



Article

The Distribution of Lectins across the Phylum Nematoda: A Genome-Wide Search

Lander Bauters, Diana Naalden and Godelieve Gheysen *

Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Ghent, Belgium; lander.bauters@ugent.be (L.B.); diana.naalden@ugent.be (D.N.)

* Correspondence: godelieve.gheysen@ugent.be; Tel.: +32-9-264-5888

Academic Editor: Cheorl-Ho Kim

Received: 18 November 2016; Accepted: 28 December 2016; Published: 4 January 2017

Abstract: Nematodes are a very diverse phylum that has adapted to nearly every ecosystem. They have developed specialized lifestyles, dividing the phylum into free-living, animal, and plant parasitic species. Their sheer abundance in numbers and presence in nearly every ecosystem make them the most prevalent animals on earth. In this research nematode-specific profiles were designed to retrieve predicted lectin-like domains from the sequence data of nematode genomes and transcriptomes. Lectins are carbohydrate-binding proteins that play numerous roles inside and outside the cell depending on their sugar specificity and associated protein domains. The sugar-binding properties of the retrieved lectin-like proteins were predicted in silico. Although most research has focused on C-type lectin-like, galectin-like, and calreticulin-like proteins in nematodes, we show that the lectin-like repertoire in nematodes is far more diverse. We focused on C-type lectins, which are abundantly present in all investigated nematode species, but seem to be far more abundant in free-living species. Although C-type lectin-like proteins are omnipresent in nematodes, we have shown that only a small part possesses the residues that are thought to be essential for carbohydrate binding. Curiously, hevein, a typical plant lectin domain not reported in animals before, was found in some nematode species.

Keywords: nematode; C-type lectin; hevein; bioinformatic analysis

1. Introduction

Lectins occur in plants, animals, and microorganisms. Although they are best known for their carbohydrate binding properties, some of them are also involved in protein–protein, protein–lipid, or protein–nucleic acid interactions [1]. Lectins are defined as proteins that have at least one non-catalytic domain that reversibly binds to a specific mono- or oligosaccharide [2]. These carbohydrate-binding proteins may be involved in a variety of functions, ranging from innate immunity to glycoprotein synthesis [3].

The first lectin was isolated from seeds of the castor bean, *Ricinus communis*, from which its name was derived: ricin. Stillmark [4] discovered that this extremely toxic protein was able to clot red blood cells. It took over 60 years to assign the agglutination activity to the ability of lectins to specifically recognize and bind carbohydrate structures on the surface of erythrocytes [5]. Since that initial find, more lectins have been reported and characterized in other plant species, and their omnipresence in the plant kingdom is supported by several genome and transcriptome analyses [6,7]. It took almost a century after the initial discovery of lectins in plants to report lectins in animals. A hepatic lectin was found in the rabbit liver in 1974 and was claimed to be the first lectin of mammalian origin [8]. However, it seems likely that the first animal lectins were revealed earlier than plant lectins since the first lectin activity was recorded by Flexner and Noguchi in 1902, but it was just not identified as a carbohydrate-binding protein [9]. Although plant and animal lectins both bind carbohydrates,

the family is very diverse, revealing a wide variety of lectin folds and specific carbohydrate-binding sites [10]. The structural diversity of the carbohydrate-binding domain, rather than carbohydrate specificity, is used to divide the lectin group into several protein families.

Animal lectins were initially classified into two major distinct classes: the C-type (Ca^{2+} -dependent) and the S-type (thiol-dependent, known as galectins) lectins [11]. Over the years it became clear that the lectin group is far more diverse than initially stated. At least 12 structural families are present in animals, although not all of them are well characterized [1,12]. Plant lectins are grouped into 12 different families according to their carbohydrate-binding domains [6]. Despite their diversity, plant and animal lectins have some carbohydrate recognition domains in common, providing evidence for a common ancestor. Two examples are the legume and ricin-type lectins from plants, which are also found in animals, where they are designated as L-type and R-type lectins, respectively [13].

Although research regarding lectins in nematodes is scarce, several reports of nematode lectins have been made previously. Most research has focused on the model organism *Caenorhabditis elegans*, but lectins have been described in other nematodes as well. Thus far, only the two biggest animal lectin families have been reported in nematodes: C-type lectins and galectins [14–17]. In this research we give an overview of the number and type of lectins present in nematodes, with a focus on C-type lectins. To our knowledge, this is the first time an attempt has been made to extract the different types of lectins from several nematode genomes. Drickamer et al. [18] already gave an overview of the different potential C-type lectins present in the genome of *Caenorhabditis elegans*, leading to the discovery of 125 proteins with one or more C-type lectin-like domains (CTLDs), grouped in nine different categories. Another study used updated techniques and found 278 proteins in *C. elegans* containing at least one CTLD [19]. Recently, Pees et al. [20] reviewed the presence and potential function of C-type lectin-like domains in invertebrates, thereby making a new classification of CTLD-containing sequences in *C. elegans*.

The availability of several nematode genomes gave us the ability to perform a thorough search across different species regarding lectins. This search shows that CTLDs are the most abundant type of lectin domain in all nematode species. In addition, the number of proteins containing a CTLD is much higher in free-living, soil-dwelling nematodes compared to species with different lifestyles, except for *Steinernema* species. Hevein, a typical chitin-binding plant lectin, was retrieved from several nematode genomes in this study. To our knowledge, this is the first report of the presence of a hevein-like domain in animals. The carbohydrate-binding properties of each type of lectin were investigated *in silico*, but still need to be confirmed experimentally. The lectin-like proteins identified in this research should be considered as potential lectins until carbohydrate binding is supported by experimental evidence. The results reported in this work give a better overview of the distribution of lectin-like domains in the phylum Nematoda.

2. Results

2.1. Selection of Lectin-Like Domains

No specific nematode-lectin domain profiles exist to date. Most lectins in nematode data are not annotated as such, so a more indirect approach was used to extract lectin domains from nematode sequence data. Several nematode genomes are available, but only a few nematode genomes were used for a thorough analysis in this study. One or more well-characterized genomes were selected, representing nematode groups with different lifestyles. The genome of *C. elegans* was chosen as a representative for nematodes with a free-living lifestyle, *Ascaris suum* and *Brugia malayi* represent the group of animal parasitic nematodes, while *Globodera pallida*, *Meloidogyne incognita*, *Meloidogyne hapla*, and *Bursaphelenchus xylophilus* represent the plant parasitic nematodes. Available Pfam profiles of plant lectin and animal lectin domains were used to query several nematode genome databases. Since not all Pfam profiles exclusively represent a lectin domain, all hits were checked manually before withholding them. All hits were blasted against Swissprot. If the top hit did not contain any

annotation regarding (potential) carbohydrate-binding activity, the sequence was discarded. Due to the fact that some profiles were not specific for some lectin-like domains, two types of lectin domains were checked in more detail before annotating them as potential lectins: I-type lectins (PF07686) and M-type lectins (PF01532).

M-type lectins are closely related to mannosidases found in the endoplasmic reticulum, but they lack mannosidase activity due to the absence of some key residues and an important disulfide bond [21,22]. A phylogenetic tree containing all potential M-type lectins was constructed using the maximum likelihood method (results not shown). Sequences lacking the two cysteine residues needed for the important disulfide bond or sequences lacking a key glutamic acid (Glu) residue needed for mannosidase activity clustered together in one branch of the tree. A total of 35 sequences lacking key residues needed for mannosidase activity clustered together in this branch and were considered potential M-type lectins.

I-type lectins owe their name to the presence of the immunoglobulin-like domain in their structure that mediates carbohydrate recognition. Siglecs are the major subfamily of I-type lectins and are characterized by a single N-terminal V-set domain (PF07686) and a single transmembrane domain [23]. V-set domains are common in diverse protein families, like immunoglobulins and T-cell receptors. Hence, a detailed analysis was needed to extract potential I-type lectins from all the hits obtained by a profile hidden Markov model analysis (or HMMer-analysis) using the V-set profile. Potential I-type lectins had to meet three criteria [23]: (1) contain one or two N-terminal V-set domains; (2) have a single transmembrane domain; and (3) have a V-set domain containing an odd number of cysteine residues (typically 3). None of the initially found sequences containing a V-set domain met all of these criteria. These results are in agreement with previous findings that I-type lectins are confined to vertebrates [24].

After a careful selection of all potential lectin-like proteins, lectin-like domains were extracted from the nematode sequences to design a new HMM profile for each type of lectin. The new profiles were used to query the nematode databases a second time. In addition to the genome data, transcriptome data of six plant parasitic nematodes were searched as well. New hits were checked manually, as previously described. Results originating from the transcriptome data should be considered with caution since these datasets only present a partial image and are not as reliable as genome data. Six plant-type lectin domains were not present in any of the studied nematode species: *Agaricus bisporus* agglutinins, Amaranthins, Cyanovirins, *Galanthus nivalis* agglutinins, Jacalins, and Nictaba lectin. Three animal lectins were not reported as well: Intelectins, P- and I-type lectins. P-type lectins are also known as mannose-6-phosphate receptors, recognize phosphorylated mannose, and play a role in the generation of functional lysosomes. P-type lectins are conserved in vertebrates but seem to be absent in invertebrates [24]. Sequences with some similarity to the P-type lectins found in invertebrates lack most of the residues found in the sugar-binding sites. Hence, these sequences are not classified as P-type lectins [12]. Intelectins are lectins with a highly conserved protein sequence involved in early embryogenesis, host–pathogen interactions, and iron metabolism. Intelectins are absent from invertebrates, although they have been found in lancelets (*Branchiostoma lanceolatum*), which are considered the transition state between vertebrates and invertebrates [25,26]. None of the bacterial lectins was present in any of the queried nematode species. F-type lectin domains, or fuclectins, have been retrieved from bacteria, vertebrates, and invertebrates, but so far they have not been reported in the phylum Nematoda [27]. In this study we identified a sequence with a possible F-type lectin domain in *A. suum* and *C. elegans*. In *C. elegans*, the F-type lectin-like domain is coupled to a C-type lectin-like domain.

