Genomics 97 (2011) 29–36

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Data-mining of the *Meloidogyne incognita* degradome and comparative analysis of proteases in nematodes

Philippe Castagnone-Sereno *, Emeline Deleury, Etienne G.J. Danchin, Laetitia Perfus-Barbeoch, Pierre Abad

INRA UMR1301/UNSA/CNRS UMR6243, 400 route des Chappes, BP167, 06903 Sophia Antipolis cedex, France

ARTICLE INFO

Article history:
Received 29 July 2010
Accepted 7 October 2010
Available online 14 October 2010

Keywords:
Comparative genomics
Degradome
Proteases
Root-knot nematode

ABSTRACT

Proteases perform essential physiological functions in all living organisms. In parasitic helminths, they are of particular importance for tissue penetration, digestion of host tissues for nutrition, and evasion of host immune responses. The recent availability of the genome sequence of the nematode *Meloidogyne incognita* has allowed the analysis of the protease repertoire of this major crop pathogen. The *M. incognita* degradome consists of at least 334 proteases that are distributed into 43 families of the five known catalytic classes. Expression profiling identified protease genes with a differential transcript level between eggs and infective juveniles. Comparing the *M. incognita* degradome with those of five other nematodes showed discrepancies in the distribution of some protease families, including large expansion in some families, that could reflect specific aspects of the parasitic lifestyle of this organism. This comparative study should provide a framework for deciphering the diversity of protease-mediated functions in nematodes.

1. Introduction

Proteases comprise a large group of structurally and functionally diverse enzymes that share the common ability to catalyze the hydrolysis of peptide bonds for either recycling polypeptides into their constitutive amino acids or mediating selective polypeptide cleavage for post-translational modification [1]. The MEROPS database provides a catalogue and classification of peptidases (i.e. proteolytic enzymes or proteases) [2]. Based on their catalytic substrate and structural similarity, proteases are divided into five distinct major clans, i.e. aspartic, cysteine, metallo, serine and threonine proteases. Families divide clans by common ancestry, while subfamilies have common structure yet unclear ancestry. This classification provides a comprehensive support to decipher peptidase function in a wide range of organisms. The extensive biological implications of this large set of proteins with a common biochemical function led to the concept of proteases as a distinct subset of the proteome. Thus, the degradome of an organism was defined as the complete set of proteases in that organism [3].

The phylum Nematoda comprises over 25,000 described species (with perhaps 10 million undescribed), many of which are parasites of animals or plants [4]. Plant-parasitic nematodes are responsible for an estimated 100 billion euros annually in crop damage worldwide. Among them, the Southern root-knot nematode, *Meloidogyne incognita*, is able to infect the roots of almost all cultivated plants, which possibly renders this species the most damaging crop pathogen in the world [5]. *M. incognita* is an obligatory, sedentary endoparasite which has evolved an intimate interaction with its hosts. Second-stage juveniles invade the root in the zone of elongation, and migrate intercellularly, to the vascular cylinder, where permanent feeding sites are established. After three additional moults, adult females induce the redifferentiation of these root cells into specialized cells, upon which they feed continuously. *M. incognita* can infect *Arabidopsis thaliana*, making it a key model system for the understanding of adaptations to plant parasitism [6]. In nematodes, proteases have important roles in a variety of physiological processes, e.g. embryogenesis [7] or cuticle remodeling during larval moulting [8]. In addition, proteases are involved in several aspects of the parasitic life style in parasitic helminths, among which tissue penetration, digestion of host tissues for nutrition and evasion of host immune responses [9,10]. In plant-parasitic nematodes, studies have been focused mainly on cysteine and serine digestive proteases [11–14]. Only very recently an aspartic protease has been characterized in the root-knot nematode species *M. incognita* [15], but its function remains unknown. So far, a comprehensive view of the degradome of a plant-parasitic nematode is not available.

The recent completion of large-scale genome-sequencing projects has provided new opportunities to evaluate and compare the complexity of degradomes in nematodes. Besides the genomes of the model free-living species *Caenorhabditis elegans* and *Caenorhabditis briggsae* [16,17], additional whole-genome data are now available for nematodes with very different life styles, i.e. the necromenic species *Pristionchus pacificus* [18], the plant-parasitic species *M. incognita* and *M. hapla* [19,20], and the animal-parasitic species *Brugia*

* Corresponding author. Fax: +33 492386587.
E-mail address: philippe.castagnone@sophia.inra.fr (P. Castagnone-Sereno).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.10.002

malayi [21]. Based on these genomic resources, we report here the first comprehensive survey of proteases in a plant-parasitic nematode, with the analysis of a total of 334 putative protease-coding genes from the root-knot nematode *M. incognita*. We also report the preliminary results of a comparative analysis of the *M. incognita* degradome with those of other nematodes with different life styles. We further performed an expression profiling of *M. incognita* proteases based on public and home-made EST resources, to confirm active transcription and to examine their expression patterns. Finally, we discuss the potential evolutionary relevance of these studies in relation to the particular life style of plant-parasitic nematodes.

2. Materials and methods

2.1. Bioinformatic analyses

The complete set of predicted gene models encoding proteases that resulted from the previous annotation of the *M. incognita genome* [19] was manually checked, and predicted proteins that were shorter than 100 residues or lacked entire conserved domains identified by searching the PFAM database (http://pfam.sanger.ac.uk/) [22] were not retained. The remaining sequences were classified into families according to the MEROPS database [2], and characterized according to a number of further analyses. Putative protease sequences were analyzed for the presence of a N-terminal secretion signal peptide using the program SignalP with the HMM method (probability cutoff = 90%) [23]. The non redundant database at National Center for Biotechnology Information (NCBI at http://www.ncbi.nlm.nih.gov) was searched for homologs using each single *M. incognita* protease as query sequence in a BLASTP analysis [24] with default parameters.

Complete predicted protein sets from *C. elegans*, *C. briggsae*, *B. malayi*, *P. pacificus* and *M. hapla* were downloaded from WormBase (http://www.wormbase.org) and the Plant Nematode Genomics Group website (http://www.pngg.org). To perform an automatic bioinformatic search of proteases in these proteomes, we used HMM profiles retrieved from the PFAM database, and HMM searches were performed on the different predicted protein sets using the HMMER 2.3.2 package (http://hmmer.janelia.org). In addition, the *M. incognita* and *C. elegans* degradomes were further compared using the InterProScan program [25]. To homogenize the granularity level of annotation between both nematodes, for each non-overlapping set of domains found, we only kept the root domain, using the hierarchical organization of domains proposed in the 'Parent–Child' description available on the EBI public ftp server (ftp://ftp.ebi.ac.uk/).

