novel cell wall modifying enzymes were discovered (e.g. arabinase) (Abad et al., 2008). Next generation sequencing techniques nowadays make it much cheaper to generate lots of sequences independent of the construction of a cDNA library. It can be expected that in the next few years, even more nematode sequence data will become available: on the one hand to assemble new genomes and on the other hand to compare expression levels of both nematodes and plant responses. The comparison between animal-parasitic and plant-parasitic nematode sequence data can reveal similar mechanisms of infection either conserved during evolution or originated from convergent evolution. Moreover, comparisons between migratory and sedentary plant-parasitic nematodes could provide useful insights in genes required for feeding site formation. The challenge will be to handle all these large sequence datasets and distilling the useful information without loosing track.

The identification of new genes is only the first step, the next is to investigate the function of these genes. For example the question if the identified pseudo-endoglucanases of *D. africanus* still have a possible function is intriguing. For the *R. similis* endoxylanase, a first step towards a functional characterisation was done by reducing the expression of the gene by soaking the nematodes in dsRNA and examining the effect on infection. It appears that the silencing effect can vary by the position of the target sequence of the gene, but also by the replica of the experiment itself. This demonstrates that soaking nematodes in dsRNA has its shortcomings; it largely depends on the uptake of the dsRNA of the nematodes. A more efficient way to force the nematode to feed on dsRNA is making transgenic plants producing dsRNA, which also implies a constant dsRNA supply to the nematodes (Gheysen & Vanholme, 2007). This promising technique was shown to be effective in transgenic *Arabidopsis*, tobacco and soybean plants against sedentary nematodes (Steeves et al., 2006; Yadav et al., 2006; Huang et al., 2006a; Fairbairn et al., 2007; Sindhu et al., 2009). In future experiments it would be interesting to test different transgenic plant species in combination with migratory nematodes. In the long run, this could lead to transgenic crops resistant to multiple nematode species.

Summary

Plant-parasitic nematodes cause a significant amount of crop damage worldwide. Although the knowledge about nematodes has increased greatly during the last decade, we are only beginning to understand the complex molecular plant-nematode interaction. Plant-parasitic nematodes all have a stylet which they use to mechanically wound the plant cells but also to secrete proteins important for parasitism. These proteins are encoded by so-called parasitism genes. Most studies on parasitism genes have been conducted on sedentary endoparasitic nematodes, which manipulate the plant to synthesise a specialised nematode feeding site. In this thesis, parasitism genes from the less specialised migratory endoparasites were investigated. The focus was on plant cell wall modifying enzymes, which nematodes secrete to facilitate the migration through the rigid plant cell wall.

The major constituent of the plant cell wall is cellulose, the most abundant sugar polymer on earth. Only very few organisms are able to degrade cellulose: bacteria, fungi, nematodes and a few beetles. In nematodes, cellulases or endo-1,4-β-glucanases have been found predominantly in sedentary species, and within the nematode superfamily Tylenchoidea (order Rhabditida, infraorder Tylenchomorpha), all endoglucanases belong to glycosyl hydrolase family 5 (GHF5). A large part of this thesis deals with the characterisation of GHF5 endoglucanases in migratory endoparasitic nematodes, namely in the species *Radopholus similis*, *Ditylenchus africanus* and *Pratylenchus coffeae*. In the burrowing nematode *R. similis* four endoglucanases were cloned, consisting of a signal peptide, a catalytic domain and in two cases a carbohydrate binding module (CBM) (chapter 2). The signal peptide makes sure the mature protein is secreted from the nematode, the catalytic domain holds the actual enzymatic activity while the CBM is important for the enzyme to bind its substrate. All genes are expressed in the gland cell area of the nematode, pointing towards the secretion of the proteins through the stylet. Remarkably, three of the four endoglucanases showed a reduced expression in males as opposed to females. This could be explained by the observation that adult males do not feed and are thought to be non-parasitic, in which case they need less cellulase activity. In the peanut pod nematode *D. africanus* the first GHF5 endoglucanase was found in the superfamily of the Sphaerularioidea (chapter 3). Furthermore, four pseudo-endoglucanases were cloned containing deletions in the catalytic domain probably due to homologous recombination. In two of these four pseudogenes, the deletion causes a frameshift leading to an erroneous protein, while in the other two cases, the CBM remains intact suggesting that these latter genes could still be partly functional. An additional endoglucanase from *P. coffeae* was identified and a detailed model for the evolution of the gene structure of GHF5 endoglucanases within nematodes combining sequence and intron data was proposed, partially based on a previous model (chapter 4). The model implies an early gene duplication of the ancestral endoglucanase, followed by more recent duplications. Moreover, it proposes a detailed overview of intron gain, loss and sliding events.

The second most abundant sugar polymer in plant cell walls is xylan. An endo-1,4-β-xylanase in *R. similis* probably able to degrade xylan, was characterised (chapter 5). The protein consists of a signal peptide, a catalytic domain and a putative CBM, the first CBM associated with an animal endoxylanase. Homologs in the genomes of *Meloidogyne* spp. revealed that root-knot nematode endoxylanases are probably also associated with a CBM. Moreover, intron positions were conserved in these homologs, suggesting a common origin for root-knot nematode and *R. similis* endoxylanases. Different approaches showed that the catalytic domain does not belong to GHF5 as previously suggested for similar endoxylanases, but rather forms a subclass within GHF30. Expression of the gene was again located in the gland cell area of the nematode, and in this case no difference in expression level could be observed between males and females. Silencing the gene by RNAi resulted in a decrease in infection of 60%, indicating that the endoxylanase gene is important for the nematode to infect its host plant.

Next to cellulose and xylan degrading enzymes, nematodes possess other proteins that can modify the plant cell wall. An example is an expansin-like protein, which was found in *D. africanus* (chapter 3). Interestingly, the expansin-like protein contains a CBM at the C-terminal end of the protein, similar to the CBM of an endoglucanase. This expansin-like protein domain structure is unique so far in nematodes, and is possibly the result of a recent domain shuffling event.

A high-throughput technique to identify possible parasitism genes is expressed sequence tag (EST) analysis, which is basically random sequencing of pieces of genes that are being expressed by the nematode. This approach was used for *D. africanus*, and 4847 ESTs were analysed (chapter 6). A large part of the dataset (43%) did not show any homology to sequence databases, partly due to the relatively low quality of the cDNA library, but possibly also due to the lack of sequence data of related nematodes in public databases. Genes involved in oxidative stress and anhydrobiosis were identified, as well as genes putatively involved in parasitism (expansin, SEC-2, calreticulin, 14-3-3b and various allergen proteins). Furthermore, some putative novel parasitism genes were selected, in addition to good target candidates for RNAi experiments. These genes are interesting for further research and functional characterisation.

Scientific research can sometimes lead to unexpected discoveries. During an EST analysis of *R. similis*, several ESTs were found with similarity to sequences from a *Wolbachia* endosymbiotic bacterium. In this thesis, this finding was further investigated and the intracellular presence of *Wolbachia* in the nematode was confirmed (chapter 7). Although widely present in filarial nematodes, it is the first time this endosymbiont is discovered in a plant-parasitic nematode, which sheds new light on the evolution of *Wolbachia*. This *Wolbachia* strain is indeed distantly related to all known *Wolbachia* supergroups and was

therefore designated as a new supergroup. Further studies will have to elucidate what function it has within the nematode.

With this thesis, the molecular knowledge of some migratory endoparasitic nematodes has significantly progressed. But as plant nematology enters the era of functional and comparative genomics, the challenge to find new routes to effective and safe plant parasite control has just begun.

SAMENVATTING

Plantenparasitaire nematoden of aaltjes veroorzaken wereldwijd een aanzienlijke hoeveelheid schade aan verschillende gewassen. Hoewel de kennis van nematoden gedurende de laatste tien jaar sterk is toegenomen, staan we slechts aan het begin van het doorgronden van de complexe plant-nematode interactie. Plantenparasitaire nematoden hebben allen een stylet waarmee ze de plantencellen mechanisch kunnen verwonden, maar waardoorheen ze ook eiwitten uitscheiden die belangrijk zijn voor hun infectieproces. Deze eiwitten worden gecodeerd door zogenaamde parasitismegenen. De meeste studies die handelen over parasitismegenen werden uitgevoerd op sedentaire endoparasitaire nematoden, die de plant manipuleren om een gespecialiseerde nematodenvoedingsplaats aan te maken. In deze thesis werden parasitismegenen van de minder gespecialiseerde migratorische endoparasitaire nematoden onderzocht. De focus lag op plantencelwandmodificerende enzymen die nematoden secreteren om de migratie doorheen de rigide plantencelwand mogelijk te maken.
Het hoofdbestanddeel van de plantencelwand is cellulose, het meest voorkomende suikerpolymeer op aarde. Slechts enkele organismen zijn in staat om cellulose af te breken: bacteriën, schimmels, nematoden en enkele kevers. In nematoden werden cellulases of endo-1,4-β-glucanases vooral geïdentificeerd in sedentaire soorten, en binnen de nematodensuperfamilie Tylenchoidea (orde Rhabditida, infraorde Tylenchomorpha) behoren alle endoglucanases tot glycosylhydrolase-familie 5 (GHF5). Een groot deel van deze thesis handelt over de karakterisatie van GHF5-endoglucanases in migratorische endoparasitaire nematoden, namelijk in de soorten *Radopholus similis*, *Ditylenchus africanus* en *Pratylenchus coffeae*. In *R. similis* werden vier endoglucanases gevonden die bestaan uit een signaalpeptide, een katalytisch domein en in twee gevallen een carbohydraatbindingsmodule (CBM) (hoofdstuk 2). Het signaalpeptide zorgt ervoor dat het mature eiwit wordt uitgescheiden door de nematode, het katalytisch domein bezit de eigenlijke enzymatische activiteit terwijl de CBM belangrijk is voor het binden van het enzym aan het substraat. Alle genen komen tot expressie in de klierregio van de nematode, wat erop wijst dat de eiwitten doorheen het stylet uitgescheiden worden. Drie van de vier genen vertonen een lagere expressie in mannetjes ten opzichte van vrouwtjes. Dit zou kunnen verklaard worden doordat volwassen mannetjes zich niet voeden en niet parasitair zijn, waardoor deze minder cellulase-activiteit nodig hebben. In *D. africanus* werd het eerste GHF5-endoglucanase gevonden binnen de superfamilie Sphaerularioidea (hoofdstuk 3). Daarnaast werden vier pseudo-endoglucanases gekloneerd met deleties in het katalytisch domein, waarschijnlijk ontstaan door homologe recombinatie. In twee van de vier pseudogenen veroorzaakt de deletie een leesraamverschuiving resulterend in een fout eiwit, terwijl in de andere twee pseudogenen de CBM intact is waardoor deze laatste mogelijks nog deels functioneel zijn. Een bijkomend endoglucanase van *P. coffeae* werd geïdentificeerd en een gedetailleerd model voor de evolutie van de genstructuur van GHF5-endoglucanases binnen nematoden gebaseerd op sequentie- en introndata werd opgesteld, deels gebaseerd op een eerder model (hoofdstuk 4). Het model gaat uit van een

vroege genduplicatie van het ancestrale endoglucanase, gevolgd door recentere duplicaties. Het model geeft eveneens een gedetailleerd overzicht van intronінserties, -verliezen en -verschuivingen.