Results of the second round of profile searching are summarized in Table 1. Lectin domains that are not present in any of the nematode species are not shown in this table. The numbers in this table reflect the general idea that most animal lectins belong to two families: the C-type lectins and the galectins [11]. Looking at CTLDs, *C. elegans* has on average eight times more genes containing this domain compared to other nematode genomes. In addition, *C. elegans* is the only nematode of the seven investigated genomes that has genes containing hevein-like domains.

Table 1. Results of the profile hidden Markov model (HMMer) search. Numbers indicate the number of sequences with one or more domains of a type of lectin domain in each nematode species. The total number of sequences per lectin domain is indicated in the bottom row. Numbers in brackets give an indication of the number of potential carbohydrate-binding lectins, a question mark (?) indicates that we are unable to draw conclusions regarding carbohydrate binding properties. CE: *Caenorhabditis elegans*, AS: *Ascaris suum*, BM: *Brugia malayi*, BX: *Bursaphelenchus xylophilus*, GP: *Globodera pallida*, HA: *Heterodera avenae*, MI: *Meloidogyne incognita*, MH: *Meloidogyne hapla*, MG: *Meloidogyne graminicola*, HO: *Hirschmanniella oryzae*, HG: *Heterodera glycines*, PC: *Pratylenchus coffeae*, PT: *Pratylenchus thornei*, #Seq: the number of protein sequences that was scanned for potential lectin domains. # indicates the total number of potential lectins for each class.

Species	#Seq	Class V Chitinase	Hevein	Legume	LysM	F-Type Lectin	Ricin-B	Calnexin/Calreticulin	M-Type Lectin	C-Type Lectin	Galectin	
Genomes	CE	26150	34 (30)	3	2 (2)	8 (?)	1 (0)	14 (6)	2 (2)	3	252 (52)	38 (18)
	AS	18542	8 (4)	0	3 (2)	2 (?)	1 (0)	9 (8)	3 (2)	5	36 (8)	11 (6)
	BM	17938	12 (7)	0	2 (2)	4 (?)	0	11 (10)	8 (5)	4	18 (1)	13 (10)
	BX	18074	7 (1)	0	2 (2)	1 (?)	0	11 (7)	3 (1)	4	15 (0)	14 (4)
	GP	16417	9 (6)	0	3 (2)	0	0	9 (6)	3 (2)	5	22 (4)	38 (7)
	MI	20359	2 (0)	0	1 (1)	3 (?)	0	7 (5)	4 (3)	6	57 (8)	70 (9)
	MH	14421	4 (3)	0	2 (2)	4 (?)	0	8 (5)	1 (0)	4	29 (2)	53 (3)
Transcriptomes	MG	10973	3	0	0	0	0	1	4	2	48	13
	HO	17061	5	0	2	2	0	2	5	2	33	28
	HA	10070	3	0	2	2	0	6	1	4	1	9
	HG	12262	4	0	0	1	0	0	5	0	4	9
	PC	11507	0	0	0	1	0	2	1	0	3	7
	PT	19533	0	0	0	1	0	4	2	1	0	5
#		91	3	19	29	2	84	42	40	518	308	

2.2. Predicting Carbohydrate Binding Properties

Although the identified domains are similar to previously characterized lectin domains, it does not guarantee carbohydrate binding. Identifying potential lectins can be considered a two-step process in which lectin-like sequences are identified by a profile analysis, after which candidates are screened for the necessary sugar-binding residues [12]. Numbers in Table 1 give an indication of the prevalence of several lectin-like domains in nematodes, but their carbohydrate binding activity still needs to be investigated. In the following paragraphs we will further investigate the carbohydrate-binding properties of potential lectin domains retrieved from the nematode genome data and make predictions about their carbohydrate-binding abilities.

L-fucose binding of the F-type lectin domain was assigned to a specific motif: HX(26)RXDX(4)R/K [28]. Previous studies have shown that mutating one residue might be sufficient to abrogate the carbohydrate-binding properties of the domain [29]. The two identified F-type lectin-like domains in *C. elegans* and *A. suum* do not possess the consensus fucose-binding motif. Bishnoi et al. [27] showed that several variations are possible in this motif, but their carbohydrate-binding capabilities have not yet been investigated. Further research is necessary to annotate the two F-type lectin-like domains as true lectins.

Calnexins/calreticulins interact with carbohydrates through six specific amino acids in their sequence: two tyrosines (Tyr), a lysine (Lys), a methionine (Met) and two glutamates/aspartates (Glu/Asp) [30,31]. Both calnexin and calreticulin are important lectins aiding the process of protein folding in the endoplasmic reticulum. All potential calnexins/calreticulins were aligned with ClustalO to study the conservation of these important residues. Fifteen out of 24 sequences contained all residues that were shown to be essential for carbohydrate binding, hence these proteins might be considered true lectins (Supplementary Figure S1).

The carbohydrate-binding domain in legume lectins is formed by four loops in the tertiary structure of the protein. Legume lectins display a large divergence in their carbohydrate specificity, and hence the residues in the carbohydrate binding domain are very variable. L-type lectins are involved in protein sorting in animals; their role in plants is not completely clear but they might be involved in immunity [32]. Despite the large variability, a few key amino acid residues are largely conserved among these lectins: an aspartate in the first loop, a glycine in the second loop, and an asparagine combined with an aromatic amino acid in the third loop [33,34]. In addition, a histidine residue was shown to be essential for carbohydrate binding [35]. Potential legume lectins were aligned and scanned for the presence of key residues. Ten sequences contained all of the described residues. Except for a protein originating from *A. suum* and *G. pallida*, the remaining sequences contained all but one of the key residues (Supplementary Figure S2). Although experimental testing is necessary, results indicate that 13 out of 15 proteins are likely true carbohydrate-binding legume lectins.

Ricin-B-type lectin (R-type) domains contain three tandemly repeated subdomains, designated as α , β , and γ . Each subdomain has a characteristic aspartate (D) residue and a glutamine-x-tryptophan (QXW) motif that might be involved in carbohydrate binding. Previous mutation studies have shown that these three residues occur in each subdomain and are involved in carbohydrate binding [36–38]. Experimental data suggested that a single complete binding triad is sufficient to bind a carbohydrate structure [39]. All potential Ricin-B-type lectin domains identified by a profile analysis were aligned and sequences having at least one full binding triad were considered as carbohydrate-binding. Forty-seven out of 69 sequences met this criterion (Supplementary Figure S3). Since Ricin-B lectin domains with substitutions in the triad might also bind carbohydrates, these results should be considered with caution. For instance, the tryptophan might be replaced by phenylalanine without changes in protein function [37,39].

So far, the data regarding the identification of key residues in carbohydrate binding of LysM domains are scarce. Mutagenesis studies showed that a key aromatic amino acid is needed for successful carbohydrate binding in a eukaryotic LysM domain [40]. In addition, a recent study determined the solution structure of a bacterial LysM domain to identify residues involved in

interaction with carbohydrate molecules [41]. To our knowledge, no similar attempts have been made for eukaryotic LysM domains. Upon alignment of the nematode-derived LysM domains, a clear separation into two groups was seen: a group with three conserved cysteine residues and a group without these residues (Supplementary Figure S4). It was suggested before that the cysteine residues might form disulfide bonds aiding in stabilizing a eukaryotic LysM domain [40]. The prominent cysteine residues are absent in the Pfam profile (PF01476), which is predominantly constructed with bacterial sequences. The key aromatic amino acid described by Ohnuma and colleagues [40] is only present in the cysteine-rich sequences together with one (T13) of the described key residues in bacteria [41]. The sequences lacking conserved cysteine residues contained some (T13, L14, and F40) of the described bacterial key residues (Supplementary Figure S4). It is difficult to predict carbohydrate-binding properties from these results; hence, only experimental data can give a decisive answer.

Sequences with similarity to M-type lectins have gone through a thorough analysis, as described above. An alignment shows the potential M-type lectins in nematodes, indicating the residues discriminating them from mannosidases (Supplementary Figure S5).

A total of 76 sequences with similarity to class V chitinases were identified in the queried nematode genomes. To annotate them as potential lectins, they should not have any catalytic activity. Three acidic residues, grouped in the DxExE motif in the catalytic site, are responsible for hydrolyzing chitin. Chitinase-like proteins lacking catalytic activity are known to have at least one non-conservative mutation in the key acidic residues [42]. Twenty-five sequences possess all three key residues, ruling them out as potential chitinase-like lectins (Supplementary Figure S6). Although the remaining 51 sequences are most probably devoid of any chitin-hydrolyzing activities, their chitin-binding properties should still be confirmed.

A large group of potential galectins (galactoside-binding lectins) was found in the nematode genome data. A total of 237 sequences contained one or more galectin-like domains. All galectin-like domains (366 in total) were searched for a consensus carbohydrate-binding domain, confirming their lectin characteristics. The eight residues thought to be important in carbohydrate binding, in vertebrate as well as invertebrate organisms, are His44, Asn46, Arg48, Val59, Asn61, Trp68, Glu71, and Arg73 (the numbers correspond to those of human galectin-1) [43]. An alignment of the galectin-like domains was searched for these residues and their described variants in *C. elegans* [43]. A total of 74 galectin-like domains contained the conserved residues; these domains were present in 57 individual protein sequences. This number is probably a large underestimation. Mutation analyses on the eight key residues showed that not all of them are essential for carbohydrate binding, although they can influence sugar specificity. The alignment of the 74 selected domains can be seen in Supplementary Figure S7.