2.2. EST resources

To perform expression profiling of the *M. incognita* predicted proteases, we generated a specific resource of about 40,000 ESTs from two development stages (infective J2s and eggs) which complement the *M. incognita* ESTs available at NCBI-dbEST and additional ESTs previously generated at our laboratory. After base-calling and trimming, a total of 47,377 EST sequences were validated, including 17,162 ESTs from the J2 library and 6790 ESTs from the egg library. After clustering, these 47,377 ESTs represent 11,644 unique sequences, i.e. 5943 contigs and 5701 singlets. These sequences were mapped on the genome using Genomethreader [26], and used in a local BLASTN comparison using the predicted *M. incognita* proteases as query sequences. The level of expression of the predicted proteases was deduced from the corresponding number of ESTs, the basic assumption being that the larger the number of ESTs reported per gene model, the more actively transcribed the corresponding gene. Accordingly, highly expressed genes have >3 ESTs, moderately expressed genes 2–3 ESTs, and weakly expressed genes 1 EST [27]. The significance of the differential expression of protease genes between J2 and eggs was evaluated according to the Bayesian method developed by Audric and Claverie [28].

2.3. Phylogenetic analysis

Multiple alignments were performed with the MUSCLE program [29], using standard parameters. Alignment were visually examined and edited using ClustalX [30], and were subject to phylogenetic analysis using the Maximum Parsimony approach as implemented in PAUP* [31]. All characters were run unordered and of equal weight, according to the tree-bisection-reconnection algorithm. Bootstrap resampling [32] with 1000 pseudoreplicates was carried out to assess support for each individual branch. Bootstrap values <50% were collapsed and treated as unresolved polytomies.

3. Results

3.1. Whole-genome analysis of the M. incognita degradome

By using primary information retrieved from the sequencing project of the *M. incognita* genome [19] and further manual annotation steps, we have characterized a total of 334 genes that encode proteases in this nematode (Additional file 1). In addition, several sequences related to aspartic proteases that are embedded in endogenous retroviral elements have not been included in this repertoire (data not shown). The 334 proteases identified here cover 1.74% of the 19,212 protein-coding genes predicted from the *M. incognita* genome, and belong to 43 families of all five catalytic classes. Analysis of the distribution of the proteases in these catalytic classes is shown in Table 1. Metallo and cysteine proteases are the most abundant proteolytic enzymes in *M. incognita*, with 136 and 106 members, respectively. Serine proteases also contain numerous members (52), while there are only 26 aspartic and 14 threonine proteases in the nematode. The overall abundance and diversity likely reflect the various highly specialized roles these enzymes play. Within each catalytic class, the number of members per family also is highly variable, ranging from one to 55 (i.e., the M12 family of metallo proteases; Additional file 1). Several families of cysteine proteases exhibit a large number of representatives, e.g. the C1, C19 and C48 families with 36, 31 and 20 members, respectively (Additional file 1). In contrast, there are a few families with one single member in the *M. incognita* genome (i.e., C14, C46, S14 and S59; Additional file 1). Overall, 64 proteases harbour a signal peptide at their N terminus (i.e., 19.2% of the whole degradome), indicating that they may be involved in the nematode secretory pathway. Such putatively secreted proteases belong to all catalytic classes, except threonine proteases (Table 1).

We next examined the *M. incognita* genome for the presence of local, tandem gene duplications in the degradome. Globally, 41 gene models (i.e., 12.3% of the whole degradome) were found to be involved in 19 blocks of two, three or four tandem duplications,

Table 1
Summary of the main characteristics of the putative proteases annotated from the *Meloidogyne incognita* genome.

	Catalytic class					Total
	Aspartic	Cysteine	Metallo	Serine	Threonine	
No. gene models	26	106	136	52	14	334
No. protease families	3	11	16	10	3	43
Secreted proteases[a]	12	23	15	14	0	64
EST-supported proteases	11	53	70	21	10	165
Proteases with nematode best hit[b]	16	66	124	41	12	259

[a] Based on SignalP analysis.
[b] According to BlastP search against nr.

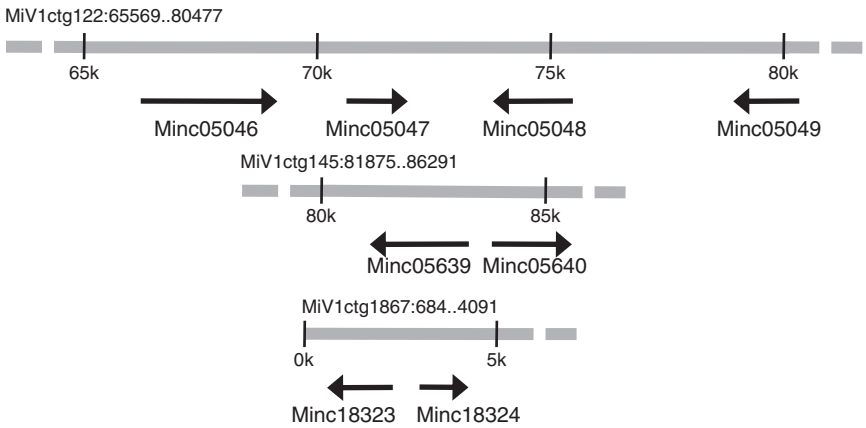

Fig. 1. Genomic distribution of the tandemly-arranged *Meloidogyne incognita* genes belonging to the C1 papain-like family. Above the grey lines are indicated the contig and the physical interval encompassing the considered genes.

belonging to all catalytic classes, except threonine proteases (Additional file 1). For example, in the cysteine class, among the 36 members of the C1 papain-like family, eight genes are involved in perfect tandem duplications (either direct or inverted) on three independent contigs (Fig. 1).

3.2. Comparative analysis with other nematode genomes

The recent availability of new nematode genomes made it possible to compare the degradomes of two free-living (*C. elegans* and *C. briggsae*), one necromenic (*P. pacificus*), one animal-parasitic (*B. malayi*) and two plant-parasitic species (*M. hapla* and *M. incognita*), although it should be emphasized here that verification and basic delineation of these newly characterized genomes are still running. From a qualitative point of view, it is interesting to note that 77.5% of the proteases from the *M. incognita* degradome have a nematode protein as best BlastP hit against the NCBI nr database (Table 1; Additional file 1). Overall, and although the number of annotated proteases is variable between the five nematode genomes (with the restriction that the *B. malayi* annotation is probably incomplete; Ghedin et al. [21]), the relative proportions of each catalytic class appear globally equivalent (Fig. 2). However, we noticed a slight expansion of cysteine proteases, and a slight depletion of metalloproteases in the *M. incognita* degradome compared to *C. elegans* (Fig. 2).