Het op een na meest voorkomende suikerpolymeer in een plantencelwand is xylaan. Een endo-1,4-β-xylanase van *R. similis* vermoedelijk in staat om xylaan af te breken, werd gekarakteriseerd (hoofdstuk 5). Het eiwit bestaat uit een signaalpeptide, een katalytisch domein en vermoedelijk een CBM, de eerste CBM geassocieerd met een dierlijk endoxylanase. Homologen in de genomen van *Meloidogyne* spp. bezitten waarschijnlijk een gelijkaardig CBM. De intronposities van deze homologen zijn identiek als voor het *R. similis* endoxylanase, wat wijst op een zelfde oorsprong. Via verschillende methoden werd aangetoond dat het katalytisch domein niet tot GHF5 behoort zoals eerder beschreven voor gelijkaardige endoxylanases, maar eerder een subklasse binnen GHF30 vormt. De expressie van het gen is gelokaliseerd in de klierregio van de nematode, en in dit geval kon geen verschil in expressieniveau tussen mannetjes en vrouwtjes aangetoond worden. Wanneer het gen via RNAi uitgeschakeld werd, verminderde de infectie op de plant met 60%, wat aantoont dat het endoxylanase gen belangrijk is voor de nematode om zijn waardplant te infecteren.

Naast cellulose- en xylaan-afbrekende enzymen bezitten nematoden ook nog andere eiwitten die de plantencelwand modificeren. Een voorbeeld hiervan is een expansine-eiwit, dat werd teruggevonden in *D. africanus* (hoofdstuk 3). Dit expansine-eiwit bevat een CBM aan het C-terminaal deel van het eiwit, die sterk gelijkt op de CBM van een endoglucanase. Deze expansine-domeinstructuur is tot dusver uniek in nematoden en is mogelijk het gevolg van domeinshuffling.

Een andere techniek om nieuwe parasitismegenen op te sporen is via “expressed sequence tags” (ESTs), gebaseerd op het willekeurig sequeneren van stukjes genen die door de nematode tot expressie worden gebracht. Deze aanpak werd toegepast op *D. africanus*, en 4847 ESTs werden geanalyseerd (hoofdstuk 6). Een groot deel van de dataset (43%) vertoonde geen enkele homologie met een van de sequentiedatabanken, deels als gevolg van de lage kwaliteit van de cDNA bibliotheek, maar waarschijnlijk ook deels als gevolg van het ontbreken van sequentiedata van verwante nematoden in de publieke databanken. Genen betrokken bij oxidatieve stress en anhydrobiose waren aanwezig in de ESTs, alsook genen die mogelijk belangrijk zijn voor parasitisme (expansine, SEC-2, calreticuline, 14-3-3b en verschillende allergene eiwitten). Daarnaast werden enkele nieuwe vermoedelijke parasitismegenen en goeie kandidaten voor RNAi-experimenten geselecteerd. Deze genen zijn interessant voor verder onderzoek en functionele karakterisatie.

Wetenschappelijk onderzoek leidt soms tot onverwachte ontdekkingen. Tijdens een EST-analyse van *R. similis* werden verschillende ESTs gevonden die sterk gelijken op sequenties van een *Wolbachia* endosymbionte bacterie. In deze thesis werd deze bevinding verder onderzocht en kon de aanwezigheid van een intracellulaire *Wolbachia*-bacterie in de

nematode aangetoond worden (hoofdstuk 7). Hoewel de bacterie vaak voorkomt in filariale nematoden is het de eerste keer dat deze endosymbiont ontdekt wordt in een plantenparasitaire nematode, wat een andere kijk op de evolutie van de *Wolbachia*-bacterie oplevert. Deze *Wolbachia*-stam is inderdaad slechts weinig verwant aan alle tot nu toe gekende *Wolbachia*-supergroepen, en daarom werd de stam toegekend aan een nieuwe supergroep. Verder onderzoek zal moeten uitmaken wat de functie van de symbiont is binnen de nematode.

Met deze thesis werd een aanzienlijke vooruitgang geboekt aan moleculaire kennis van migratorische endoparasitaire nematoden. Aangezien nematologie nu het tijdperk van functionele en vergelijkende genoomanalyse binnentreedt, is de eigenlijke uitdaging om nieuwe effectieve en veilige bestrijdingsmaatregelen tegen plantenparasitaire nematoden te vinden nog maar pas begonnen.

References

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchi, E. G. J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V. et al. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nature Biotechnology, 26, 909-915.

Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. (2004). Image processing with ImageJ. Biophotonics International, 11, 36-42.

Adam, M. A. M., Phillips, M. S., Jones, J. T. & Blok, V. C. (2008). Characterisation of the cellulose-binding protein *Mj-cbp-1* of the root knot nematode, *Meloidogyne javanica*. Physiological and Molecular Plant Pathology, 72, 21-28.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic Local Alignment Search Tool. Journal of Molecular Biology, 215, 403-410.

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

Babenko, V. N., Rogozin, I. B., Mekhedov, S. L. & Koonin, E. V. (2004). Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic Acids Research, 32, 3724-3733.

Bailey, L. C., Searls, D. B. & Overton, G. C. (1998). Analysis of EST-driven gene annotation in human genomic sequence. Genome Research, 8, 362-376.

Baldo, L., Hotopp, J. C. D., Jolley, K. A., Bordenstein, S. R., Biber, S. A., Choudhury, R. R., Hayashi, C., Maiden, M. C. J. et al. (2006). Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Applied and Environmental Microbiology, 72, 7098-7110.

Baldo, L., Prendini, L., Corthals, A. & Werren, J. H. (2007). *Wolbachia* are present in Southern African scorpions and cluster with supergroup F. Current Microbiology, 55, 367-373.

Baldo, L. & Werren, J. H. (2007). Revisiting *Wolbachia* supergroup typing based on WSP: Spurious lineages and discordance with MLST. Current Microbiology, 55, 81-87.

Baldwin, J. G., Nadler, S. A. & Adams, B. J. (2004). Evolution of plant parasitism among nematodes. Annual Review of Phytopathology, 42, 83-105.

Bandi, C., Anderson, T. J. C., Genchi, C. & Blaxter, M. L. (1998). Phylogeny of *Wolbachia* in filarial nematodes. Proceedings of the Royal Society of London Series B-Biological Sciences, 265, 2407-2413.

Barthelmes, J., Ebeling, C., Chang, A., Schomburg, I. & Schomburg, D. (2007). BRENDA, AMENDA and FRENDA: the enzyme information system in 2007. Nucleic Acids Research, 35, D511-D514.

Basson, S., De Waele, D. & Meyer, A. J. (1992). Effect of host plant age on population development and pathogenicity of *Ditylenchus destructor* on peanut. Journal of Nematology, 24, 310-314.

Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. (2004). The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. Annual Review of Microbiology, 58, 521-554.

Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. Current Opinion in Structural Biology, 8, 548-557.

Bekal, S., Niblack, T. L. & Lambert, K. N. (2003). A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. Molecular Plant-Microbe Interactions, 16, 439-446.

Belien, T., Van Campenhout, S., Robben, J. & Voickaert, G. (2006). Microbial endoxylanases: Effective weapons to breach the plant cell-wall barrier or, rather, triggers of plant defense systems? Molecular Plant-Microbe Interactions, 19, 1072-1081.

Bellafiore, S., Shen, Z. X., Rosso, M. N., Abad, P., Shih, P. & Briggs, S. P. (2008). Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. Plos Pathogens, 4, e1000192.

Belle, E. M. S., Smith, N. & Eyre-Walker, A. (2002). Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. Journal of Molecular Evolution, 55, 356-363.

Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology, 340, 783-795.

Bera-Maillet, C., Arthaud, L., Abad, P. & Rosso, M. N. (2000). Biochemical characterization of MI-ENG1, a family 5 endoglucanase secreted by the root-knot nematode *Meloidogyne incognita*. European Journal of Biochemistry, 267, 3255-3263.

Bernardi, G. (2000). Isochores and the evolutionary genomics of vertebrates. Gene, 241, 3-17.

Bert, W., Leliaert, F., Vierstraete, A. R., Vanfleteren, J. R. & Borgonie, G. (2008). Molecular phylogeny of the Tylenchina and evolution of the female gonoduct (Nematoda : Rhabditida). Molecular Phylogenetics and Evolution, 48, 728-744.

Bird, D. M. & Kaloshian, I. (2003). Are roots special? Nematodes have their say. Physiological and Molecular Plant Pathology, 62, 115-123.

Bird, D. M. & Koltai, H. (2000). Plant parasitic nematodes: Habitats, hormones, and horizontally-acquired genes. Journal of Plant Growth Regulation, 19, 183-194.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y. et al. (1998). A molecular evolutionary framework for the phylum Nematoda. Nature, 392, 71-75.

Blumenthal, T. & Steward, K. (1997). RNA processing and gene structure. In: *C. elegans* II, eds. Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess J.R. (Cold Spring Harbor Laboratory Press, New York).

Bolla, R. I., Weaver, C. & Winter, R. E. K. (1988). Genomic differences among pathotypes of *Bursaphelenchus xylophilus*. Journal of Nematology, 20, 309-316.

Bolton, C., De Waele, D. & Basson, S. (1990). Comparison of two methods of extracting *Ditylenchus destructor* from hulls and seeds of groundnuts. Revue Nématologie, 13, 233-235.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochemical Journal, 382, 769-781.