Not all proteins containing a C-type lectin-like fold have carbohydrate-binding properties; a large subgroup is known to bind protein ligands instead, whether or not they are Ca²⁺-dependent. Carbohydrate-binding properties have been associated with five key amino acid residues. These five residues (Glu185, Asn187, Glu193, Asn205, and Asp206; the numbers correspond to rat mannose-binding protein) are part of a Ca²⁺ binding site in the C-type fold [44]. The first two residues are known to be involved in the specificity of sugar binding. Mutating these residues to Gln and Asp, respectively, switches the binding specificity from mannose/glucose to galactose [45], although variations are possible [46,47]. All C-type lectin-like domains were searched for the presence of key residues. If three out of five residues were present, the domains were considered as carbohydrate binding. This approach was also used in previous studies [18,48]. Seventy-seven domains met the criteria, corresponding to 75 CTLD-containing sequences (Supplementary Figure S8).

Proteins with a hevein-like domain are discussed in detail in the following section.

2.3. Hevein-Type Lectins

Hevein was discovered in the rubber tree (*Hevea brasiliensis*), where it is involved in the coagulation of latex. Hevein-like peptides bind to chitin, a polymer of *N*-acetylglucosamine, in a calcium-dependent

manner [49]. These 29–43 amino acid chitin-binding domains are rich in cysteine, glycine, and several aromatic amino acids. Based upon the number of cysteine residues, they can be classified into three groups: 6C-, 8C-, and 10C hevein-like peptides [50]. To our knowledge, no such domains have been reported outside of the kingdoms of plants and fungi [6]. Three sequences, originating from *C. elegans*, appear to contain a domain similar to a hevein domain. Two of them (T01C4.1 and F07G11.9) contain two hevein-like domains, while T13H5.3 has three (Figure 1). To exclude any possible contamination, additional available genomes of other nematode species were queried for this domain (see Table S1). Several species seem to have genes containing one to three hevein-like domains. In particular, all free-living nematode species investigated in this study possess such genes. In addition, hevein-like domains were also detected in some insect and animal parasitic nematodes assigned to clade 8a, 9b, 10a, and 10b (classification according to [51]). No hevein-like domains were found in nematode species from clades 2a, 8b, and 9c. A total of 136 hevein-like domains were found in 32 species of the additional queried genomes. Only nine of these domains did not contain the eight conserved cysteine residues that are present in the original discovered hevein and hevein-like peptides. The eight cysteine residues form four disulfide bridges to stabilize the peptide [52]. Hevein-like peptides show three conserved aromatic amino acid residues, responsible for chitin binding (residues 21 (Trp), 23 (Trp), and 30 (Tyr)) and a conserved serine residue (S19) [53]. Two of the aromatic residues are conserved in all hevein-like domains of *C. elegans*, while at position 21 only T13H5.3_1 shows an aromatic amino acid (Figure 1). All three *C. elegans* proteins containing a hevein-like domain also have a glycoside hydrolase family 18 (GHF18) domain (PF00704) and a predicted N-terminal signal peptide in their sequence (Figure 1). In addition, the two longest proteins both have a stretch of LysM domains preceding the GHF18 domain. In accordance with these three *C. elegans* genes, the hevein containing genes in the additional investigated nematode genomes were also combined with a GHF18 domain and/or LysM domains. Furthermore, 16 sequences contained a domain with homology to the plant self-incompatibility protein S1 domain (PF05938). This domain was initially discovered in Papaver and has only been found in plants and some nematodes [54].

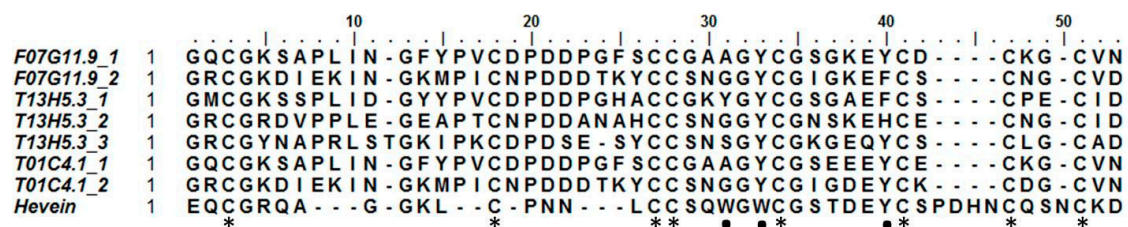


Figure 1. Alignment of the eight discovered hevein domains in *Caenorhabditis elegans* with the original hevein peptide from *Hevea brasiliensis*. Asterisks indicate the eight conserved cysteine residues forming four disulfide bridges; dots show the aromatic amino acids and serine residue involved in chitin binding.

A profile HMM, constructed from the seven hevein-like domains in *C. elegans*, was used to query the UniprotKB database. The search was restricted to Animalia. All nematode hits (46) originated from free-living species or animal parasites, divided between clades 8A, 9, and 10 [51]. In addition, 43 hits originated from Branchiostoma (lancelets) and eight from Arthropoda (including the pea aphid, the red spider mite, the lygus bug, the sea louse, and the salmon louse). These results, together with the results in Table S1, indicate that the hevein-like domain is not restricted to plants and fungi but is also present in animal species.

2.4. C-Type Lectin-Like Domains

Remarkably, *C. elegans* has 5- to 10-fold more genes containing at least one CTLD compared to the investigated animal and plant parasitic nematodes. The number of CTLD-containing genes in

C. elegans found in this research was twice the number predicted by Drickamer et al. [18], but was comparable to the number of CTLD-containing genes predicted by Pees et al. [20]. Next to the seven genomes that were thoroughly scanned in this study for potential lectin-like proteins, another 63 genomes of free-living (11), entomopathogenic (7), and animal parasitic (45) nematodes were queried to look for CTLDs (Table S1). Results showed that not only *Caenorhabditis* species, but also other free-living nematodes (for instance, *Pristionchus* spp.) have a high number of genes containing CTLDs. The only exception was the free-living nematode *Rhabditophanes* sp. KR3021, which only contained 34 genes with a CTLD in its genome. Interestingly, all *Steinernema* species contained high numbers of CTLD-containing genes as well. *Steinernema* species are parasitic nematodes capable of infecting a broad range of insect species. In contrast, *Heterorhabditis bacteriophora*, another entomopathogenic nematode, has a very low number of CTLD-containing genes in its genome. It can be concluded that free-living nematodes and *Steinernema* species have acquired a large collection of genes containing one or multiple CTLDs, while plant and animal parasitic nematodes have much smaller collections. A Pearson's correlation test (p -value = 0.692 > 0.05) showed that this observation could not be attributed to a difference in genome size between different nematode species.

Drickamer et al. [18] organized 125 proteins from *C. elegans* containing a CTLD into nine different families according to the number of CTLDs and the presence of other protein domains. In this research we used the more recent classification by Pees et al. [20], which divides CTLD containing genes into six different classes according to the presence of protein domains. We added one group to this classification for genes containing four CTLDs. This additional group consisted of only a single sequence from *A. suum* (Table 2).

Table 2. Organization of proteins from the investigated nematode species containing CTLDs. An asterisk (*) indicates a group that was not previously characterized [20]. CTLD: C-type lectin-like domain, CE: *C. elegans*, AS: *A. suum*, BM: *B. malayi*, GP: *G. pallida*, MI: *M. incognita*, MH: *M. hapla*, BX: *B. xylophilus*. CUB: Complement C1r/C1s, Uegf, Bmp1 domain, CW: conserved cysteine and tryptophan residues, also known as PAN-3 domain, VWA: Von Willebrand factor type A domain. Complex structure: CTLD-containing gene with multiple domains where CTLD might not determine the actual function of the protein.

Class	Description	CE	AS	BM	GP	MI	MH	BX
I	1 CTLD	141	24	15	17	27	17	9
II	2–3 CTLD	46	0	0	1	24	10	1
III	1–3 CTLD + 1–3 CUB	28	0	1	0	0	1	1
IV	1–2 CTLD + 1–2 CW	16	0	0	0	0	0	0
V	1 CTLD + 1–2 VWA	14	9	0	0	0	0	0
VI	Complex structure	7	2	1	4	6	1	4
VII*	4 CTLD	0	1	0	0	0	0	0

A reciprocal blast was performed to look for homologous CTLD sequences present in all seven nematode species under study. Results showed that three CTLD-containing proteins are present in all investigated species. The first sequence was classified into class I in *C. elegans* (F25B4.9) according to the classification of Pees et al. [20]. This group of sequences only contains a single CTLD. The sequence of *G. pallida* is the only sequence that is not predicted to have a secretion signal and where the CTLD is preceded by a protein kinase domain (PF00069) and followed by a small helix-loop-helix DNA-binding domain (PF00010). The second protein was also classified in class I (F32E10.3), while the other was classified in class VI (F47C12.2) in *C. elegans*. All F32E10.3 homologues have a single CTLD, followed by a large C-terminal domain without any predicted protein domains, except for the sequence of *A. suum*, which only consists of a short N-terminal domain followed by a CTLD. The sequences of the two *Meloidogyne* species and of *A. suum* are the only ones without a predicted N-terminal secretion signal. Although results of the reciprocal blasts with F47C12.2 indicated homology between the sequences, the homologous sequence originating from *B. malayi* was classified in class I instead of class VI. All of

the proteins are predicted to have an N-terminal signal peptide, except for the proteins of the two *Meloidogyne* species and *A. suum*.