To carry out a more detailed evaluation of possible differences between the five nematode degradomes, we next performed a comparative analysis of members of each family, named according to the classification of the MEROPS database. A global view of the comparison of *M. incognita* degradome with those from the four other nematode species is presented in Fig. 3. Overall, 27 protease families appear common to the five species investigated. One single family (M50) is found in all species except in *M. incognita*, while no *M. incognita*-specific family was detected. Based on the presence–absence of proteases families, the *C. elegans* and *C. briggsae* degradomes appear strictly identical. In contrast, 11 differences were noticed between the protease family sets of *M. hapla* and *M. incognita.*

Since *C. elegans* is considered as the nematode model species, we next performed a detailed comparison, at the family level, between the degradomes of *C. elegans* and *M. incognita* (Fig. 4). Globally, a very high proportion (87.7%) of proteases families is shared between both species. However, no representatives of seven families belonging to cysteine proteases and metalloproteases could be detected in the *M. incognita* genome (C39, C50, C54, M8, M23, M49 and M50). At the quantitative level, significant differences in gene numbers were detected for some families, either over- or under-represented in *M. incognita* compared to *C. elegans* (A2, C48 and S16, and C2, M1, M13, M20, M28 and S28, respectively). When comparing further the relative abundance of the InterPro (IPR) domains in the *C. elegans* and *M. incognita* genomes, two IPR domains appear into the list of the most over-represented domains in *M. incognita*, that correspond to protease families C48 (IPR003653) and S16 (IPR008269), respectively (Table 2). After manual annotation, we validated 20 and nine genes encoding C48 and S16 proteases, respectively, in *M. incognita* (Additional file 1), compared to five and three genes, respectively, in *C. elegans*. Because of its large size and striking difference with *C. elegans*, the C48 family of sentrin/SUMO proteases in *M. incognita* deserves a particular analysis which is presented later.

3.3. Specific expansion of the C48 family of sentrin/SUMO proteases in M. incognita

With 20 genes encoding C48 proteases, *M. incognita* exhibited a significant expansion of this family compared to the other nematodes species whose genome has been sequenced, i.e. *C. elegans*, *C. briggsae*, *P. pacificus*, *B. malayi* and *M. hapla* (Fig. 5A). A phylogenetic analysis of all these nematode members of the C48 family has been performed, and resulted in a lack of resolution in deeper nodes of some clades, making it difficult to ascertain any level of functional clustering. Despite this, strongly supported clusters with unambiguous homologues in all species can be identified. In such clusters, the occurrence of one homologue for each species is observed, except for *M. incognita* where two copies are often found associated. The peculiar genome

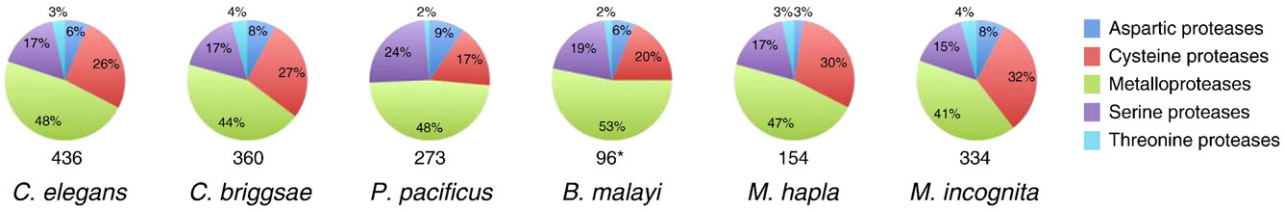

Fig. 2. Relative distribution of the five protease catalytic classes in nematode genomes. For each nematode, the total number of proteases is indicated. * indicates that the annotation of the *Brugia malayi* genome is probably incomplete [21].

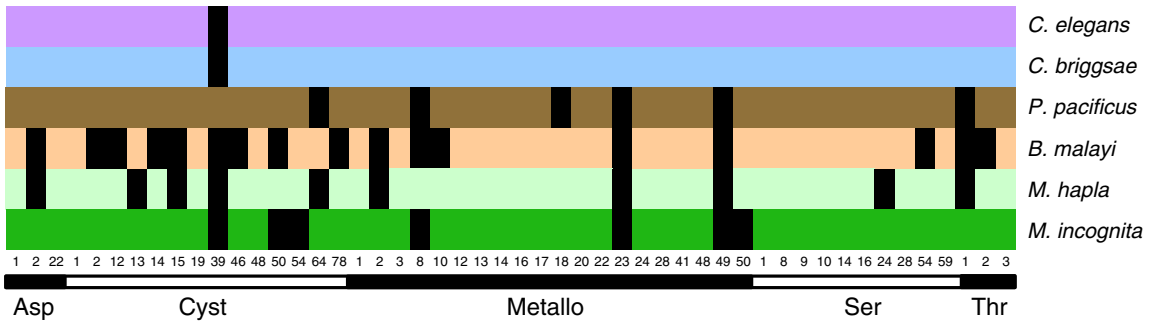

Fig. 3. Comparative view of nematode degradomes. The complete set of proteases from each species is shown, and proteases absent in one or more species are represented as black bars.

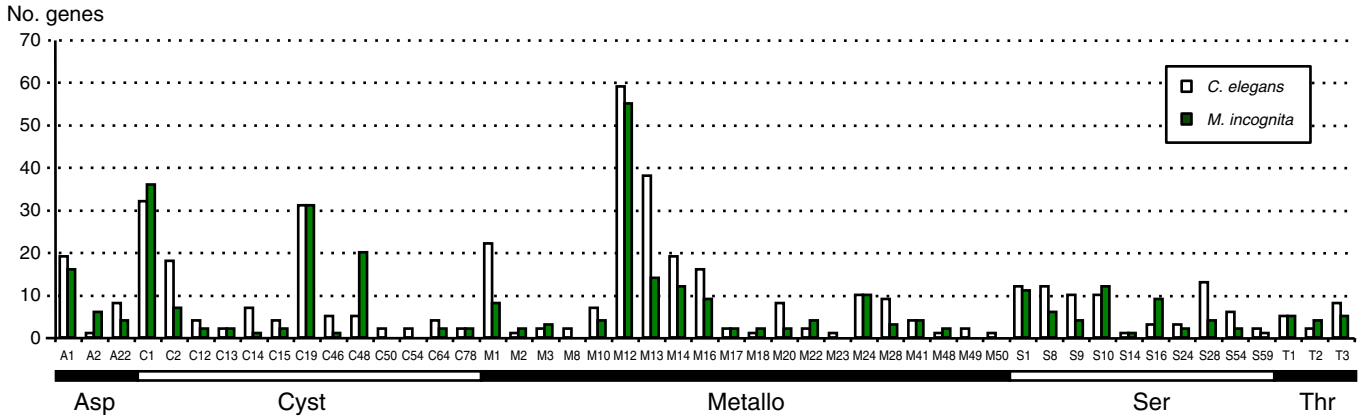

Fig. 4. Comparative distribution of proteases in the genomes of *Meloidogyne incognita* and *Caenorhabditis elegans*.

structure previously described in this species (Abad et al. [19]) probably accounts for the occurrence of these two copies. The *M. incognita* C48 protease copies in these clusters exhibited high number of exons, from 7 to 15. Among the highly bootstrap-supported clades, one is constituted exclusively by 12 *M. incognita* C48 genes, suggesting an independent expansion of one member of this family in the genome with respect to the other nematodes analyzed. All of the protease members of this cluster are characterized by the presence of a unique exon and included a number of duplicated copies. A further insight into the genesis of this specific expansion has been obtained by the analysis of the protease gene localization in genome contigs, which clearly showed the partition of these 12 copies in four different contigs (Fig. 5B). In addition, the distribution of these expansion members showed no strict correlation between the phylogenetic relationship of paralogues and their location on the different contigs. This suggests a complex evolutionary history with likely recombination events between the different contigs. Overall, the C48 proteases tree topology supports the idea that duplication events started before the separation of the different nematode species and pursued independently in root-knot nematodes with a particular emphasis in the *M. incognita* genome.