Bordenstein, S. & Rosengaus, R. B. (2005). Discovery of a novel *Wolbachia* supergroup in isoptera. Current Microbiology, 51, 393-398.

Bordenstein, S. R., Fitch, D. H. A. & Werren, J. H. (2003). Absence of *Wolbachia* in nonfilariid nematodes. Journal of Nematology, 35, 266-270.

Bordenstein, S. R., Paraskevopoulos, C., Hotopp, J. C., Sapountzis, P., Lo, N., Bandi, C., Tettelin, H., Werren, J. H. et al. (2009). Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say. Molecular Biology and Evolution, 26, 231-241.

Boudet, N., Aubourg, S., Toffano-Nioche, C., Kreis, M. & Lecharny, A. (2001). Evolution of intron/exon structure of DEAD helicase family genes in *Arabidopsis*, *Caenorhabditis*, and *Drosophila*. Genome Research, 11, 2101-2114.

Braig, H. R., Zhou, W. G., Dobson, S. L. & O'Neill, S. L. (1998). Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. Journal of Bacteriology, 180, 2373-2378.

Brennan, Y., Callen, W. N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., Hernandez, M., Keller, M. et al. (2004). Unusual microbial xylanases from insect guts. Applied and Environmental Microbiology, 70, 3609-3617.

Brenner, S. (1974). Genetics of *Caenorhabditis elegans*. Genetics, 77, 71-94.

Brooks, T. L. & Perry, V. G. (1962). Apparent parthenogenetic reproduction of the burrowing nematode *Radopholus similis* (Cobb) Thorne. Soil and Crop Science Society of Florida Proceedings, 22, 160-162.

Browne, J., Tunnacliffe, A. & Burnell, A. (2002). Anhydrobiosis - Plant desiccation gene found in a nematode. Nature, 416, 38.

Byrd, D. W., Kirkpatrick, T. & Barker, K. R. (1983). An improved technique for clearing and staining plant-tissues for detection of nematodes. Journal of Nematology, 15, 142-143.

C.elegans sequencing consortium. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. Science, 282, 2012-2018.

Casiraghi, M., Bordenstein, S. R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J. J., Werren, J. H. & Bandi, C. (2005). Phylogeny of *Wolbachia pipientis* based on *gltA*, *groEL* and *ftsZ* gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. Microbiology, 151, 4015-4022.

Casiraghi, M., Favia, G., Cancrini, G., Bartoloni, A. & Bandi, C. (2001). Molecular identification of *Wolbachia* from the filarial nematode *Mansonella ozzardi*. Parasitology Research, 87, 417-420.

Casiraghi, M., Mccall, J. W., Simoncini, L., Kramer, L. H., Sacchi, L., Genchi, C., Werren, J. H. & Bandi, C. (2002). Tetracycline treatment and sex-ratio distortion: a role for *Wolbachia* in the moulting of filarial nematodes? International Journal for Parasitology, 32, 1457-1468.

Chabrier, C. & Quénéhervé, P. (2003). Control of the burrowing nematode (*Radopholus similis* Cobb) on banana: impact of the banana field destruction method on the efficiency of the following fallow. Crop Protection, 22, 121-127.

Chabrier, C. & Quénéhervé, P. (2008). Preventing nematodes from spreading: A case study with *Radopholus similis* (Cobb) Thorne in a banana field. Crop Protection, 27, 1237-1243.

Charlesworth, D., Liu, F. L. & Zhang, L. (1998). The evolution of the alcohol dehydrogenase gene family by loss of introns in plants of the genus *Leavenworthia* (Brassicaceae). Molecular Biology and Evolution, 15, 552-559.

Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. (2006). Host-microbe interactions: Shaping the evolution of the plant immune response. Cell, 124, 803-814.

Collins, T., Gerday, C. & Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. Fems Microbiology Reviews, 29, 3-23.

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

Cosgrove, D. J. (2000a). Loosening of plant cell walls by expansins. Nature, 407, 321-326.

Cosgrove, D. J. (2000b). New genes and new biological roles for expansins. Current Opinion in Plant Biology, 3, 73-78.

Cosgrove, D. J. (2005). Growth of the plant cell wall. Nature Reviews Molecular Cell Biology, 6, 850-861.

Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In: Recent Advances in Carbohydrate Bioengineering, eds. Gilbert, H. J., Davies, G., Henrissat, B. & Svensson, B. (The Royal Society of Chemistry, Cambridge).

Craig, J. P., Bekal, S., Hudson, M., Domier, L., Niblack, T. & Lambert, K. N. (2008). Analysis of a horizontally transferred pathway involved in vitamin B-6 biosynthesis from the soybean cyst nematode *Heterodera glycines*. Molecular Biology and Evolution, 25, 2085-2098.

Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: A sequence logo generator. Genome Research, 14, 1188-1190.

Cutter, A. D., Wasmuth, J. D. & Blaxter, M. L. (2006). The evolution of biased codon and amino acid usage in nematode genomes. Molecular Biology and Evolution, 23, 2303-2315.

Czarnetzki, A. B. & Tebbe, C. C. (2004). Detection and phylogenetic analysis of *Wolbachia* in Collembola. Environmental Microbiology, 6, 35-44.

D'Errico, I., Gadaleta, G. & Saccone, C. (2004). Pseudogenes in metazoa: Origin and features. Briefings in Functional Genomics and Proteomics, 3, 157-167.

Dautova, M., Rosso, M. N., Abad, P., Gommers, F. J., Bakker, J. & Smant, G. (2001). Single pass cDNA sequencing - a powerful tool to analyse gene expression in preparasitic juveniles of the southern root-knot nematode *Meloidogyne incognita*. Nematology, 3, 129-139.

Davis, E. L., Hussey, R. S. & Baum, T. J. (2004). Getting to the roots of parasitism by nematodes. Trends in Parasitology, 20, 134-141.

Davis, E. L., Hussey, R. S., Mitchum, M. G. & Baum, T. J. (2008). Parasitism proteins in nematode-plant interactions. Current Opinion in Plant Biology, 11, 360-366.

Davison, A. & Blaxter, M. (2005). Ancient origin of glycosyl hydrolase family 9 cellulase genes. Molecular Biology and Evolution, 22, 1273-1284.

de Boer, J. M., McDermott, J. P., Davis, E. L., Hussey, R. S., Popeijus, H., Smant, G. & Baum, T. J. (2002). Cloning of a putative pectate lyase gene expressed in the subventral esophageal glands of *Heterodera glycines*. Journal of Nematology, 34, 9-11.

De Ley, P. & Blaxter, M. (2002). Systematic position and phylogeny. In: The Biology of Nematodes, ed. Lee, D. L. (Taylor & Francis, London).

De Meutter, J., Vanholme, B., Bauw, G., Tytgat, T., Gheysen, G. & Gheysen, G. (2001). Preparation and sequencing of secreted proteins from the pharyngeal glands of the plant parasitic nematode *Heterodera schachtii*. Molecular Plant Pathology, 2, 297-301.

De Waele, D., Jones, B. L., Bolton, C. & van den Berg, E. (1989). *Ditylenchus destructor* in hulls and seeds of peanut. Journal of Nematology, 21, 10-15.

De Waele, D., Jordaan, E. M. & Basson, S. (1990). Host status of 7 weed species and their effects on *Ditylenchus destructor* infestation of peanut. Journal of Nematology, 22, 292-296.

De Waele, D., Venter, C. & McDonald, A. H. (1997). The peanut pod nematode, *Ditylenchus africanus*. Nematology Circular, 218, 1-6.

De Waele, D., Willken, R. & Lindeque, J. M. (1991). Response of potato cultivars to *Ditylenchus destructor* from groundnut. Revue Nématologie, 14, 123-126.

Decraemer, W. & Hunt, D. J. (2006). Structure and Classification. In: Plant Nematology, eds. Perry, R. N. & Moens, M. (CABI Publishing, Oxfordshire, UK).

Deutsch, M. & Long, M. (1999). Intron-exon structures of eukaryotic model organisms. Nucleic Acids Research, 27, 3219-3228.

Dieterich, C., Clifton, S. W., Schuster, L. N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R. et al. (2008). The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. Nature Genetics, 40, 1193-1198.

Ding, X., Shields, J., Allen, R. & Hussey, R. S. (1998). A secretory cellulose-binding protein cDNA cloned from the root-knot nematode (*Meloidogyne incognita*). Molecular Plant-Microbe Interactions, 11, 952-959.

Ding, X., Shields, J., Allen, R. & Hussey, R. S. (2000). Molecular cloning and characterisation of a venom allergen AG5-like cDNA from *Meloidogyne incognita*. International Journal for Parasitology, 30, 77-81.

Doyle, E. A. & Lambert, K. N. (2002). Cloning and characterization of an esophageal-gland-specific pectate lyase from the root-knot nematode *Meloidogyne javanica*. Molecular Plant-Microbe Interactions, 15, 549-556.

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

Dumas, B., Bottin, A., Gaulin, E. & Esquerre-Tugaye, M. T. (2008). Cellulose-binding domains: cellulose associated-defensive sensing partners? Trends in Plant Science, 13, 160-164.

Duncan, L. W. & Moens, M. (2006). Migratory endoparasitic nematodes. In: Plant Nematology, eds. Perry, R. N. & Moens, M. (CABI publishing, Oxfordshire).

Dunning-Hotopp, J. C., Clark, M. E., Oliveira, D. C. S. G., Foster, J. M., Fischer, P., Torres, M. C., Giebel, J. D., Kumar, N. et al. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science, 317, 1753-1756.

Edwards, U., Rogall, T., Blocker, H., Emde, M. & Bottger, E. C. (1989). Isolation and direct complete nucleotide determination of entire genes - characterization of a gene coding for 16S ribosomal RNA. Nucleic Acids Research, 17, 7843-7853.

Elling, A. A., Davis, E. L., Hussey, R. S. & Baum, T. J. (2007). Active uptake of cyst nematode parasitism proteins into the plant cell nucleus. International Journal for Parasitology, 37, 1269-1279.

Elling, A. A., Mitreva, M., Gai, X. W., Martin, J., Recknor, J., Davis, E. L., Hussey, R. S., Nettleton, D. et al. (2009). Sequence mining and transcript profiling to explore cyst nematode parasitism. BMC Genomics, 10, 58.