One of the best known characteristics of CTLDs is the “WIGL” motif, which is highly conserved in CTLDs. The “WIGL” motif is involved in the formation of hydrophobic cores in the tertiary structure of the C-type lectin fold [55]. The tertiary structure is stabilized by two highly conserved disulfide bridges. According to Zelensky and Gready [55], there are five more highly conserved residues next to the cysteines involved in the formation of the two disulfide bridges. Three of them (W, G, and L) are residues from the “WIGL” motif. All nine residues seem to be conserved across different species [55]. Table 3 shows the presence of the nine residues in nematode CTLDs, all of which display a high degree of conservation. Almost half of the investigated sequences contain all four cysteine residues to form both disulfide bridges. About 25% only contain the first (C1) and last (C4) conserved cysteine residue to form the outer disulfide bridge (Table 3).

A phylogenetic tree was constructed using all 77 CTLDs predicted to have carbohydrate-binding properties. The tree shows that CTLDs originating from plant-parasitic species cluster together, while there is also a large cluster containing *C. elegans* CTLDs (indicated in Figure 2). The remaining cluster is a mixture of sequences originating from *C. elegans*, *B. malayi*, and *A. suum*.

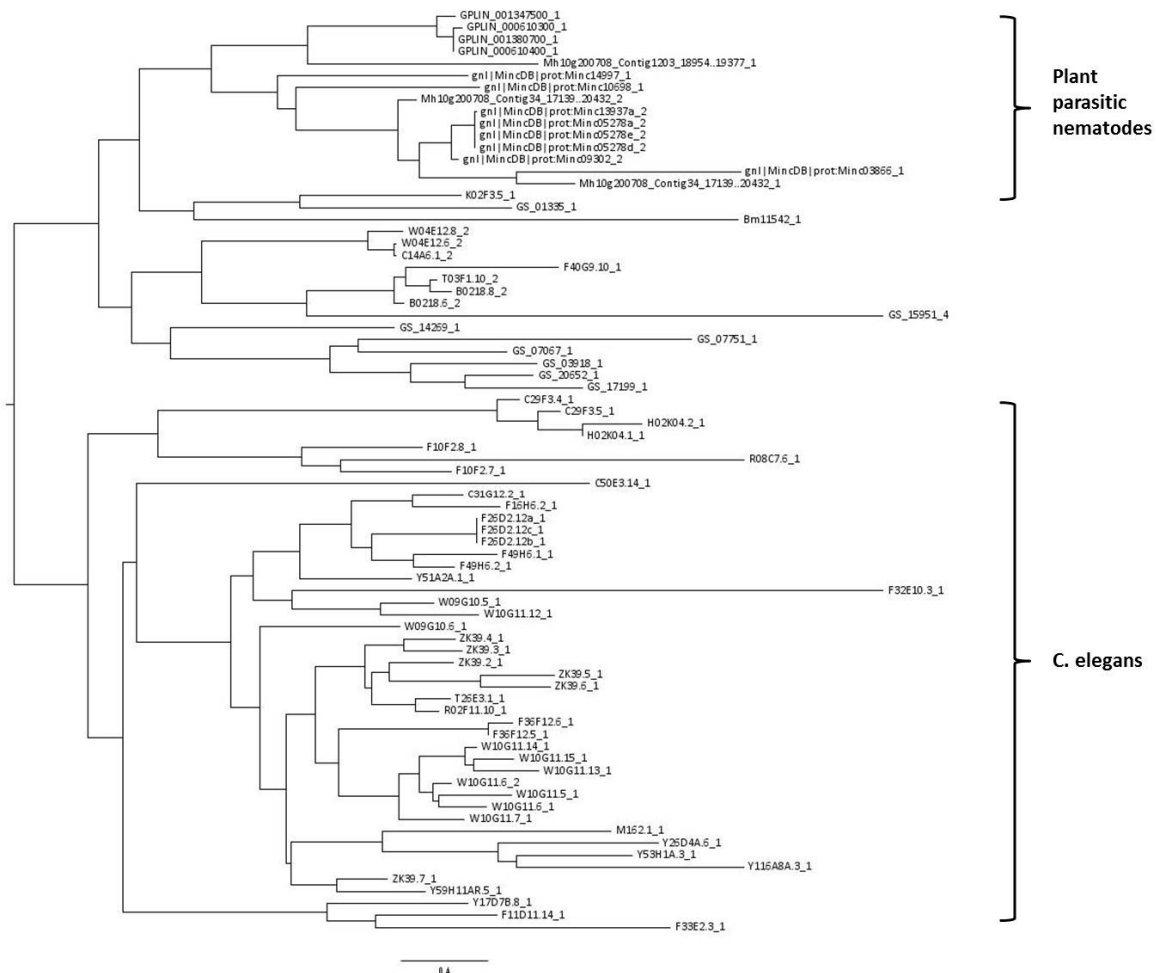


Figure 2. Midpoint rooted phylogenetic tree of 77 nematode C-type lectin-like domains (CTLDs) predicted with the maximum likelihood algorithm. Taxon labels give the full genome accession code of the gene containing the CTLD, followed by a number indicating which CTLD of that particular gene is used in this analysis. Explanation of prefixes: Bm: *Brugia malayi*, GPLIN: *Globodera pallida*, Mh10: *Meloidogyne hapla*, Mincdb: *Meloidogyne incognita*, GS: *Ascaris suum*. Taxon labels without any of the previous prefixes originate from *Caenorhabditis elegans*.

Table 3. Presence of conserved residues in nematode CTLDs. The first column shows the nine different conserved residues. The second column indicates the share of sequences that contain the respective residue. The last three rows show the share of the different combinations of disulfide bridges in the investigated nematode sequences. C1-C4: conserved cysteine residues forming the disulfide bridges (C1-C4 and C2-C3 bridge). W, G, and L: conserved residues from the “WIGL” motif. α 1A: conserved alanine residue in the first α -helix of the C-type lectin fold. β 1’L: conserved leucine residue in the first β -sheet of the C-type lectin fold.

Conserved Residue	% Residue
C1	93.17
C2	68.45
C3	70.30
C4	77.49
W	70.66
G	90.59
L	42.80
α 1A	80.07
β 1’L	64.94
Both C bridges	48.71
Only C1-C4 bridge	24.35
Only C2-C3 bridge	11.99

2.5. General Overview

All lectin-like protein sequences containing key amino acid residues indicating carbohydrate-binding properties were investigated in more detail. The Pfam database of protein domains was queried to study the domain organization of the potential lectin domain-containing proteins (Figure 3). Thirty-four percent of the lectin domain-containing proteins consist of a lectin domain flanked by regions without a known domain assignment. Only 20% of the protein sequences consist of a lectin domain combined with one or more other protein domains other than a predicted N-terminal secretion signal. Although a minority, some protein domains are interesting and seem to be related to the function of the lectin domain. The combination of a C-type lectin domain combined with a PAN module has been found in nematodes before. The PAN module is a domain found in the N-terminal domains of the plasminogen/hepatocyte growth factor family, the apple domains of the plasma prekallikrein/coagulation factor XI family and in various nematode proteins. The PAN module (also known as the CW domain) is thought to be involved in protein- or carbohydrate-binding interactions [56]. The CX domain, which is combined with a C-type lectin domain, seems to be nematode-specific, but so far no function has been assigned to this domain. R-type lectin domains are preceded by a glycosyltransferase domain in almost all cases. The lectin domain may play an important role here in binding the substrate needed by the glycosyltransferase. Two of the probable chitinase-like lectin domains are followed by carbohydrate-binding module 14 (CBM14), which is known to bind chitin [57]. Three potential M-type lectins have a Protease-Associated (PA) domain. This domain is normally associated with proteases [58], but has been found in combination with α -mannosidases as well, although its function remains elusive [22]. All hevein-like domains are associated with a chitinase domain (GHF18), strengthening the hypothesis of the chitin-binding properties of the hevein-like domain. In addition, two of the hevein domains are combined with a stretch of LysM domains, which are known for their peptidoglycan- and chitin-binding properties [59].