Table 2
The twenty protease InterPro (IPR) domains for which the abundance in *Meloidogyne incognita* is higher than five times the abundance in *Caenorhabditis elegans* ($n_{Mi} \geq (5\ n_{Ce}) + 1$, with n_{Mi} and n_{Ce} = number of gene models containing the IPR domain of interest in *M. incognita* and *C. elegans*, respectively).

IPR domain	*M. incognita*	*C. elegans*	IPR description
IPR004868	146 (13)	3 (124)	DNA polymerase type B
IPR011335	69 (38)	3 (124)	Restriction endonuclease, type II-like
IPR008906	66 (40)	10 (117)	HAT dimerisation
IPR012816	47 (52)	6 (121)	Conserved hypothetical protein CHP02464
IPR007012	46 (53)	3 (124)	Poly(A) polymerase
IPR007588	45 (54)	6 (121)	Zinc finger, FLYWCH-type
IPR011050	41 (58)	4 (123)	Pectin lyase fold/virulence factor
IPR006172	41 (58)	5 (122)	DNA-directed DNA polymerase B
IPR008042	40 (59)	3 (124)	Retrotransposon, Pao
IPR001547	35 (63)	1 (126)	Glycoside hydrolase, family 5
IPR011068	33 (65)	3 (124)	Nucleotidyltransferase, class I, C-terminal-like
IPR001503	32 (66)	5 (122)	Glycosyl transferase, family 10
IPR003653	26 (72)	5 (122)	Peptidase C48
IPR005312	22 (76)	4 (123)	Protein of unknown function DUF1759
IPR007177	13 (85)	1 (126)	Protein of unknown function DUF367
IPR008269	13 (85)	2 (125)	Peptidase S16
IPR007209	11 (87)	2 (125)	RNase L inhibitor
IPR007854	10 (88)	1 (126)	Fip1
IPR011526	7 (91)	1 (126)	Helix–turn–helix, Psq-like
IPR015463	6 (92)	1 (126)	RNA recognition motif, SEB4-related

3.4. Expression profiling of *M. incognita* proteases

Half of the 334 protease genes identified in the genome of *M. incognita* are supported by ESTs (Table 1; Additional file 1). An estimation of the level of expression of these genes during nematode development was carried out based on public and home-made EST resources from two different nematode life stages, i.e., eggs and infective J2s. The level of expression of each gene was quantitatively estimated by the number of ESTs corresponding to it, the basic assumption being that the number of detected ESTs per gene is a function of the transcript frequency in the populations of mRNAs. An overview of this analysis is shown in Table 3. Overall, we found 28 highly expressed genes, 57 moderately expressed ones, and 81 weakly expressed genes, that represented about half of the total repertoire. The results show that genes from the metallo and cysteine protease families represent a large majority of the expressed degradome, with 71 and 53 members, respectively. On the contrary, those from the aspartic and threonine protease families are in the