Elsen, A., Declerck, S. & De Waele, D. (2000). Reproduction of the burrowing nematode (*Radopholus similis*) on Ri T-DNA transformed carrot roots. Nematology, 2, 247-249.

EPPO diagnostics report (2008). *Radopholus similis*. EPPO Bulletin, 38, 374-378.

Fairbairn, D. J., Cavallaro, A. S., Bernard, M., Mahalinga-Iyer, J., Graham, M. W. & Botella, J. R. (2007). Host-delivered RNAi: an effective strategy to silence genes in plant parasitic nematodes. Planta, 226, 1525-1533.

Farris, J. S., Kallersjo, M., Kluge, A. G. & Bult, C. (1994). Testing significance of incongruence. Cladistics, 10, 315-319.

Fenn, K. & Blaxter, M. (2006). *Wolbachia* genomes: revealing the biology of parasitism and mutualism. Trends in Parasitology, 22, 60-65.

Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M. et al. (2006). Pfam: clans, web tools and services. Nucleic Acids Research, 34, D247-D251.

Fogain, R. (2000). Effect of *Radopholus similis* on plant growth and yield of plantains (*Musa*, AAB). Nematology, 2, 129-133.

Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya, A., Kapatral, V. et al. (2005). The *Wolbachia* genome of *Brugia malayi*: Endosymbiont evolution within a human pathogenic nematode. Plos Biology, 3, 599-614.

Foster, J. M., Kumar, S., Ford, L., Johnston, K. L., Ben, R., Graeff-Teixeira, C. & Taylor, M. J. (2008). Absence of *Wolbachia* endobacteria in the non-filariid nematodes *Angiostrongylus cantonensis* and *A. costaricensis*. Parasites & Vectors, 1, 31.

Furlanetto, C., Cardle, L., Brown, D. J. F. & Jones, J. T. (2005). Analysis of expressed sequence tags from the ectoparasitic nematode *Xiphinema index*. Nematology, 7, 95-104.

Gao, B., Allen, R., Davis, E. L., Baum, T. J. & Hussey, R. S. (2004a). Molecular characterisation and developmental expression of a cellulose-binding protein gene in the soybean cyst nematode *Heterodera glycines*. International Journal for Parasitology, 34, 1377-1383.

Gao, B., Allen, R., Maier, T., Davis, E. L., Baum, T. J. & Hussey, R. S. (2001). Molecular characterisation and expression of two venom allergen-like protein genes in *Heterodera glycines*. International Journal for Parasitology, 31, 1617-1625.

Gao, B., Allen, R., Maier, T., Davis, E. L., Baum, T. J. & Hussey, R. S. (2002). Identification of a new beta-1,4-endoglucanase gene expressed in the esophageal subventral gland cells of *Heterodera glycines*. Journal of Nematology, 34, 12-15.

Gao, B. L., Allen, R., Davis, E. L., Baum, T. J. & Hussey, R. S. (2004b). Developmental expression and biochemical properties of a beta-1,4-endoglucanase family in the soybean cyst nematode, *Heterodera glycines*. Molecular Plant Pathology, 5, 93-104.

Gao, B. L., Allen, R., Maier, T., Davis, E. L., Baum, T. J. & Hussey, R. S. (2003). The parasitome of the phytonematode *Heterodera glycines*. Molecular Plant-Microbe Interactions, 16, 720-726.

Ghedin, E., Wang, S. L., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J. E., Delcher, A. L. et al. (2007). Draft genome of the filarial nematode parasite *Brugia malayi*. Science, 317, 1756-1760.

Gheysen, G. & Fenoll, C. (2002). Gene expression in nematode feeding sites. Annual Review of Phytopathology, 40, 191-219.

Gheysen, G. & Jones, J. T. (2006). Molecular aspects of plant-nematode interactions. In: Plant Nematology, eds. Perry, R. N. & Moens, M. (CABI Publishing, Oxfordshire, UK).

Gheysen, G. & Vanholme, B. (2007). RNAi from plants to nematodes. Trends in Biotechnology, 25, 89-92.

Gilbert, W. (1978). Why Genes in Pieces. Nature, 271, 501.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. & Warren, R. A. J. (1991). Domains in microbial beta-1,4-glycanases - Sequence conservation, function and enzyme families. Microbiological Reviews, 55, 303-315.

Girard, C. & Jouanin, L. (1999). Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, *Phaedon cochleariae*. Insect Biochemistry and Molecular Biology, 29, 1129-1142.

Goellner, M., Smant, G., de Boer, J. M., Baum, T. J. & Davis, E. L. (2000). Isolation of beta-1,4-endoglucanase genes from *Globodera tabacum* and their expression during parasitism. Journal of Nematology, 32, 154-165.

Göhre, V. & Robatzek, S. (2008). Breaking the barriers: Microbial effector molecules subvert plant immunity. Annual Review of Phytopathology, 46, 189-215.

Goverse, A., Davis, E. L. & Hussey, R. S. (1994). Monoclonal antibodies to the esophageal glands and stylet secretions of *Heterodera glycines*. Journal of Nematology, 26, 251-259.

Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis, 18, 2714-2723.

Gupta, B. P. & Sternberg, P. W. (2003). The draft genome sequence of the nematode *Caenorhabditis briggsae*, a companion to *C. elegans*. Genome Biology, 4, 238.

Hall, T. A. (1999). BioEdit: A biological sequence alignment editor for Windows 95/98/NT. Nucleic Acids Symposium Series, 41, 95-98.

Hasegawa, M., Kishino, H. & Yano, T. A. (1985). Dating of the human ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution, 22, 160-174.

Hedges, L. M., Brownlie, J. C., O'Neill, S. L. & Johnson, K. N. (2008). *Wolbachia* and virus protection in insects. Science, 322, 702.

Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochemical Journal, 293, 781-788.

Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. Biochemical Journal, 316, 695-696.

Hertig, M. & Wolbach, S. B. (1924). Studies on *Rickettsia*-like microorganisms in insects. Journal of Medical Research, 44, 329-374.

Hewezi, T., Howe, P., Maier, T. R., Hussey, R. S., Goellner Mitchum, M., Davis, E. L. & Baum, T. J. (2008). Cellulose binding protein from the parasitic nematode *Heterodera schachtii* interacts with *Arabidopsis* pectin methylesterase: Cooperative cell wall modification during parasitism. The Plant Cell, 20, 3080-3093.

Holterman, M., Karssen, G., van den Elsen S., van Megen H., Bakker, J. & Helder, J. (2009). Small subunit rDNA-based phylogeny of the Tylenchida sheds light on relationships among some high-impact plant-parasitic nematodes and the evolution of plant feeding. Phytopathology, 99, 227-235.

Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., Bongers, T., Holovachov, O., Bakker, J. & Helder, J. (2006). Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. Molecular Biology and Evolution, 23, 1792-1800.

Hotopp, J. C. D., Clark, M. E., Oliveira, D. C. S. G., Foster, J. M., Fischer, P., Torres, M. C., Giebel, J. D., Kumar, N. et al. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science, 317, 1753-1756.

Huang, G. Z., Allen, R., Davis, E. L., Baum, T. J. & Hussey, R. S. (2006a). Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. Proceedings of the National Academy of Sciences of the United States of America, 103, 14302-14306.

Huang, G. Z., Dong, R. H., Allen, R., Davis, E. L., Baum, T. J. & Hussey, R. S. (2005). Developmental expression and molecular analysis of two *Meloidogyne incognita* pectate lyase genes. International Journal for Parasitology, 35, 685-692.

Huang, G. Z., Dong, R. H., Allen, R., Davis, E. L., Baum, T. J. & Hussey, R. S. (2006b). A root-knot nematode secretory peptide functions as a ligand for a plant transcription factor. Molecular Plant-Microbe Interactions, 19, 463-470.

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

Huettel, R. N., Dickson, D. W. & Kaplan, D. T. (1984). *Radopholus citrophilus* sp. n. (Nematoda), a sibling species of *Radopholus similis*. Proceedings of the helminthological society of Washington, 51, 35.

Hurlbert, J. C. & Preston, J. F. (2001). Functional characterization of a novel xylanase from a corn strain of *Erwinia chrysanthemi*. Journal of Bacteriology, 183, 2093-2100.

Hurst, L. D. & Merchant, A. R. (2001). High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. Proceedings of the Royal Society of London Series B-Biological Sciences, 268, 493-497.

Irimia, M., Penny, D. & Roy, S. W. (2007). Coevolution of genomic intron number and splice sites. Trends in Genetics, 23, 321-325.

Jacob, J., Mitreva, M., Vanholme, B. & Gheysen, G. (2008). Exploring the transcriptome of the burrowing nematode *Radopholus similis*. Molecular Genetics and 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.

Janda, J. M. & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. Journal of Clinical Microbiology, 45, 2761-2764.

Jasmer, D. P., Goverse, A. & Smant, G. (2003). Parasitic nematode interactions with mammals and plants. Annual Review of Phytopathology, 41, 245-270.

Jaubert, S., Laffaire, J. B., Abad, P. & Rosso, M. N. (2002a). A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. Febs Letters, 522, 109-112.

Jaubert, S., Laffaire, J. B., Ledger, T. N., Escoubas, P., Amri, E. Z., Abad, P. & Rosso, M. N. (2004). Comparative analysis of two 14-3-3 homologues and their expression pattern in the root-knot nematode *Meloidogyne incognita*. International Journal for Parasitology, 34, 873-880.

Jaubert, S., Ledger, T. N., Laffaire, J. B., Piotte, C., Abad, P. & Rosso, M. N. (2002b). Direct identification of stylet secreted proteins from root-knot nematodes by a proteomic approach. Molecular and Biochemical Parasitology, 121, 205-211.

Jaubert, S., Milac, A. L., Petrescu, A. J., de Almelda-Engler, J., Abad, P. & Rosso, M. N. (2005). In planta secretion of a calreticulin by migratory and sedentary stages of root-knot nematode. Molecular Plant-Microbe Interactions, 18, 1277-1284.

John, F. J. S., Rice, J. D. & Preston, J. F. (2006). Characterization of XynC from *Bacillus subtilis* subsp. *subtilis* strain 168 and analysis of its role in depolymerization of glucuronoxylan. Journal of Bacteriology, 188, 8617-8626.