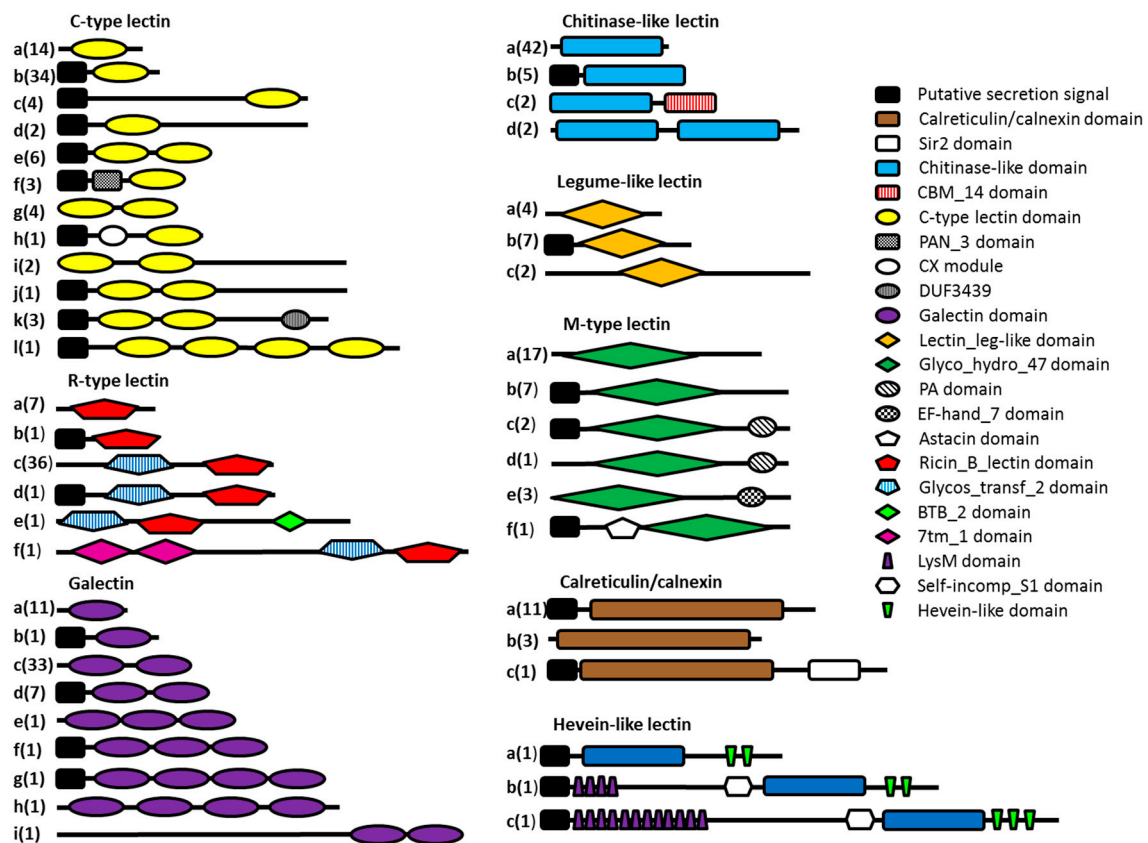


Figure 3. Schematic overview of the different types of potential lectins present in the seven investigated nematode genomes. Eight lectin classes are represented here, each divided into several subclasses (denoted by letters) according to the presence of other predicted protein domains. The number of sequences present in each subclass is indicated in brackets.

3. Discussion

Lectins have been the topic of numerous research projects in plants and animals, but information regarding this topic in nematodes is rather scarce. Lectin research involving nematodes started about 30 years ago, elaborating on the effect of non-nematode-derived lectins binding on the nematode surface [60–62]. Since the publication of the genome of *C. elegans* in 1998, a platform was created enabling scientists to investigate nematode lectins [63]. Numerous nematode genomes and transcriptomes have become publicly available since then. Research regarding lectins has mainly focused on C-type lectins [18,64], but other lectins were also investigated in more detail, like galectins [65] and calreticulin [66]. To our knowledge, this is the first attempt to make a summary of the types of lectin-like proteins present in nematodes, using the available nematode genomes. This research showed that the numbers and types of lectin-like proteins in different nematode species are quite diverse and confirmed the earlier statement of Drickamer [11] that most animal lectins belong either to the group of C-type lectins or galectins. The investigated *Heterodera* and *Pratylenchus* species seem to be an exception to this rule, but here we only queried the transcriptome data of these species. Lectin domain-containing genes were retrieved from nematode genomes using an HMM-based search. Not all animal lectin-like domains have a specific HMM-profile (e.g., ficolins, intelectins, F-type lectins, etc.), hence only the most important lectin-like domains were included in this search.

Experimental evidence in *C. elegans* suggests that a significant share of C-type lectin-like proteins is involved in the immune response upon bacterial invasion. Since there are only a few genes that are upregulated by more than two pathogen species, it seems that the action of C-type lectin-like proteins is specific to a certain type of pathogen. It was found that approximately one-fifth of the genes with

a CTLD present in the genome of *C. elegans* are induced by pathogen infection (summarized by [19] and [20]). Proteins with a galectin-like domain are involved in immune responses as well, since the expression of galectin-like encoding genes is induced upon pathogen infection in *C. elegans* [67,68]. Next to a role in innate immunity, both CTLDs and galectin-like domains from parasitic species might have a role in suppressing host immune systems. CTLDs have been found in the secretome of several parasitic nematodes and were reported to bind to the nematode cuticle or to host glycans, or expression was elevated during parasitism. These observations are an indication of a possible role in interfering with recognition or host defenses [17,69,70]. Also, galectin-like proteins were found in the secretions of several parasitic nematodes [71–73]. In addition, some of these galectin-like proteins were found to inhibit induced inflammation reactions in host organisms and are able to bind host immune cells [74,75].

Calreticulins have been reported in both plant and animal parasitic nematodes, where they are able to modulate the defense responses of the host [66,76,77]. Besides that, calreticulin is also involved in reproduction and abiotic stress responses [78,79]. To our knowledge, C-type lectins, galectins, and calreticulins are the only types of lectins that have been studied in nematodes so far. We have shown that these types of lectins are present in all investigated nematode species, but that there are also other interesting lectins awaiting further characterization.

The large variety of lectin-like proteins in nematodes may indicate their involvement in different processes, varying from immune responses to cellular growth and cell–cell interaction [1]. We have included lectins related to class V chitinase homologues (CRA) in our search. This type of lectin has a high sequence and structural similarity to true class V chitinases. In addition, although no chitinase activity is present in CRA lectins, the catalytic motif might be preserved [80]. Only experimental evidence showing carbohydrate-binding activity and lack of chitinase activity can give a decisive answer regarding the true nature of the protein: lectin or chitinase?

Potential M-type lectins were subjected to a thorough analysis before identifying them as lectins. M-type lectins are members of the glycoside hydrolase family 47 and are highly similar to α -mannosidases. Although M-type lectins can bind mannose glycans, they lack the catalytic residues and disulfide bond needed to trim the glycans. M-type lectins work in association with mannosidases and lectins from the calnexin group to supervise the correct folding of newly synthesized glycoproteins in the endoplasmic reticulum [81,82]. This process has been investigated in mammals, but key components of the machinery are present in *C. elegans* as well [83]. Mammals are known to have three M-type lectins, annotated as EDEM1-3 (ER degradation enhancing α -mannosidase-like protein) [84]. *C. elegans* also has three M-type lectins as homologues for EDEM1-3, but it seems that the investigated parasitic nematode species can contain up to six EDEM homologues. The discrepancy in the number of EDEM homologues compared to mammals and *C. elegans* seems not to be that rare, since budding yeast and *Drosophila* have one and two EDEM homologues, respectively [85,86].

A curious observation was the discovery of proteins with a hevein-like domain in *C. elegans*. This small chitin-binding monomeric protein is involved in the coagulation of latex and has antifungal properties [49,87]. Hevein-like carbohydrate-binding domains seem to be omnipresent in the plant kingdom and have been reported in fungi as well [6]. Although animal proteins with a similar three-dimensional structure have been described, this is the first time hevein-like peptides have been reported in animals [88,89]. Nematode hevein-like domains contain eight conserved cysteine residues. Two out of three aromatic amino acid residues involved in chitin binding are conserved in nematode sequences; only the first aromatic residue is replaced by alanine or glycine. It should be noted that the second conserved aromatic residue is changed from tryptophan to tyrosine in nematodes. NMR investigations have shown that a Tryptophan/Tyrosine substitution does not necessarily affect the carbohydrate-binding properties of the hevein-like domain [90], but experimental studies are needed to investigate the chitin-binding abilities. The antifungal properties of the original hevein peptide were described in 1991 [87]. Since then, several hevein-like peptides have been isolated from other plant species and their antifungal properties have been confirmed [91]. Also, oomycete growth

can be inhibited by hevein-like peptides [92]. Since most oomycetes do not contain chitin in their cell wall, the antifungal activity may not be solely based upon chitin binding. This was confirmed for hevein-like peptides from wheat, which were shown to have proteinase inhibitor properties, thereby inhibiting fungal growth [93]. The function of hevein-like domain-containing proteins remains to be elucidated. It is possible that nematodes use these proteins as some sort of protection against harmful fungi or bacteria. This hypothesis is supported by the fact that two hevein-like domain-containing genes from *C. elegans* (T01C4.1 and T13H5.3) were found to be differentially regulated upon bacterial or fungal infection [94].

A reciprocal blast of all retrieved CTLDs revealed that three domains are present in all seven investigated nematode genomes. This implies that these domains are not related to a specific lifestyle of the nematode, but rather that they have a more general function. Previous studies have shown that all three domains are differentially regulated upon bacterial or fungal infection of *C. elegans*. F32E10.3 is upregulated after infection with *Photorhabdus luminescens* and *Serratia marcescens* [94,95], but the abundance of the protein is lower after infection with *Bacillus thuringiensis* [96]. F47C12.2 is upregulated by infection with *P. luminescens* and *Enterococcus faecalis*, but downregulated by *S. marcescens* and *Drechmeria coniospora* [94]. The expression of F25B4.9, on the other hand, was downregulated upon infection with four different pathogens: *S. marcescens*, *D. coniospora*, *Xenorhabdus nematophila*, and *Pseudomonas aeruginosa* [97–99]. In addition, the corresponding protein product was less abundant upon *B. thuringiensis* infection [96]. All these results point towards a link with the innate immune system for these three common C-type lectins.

The C-type lectin family seems to be the most abundant lectin family in all types of nematodes. The C-type lectin family in *C. elegans* and other free-living nematodes is much larger compared to nematodes with a parasitic lifestyle. The number of 252 predicted members of this family is much higher than the 125 initially predicted family members [18], but it resembles the number of CTLD coding genes described by Schulenburg et al. [19] and Pees et al. [20]. While Drickamer et al. [18] identified CTLDs in *C. elegans* using a general profile, Schulenburg et al. [19] constructed a *C. elegans* specific profile, hence identifying more potential CTLD containing genes. In this research a general nematode profile was predicted, causing a small difference in number of CTLDs retrieved from the *C. elegans* genome. Taxon-specific profiles can be designed to get a more precise picture of lectin-like proteins in separate nematode species. Looking at CTLDs with predicted carbohydrate-binding properties, the ratio of *C. elegans* sequences compared to non-free-living nematodes was not altered. Genome sizes could not explain the difference in CTLD content between the different nematodes. Table S1 also shows genome sizes for each nematode. A Pearson's correlation test indicated that there is no correlation between genome size and number of CTLDs present.