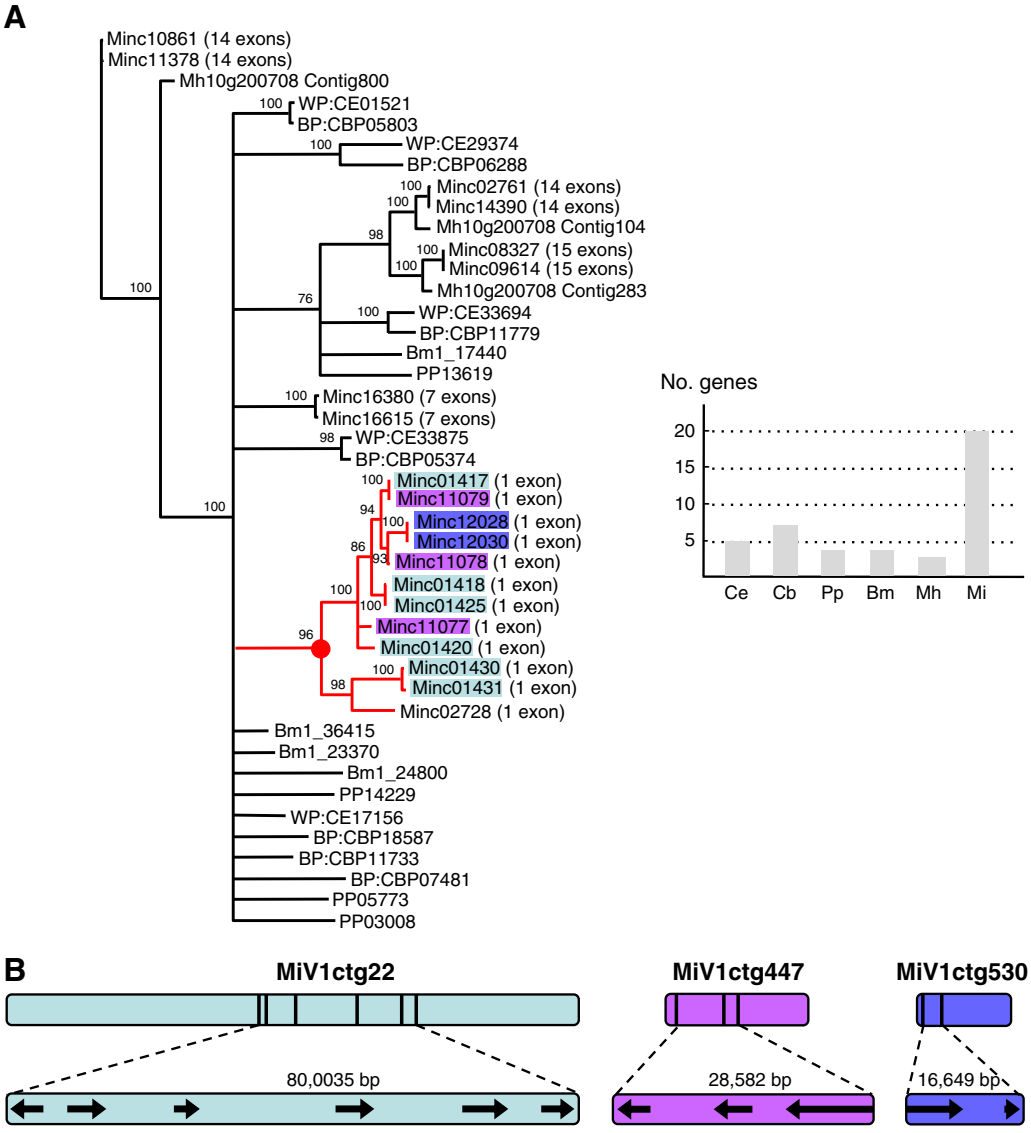

Fig. 5. Phylogenetic analysis of genes encoding proteases of the C48 family. (A) Parsimony-based tree including the corresponding genes from *Meloidogyne incognita*, *M. hapla*, *Caenorhabditis elegans*, *C. briggsae*, *Brugia malayi* and *Pristionchus pacificus*. The prefixes Minc, Mh, WP:CE, BP:CBP, Bm and PP represent the six nematode species, respectively. Numbers at nodes correspond to bootstrap values. The cluster resulting from a putative independent expansion is shown in red. The coloured boxes represent the clustered *M. incognita* genes, with colours relative to the contigs as shown in panel B. (B) Genomic organization of the *M. incognita* C48 protease genes clustered on independent contigs.

minority, with 11 and 10 members respectively. The family of serine protease occupies an intermediate position with 21 genes expressed. These values are in good agreement with the total numbers of genes per catalytic class, independently from their expression (Table 1). In four of the five catalytic classes, 40% to 53% of genes are weakly expressed. The only exception is the aspartic catalytic class, for which the highly expressed genes represent ~45% of the total. However, a great disparity is observed for the highly expressed proteases in terms of their relative contribution in the nematode life cycle. The cysteine protease family contains the greatest number of highly expressed genes (9), while two genes only are observed for the threonine protease class. Considering the relative contribution of expressed genes within a protease class, the aspartic proteases are considered the most expressed, while the metallo proteases are considered the less expressed, respectively. A more detailed analysis, conducted at the gene level, indicated that members from the five catalytic classes belong to the top twenty list of most frequently expressed proteases-encoding genes in *M. incognita* (Table 4). In particular, the S8 family of subtilisin proteases contains the top three mostly abundant expressed genes, with 28, 27 and 25 ESTs identified, respectively. These three genes alone account for more than 11% of the total number of ESTs corresponding to the *M. incognita* degradome. The C1 family of papain proteases is also highly

Table 3
Distribution of *Meloidogyne incognita* protease-encoding genes according to their expression level.

Catalytic class	Highly expressed	Moderately expressed	Weakly total expressed	
Aspartic	5 (45.45%)	2 (18.18%)	4 (36.36%)	11
Cysteine	9 (16.98%)	19 (35.85%)	25 (45.17%)	53
Metallo	6 (8.45%)	27 (38.03%)	38 (53.52%)	71
Serine	6 (28.57%)	5 (23.81%)	10 (47.62%)	21
Threonine	2 (20.0%)	4 (40.0%)	4 (40.0%)	10
Total	28	57	81	166

Only genes for which ESTs were found in the considered data sets were analyzed. Highly, moderately and weakly expressed genes were categorized according to Bortoluzzi et al. [27].

Table 4
List of the twenty most frequently expressed protease-encoding genes in *Meloidogyne incognita*.

Gene model	MEROPS family	Total no. of ESTs	% ESTs[a]	ESTs present in			Probability (*P*) of differential expression between eggs and J2[c]
				Egg[b]	J2[b]	dbEST	
Minc08696	S8	28	3.88	1	20	7	$0.98<P<0.99$
Minc14360	S8	27	3.74	1	20	6	$0.98<P<0.99$
Minc04329	S8	25	3.47	1	20	4	$0.98<P<0.99$
Minc01151	C19	22	3.05	1	3	18	–
Minc13497	C1	21	2.91	13	1	7	$P>0.999$
Minc17742	M14	18	2.50	18	0	0	$P>0.999$
Minc09945	C1	17	2.36	0	8	9	$0.90<P<0.91$
Minc00916	A1	16	2.22	6	4	6	–
Minc13862	S16	16	2.22	3	10	3	–
Minc18323	C1	15	2.08	0	10	5	$0.94<P<0.95$
Minc00503	C2	15	2.08	8	5	2	–
Minc02197	C2	15	2.08	8	5	2	–
Minc01641	A22	14	1.94	5	8	1	–
Minc16905	S16	13	1.80	3	9	1	–
Minc03305	T1	13	1.80	3	4	6	–
Minc10308	T1	13	1.80	3	5	5	–
Minc01640	A22	11	1.53	5	6	0	–
Minc03134	C1	11	1.53	1	1	9	–
Minc05000	C1	11	1.53	1	1	9	–
Minc07562	M24	10	1.39	1	2	7	–

[a] Expressed as the ratio between the number of ESTs identified for the gene model and the total number of ESTs identified for *M. incognita* proteases.
[b] Home-made ESTs libraries from *M. incognita* eggs and J2, respectively.
[c] Only the probabilities of differential expression between eggs and J2 >0.90 are indicated.

expressed, with five genes in the top twenty list, together totalling more than 10% of the total number of ESTs.

The life cycle of this nematode exhibits five different stages, i.e., four juvenile stages, among which the infective J2 stage which penetrates the plant root, and an adult stage. The availability of a large set of EST from different infective and non-infective stages allowed exploring the relative expression of the different protease catalytic classes during the parasitism process. In order to avoid any distortion in the data, we have chosen to exploit only data from stages with large representative libraries. In this respect, since transcriptomic data on the parasitic stages inside the plant were much less abundant comparing to the free-living stages, we omitted these stages in the analysis, focusing our study on the latter stages. Indeed, stage-specific expression profiling of proteases was conducted on a large set of more than 24,000 home-made ESTs from non-infective (egg; 6790 ESTs) and infective stages (J2; 17,612 ESTs), and revealed contrasting features (Fig. 6). Among the ~450 ESTs from both libraries that correspond to proteases (i.e., 134 gene models), 2/3 are from the J2 stage. In terms of stage-specific expression, data showed a majority of ESTs commonly expressed in the two stages. However, ESTs specifically expressed in the J2 stage only constitute an important part and are 3 fold times more abundant than ESTs specific of the egg stage. In terms of gene models, the J2-specific representatives represent more than half of the diversity of the proteases (i.e., 74 are J2-specific, 24 are egg-specific and 30 are commonly expressed in the two stages). As illustrated in Table 4, when the 20 most frequently expressed protease-encoding genes are considered, strong contrasting distributions of ESTs are observed in the analyzed stages, indicating differential expression. This is the case for each of three different genes encoding proteases from the S8 family, with 20 ESTs present in the J2 infective stage, while only one EST is detected in the eggs. The same type of very differentiated expression in favour of J2s is also observed for two genes encoding members of C1 family. Conversely, for other genes encoding members of the C1 and M14 families, a significant distribution of ESTs occurs in the opposite direction.