Johnson, L. A. & Soltis, D. E. (1998). Assessing congruence: Empirical examples from molecular data. In: Molecular Systematics of Plants., eds. Soltis, D. E., Soltis, P. & Doyle, J. J. (Kluwer Academic Publishers, Boston, MA).

Jones, J. T., 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*. Molecular Plant Pathology, 4, 43-50.

Jones, J. T., Furlanetto, C. & Kikuchi, T. (2005). Horizontal gene transfer from bacteria and fungi as a driving force in the evolution of plant parasitism in nematodes. Nematology, 7, 641-646.

Juge, N. (2006). Plant protein inhibitors of cell wall degrading enzymes. Trends in Plant Science, 11, 359-367.

Kaessmann, H., Zollner, S., Nekrutenko, A. & Li, W. H. (2002). Signatures of domain shuffling in the human genome. Genome Research, 12, 1642-1650.

Kaplan, D. T. & Opperman, C. H. (2000). Reproductive strategies and karyotype of the burrowing nematode, *Radopholus similis*. Journal of Nematology, 32, 126-133.

Kaplan, D. T. & Oppermann, C. H. (1997). Genome similarity implies that citrus-parasitic burrowing nematodes do not represent a unique species. Journal of Nematology, 29, 430-440.

Kaplan, D. T., Thomas, W. K., Frisse, L. M., Sarah, J. L., Stanton, J. M., Speijer, P. R., Marin, D. H. & Opperman, C. H. (2000). Phylogenetic analysis of geographically diverse *Radopholus similis* via rDNA sequence reveals a monomorphic motif. Journal of Nematology, 32, 134-142.

Keen, N. T., Boyd, C. & Henrissat, B. (1996). Cloning and characterization of a xylanase gene from corn strains of *Erwinia chrysanthemi*. Molecular Plant-Microbe Interactions, 9, 651-657.

Keen, N. T. & Roberts, P. A. (1998). Plant parasitic nematodes: Digesting a page from the microbe book. Proceedings of the National Academy of Sciences of the United States of America, 95, 4789-4790.

Keiser, P. B., Coulibaly, Y., Kubofcik, J., Diallo, A. A., Klion, A. D., Traore, S. F. & Nutman, T. B. (2008). Molecular identification of *Wolbachia* from the filarial nematode *Mansonella perstans*. Molecular and Biochemical Parasitology, 160, 123-128.

Kerff, F., Amoroso, A., Herman, R., Sauvage, E., Petrella, S., Filee, P., Charlier, P., Joris, B. et al. (2008). Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. Proceedings of the National Academy of Sciences of the United States of America, 105, 16876-16881.

Kikuchi, T., Aikawa, T., Kosaka, H., Pritchard, L., Ogura, N. & Jones, J. T. (2007). Expressed sequence tag (EST) analysis of the pine wood nematode *Bursaphelenchus xylophilus* and *B. mucronatus*. Molecular and Biochemical Parasitology, 155, 9-17.

Kikuchi, T., Jones, J. T., 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., Aikawa, T. & Jones, J. T. (2006). Cloning and characterization of pectate lyases expressed in the esophageal gland of the pine wood nematode *Bursaphelenchus xylophilus*. Molecular Plant-Microbe Interactions, 19, 280-287.

Kikuchi, T., Shibuya, H. & Jones, J. T. (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. Biochemical Journal, 389, 117-125.

Kleynhans, K. P. N. (1999). Collecting and preserving nematodes - A manual for nematology. (Ultra Litho Ltd, Johannesburg, South Africa).

Kloek, A. P., Sherman, D. R. & Goldberg, D. E. (1993). Novel gene structure and evolutionary context of *Caenorhabditis elegans* globin. Gene, 129, 215-221.

Kramer, L. H., Passeri, B., Corona, S., Simoncini, L. & 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). Parasitology Research, 89, 381-386.

Kudla, U., Qin, L., Milac, A., Kielak, A., Maissen, C., Overmars, H., Popeijus, H., Roze, E. et al. (2005). Origin, distribution and 3D-modeling of Gr-EXPB1, an expansin from the potato cyst nematode *Globodera rostochiensis*. Febs Letters, 579, 2451-2457.

Lambert, K. N., Allen, K. D. & Sussex, I. M. (1999). Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*. Molecular Plant-Microbe Interactions, 12, 328-336.

Larson, S. B., Day, J., de la Rosa, A. P. B., Keen, N. T. & McPherson, A. (2003). First crystallographic structure of a xylanase from glycoside hydrolase family 5: Implications for catalysis. Biochemistry, 42, 8411-8422.

Ledger, T. N., Jaubert, S., Bosselut, N., Abad, P. & Rosso, M. N. (2006). Characterization of a new beta-1,4-endoglucanase gene from the root-knot nematode *Meloidogyne incognita* and evolutionary scheme for phytonematode family 5 glycosyl hydrolases. Gene, 382, 121-128.

Lee, S. J., Kim, S. R., Yoon, H. J., Kim, I., Lee, K. S., Je, Y. H., Lee, S. M., Seo, S. J. et al. (2004). cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, *Apriona germari*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 139, 107-116.

Liaud, M. F., Brinkmann, H. & Cerff, R. (1992). The beta-tubulin gene family of pea - Primary structures, genomic organization and intron-dependent evolution of genes. Plant Molecular Biology, 18, 639-651.

Lilley, C. J., Bakhetia, M., Charlton, W. L. & Urwin, P. E. (2007). Recent progress in the development of RNA interference for plant parasitic nematodes. Molecular Plant Pathology, 8, 701-711.

Lo, N., Casiraghi, M., Salati, E., Bazzocchi, C. & Bandi, C. (2002). How many *Wolbachia* supergroups exist? Molecular Biology and Evolution, 19, 341-346.

Lo, N., Paraskevopoulos, C., Bourtzis, K., O'Neill, S. L., Werren, J. H., Bordenstein, S. R. & Bandi, C. (2007). Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. International Journal of Systematic and Evolutionary Microbiology, 57, 654-657.

Lo, N., Watanabe, H. & Sugimura, M. (2003). Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals. Proceedings of the Royal Society of London Series B-Biological Sciences, 270, S69-S72.

Lodish, H., Berk, A., Zipursky, L. S., Matsudaira, P., Baltimore, D. & Darnell, J. (2000). Molecular Cell Biology. (W. H. Freeman and Company, New York, USA).

Logsdon, J. M., Tyshenko, M. G., Dixon, C., Jafari, J. D., Walker, V. K. & Palmer, J. D. (1995). 7 Newly discovered intron positions in the triose-phosphate isomerase gene - Evidence for the introns-late

theory. Proceedings of the National Academy of Sciences of the United States of America, 92, 8507-8511.

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 Pathology, 55, 559-563.

Lu, S., Yu, H. & Wang, X. (2007). Molecular characterization and functional analysis of venom allergen-like protein genes in the potato cyst nematode, *Globodera rostochiensis*. Phytopathology, 97, S67-S68.

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

Lynch, M. & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. Science, 290, 1151-1155.

MacGowan, J. B. (1977). The burrowing nematode *Radopholus similis* (Cobb 1893) Thorne 1949. Nematology Circular, 27.

McCarter, J. P., Mitreva, M. D., Martin, J., Dante, M., Wylie, T., Rao, U., Pape, D., Bowers, Y. et al. (2003). Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*. Genome Biology, 4, R26.

Min, X. J., Butler, G., Storms, R. & Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Research, 33, W677-W680.

Mitreva, M., Blaxter, M. L., Bird, D. M. & McCarter, J. P. (2005). Comparative genomics of nematodes. Trends in Genetics, 21, 573-581.

Mitreva, M., Elling, A. A., Dante, M., Kloek, A. P., Kalyanaraman, A., Aluru, S., Clifton, S. W., Bird, D. M. et al. (2004). A survey of SL1-spliced transcipts from the root-lesion nematode *Pratylenchus penetrans*. Molecular Genetics and Genomics, 272, 138-148.

Mitreva, M., Smant, G. & Helder, J. (2009). Role of horizontal gene transfer in the evolution of plant parasitism among nematodes. In: Horizontal gene transfer - Genomes in flux, eds. Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. C. (Humana Press, New York, NY).

Mitreva, M., Wendl, M. C., Martin, J., Wylie, T., Yin, Y., Larson, A., Parkinson, J., Waterston, R. H. et al. (2006). Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. Genome Biology, 7, R75.

Mitreva-Dautova, M., Roze, E., Overmars, H., de Graaff, L., Schots, A., Helder, J., Goverse, A., Bakker, J. et al. (2006). A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode *Meloidogyne incognita*. Molecular Plant-Microbe Interactions, 19, 521-529.

Moody, E. H., Lownsbery, B. F. & Ahmed, J. M. (1973). Culture of root-lesion nematode *Pratylenchus vulnus* on carrot disks. Journal of Nematology, 5, 225-226.

Munoz, E. T., Bogarad, L. D. & Deem, M. W. (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, J., Gregory, W. F., Gomez-Escobar, N., Atmadja, A. K. & Maizels, R. M. (2001). Expression and immune recognition of *Brugia malayi* VAL-1, a homologue of vespid venom allergens and *Ancylostoma* secreted proteins. Molecular and Biochemical Parasitology, 118, 89-96.

Mutwil, M., Debolt, S. & Persson, S. (2008). Cellulose synthesis: a complex complex. Current Opinion in Plant Biology, 11, 252-257.

Noel, G. R. & Atibalentja, N. (2006). '*Candidatus Paenicardinium endonii*', an endosymbiont of the plant-parasitic nematode *Heterodera glycines* (Nemata : Tylenchida), affiliated to the phylum Bacteroidetes. International Journal of Systematic and Evolutionary Microbiology, 56, 1697-1702.

Olsen, A. N. & Skriver, K. (2003). Ligand mimicry? Plant-parasitic nematode polypeptide with similarity to CLAVATA3. Trends in Plant Science, 8, 55-57.

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S. et al. (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. Proceedings of the National Academy of Sciences of the United States of America, 105, 14802-14807.

Page, R. D. M. (1996). TreeView: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences, 12, 357-358.

Park, J. E., Lee, K. Y., Lee, S. J., Oh, W. S., Jeong, P. Y., Woo, T., Kim, C. B., Paik, Y. K. et al. (2008). The efficiency of RNA interference in *Bursaphelenchus xylophilus*. Molecules and Cells, 26, 81-86.