It was noticed that all free-living nematodes possess more CTLD-containing genes compared to nematode species with a different lifestyle. The only exception is the free-living nematode *Rhabditophanes*. It should be noted here that *Rhabditophanes* is classified as a free-living nematode, but it is placed within a clade of animal parasitic nematodes. This suggests that it has reverted to a non-parasitic lifestyle, which may explain the difference in CTLD content [100]. Besides the free-living species, some entomopathogenic species (*Steinernema*) also appear to have a large number of CTLD-containing genes in their genome. Curiously, only 15 CTLD-containing genes were found in the genome of the entomopathogenic nematode *Heterorhabditis bacteriophora*. Both free-living and insect-parasitic nematodes spend a large share of their lifecycle in soil or decaying organic matter, so they are prone to infections from the vast bacterial and fungal communities in those habitats. The large number of CTLD-containing proteins might help in giving the innate immune system an extra boost as protection against a harmful environment. Animal and plant parasitic nematodes are partially protected by the host environment, thus may need fewer CTLD-containing proteins supporting an adequate immune system. The important role of some of these proteins in the innate immune systems has already been described in *C. elegans* [19]. It should be noted that this hypothesis is based upon differential expression of CTLD-containing genes upon bacterial infection.

Since experimental validation is lacking, the exact function in innate immunity remains unclear. This issue is raised and described by Pees et al. [20]. While *Steinernema* spp. and *Heterorhabditis* spp. occupy the same habitat and share similarities in morphology, lifecycle, and bacterial symbiosis, there is a large difference in the number of CTLD-containing genes. Although there are some contradicting studies, these similarities can probably be attributed to convergent evolution (reviewed by [101]). This hypothesis is supported by more recent phylogenetic analyses putting the two genera in different clades [51]. Convergent evolution might be an explanation for the discrepancy in CTLD content in these two genera. The genome of only one *Heterorhabditis* species was investigated. Results should be confirmed with the genome data of additional *Heterorhabditis* species (e.g., *Heterorhabditis megidis*, *Heterorhabditis indica*, and *Heterorhabditis sonorensis*) which might be available in the near future.

This research showed that several lectin folds are present over different nematode species. Although these typical lectin folds point towards carbohydrate-binding properties, this assumption cannot be made without any further investigations. Some lectin-like domains do not bind sugars or are known to bind other ligands instead. For instance, protein domains with a C-type lectin-like fold appear to be protein binding [102], while others are shown to bind lipids [103] or might bind inorganic compounds [104]. This encouraged us to subject the retrieved lectin-like domains to a more detailed investigation. Protein sequences were examined for the presence of certain key amino acid residues involved in carbohydrate binding in each type of lectin domain. Although this analysis gives an indication about the carbohydrate-binding properties of the respective lectin-like domains, experimental evidence is needed as confirmation. The domain search of potential carbohydrate-binding lectins indicated that most potential carbohydrate-binding lectin domains occur as sole domain or as tandem repeat in a protein. The fact that the remaining part of potential carbohydrate-binding domains is combined with a variety of other protein domains is in accordance with the observation that lectin-like proteins are involved in a multitude of processes [3].

This research has provided new information regarding the distribution and conservation of lectin domain-containing genes in the phylum Nematoda. Next to animal lectins, a typical plant lectin domain was also retrieved from the genome of some nematode species: the hevein domain. Functional characterization is needed to elucidate the role of the hevein domain in this phylum. Results have also confirmed that the C-type lectin family is the most abundant lectin family in nematodes and that it is much larger in free-living and some entomopathogenic nematodes. Further research is needed to characterize these C-type lectins regarding their function and carbohydrate-binding properties.

4. Materials and Methods

4.1. Lectin-Like Domain Discovery

Available Pfam profiles were downloaded for 11 plant lectins: *Agaricus bisporus* agglutinin homologs (PF07367), Amaranthins (PF07468), Class V chitinases (PF00704), Cyanovirin (PF08881), *Galanthus nivalis* agglutinin (PF01453), Hevein (PF00187), Jacalins (PF01419), Legume-like lectin (PF00139), LysM (PF01476), Nictaba (PF14299), and Ricin-B (PF00652). Six Pfam profiles of animal lectins were downloaded as well: Calnexin/calreticulin (PF00262), M-type lectins (PF01532), P-type lectins (PF02157), C-type lectins (PF00059), galectins (PF00337), and I-type lectins (PF07686). Two additional animal lectins share the same Pfam profile with plant lectins: R-type lectins (PF00652) and L-type lectins (PF00139). Intelectins and F-type lectins are two animal-type lectins for which no Pfam profile is available. Characterized sequences per lectin type were downloaded from www.imperial.ac.uk/research/animalleclectins, serving as seeds to design a new hmm profile for both lectin types. Profiles were built with the Hmmbuild function of HMMER v3.0 (<http://hmmer.org>) using a Stockholm alignment of the seed sequences as input. Data were scanned with profiles of bacterial lectins as well: LecA (PF07828), LecB (PF07472), FimH (PF09160), PapG (PF03627), F17G (PF09222), Verotoxin (PF02258), Hemolysin (PF16458), and Intimin (PF07979).

Hmmsearch (HMMER v3.0) was used to query predicted protein sequences from nematode genome databases with the 27 profiles (bitscore > 15). Sequence data were used from *Caenorhabditis elegans*, [63], *Brugia malayi* [105], *Ascaris suum* [106], *Globodera pallida* [107], *Bursaphelenchus xylophilus* [108], *Meloidogyne hapla* [109], and *Meloidogyne incognita* [110]. All hits were blasted against the Swissprot database (blastp) to help with the selection of true hits. All hits were checked manually to reduce the number of false positives. A phylogenetic tree of the M-type lectin domains was constructed with the maximum likelihood method using BioEdit (Ibis Biosciences, Carlsbad, CA, USA) [111]. Potential transmembrane regions were predicted with the TMHMM Server v. 2.0 [112].

Lectin-like domains in nematode hits were extracted and aligned with ClustalO [113] to construct a new HMM profile for each type of lectin with HMMER. These new profiles were used to query the different nematode genome databases again (bitscore > 20). In addition, transcriptome databases of some plant parasitic nematodes were searched for lectin-like domains as well: *Hirschmanniella oryzae* [114], *Meloidogyne graminicola* [115], *Heterodera glycines* (<http://nematode.net>), *Heterodera avenae* [116], *Pratylenchus coffeae* [117], and *Pratylenchus thornei* [118]. All new hits were checked separately to reduce the number of false positives. The nematode-specific profile was used on additional nematode genomes downloaded from Wormbase Parasite [119].

4.2. Structural Classification

Nematode proteins predicted to have a CTLD were used in a HMMer search against the Pfam database to predict protein domains. The presence of the different domains was used to make a structural classification based upon the organization of Drickamer et al. [18].

4.3. Conserved Sequences

A reciprocal blast (blastp, bitscore 50) was performed to search for protein sequences containing a CTLD that is conserved among several species. The reciprocal blast was carried out using the full protein sequences.

4.4. Phylogenetic Analysis

Potential carbohydrate-binding CTLDs were aligned with ClustalO [113] Model selection was done in TOPALi v2.5 [120]. According to the BIC criterion, a WAG substitution model of protein evolution was used. A phylogenetic tree was estimated using the RAxML webserver program [121]. A maximum likelihood search was performed with a total of 100 bootstrap replications. The predicted tree was visualized with Figtree (University of Edinburgh, Edinburgh, UK).

5. Conclusions

A thorough research of several nematode genomes provides us with a clearer image of the presence of several classes of potential lectins. Results from this analysis showed that C-type lectin-like proteins as well as galectin-like proteins are more prevalent compared to other lectin-like proteins in nematodes. Free living nematodes possess more C-type lectin-like proteins compared to nematodes with an alternative life style. Not only the typical animal lectin classes are present, but also some plant-type lectins, one of them being hevein. So far, hevein-type lectins have not been reported in animals, which makes it an interesting topic for further investigation. Although we have provided a list with possible lectin proteins in nematodes, their potential carbohydrate-binding properties still await further investigation and experimental confirmation.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/1422-0067/18/1/91/s1.

Acknowledgments: The authors would like to thank the Parasite Genomic group at the Wellcome Trust Sanger Institute, the Blaxter laboratory at the University of Edinburgh, the Mitreva laboratory at the Genome Institute of Washington University, and the Sommer group at the Max Planck Institute for Developmental Biology for making

some unpublished nematode genomes available at WormBase-ParaSite. This work was supported by an FWO project (Project number G.0107.12N). The authors would like to thank Els Van Damme for a critical reading of the manuscript and helpful suggestions.