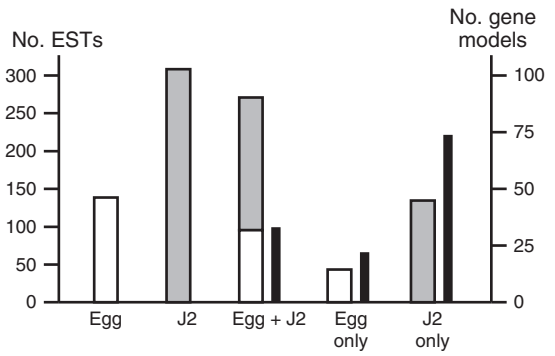

Fig. 6. Expression profiling of *Meloidogyne incognita* proteases. The white and grey bars correspond to EST found in nematode eggs and/or second-stage juveniles (J2), respectively. The black bars represent the corresponding gene models found in the nematode genome.

4. Discussion

In this work, we have performed a genomic analysis of the complete set of proteases of the plant-parasitic nematode *M. incognita*, which is considered as a model organism for plant nematologists. Based on the recent sequencing of the nematode genome [19], this work provides the first overview of the complete degradome of a plant-parasitic nematode. Our results indicated that the degradome of this nematode comprises a minimum of 334 protease-coding genes, and the putative encoded proteases are distributed into 43 families of the five catalytic classes, i.e., 26 aspartic, 106 cysteine, 136 metallo, 52 serine and 14 threonine proteases. However, these numbers should not be considered as definitive and may require refinements, since the *M. incognita* genome characterization is new and still subject to verification and basic delineation. Indeed, a new round of genome-sequencing and assembly is ongoing in the laboratory, that should provide improved data about the gene repertoire in the *M. incognita* genome, including members of the degradome. Also, the proteins identified here result from bioinformatic predictions, and their functional relevance as proteolytic enzymes remains to be confirmed. To our knowledge, very few data are currently available concerning the biochemical characterization of proteases in root-knot nematodes, although recent studies detected global aspartic protease activity and cysteine protease activity in protein extracts from *M. incognita* females and J2s, respectively [14,15]. Yet, the availability of the repertoire of *M. incognita* proteases will no doubt constitute a valuable resource and starting point for further experimental studies on the role of these enzymes in interactions between parasitic nematodes and their hosts.

The comparative analysis of the *M. incognita* degradome with those from other nematodes has provided evidence of common *versus* divergent features between them. Although some slight modifications could be noticed between the analyzed data sets, the global distribution of *M. incognita* protease classes closely resembles those in the other nematode genomes, which supports the current hypothesis that a core protease system is conserved throughout evolution [33,34]. Conversely, some discrepancies have been observed in the distribution of protease families, that could reflect some specificity among nematodes in the biological pathways they catalyze. However, a nearly ubiquitous set of 16 protease families has been identified in the genomes of all forms of life, that encompasses all the requirements for complex proteolysis capable of digestion and protein processing [35], and most of these families are found in the six nematode genomes investigated here. One remarkable feature of the *M. incognita* genome is that most of it is composed of pairs of homologous segments that may denote former diverged alleles [19]. Taking this peculiarity into account, it is reasonable to consider that the number of independent protease-coding genes identified here has been largely over-estimated. Together with the reduced number of proteases observed in the other parasitic nematode

species, this observation is congruent with the hypothesis that genome compaction could be an attribute of the parasitic lifestyle, as proposed for the root-knot nematode species *M. hapla*, considered as the smallest metazoan genome yet completed [20], and the human parasite *B. malayi* [21]. In the case of *B. malayi*, the important caveat that the actual genome sequence is probably incomplete should nevertheless be considered. Interestingly, our analysis identified twice more proteases in *M. incognita* compared to *M. hapla* (334 *versus* 154 genes). Recently, a preliminary comparative analysis of carbohydrate-active enzymes (CAZymes) in *M. hapla* and *M. incognita* revealed the same trend for some gene families, e.g., chitinases of the GH19 and GH20 family, as well as fucosyltransferases of families GT10 and GT23 [36]. Together with the hypothesis of genome duplication in *M. incognita* [19], it is tempting to correlate the larger protease gene set in this species with its strictly mitotic parthenogenetic reproduction mode.

Significant expansion in some protease families was detected in the genome of *M. incognita* compared to *C. elegans*, that are independent from the genome duplication mentioned earlier. For example, 20 cysteine proteases of the C48 sub-family, predicted to encode SUMO (small ubiquitin-like modifier) deconjugating enzymes were identified in *M. incognita*, compared to five in *C. elegans* and three in *M. hapla*. The evidence that some effectors of phytopathogenic bacteria involved in virulence and the activation of plant immunity are SUMO proteases [37,38] suggests that the proteolysis of specific sumoylated host substrates by these enzymes may be a general strategy used by diverse pathogens to manipulate host plant signal transduction. Also, serine proteases of the S16 family (Lon proteases) were more abundant in *M. incognita* compared to *C. elegans*, and these proteases are known to regulate type III protein secretion in phytopathogenic bacteria [39]. The relative abundance of these two protein families in *M. incognita* compared to the free-living nematode *C. elegans* may reflect specific aspects of its parasitic lifestyle, i.e. intimate interaction with host plant tissues. All together, these data reinforce the hypothesis that members of the nematode degradome may play a direct role in the host–parasite interaction.

Expression profiling of *M. incognita* proteases during nematode development revealed striking differences. In that respect, the comparison between eggs and infective juveniles suggests that gene overexpression in the latter stage is hypothesized to be linked to specific steps of the parasitic process. The three most frequently expressed protease-encoding genes in *M. incognita* belong to the S8 family and the corresponding ESTs were quite exclusively distributed in the J2 transcriptome. It is known that proteolytic cleavage or activation of proteins by subtilisin-like proteases (S8 family) plays a major role in multiple processes of the nematode biology, such as the construction and maintenance of the cuticle, neural signaling and nematode development [40]. In addition, recent functional studies showed that protein degradation and amino acid metabolism by subtilisin-like proteases are essential for infection by the plant-pathogenic fungus *Magnaporthe oryzae* [41,42]. The same relative distribution of ESTs in favour of infective stages was also noticed for a gene encoding a S16 protease, a family for which a direct role in plant pathogenesis has been demonstrated (see earlier). These data suggest that similar (general) roles in pathogenicity may be considered for nematodes. A more complex situation was noticed for genes encoding proteases from the C1 family, some of which appear more expressed in eggs compared to J2s, while the opposite is observed for others. Although functional data about proteases in plant-parasitic nematodes are scarce, a noticeable exception is constituted by *Mi-cpl1*, a gene encoding a cathepsin L cystein protease (i.e., belonging to the C1 family) in *M. incognita.* This gene was shown to be specifically expressed in the intestinal cells of nematodes, in the developmental stages which are in close interaction with the root tissues (i.e. J2s and females) [12]. The effects of knocking-out *Mi-cpl-1* were further shown to reduce feeding efficiency and parasitic success of *M. incognita* [14]. This may indicate that some function of the cathepsin L cysteine protease in *M. incognita* is more directly related to the parasitic aspects of the plant–nematode relationship, e.g. pathogenicity and/or evasion of primary host plant defence systems. In blood-feeding nematodes that attack either humans or animals, a battery of proteolytic digestive enzymes is involved in the degradation of substrates for nutrient uptake, among which some cathepsin-like cysteine proteases secreted in gut, in a way contributing to the specificity of the host–nematode interaction [10]. Additional comparative studies will offer the opportunity to strengthen the current view that animal and plant-parasitic nematodes may share common basic strategies of parasitism [43].