Parkinson, J., Mitreva, M., Whitton, C., Thomson, M., Daub, J., Martin, J., Schmid, R., Hall, N. et al. (2004). A transcriptomic analysis of the phylum Nematoda. Nature Genetics, 36, 1259-1267.

Patthy, L. (1999). Genome evolution and the evolution of exon-shuffling - a review. Gene, 238, 103-114.

Pearl, F. M. G., Bennett, C. F., Bray, J. E., Harrison, A. P., Martin, N., Shepherd, A., Sillitoe, I., Thornton, J. et al. (2003). The CATH database: an extended protein family resource for structural and functional genomics. Nucleic Acids Research, 31, 452-455.

Pertea, G., Huang, X. Q., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J. et al. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics, 19, 651-652.

Pfarr, K., Foster, J., Slatko, B., Hoerauf, A. & Eisen, J. A. (2007). On the taxonomic status of the intracellular bacterium *Wolbachia pipientis*: should this species name include the intracellular bacteria of filarial nematodes? International Journal of Systematic and Evolutionary Microbiology, 57, 1677-1678.

Picoult-Newberg, L., Ideker, T. E., Pohl, M. G., Taylor, S. L., Donaldson, M. A., Nickerson, D. A. & Boyce-Jacino, M. (1999). Mining SNPs from EST databases. Genome Research, 9, 167-174.

Poinar, G., Kerp, H. & Hass, H. (2008). *Palaeonema phyticum* gen. n., sp n. (Nematoda : Palaeonematidae fam. n.), a Devonian nematode associated with early land plants. Nematology, 10, 9-14.

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

Popeijus, M., Blok, V. C., Cardle, L., Bakker, E., Phillips, M. S., Helder, J., Smant, G. & Jones, J. T. (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.

Popper, Z. A. (2008). Evolution and diversity of green plant cell walls. Current Opinion in Plant Biology, 11, 286-292.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. Bioinformatics, 14, 817-818.

Prior, A., Jones, J. T., Blok, V. C., Beauchamp, J., McDermott, L., Cooper, A. & Kennedy, M. W. (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. Biochemical Journal, 356, 387-394.

Qin, L., Kudla, U., Roze, E. H. A., Goverse, A., Popeijus, H., Nieuwland, J., Overmars, H., Jones, J. T. et al. (2004). Plant degradation: A nematode expansin acting on plants. Nature, 427, 30.

Qin, L., Overmars, B., Helder, J., Popeijus, H., van der Voort, J. R., Groenink, W., van Koert, P., Schots, A. et al. (2000). An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*. Molecular Plant-Microbe Interactions, 13, 830-836.

Quénéhervé, P., Salmon, F., Topart, P. & Horry, J. P. (2009a). Nematode resistance in bananas: screening results on some new *Mycosphaerella* resistant banana hybrids. Euphytica, 165, 137-143.

Quénéhervé, P., Valette, C., Topart, P., du Montcel, H. T. & Salmon, F. (2009b). Nematode resistance in bananas: screening results on some wild and cultivated accessions of *Musa* spp. Euphytica, 165, 123-136.

Quillet, L., Barray, S., Labedan, B., Petit, F. & Guespinmichel, J. (1995). The gene encoding the beta-1,4-endoglucanase (Cela) from *Myxococcus xanthus* - Evidence for independent acquisition by horizontal transfer of binding and catalytic domains from Actinomycetes. Gene, 158, 23-29.

Rasgon, J. L. & Scott, T. W. (2004). Phylogenetic characterization of *Wolbachia* symbionts infecting *Cimex lectularius* L. and *Oeciacus vicarius* Horvath (Hemiptera : Cimicidae). Journal of Medical Entomology, 41, 1175-1178.

Rehman, S., Butterbach, P., Popeijus, H., Overmars, H., Davis, E. L., Jones, J. T., Goverse, A., Bakker, J. et al. (2009a). Identification and characterization of the most abundant cellulases in stylet secretions from *Globodera rostochiensis*. Phytopathology, 99, 194-202.

Rehman, S., Postma, W., Tytgat, T., Prins, P., Qin, L., Overmars, H., Vossen, J., Spiridon, L. N. et al. (2009b). 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. Molecular Plant-Microbe Interactions, 22, 330-340.

Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: The European molecular biology open software suite. Trends in Genetics, 16, 276-277.

Rogers, A., Antoshechkin, I., Bieri, T., Blasiar, D., Bastiani, C., Canaran, P., Chan, J., Chen, W. J. et al. (2008). WormBase 2007. Nucleic Acids Research, 36, D612-D617.

Rogozin, I. B., Lyons-Weiler, J. & Koonin, E. V. (2000). Intron sliding in conserved gene families. Trends in Genetics, 16, 430-432.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19, 1572-1574.

Rosso, M. N., Favery, B., Piotte, C., Arthaud, L., de Boer, J. M., Hussey, R. S., Bakker, J., Baum, T. J. et al. (1999). Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. Molecular Plant-Microbe Interactions, 12, 585-591.

Rowley, S. M., Raven, R. J. & Mcgraw, E. A. (2004). *Wolbachia pipientis* in Australian spiders. Current Microbiology, 49, 208-214.

Sampedro, J. & Cosgrove, D. J. (2005). The expansin superfamily. Genome Biology, 6, 242-250.

Sarah, J. L., Pinochet, J. & Stanton, J. (1996). The burrowing nematode of bananas, *Radopholus similis* Cobb, 1913. International Network for the Improvement of Banana and Plantain, Musa Pest Fact Sheet, 1.

Sato, Y., Niimura, Y., Yura, K. & Go, M. (1999). Module-intron correlation and intron sliding in family F/10 xylanase genes. Gene, 238, 93-101.

Schafer, U. A., Reed, D. W., Hunter, D. G., Yao, K., Weninger, A. M., Tsang, E. W. T., Reaney, M. J. T., MacKenzie, S. L. et al. (1999). An example of intron junctional sliding in the gene families encoding squalene monooxygenase homologues in *Arabidopsis thaliana* and *Brassica napus*. Plant Molecular Biology, 39, 721-728.

Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics, 18, 502-504.

Scholl, E. H., Thorne, J. L., McCarter, J. P. & Bird, D. M. (2003). Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. Genome Biology, 4, R39.

Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America, 95, 5857-5864.

Semblat, J. P., Rosso, M. N., Hussey, R. S., Abad, P. & Castagnone-Sereno, P. (2001). Molecular cloning of a cDNA encoding an amphid-secreted putative avirulence protein from the root-knot nematode *Meloidogyne incognita*. Molecular Plant-Microbe Interactions, 14, 72-79.

Sharp, P. M. & Li, W. H. (1987). The Codon Adaptation Index - A measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research, 15, 1281-1295.

Shin, H., Hirst, M., Bainbridge, M. N., Magrini, V., Mardis, E., Moerman, D. G., Marra, M. A., Baillie, D. L. et al. (2008). Transcriptome analysis for *Caenorhabditis elegans* based on novel expressed sequence tags. BMC Biology, 6,

Sijmons, P. C., Grundler, F. M. W., Vonmende, N., Burrows, P. R. & Wyss, U. (1991). *Arabidopsis thaliana* as a new model host for plant-parasitic nematodes. Plant Journal, 1, 245-254.

Sindhu, A. S., Maier, T. R., Mitchum, M. G., Hussey, R. S., Davis, E. L. & Baum, T. J. (2009). Effective and specific in planta RNAi in cyst nematodes: expression interference of four parasitism genes reduces parasitic success. Journal of Experimental Botany, 60, 315-324.

Smant, G., Stokkermans, J. P. W. G., Yan, Y. T., de Boer, J. M., Baum, T. J., Wang, X. H., Hussey, R. S., Gommers, F. J. et al. (1998). Endogenous cellulases in animals: Isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. Proceedings of the National Academy of Sciences of the United States of America, 95, 4906-4911.

Somerville, C. (2006). Cellulose synthesis in higher plants. Annual Review of Cell and Developmental Biology, 22, 53-78.

Spurr, A. R. (1969). A low-viscosity epoxy resin embedding medium for electron microscopy. Journal of Ultrastructure Research, 26, 31-43.

Steeves, R. M., Todd, T. C., Essig, J. S. & Trick, H. N. (2006). Transgenic soybeans expressing siRNAs specific to a major sperm protein gene suppress *Heterodera glycines* reproduction. Functional Plant Biology, 33, 991-999.

Stoltzfus, A., Logsdon, J. M., Palmer, J. D. & Doolittle, W. F. (1997). Intron "sliding" and the diversity of intron positions. Proceedings of the National Academy of Sciences of the United States of America, 94, 10739-10744.

Subbotin, S. A., Krall, E. L., Riley, I. T., Chizhov, V. N., Staelens, A., De Loose, M. & Moens, M. (2004). Evolution of the gall-forming plant parasitic nematodes (Tylenchida : Anguinidae) and their relationships with hosts as inferred from Internal Transcribed Spacer sequences of nuclear ribosomal DNA. Molecular Phylogenetics and Evolution, 30, 226-235.

Subbotin, S. A., Sturhan, D., Chizhov, V. N., Vovlas, N. & Baldwin, J. G. (2006). Phylogenetic analysis of Tylenchida Thorne, 1949 as inferred from D2 and D3 expansion fragments of the 28S rRNA gene sequences. Nematology, 8, 455-474.

Subramaniyan, S. & Prema, P. (2002). Biotechnology of microbial xylanases: Enzymology, molecular biology, and application. Critical Reviews in Biotechnology, 22, 33-64.

Sugimura, M., Watanabe, H., Lo, N. & Saito, H. (2003). Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle, *Psacothea hilaris*. European Journal of Biochemistry, 270, 3455-3460.

Sukno, S. A., McCuiston, J., Wong, M. Y., Wang, X. H., Thon, M. R., Hussey, R., Baum, T. & Davis, E. (2007). Quantitative detection of double-stranded RNA-mediated gene silencing of parasitism genes in *Heterodera glycines*. Journal of Nematology, 39, 145-152.

Suzuki, T., Ibata, K., Hatsu, M., Takamizawa, K. & Kawai, K. (1997). Cloning and expression of a 58-kDa xylanase VI gene (xynD) of *Aeromonas caviae* ME-1 in *Escherichia coli* which is not categorized as a family F or family G xylanase. Journal of Fermentation and Bioengineering, 84, 86-89.