Author Contributions: Lander Bauters, Diana Naalden, and Godelieve Gheysen conceived and designed the experiments; Lander Bauters analyzed the data and wrote the paper. All authors proofread and corrected the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CTLD	C-type lectin-like domain
CE	<i>C. elegans</i>
BM	<i>B. malayi</i>
AS	<i>A. suum</i>
MI	<i>M. incognita</i>
MH	<i>M. hapla</i>
MG	<i>M. graminicola</i>
BX	<i>B. xylophilus</i>
GP	<i>G. pallida</i>
HA	<i>H. avenae</i>
HG	<i>H. glycines</i>
HO	<i>H. oryzae</i>
PC	<i>P. coffeae</i>
PT	<i>P. thornei</i>
Trp	Tryptophan
Tyr	Tyrosine
GHF	Glycosyl hydrolase family
HMM	Hidden Markov model
EDEM	ER degradation-enhancing α -mannosidase-like protein

References

1. Kilpatrick, D.C. Animal lectins: A historical introduction and overview. *Biochim. Biophys. Acta* **2002**, *1572*, 187–197. [[CrossRef](#)]
2. Peumans, W.J.; van Damme, E.J. Lectins as plant defense proteins. *Plant Physiol.* **1995**, *109*, 347–352. [[CrossRef](#)] [[PubMed](#)]
3. Sharon, N.; Lis, H. History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology* **2004**, *14*, 53R–62R. [[CrossRef](#)] [[PubMed](#)]
4. Stillmark, H. *About Ricin, a Toxic Enzyme from the Seeds of Ricinus communis L. and Some Other Euphorbiaceae*; Medical School of Dorpat: Tartu, Estonia, 1888.
5. Watkins, W.M.; Morgan, W.T. Neutralization of the anti-H agglutinin in eel serum by simple sugars. *Nature* **1952**, *169*, 825–826. [[CrossRef](#)] [[PubMed](#)]
6. Van Damme, E.J.; Lannoo, N.; Peumans, W.J. Plant lectins. *Adv. Bot. Res.* **2008**, *48*, 107–209.
7. Van Damme, E.J. History of plant lectin research. *Methods Mol. Biol.* **2014**, *1200*, 3–13. [[PubMed](#)]
8. Stockert, R.J.; Morell, A.G.; Scheinberg, I.H. Mammalian hepatic lectin. *Science* **1974**, *186*, 365–366. [[CrossRef](#)] [[PubMed](#)]
9. Flexner, S.; Noguchi, H. Snake venom in relation to haemolysis, bacteriolysis and toxicity. *J. Exp. Med.* **1902**, *6*, 277–301. [[CrossRef](#)] [[PubMed](#)]
10. Loris, R. Principles of structures of animal and plant lectins. *Biochim. Biophys. Acta* **2002**, *1572*, 198–208. [[CrossRef](#)]
11. Drickamer, K. Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. Biol. Chem.* **1988**, *263*, 9557–9560. [[PubMed](#)]
12. Dodd, R.B.; Drickamer, K. Lectin-like proteins in model organisms: Implications for evolution of carbohydrate-binding activity. *Glycobiology* **2001**, *11*, 71R–79R. [[CrossRef](#)] [[PubMed](#)]
13. Taylor, M.E.; Drickamer, K. Convergent and divergent mechanisms of sugar recognition across kingdoms. *Curr. Opin. Struct. Biol.* **2014**, *28*, 14–22. [[CrossRef](#)] [[PubMed](#)]
14. Hirabayashi, J.; Ubukata, T.; Kasai, K. Purification and molecular characterization of a novel 16-kDa galectin from the nematode *Caenorhabditis elegans*. *J. Biol. Chem.* **1996**, *271*, 2497–2505. [[CrossRef](#)] [[PubMed](#)]

15. Loukas, A.; Mullin, N.P.; Tetteh, K.K.; Moens, L.; Maizels, R.M. A novel C-type lectin secreted by a tissue-dwelling parasitic nematode. *Curr. Biol.* **1999**, *9*, 825–828. [[CrossRef](#)]
16. Newton, S.E.; Monti, J.R.; Greenhalgh, C.J.; Ashman, K.; Meeusen, E.N. cDNA cloning of galectins from third stage larvae of the parasitic nematode *Teladorsagia circumcincta*. *Mol. Biochem. Parasitol.* **1997**, *86*, 143–153. [[CrossRef](#)]
17. Harcus, Y.; Nicoll, G.; Murray, J.; Filbey, K.; Gomez-Escobar, N.; Maizels, R.M. C-type lectins from the nematode parasites *Heligmosomoides polygyrus* and *Nippostrongylus brasiliensis*. *Parasitol. Int.* **2009**, *58*, 461–470. [[CrossRef](#)] [[PubMed](#)]
18. Drickamer, K.; Dodd, R.B. C-Type lectin-like domains in *Caenorhabditis elegans*: Predictions from the complete genome sequence. *Glycobiology* **1999**, *9*, 1357–1369. [[CrossRef](#)] [[PubMed](#)]
19. Schulenburg, H.; Hoepfner, M.P.; Weiner, J.; Bornberg-Bauer, E. Specificity of the innate immune system and diversity of C-type lectin domain (CTLD) proteins in the nematode *Caenorhabditis elegans*. *Immunobiology* **2008**, *213*, 237–250. [[CrossRef](#)] [[PubMed](#)]
20. Pees, B.; Yang, W.; Zárate-Potes, A.; Schulenburg, H. Dierking, K. High innate immune specificity through diversified C-type lectin-like domain proteins in invertebrates. *J. Innate. Immun.* **2016**, *8*, 129–142. [[CrossRef](#)] [[PubMed](#)]
21. Hosokawa, N.; Wada, I.; Hasegawa, K.; Yoriyuzi, T.; Tremblay, L.O.; Herscovics, A.; Nagata, K. A novel ER α -mannosidase-like protein accelerates ER-associated degradation. *EMBO Rep.* **2001**, *2*, 415–422. [[CrossRef](#)] [[PubMed](#)]
22. Hirao, K.; Natsuka, Y.; Tamura, T.; Wada, I.; Morito, D.; Natsuka, S.; Romero, P.; Sleno, B.; Tremblay, L.O.; Herscovics, A.; et al. EDEM3, a soluble EDEM homolog, enhances glycoprotein endoplasmic reticulum-associated degradation and mannose trimming. *J. Biol. Chem.* **2006**, *281*, 9650–9658. [[CrossRef](#)] [[PubMed](#)]
23. Varki, A.; Angata, T. Siglecs—The major subfamily of I-type lectins. *Glycobiology* **2006**, *16*, 1R–27R. [[CrossRef](#)] [[PubMed](#)]
24. Drickamer, K.; Fadden, A.J. Genomic analysis of C-type lectins. *Biochem. Soc. Symp.* **2002**, *69*, 59–72. [[CrossRef](#)]
25. Yan, J.; Zhang, C.; Zhang, Y.; Li, K.; Xu, L.; Guo, L.; Kong, Y.; Feng, L. Characterization and comparative analyses of two amphioxus intelectins involved in the innate immune response. *Fish Shellfish Immunol.* **2013**, *34*, 1139–1146. [[CrossRef](#)] [[PubMed](#)]
26. Yan, J.; Xu, L.; Zhang, Y.; Zhang, C.; Zhang, C.; Zhao, F.; Feng, L. Comparative genomic and phylogenetic analyses of the intelectin gene family: Implications for their origin and evolution. *Dev. Comp. Immunol.* **2013**, *41*, 189–199. [[CrossRef](#)] [[PubMed](#)]
27. Bishnoi, R.; Khatri, I.; Subramanian, S.; Ramya, T.N. Prevalence of the F-type lectin domain. *Glycobiology* **2015**, *25*, 888–901. [[CrossRef](#)] [[PubMed](#)]
28. Vasta, G.R.; Odom, E.W.; Bianchet, M.A.; Amzel, L.M.; Saito, K.; Ahmed, H. F-type lectins: A new family of recognition factors. In *Animal Lectins: A Functional View*; Vasta, G.R., Ahmed, H., Eds.; CRC Press: Boca Raton, FL, USA, 2008; pp. 465–474.
29. Farrand, S.; Hotze, E.; Friese, P.; Hollingshead, S.K.; Smith, D.F.; Cummings, R.D.; Dale, G.L.; Tweten, R.K. Characterization of a streptococcal cholesterol-dependent cytolysin with a lewis Y and B specific lectin domain. *Biochemistry* **2008**, *47*, 7097–7107. [[CrossRef](#)] [[PubMed](#)]
30. Leach, M.R.; Williams, D.B. Lectin-deficient calnexin is capable of binding class I histocompatibility molecules in vivo and preventing their degradation. *J. Biol. Chem.* **2004**, *279*, 9072–9079. [[CrossRef](#)] [[PubMed](#)]
31. Thomson, S.P.; Williams, D.B. Delineation of the lectin site of the molecular chaperone calreticulin. *Cell Stress Chaperones* **2005**, *10*, 242–251. [[CrossRef](#)] [[PubMed](#)]
32. Grandhi, N.J.; Mamidi, A.S.; Surolia, A. Pattern recognition in legume lectins to extrapolate amino acid variability to sugar specificity. *Adv. Exp. Med. Biol.* **2015**, *842*, 199–215. [[PubMed](#)]
33. Sharma, V.; Surolia, A. Analyses of carbohydrate recognition by legume lectins: Size of the combining site loops and their primary specificity. *J. Mol. Biol.* **1997**, *267*, 433–445. [[CrossRef](#)] [[PubMed](#)]
34. Itin, C.; Roche, A.C.; Monsigny, M.; Hauri, H.P. ERGIC-53 is a functional mannose-selective and calcium-dependent human homologue of leguminous lectins. *Mol. Biol. Cell* **1996**, *7*, 483–493. [[CrossRef](#)] [[PubMed](#)]