Protein inhibitors of a range of protease classes are widely expressed in plants, where they participate to natural defence strategies [44]. Therefore, the potential of disrupting proteases for plant–nematode control, via expression of protease inhibitors in transgenic plants, has been explored (reviewed in [45]). In particular, the efficacy of the defence against root-knot nematodes conferred by the transgenic expression of engineered plant cystatins (= inhibitors of cysteine proteases) has been established [46–48]. In this context, the catalogue of *M. incognita* proteases generated in this work may be helpful for the identification of new targets and for the further development of target-specific strategies to limit crop damage due to this major plant pathogen.

Acknowledgments

We acknowledge the support of the Institut National de la Recherche Agronomique (INRA), France, and of the European Cooperation in Science and Technology (COST) Action 872 'Exploiting genomics to understand plant-nematode interactions'. We thank Florence Deau and Eric Grenier for their help in data analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.ygeno.2010.10.002.

References

[1] A.J. Barrett, N.D. Rawlings, J.F. Woessner, Handbook of Proteolytic Enzymes, Second editionElsevier Academic Press, London, 2004.
[2] N.D. Rawlings, A.J. Barrett, A. Bateman, MEROPS: the peptidase database, Nucleic Acids Res. 38 (2010) D227–D233.
[3] C. Lopez-Otin, C.M. Overall, Protease degradomics: a new challenge for proteomics, Nat. Rev. Mol. Cell Biol. 3 (2002) 509–519.
[4] M.L. Blaxter, Nematoda: genes, genomes and the evolution of parasitism, Adv. Parasitol. 54 (2003) 101–195.
[5] D.L. Trudgill, V.C. Blok, Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens, Ann. Rev. Phytopathol. 39 (2001) 53–77.
[6] P. Abad, B. Favery, M.N. Rosso, P. Castagnone-Sereno, Root-knot nematode parasitism and host response: molecular basis of a sophisticated interaction, Mol. Plant Pathol. 4 (2003) 217–224.
[7] C. Britton, L. Murray, A cathepsin L protease essential for *Caenorhabditis elegans* embryogenesis is functionally conserved in parasitic nematodes, Mol. Biochem. Parasitol. 122 (2002) 21–33.
[8] H. Craig, R.E. Isaac, D.R. Brooks, Unravelling the moulting degradome: new opportunities for chemotherapy? Trends Parasitol. 23 (2007) 248–253.
[9] J. Tort, P.J. Brindley, D. Knox, K.H. Wolfe, J.P. Dalton, Proteinases and associated genes of parasitic helminths, Adv. Parasitol. 43 (1999) 161–266.
[10] A.L. Williamson, P.J. Brindley, D.P. Knox, P.J. Hotez, A. Loukas, Digestive proteases of blood-feeding nematodes, Trends Parasitol. 19 (2003) 417–423.
[11] C.J. Lilley, P.E. Urwin, H.J. Atkinson, M.J. McPherson, Characterization of cNDAs encoding serine proteases from the soybean cyst nematode *Heterodera glycines*, Mol. Biochem. Parasitol. 89 (1997) 195–207.
[12] C. Neveu, P. Abad, P. Castagnone-Sereno, Molecular cloning and characterization of an intestinal cathepsin L protease from the root-knot nematode *Meloidogyne incognita*, Physiol. Mol. Plant Pathol. 63 (2003) 159–165.
[13] R.R. Fragoso, J.A. Batista, O.B. Neto, M.F. Grossi-de-Sa, Isolation and characterization of a cDNA encoding a serine protease from the root-knot nematode *Meloidogyne incognita*, Exp. Parasitol. 110 (2005) 123–133.
[14] J. Shingles, C.J. Lilley, H.J. Atkinson, P.E. Urwin, *Meloidogyne incognita*: molecular and biochemical characterization of a cathepsin L cysteine protease and the effect on parasitism following RNAi, Exp. Parasitol. 115 (2007) 114–120.