Sverdlov, A. V., Babenko, V. N., Rogozin, I. B. & Koonin, E. V. (2004). Preferential loss and gain of introns in 3 ' portions of genes suggests a reverse-transcription mechanism of intron insertion. Gene, 338, 85-91.

Sverdlov, A. V., Csuros, M., Rogozin, I. B. & Koonin, E. V. (2007). A glimpse of a putative pre-intron phase of eukaryotic evolution. Trends in Genetics, 23, 105-108.

Sverdlov, A. V., Rogozin, I. B., Babenko, V. N. & Koonin, E. V. (2003). Evidence of splice signal migration from exon to intron during intron evolution. Current Biology, 13, 2170-2174.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony. Version 4. (Sinauer Associates, Sunderland, Massachusetts, USA).

Taylor, N. G. (2008). Cellulose biosynthesis and deposition in higher plants. New Phytologist, 178, 239-252.

Teixeira, L., Ferreira, A. & Ashburner, M. (2008). The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. Plos Biology, 6, 2753-2763.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22, 4673-4680.

Trinh, P. Q., Nguyen, C. N., Waeyenberge, L., Subbotin, S. A., Karssen, G. & Moens, M. (2004). *Radopholus arabocoffeae* sp n. (Nematoda : Pratylenchidae), a nematode pathogenic to *Coffea arabica* in Vietnam, and additional data on *R-duriophilas*. Nematology, 6, 681-693.

Tsai, K. H., Huang, C. G., Wang, L. C., Yu, Y. W., Wu, W. J. & Chen, W. J. (2007). Molecular evidence for the endosymbiont *Wolbachia* in a non-filaroid nematode, *Angiostrongylus cantonensis*. Journal of Biomedical Science, 14, 607-615.

Tyson, T., Reardon, W., Browne, J. A. & Burnell, A. M. (2007). Gene induction by desiccation stress in the entomopathogenic nematode *Steinernema carpocapsae* reveals parallels with drought tolerance mechanisms in plants. International Journal for Parasitology, 37, 763-776.

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.

Valette, C., Mounport, D., Nicole, M., Sarah, J. L. & Baujard, P. (1998). Scanning electron microscope study of two African populations of *Radopholus similis* (Nematoda : Pratylenchidae) and proposal of *R. citrophilus* as a junior synonym of *R. similis*. Fundamental and Applied Nematology, 21, 139-146.

Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E. A., Menck, C. F. M., da Silva, A. C. R., Ferro, J. A., Oliveira, M. C., Setubal, J. C. et al. (2002). Comparative genomic analysis of plant-associated bacteria. Annual Review of Phytopathology, 40, 169-189.

Vandekerckhove, T. T. M., Watteyne, S., Willems, A., Swing, J. G., Mertens, J. & Gillis, M. (1999). Phylogenetic analysis of the 16S rDNA of the cytoplasmic bacterium *Wolbachia* from the novel host *Folsomia candida* (Hexapoda, Collembola) and its implications for wolbachial taxonomy. Fems Microbiology Letters, 180, 279-286.

Vandekerckhove, T. T. M., Willems, A., Gillis, M. & Coomans, A. (2000). Occurrence of novel verrucomicrobial species, endosymbiotic and associated with parthenogenesis in *Xiphinema americanum*-group species (Nematoda, Longidoridae). International Journal of Systematic and Evolutionary Microbiology, 50, 2197-2205.

Vanholme, B., De Meutter, J., Tytgat, T., Gheysen, G. D. C., Vanhoutte, I. & Gheysen, G. D. R. (2002). An improved method for whole-mount *in situ* hybridization of *Heterodera schachtii* juveniles. Parasitology Research, 88, 731-733.

Vanholme, B., De Meutter, J., Tytgat, T., Van Montagu, M., Coomans, A. & Gheysen, G. (2004). Secretions of plant-parasitic nematodes: a molecular update. Gene, 332, 13-27.

Vanholme, B., Haegeman, A., Jacob, J., Cannoot, B. & Gheysen, G. (2009). Arabinogalactan endo-1,4-beta-galactosidase: a putative plant cell wall-degrading enzyme of plant-parasitic nematodes. Nematology, doi: 10.1163/156854109X404599.

Vanholme, B., Mitreva, M., Van Criekinge, W., Logghe, M., Bird, D., McCarter, J. P. & Gheysen, G. (2006). Detection of putative secreted proteins in the plant-parasitic nematode *Heterodera schachtii*. Parasitology Research, 98, 414-424.

Vanholme, B., Van Thuyne, W., Vanhouteghem, K., De Meutter, J., Cannoot, B. & Gheysen, G. (2007). Molecular characterization and functional importance of pectate lyase secreted by the cyst nematode *Heterodera schachtii*. Molecular Plant Pathology, 8, 267-278.

Venter, C., De Waele, D. & Meyer, A. J. (1991). Reproductive and damage potential of *Ditylenchus destructor* on peanut. Journal of Nematology, 23, 12-19.

Venter, C., De Waele, D. & Meyer, A. J. (1992). Minimizing damage by *Ditylenchus destructor* to peanut seed with early harvest. Journal of Nematology, 24, 528-532.

Venter, C., De Waele, D. & Meyer, A. J. (1993). Reproductive and damage potential of *Ditylenchus destructor* on 6 peanut cultivars. Journal of Nematology, 25, 59-62.

Venter, C., Vanaswegen, G., Meyer, A. J. & De Waele, D. (1995). Histological studies of *Ditylenchus africanus* within peanut pods. Journal of Nematology, 27, 284-291.

Veronico, P., Jones, J., Di Vito, M. & De Giorgi, C. (2001). Horizontal transfer of a bacterial gene involved in polyglutamate biosynthesis to the plant-parasitic nematode *Meloidogyne artiellia*. Febs Letters, 508, 470-474.

Wang, D. Y. C., Kumar, S. & Hedges, S. B. (1999). Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proceedings of the Royal Society of London Series B-Biological Sciences, 266, 163-171.

Wang, J. P. Z., Lindsay, B. G., Cui, L. Y., Wall, P. K., Marion, J., Zhang, J. X. & dePamphilis, C. W. (2005). Gene capture prediction and overlap estimation in EST sequencing from one or multiple libraries. BMC Bioinformatics, 6, 300.

Wang, X., Li, H. M., Hu, Y. J., Fu, P. & Xu, H. H. (2007). Molecular cloning and analysis of a new venom allergen-like protein gene from the root-knot nematode *Meloidogyne incognita*. Experimental Parasitology, 117, 133-140.

Wasmuth, J. D. & Blaxter, M. L. (2004). Prot4EST: Translating expressed sequence tags from neglected genomes. BMC Bioinformatics, 5, 187.

Wei, Y. D., Lee, K. S., Gui, Z. Z., Yoon, H. J., Kim, I., Zhang, G. Z., Guo, X. J., Sohn, H. D. et al. (2006). Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry longicorn beetle, *Apriona germari*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 145, 220-229.

Wendt, K. R., Swart, A., Vrain, T. C. & Webster, J. M. (1995). *Ditylenchus africanus* sp-N from South-Africa - A morphological and molecular characterization. Fundamental and Applied Nematology, 18, 241-250.

Werren, J. H., Baldo, L. & Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate biology. Nature Reviews Microbiology, 6, 741-751.

Werren, J. H., Zhang, W. & Guo, L. R. (1995). Evolution and phylogeny of *Wolbachia* - Reproductive parasites of arthropods. Proceedings of the Royal Society of London Series B-Biological Sciences, 261, 55-63.

Wu, M., Sun, L. V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J. C., Mcgraw, E. A., Martin, W. et al. (2004). Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: A streamlined genome overrun by mobile genetic elements. Plos Biology, 2, 327-341.

Wylie, T., Martin, J. C., Dante, M., Mitreva, M. D., Clifton, S. W., Chinwalla, A., Waterston, R. H., Wilson, R. K. et al. (2004). Nematode.net: a tool for navigating sequences from parasitic and free-living nematodes. Nucleic Acids Research, 32, D423-D426.

Xu, B. Z., Janson, J. C. & Sellos, D. (2001). Cloning and sequencing of a molluscan endo-beta-1,4-glucanase gene from the blue mussel, *Mytilus edulis*. European Journal of Biochemistry, 268, 3718-3727.

Yadav, B. C., Veluthambi, K. & Subramaniam, K. (2006). Host-generated double stranded RNA induces RNAi in plant-parasitic nematodes and protects the host from infection. Molecular and Biochemical Parasitology, 148, 219-222.

Yan, Y. T., 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. Molecular Plant-Microbe Interactions, 14, 63-71.

Yan, Y. T., Smant, G., Stokkermans, J., Qin, L., Helder, J., Baum, T., Schots, A. & Davis, E. (1998). Genomic organization of four beta-1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications. Gene, 220, 61-70.

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

Yoshihama, M., Nakao, A., Nguyen, H. D. & Kenmochi, N. (2006). Analysis of ribosomal protein gene structures: Implications for intron evolution. Plos Genetics, 2, 237-242.

Young, A. R. & Meeusen, E. N. (2002). Galectins in parasite infection and allergic inflammation. Glycoconjugate Journal, 19, 601-606.

Zdobnov, E. M. & Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics, 17, 847-848.

Zeh, D. W., Zeh, J. A. & Bonilla, M. M. (2005). *Wolbachia*, sex ratio bias and apparent male killing in the harlequin beetle riding pseudoscorpion. Heredity, 95, 41-49.

Nawoord

Vier jaar geleden kwam een einde aan mijn studententijd en begon ik met een bang hartje aan een doctoraat. Gelukkig was wetenschappelijk onderzoek wel mijn ding en kon ik goed aarden in het labo. Met al die pas afgestudeerde collega's bleek de sfeer bovendien opperbest, en leek het een beetje op een verlenging van mijn studententijd. Natuurlijk waren er ook tegenslagen en teleurstellingen, maar dan stond er altijd iemand klaar om mijn experimenten wat bij te sturen of om mij wat op te beuren. En nu is het dan zover: mijn doctoraat is klaar. Bij deze neem ik dan ook afscheid van mijn periode als doctoraatsstudent, maar niet zonder enkele mensen te bedanken.

Eerst en vooral bedank ik mijn promotor, prof. Gheysen, voor alle steun en vertrouwen. Lieve, je hebt altijd in me geloofd en me mijn ding laten doen. Ik kon ook altijd bij je terecht om te praten over wetenschappelijke resultaten of over hoe het verder moest.