[15] R.R. Fragoso, I.T. Lourenço, J.A.N. Batista, O.B. Oliveira-Neto, M.C.M. Silva, T.L. Rocha, M.V. Coutinho, M.F. Grossi-de-Sa, *Meloidogyne incognita*: molecular cloning and characterization of a cDNA encoding a cathepsin D-like aspartic proteinase, Exp. Parasitol. 121 (2009) 115–123.
[16] The *C. elegans* sequencing consortium, genome sequence of the nematode *C. elegans*: a platform for investigating biology, Science 282 (1998) 2012–2018.
[17] L.D. Stein, Z. Bao, D. Blasiar, T. Blumenthal, M.R. Brent, N. Chen, A. Chinwalla, L. Clarke, C. Clee, A. Coghlan, A. Coulson, P. D'Eustachio, D.H.A. Fitch, L.A. Fulton, R.E. Fulton, S. Griffiths-Jones, T.W. Harris, L.W. Hillier, R. Kamath, P.E. Kuwabara, E.R. Mardis, M.A. Marra, T.L. Miner, P. Minx, J.C. Mullikin, R.W. Plumb, J. Rogers, J.E. Schein, M. Sohrmann, J. Spieth, J.E. Stajich, C. Wei, D. Willey, R.K. Wilson, R. Durbin, R.H. Waterston, The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics, PLoS Biol. 1 (2003) 166–192.
[18] C. Dieterich, S.W. Clifton, L.N. Schuster, A. Chinwalla, K. Delehaunty, I. Dinkelacker, L. Fulton, R. Fulton, J. Godfrey, P. Minx, M. Mitreva, W. Roeseler, H. Tian, H. Witte, S.P. Yang, R.K. Wilson, R.J. Sommer, The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism, Nat. Genet. 40 (2008) 1193–1198.
[19] P. Abad, J. Gouzy, J.M. Aury, P. Castagnone-Sereno, E.G.J. Danchin, E. Deleury, L. Perfus-Barbeoch, V. Anthouard, F. Artiguenave, V.C. Blok, M.C. Caillaud, P.M. Coutinho, C. Dasilva, F. De Luca, F. Deau, M. Esquibet, T. Flutre, J.V. Goldstone, N. Hamamouch, T. Hewezi, O. Jaillon, C. Jubin, P. Leonetti, M. Magliano, T.R. Maier, G.V. Markov, P. McVeigh, G. Pesole, J. Poulain, M. Robinson-Rechavi, E. Sallet, B. Ségurens, D. Steinbach, T. Tytgat, E. Ugarte, C. van Ghelder Test, Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*, Nat. Biotechnol. 26 (2008) 909–915.
[20] C.H. Opperman, D.M. Bird, V.M. Williamson, D.S. Rokhsar, M. Burke, J. Cohn, J. Cromer, S. Diener, J. Gajan, S. Graham, T.D. Houfek, Q. Liu, T. Mitros, J. Schaff, R. Schaffer, E. Scholl, B.R. Sosinski, V.P. Thomas, E. Windham, Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism, Proc. Natl Acad. Sci. USA 105 (2008) 14802–14807.
[21] E. Ghedin, S. Wang, D. Spiro, E. Caler, Q. Zhao, J. Crabtree, J.E. Allen, A.L. Delcher, D.B. Guiliano, D. Miranda-Saavedra, S.V. Angiuoli, T. Creasy, P. Amedeo, B. Haas, N.M. El-Sayed, J.R. Wortman, T. Feldblyum, L. Tallon, M. Schatz, M. Shumway, H. Koo, S.L. Salzberg, S. Schobel, M. Pertea, M. Pop, O. White, G.J. Barton, C.K. Carlow, M.J. Crawford, J. Daub, M.W. Dimmic, C.F. Estes, J.M. Foster, M. Ganatra, W.F. Gregory, N.M. Johnson, J. Jin, R. Komuniecki, I. Korf, S. Kumar, S. Laney, B.W. Li, W. Li, T.H. Lindblom, S. Lustigman, D. Ma, C.V. Maina, D.M.A. Martin, J.P. McCarter, L. McReynolds, M. Mitreva, T.B. Nutman, J. Parkinson, J.M. Peregrín-Alvarez, C. Poole, Q. Ren, L. Saunders, A.E. Sluder, K. Smith, M. Stanke, T.R. Unnasch, J. Ware, A.D. Wei, G. Weil, D.J. Williams, Y. Zhang, S.A. Williams, C. Fraser-Liggett, B. Slatko, M.L. Blaxter, A.L. Scott, Draft genome of the filarial nematode parasite *Brugia malayi*, Science 317 (2007) 1756–1760.
[22] R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, J.S. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L. Sonnhammer, A. Bateman, The Pfam protein families database, Nucleic Acids Res. 36 (2008) D281–D288.
[23] O. Emanuelsson, S. Brunak, G. von Heijne, H. Nielsen, Locating proteins in the cell using TargetP, SignalP, and related tools, Nat. Protoc. 2 (2007) 953–971.
[24] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[25] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: protein domains identifier, Nucleic Acids Res. 33 (2005) W116–W120.
[26] G. Gremme, V. Brendel, M.E. Sparks, S. Kurtz, Engineering a software tool for gene structure prediction in higher organisms, Inform. Softw. Tech. 47 (2005) 965–978.
[27] S. Bortoluzzi, F. d'Alessi, C. Romuyaldi, G.A. Danieli, The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach, Genome Res. 10 (2000) 344–349.
[28] S. Audic, J.M. Claverie, The significance of digital gene expression profiles, Genome Res. 7 (1997) 986–995.
[29] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[30] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948.
[31] D.L. Swofford, PAUP*: Phylogenetic Analysis Using Parsimony* (and other methods), Version 4.0b10, Sinauer Associates, Sunderland Mass, 2002.
[32] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution 39 (1985) 783–791.
[33] S. Southan, A genomic perspective on human proteases, FEBS Lett. 498 (2001) 214–218.
[34] Y. Wu, X. Wang, X. Liu, Y. Wang, Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite, Genome Res. 13 (2003) 601–616.
[35] M.J. Page, E. Di Sera, Evolution of peptidase diversity, J. Biol. Chem. 283 (2008) 30010–30014.
[36] D.M. Bird, V.M. Williamson, P. Abad, J. McCarter, E.G.J. Danchin, P. Castagnone-Sereno, C.H. Opperman, The genomes of root-knot nematodes, Annu. Rev. Phytopathol. 47 (2009) 333–351.
[37] A. Hoston, M.B. Mudgett, Cysteine proteases in phytopathogenic bacteria: identification of plant targets and activation of innate immunity, Curr. Opin. Plant Biol. 7 (2004) 1–7.
[38] Y. Xia, Proteases in pathogenesis and plant defence, Cell. Microbiol. 6 (2004) 905–913.
[39] X. Tang, Y. Xiao, J.M. Zhou, Regulation of type III secretion system in phytopathogenic bacteria, Mol. Plant Microbe Interact. 19 (2006) 1159–1166.
[40] C.B. Poole, J.M. Jin, L.A. McReynolds, Subtilisin-like proteases in nematodes, Mol. Biochem. Parasitol. 155 (2007) 1–8.
[41] Y. Oh, N. Donofrio, H.Q. Pan, S. Coughlan, D.E. Brown, S. Meng, T. Mitchell, R.A. Dean, Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus *Magnaporthe oryzae*, Genome Biol. 9 (2008) R85.
[42] H. Saitoh, S. Fujisawa, A. Ito, C. Mitsuoka, T. Berberich, Y. Tosa, M. Asakura, Y. Takano, R. Terauchi, SPM1 encoding a vacuole-localized protease is required for infection-related autophagy of the rice blast fungus *Magnaporthe oryzae*, FEMS Microbiol. Lett. 300 (2009) 115–121.
[43] D.P. Jasmer, A. Goverse, G. Smant, Parasitic nematode interactions with mammals and plants, Annu. Rev. Phytopathol. 41 (2003) 245–270.
[44] C.A. Ryan, Protease inhibitors in plants: genes for improving defenses against insects and pathogens, Annu. Rev. Phytopathol. 2 (1990) 425–449.
[45] V.L. Fuller, C.J. Lilley, P.E. Urwin, Nematode resistance, New Phytol. 180 (2008) 27–44.
[46] P.E. Urwin, C.J. Lilley, M.J. McPherson, H.J. Atkinson, Resistance to both cyst- and root-knot nematodes conferred by transgenic *Arabidopsis* expressing a modified plant cystatin, Plant J. 12 (1997) 455–461.
[47] P. Vain, B. Worland, M.C. Clarke, G. Richard, M. Beavis, H. Liu, A. Kohli, M. Leech, J. Snape, P. Christou, H. Atkinson, Expression of an engineered cysteine proteinase inhibitor (OC-IΔD86) for nematode resistance in transgenic rice plants, Theor. Appl. Genet. 96 (1998) 266–271.
[48] C.J. Lilley, P.E. Urwin, K.A. Johnston, H.J. Atkinson, Preferential expression of a plant cystatin at nematode feeding sites confers resistance to *Meloidogyne incognita* and *Globodera pallida*, Plant Biotechnol. J. 2 (2004) 3–12.