Vooral tijdens de eerste jaren van mijn doctoraat was Bartel mijn grote voorbeeld. Gebeten door het onderzoek was hij altijd druk bezig, maar toch maakte hij graag tijd vrij om me een nieuwe techniek te leren of om te discussiëren over mislukte experimenten. Gaandeweg kon ik meer en meer mijn eigen boontjes doppen, maar toch zijn onze wetenschappelijke discussies gebleven. Bedankt Bartel, zonder jou was mijn doctoraat ongetwijfeld een stuk minder dik geweest. En ja hoor, je wordt gemist op het labo!

Gelukkig zijn er nog veel andere "nema"-collega's om op terug te vallen. Tina bijvoorbeeld, die van het papaja-team overstapte naar het nema-team. Het was zij die me tijdens mijn thesis warm maakte voor een doctoraat. Ze heeft me ook altijd geholpen en met raad en daad bijgestaan. Tina, ik heb echt genoten van onze samenwerking en evenzeer van onze gesprekken. Bedankt dat je er altijd voor me was als ik eens mijn frustraties kwijt wou.

Mijn eerste stapjes in de nematodenwereld deed ik niet alleen, maar samen met Joachim. Naast vrienden werden we nu ook collega's, en samen leerden we wortels schillen en nematoden kweken. Joachim, ik mocht je altijd lastigvallen met een labo- of bio-informaticaprobleem. Bovendien hebben we samen veel plezier gemaakt, niet alleen in het labo, maar ook op congres of in onze vrije tijd. Bedankt daarvoor!

Het nema-team zou natuurlijk niet compleet zijn zonder Bernard en Wim. Als ik het te druk had, kon ik rekenen op Bernard om de nematodenstock of een miniprep te doen. Wim kwam altijd met goede ideeën aanzetten nadat ik mijn resultaten had voorgesteld op een nema-meeting. Voor de administratie en allerlei andere praktische problemen kon ik telkens terugvallen op Sofie en Fien, terwijl Winnok altijd bereid was om me weer eens te helpen aan de microscoop. Natuurlijk mocht er ook wel eens leute gemaakt worden, en dat was nooit een probleem met Birger in de buurt. Birger, in de loop der jaren ben je echt mijn bureaumaatje geworden en samen hebben we veel gelachen, naar muziek geluisterd, geroddeld of genoten van een glaasje na het werk. 't Maakte de doorsnee werkdag een stuk aangenamer, merci! Daarnaast is er ook nog de bende "van beneden": de biobix'ers, glyco's en immuno's. 't Was altijd leuk om over de middag nog eens samen te lunchen of om een praatje te slaan in de gang. Bovendien kon ik beneden altijd wel iemand vinden die me kon bijstaan met eiwitwerk of bio-informatica (merci Maté en Tom).

Naast mijn collega's wil ik ook mijn ouders bedanken voor alle kansen die ze me gegeven hebben. En natuurlijk Stijn, bedankt voor alle steun en liefde. We vonden elkaar tijdens ons doctoraat, en konden altijd bij elkaar ons hart luchten over de (meestal gelijkaardige) problemen in onze labo's. Nu mijn doctoraat klaar is, kan ik eindelijk beginnen helpen met het jouwe ;-)

Annelies

Curriculum vitae

Personalia

Annelies Haegeman
Birth date/place: 02/11/1982, Oostende
Nationality: Belgian
Personal e-mail: annelies.haegeman@gmail.com

Ghent University, Faculty of Bioscience Engineering
Department of Molecular Biotechnology
Coupure links 653
9000 Ghent, Belgium
e-mail: annelies.haegeman@ugent.be
tel: 003292645971
fax: 003292646219

Education

Oct 2005 – Jan 2009	PhD training Ghent University, Faculty of Bioscience Engineering Coupure links 653 9000 Ghent
Oct 2000 – June 2005	Bio-engineer in cell and gene biotechnology, great distinction Ghent University, Faculty of Bioscience Engineering Coupure links 653 9000 Ghent Thesis: Microsatellite analysis of the genetic diversity and relationships within the genus *Vasconcellea* (Caricaceae) Promotor: Prof. dr. G. Gheysen
Sept 1994 – June 2000	High school, science – mathematics Sint-Pieterscollege Weststraat 86 8370 Blankenberge

Professional career

Jan 2006 – Dec 2009 IWT fellowship (Institute for the Promotion of Innovation by Science and Technology in Flanders)
Promotor: Prof. dr. G. Gheysen
Project: "Molecular analysis of migratory plant-parasitic nematodes with a focus on plant cell wall modifying enzymes"
Ghent University, Faculty of Bioscience Engineering
Department of Molecular Biotechnology
Coupure links 653
9000 Ghent, Belgium

Oct 2005 – Dec 2005 PhD student on GOA project 01G00805 (Ghent University)

Publications with peer reviewing

- **Haegeman, A.**; Jacob, J.; Vanholme, B.; Kyndt, T.; Mitreva, M. & Gheysen, G. (2009). Expressed sequence tags of the peanut pod nematode *Ditylenchus africanus*: the first transcriptome analysis of an Anguinid nematode. Molecular and Biochemical Parasitology, doi: 10.1016/j.molbiopara.2009.04.004.
- **Haegeman, A.**; Vanholme, B.; Jacob, J.; Vandekerckhove, T.T.M.; Claeys, M.; Borgonie, G. & Gheysen, G. (2009). An endosymbiotic bacterium in a plant-parasitic nematode: member of a new *Wolbachia* supergroup. International Journal for Parasitology, doi: 10.1016/j.ijpara.2009.01.006.
- Vanholme, B.; **Haegeman, A.**; Jacob, J.; Cannoot, B. & Gheysen, G. (2009). Arabinogalactan endo-1,4-β-galactosidase: a putative plant cell wall degrading enzyme of plant-parasitic nematodes. Nematology, doi: 10.1163/156854109X404599.
- **Haegeman, A.**; Vanholme, B. & Gheysen, G. (2009). Characterization of a putative endoxylanase in the migratory plant-parasitic nematode *Radopholus similis*. Molecular Plant Pathology, 10, 389-401.
- Vandegehuchte, M.B., Kyndt, T., Vanholme, B., **Haegeman, A.**, Gheysen, G. & Janssen, C. (2009). Occurrence of DNA methylation in *Daphnia magna* and influence of multigeneration Cd exposure. Environment International, 35, 700-706.
- Vanholme, B.; Kast, P.; **Haegeman, A.**; Jacob, J.; Grünewald, 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. Molecular Plant Pathology, 10, 189-200.

- Kyndt, T.; **Haegeman, A.** & Gheysen, G. (2008). Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: no evidence for an early domain shuffling event. BMC Evolutionary Biology, 8, 305.
- **Haegeman, A.**; Jacob, J.; Vanholme, B.; Kyndt, T. & Gheysen, G. (2008). A family of GHF5 endo-1,4-beta-glucanases in the migratory plant-parasitic nematode *Radopholus similis*. Plant Pathology, 57, 581-590.
- 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.
- Kyndt, T.; Van Droogenbroeck, B.; **Haegeman, A.**; Roldán-Ruiz, I. & Gheysen, G. (2006). Cross-species microsatellite amplification in *Vasconcellea* and related genera and their use in germplasm identification. Genome, 49, 786-798.
- Kyndt, T.; **Haegeman, A.**; Van Glabeke, S.; Maertens, I.; Van Droogenbroeck, B.; Roldán-Ruiz, I. & Gheysen, G. (2005). Isolation and characterisation of microsatellite loci in the highland papaya *Vasconcellea* × *heilbornii* V. Badillo (Caricaceae). Molecular Ecology Notes, 5, 590-592.
- Van Droogenbroeck, B.; Maertens, I.; **Haegeman, A.**; Kyndt, T.; O'Brien, C.; Drew, R.A. & Gheysen, G. (2005). Maternal inheritance of cytoplasmic organelles in intergeneric hybrids of *Carica papaya* L. and *Vasconcellea* spp. (Caricaceae Dumort., Brassicales). Euphytica, 143,161-168.

Posters

- Jacob, J.; **Haegeman, A.**; Kyndt, T.; Vanholme, B. & Gheysen, G. In search of parasitism genes in the burrowing nematode *Radopholus similis*. COST 872 meeting (Exploiting genomics to understand plant-nematode interactions); 09-11/05/07; La Colle-Sur-Loup, France.

Meetings with oral presentation

- **Haegeman, A.**; Vanholme, B.; Jacob, J.; Kyndt, T. & Gheysen, G. Cell-wall degrading enzymes in the migratory nematode *Radopholus similis*. COST 872 meeting (Exploiting genomics to understand plant-nematode interactions); 26-29/05/08; Postojna, Slovenia.

- **Haegeman, A.;** Vanholme, B.; Jacob, J.; Kyndt, T. & Gheysen, G. Cell-wall degrading enzymes in the migratory nematode *Radopholus similis*. 60th International Symposium on Crop Protection; 20/05/08; Ghent, Belgium.
- **Haegeman A**. GHF5 endo-1,4-beta-glucanases in migratory endoparasitic nematodes. IMPNIG Meeting; 10-12/09/07; Wageningen, the Netherlands.

Other meetings

- 28th International Symposium of the European Society of Nematologists; 05-09/06/06; Blagoevgrad, Bulgary
- 58th International Symposium on Crop Protection; 23/05/06; Ghent, Belgium
- Life, A Nobel Story; 28/04/04; Brussels Expo, Brussels, Belgium

Supervision of master students and practical exercises

- Maria Peñuelas Hortelano (Spain), "Suitability of potential host plants for *Ditylenchus africanus* and molecular analysis of parasitism genes", academic year 2007-2008.
- Practical exercises "Molecular techniques in Nematology" (Master of Nematology) & "Molecular techniques" (Master of Nutrition and Rural Development, main subject: Tropical Agriculture), academic years 2006-2007, 2007-2008, 2008-2009.

PR-activities

- open day Faculty of Bioscience Engineering, 14/03/09
- Flemish Science Week, 17-26/10/08
- Biotechnology day, Zwijnaarde technologiepark, 21/06/08
- Alumni day Ghent University: celebration of 20 years vzw Alumni UGent, 21/10/06
- Flemish Science Week, 19-29/10/06
- open day Faculty of Bioscience Engineering, 18/03/06