35. Kamiya, Y.; Yamaguchi, Y.; Takahashi, N.; Arata, Y.; Kasai, K.; Ihara, Y.; Matsuo, I.; Ito, Y.; Yamamoto, K.; Kato, K. Sugar-binding properties of VIP36, an intracellular animal lectin operating as a cargo receptor. *J. Biol. Chem.* **2005**, *280*, 37178–37182. [[CrossRef](#)] [[PubMed](#)]
36. Inoue, K.; Sobhany, M.; Transue, T.R.; Oguma, K.; Pedersen, L.C.; Negishi, M. Structural analysis by X-ray crystallography and calorimetry of a haemagglutinin component (HA1) of the progenitor toxin from *Clostridium botulinum*. *Microbiology* **2003**, *149*, 3361–3370. [[CrossRef](#)] [[PubMed](#)]
37. Tenno, M.; Saeki, A.; Kezdy, F.J.; Elhammer, A.P.; Kurosaka, A. The lectin domain of UDP-GalNAc: Polypeptide *N*-acetylgalactosaminyltransferase 1 is involved in *O*-glycosylation of a polypeptide with multiple acceptor sites. *J. Biol. Chem.* **2002**, *277*, 47088–47096. [[CrossRef](#)] [[PubMed](#)]
38. Tateno, H.; Goldstein, I.J. Partial identification of carbohydrate-binding sites of a Gal α 1,3Gal β 1,4GlcNAc-specific lectin from the mushroom *Marasmius oreades* by site-directed mutagenesis. *Arch. Biochem. Biophys.* **2004**, *427*, 101–109. [[CrossRef](#)] [[PubMed](#)]
39. Gerken, T.A.; Revoredo, L.; Thome, J.J.; Tabak, L.A.; Vester-Christensen, M.B.; Clausen, H.; Gahlay, G.K.; Jarvis, D.L.; Johnson, R.W.; Moniz, H.A.; et al. The lectin domain of the polypeptide GalNAc transferase family of glycosyltransferases (ppGalNAc Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type *O*-glycosylation. *J. Biol. Chem.* **2013**, *288*, 19900–19914. [[CrossRef](#)] [[PubMed](#)]
40. Ohnuma, T.; Onaga, S.; Murata, K.; Taira, T.; Katoh, E. LysM domains from *Pteris ryukyuensis* chitinase-A: A stability study and characterization of the chitin-binding site. *J. Biol. Chem.* **2008**, *283*, 5178–5187. [[CrossRef](#)] [[PubMed](#)]
41. Mesnage, S.; Dellarole, M.; Baxter, N.J.; Rouget, J.B.; Dimitrov, J.D.; Wang, N.; Fujimoto, Y.; Hounslow, A.M.; Lacroix-Desmazes, S.; Fukase, K.; et al. Molecular basis for bacterial peptidoglycan recognition by LysM domains. *Nat. Commun.* **2014**, *5*, 4269. [[CrossRef](#)] [[PubMed](#)]
42. Sulzenbacher, G.; Roig-Zamboni, V.; Peumans, W.J.; Henrissat, B.; van Damme, E.J.; Bourne, Y. Structural basis for carbohydrate binding properties of a plant chitinase-like agglutinin with conserved catalytic machinery. *J. Struct. Biol.* **2015**, *190*, 115–121. [[CrossRef](#)] [[PubMed](#)]
43. Nemoto-Sasaki, Y.; Hayama, K.; Ohya, H.; Arata, Y.; Kaneko, M.K.; Saitou, N.; Hirabayashi, J.; Kasai, K. *Caenorhabditis elegans* galectins LEC-1-LEC-11: Structural features and sugar-binding properties. *Biochim. Biophys. Acta* **2008**, *1780*, 1131–1142. [[CrossRef](#)] [[PubMed](#)]
44. Weis, W.I.; Drickamer, K.; Hendrickson, W.A. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* **1992**, *360*, 127–134. [[CrossRef](#)] [[PubMed](#)]
45. Drickamer, K. Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature* **1992**, *360*, 183–186. [[CrossRef](#)] [[PubMed](#)]
46. Childs, R.A.; Wright, J.R.; Ross, G.F.; Yuen, C.T.; Lawson, A.M.; Chai, W.; Drickamer, K.; Feizi, T. Specificity of lung surfactant protein SP-A for both the carbohydrate and the lipid moieties of certain neutral glycolipids. *J. Biol. Chem.* **1992**, *267*, 9972–9979. [[PubMed](#)]
47. Nagae, M.; Ikeda, A.; Hanashima, S.; Kojima, T.; Matsumoto, N.; Yamamoto, K.; Yamaguchi, Y. Crystal structure of human dendritic cell inhibitory receptor C-type lectin domain reveals the binding mode with *N*-glycan. *FEBS Lett.* **2016**, *590*, 1280–1288. [[CrossRef](#)] [[PubMed](#)]
48. Wood-Charlson, E.M.; Weis, V.M. The diversity of C-type lectins in the genome of a basal metazoan, *Nematostella vectensis*. *Dev. Comp. Immunol.* **2009**, *33*, 881–889. [[CrossRef](#)] [[PubMed](#)]
49. Gidrol, X.; Chrestin, H.; Tan, H.L.; Kush, A. Hevein, a lectin-like protein from *Hevea brasiliensis* (rubber tree) is involved in the coagulation of latex. *J. Biol. Chem.* **1994**, *269*, 9278–9283. [[PubMed](#)]
50. Tam, J.P.; Wang, S.; Wong, K.H.; Tan, W.L. Antimicrobial peptides from plants. *Pharmaceuticals* **2015**, *8*, 711–757. [[CrossRef](#)] [[PubMed](#)]
51. Van Megen, H.; van den Elsen, S.; Holterman, M.; Karssen, G.; Mooyman, P.; Bongers, T.; Holovachov, O.; Bakker, J.; Helder, J. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* **2009**, *11*, 927–950. [[CrossRef](#)]
52. Kini, S.G.; Nguyen, P.Q.; Weissbach, S.; Mallagaray, A.; Shin, J.; Yoon, H.S.; Tam, J.P. Studies on the chitin binding property of novel cysteine-rich peptides from *Alternanthera sessilis*. *Biochemistry* **2015**, *54*, 6639–6649. [[CrossRef](#)] [[PubMed](#)]

53. Asensio, J.L.; Canada, F.J.; Siebert, H.C.; Laynez, J.; Poveda, A.; Nieto, P.M.; Soedjanaamadja, U.M.; Gabius, H.J.; Jimenez-Barbero, J. Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains. *Chem. Biol.* **2000**, *7*, 529–543. [[CrossRef](#)]
54. Foote, H.C.; Ride, J.P.; Franklin-Tong, V.E.; Walker, E.A.; Lawrence, M.J.; Franklin, F.C. Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2265–2269. [[CrossRef](#)] [[PubMed](#)]
55. Zelensky, A.N.; Gready, J.E. Comparative analysis of structural properties of the C-type-lectin-like domain (CTLD). *Proteins* **2003**, *52*, 466–477. [[CrossRef](#)] [[PubMed](#)]
56. Tordai, H.; Banyai, L.; Patthy, L. The PAN module: The N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. *FEBS Lett.* **1999**, *461*, 63–67. [[CrossRef](#)]
57. Shen, Z.; Jacobs-Lorena, M. A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin. *J. Biol. Chem.* **1998**, *273*, 17665–17670. [[CrossRef](#)] [[PubMed](#)]
58. Luo, X.; Hofmann, K. The protease-associated domain: A homology domain associated with multiple classes of proteases. *Trends Biochem. Sci.* **2001**, *26*, 147–148. [[CrossRef](#)]
59. Buist, G.; Steen, A.; Kok, J.; Kuipers, O.P. LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol. Microbiol.* **2008**, *68*, 838–847. [[CrossRef](#)] [[PubMed](#)]
60. Davis, E.L.; Kaplan, D.T.; Dickson, D.W.; Mitchell, D.J. Root tissue response of two related soybean cultivars to infection by lectin-treated *Meloidogyne* spp. *J. Nematol.* **1989**, *21*, 219–228. [[PubMed](#)]
61. McClure, M.A.; Stynes, B.A. Lectin binding sites on the amphidial exudates of *Meloidogyne*. *J. Nematol.* **1988**, *20*, 321–326. [[PubMed](#)]
62. Spiegel, Y.; Cohn, E. Lectin binding to *Meloidogyne javanica* eggs. *J. Nematol.* **1982**, *14*, 406–407. [[PubMed](#)]
63. Sequencing Consortium. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **1998**, *282*, 2012–2018.
64. Takeuchi, T.; Sennari, R.; Sugiura, K.; Tateno, H.; Hirabayashi, J.; Kasai, K. A C-type lectin of *Caenorhabditis elegans*: Its sugar-binding property revealed by glycoconjugate microarray analysis. *Biochem. Biophys. Res. Commun.* **2008**, *377*, 303–306. [[CrossRef](#)] [[PubMed](#)]
65. Arata, Y.; Hirabayashi, J.; Kasai, K. The two lectin domains of the tandem-repeat 32-kDa galectin of the nematode *Caenorhabditis elegans* have different binding properties. Studies with recombinant protein. *J. Biochem.* **1997**, *121*, 1002–1009. [[CrossRef](#)] [[PubMed](#)]
66. Jaouannet, M.; Magliano, M.; Arguel, M.J.; Gourgues, M.; Evangelisti, E.; Abad, P.; Rosso, M.N. The root-knot nematode calreticulin Mi-CRT is a key effector in plant defense suppression. *Mol. Plant Microbe Interact.* **2013**, *26*, 97–105. [[CrossRef](#)] [[PubMed](#)]
67. Mallo, G.V.; Kurz, C.L.; Couillault, C.; Pujol, N.; Granjeaud, S.; Kohara, Y.; Ewbank, J.J. Inducible antibacterial defense system in *C. elegans*. *Curr. Biol.* **2002**, *12*, 1209–1214. [[CrossRef](#)]
68. Ideo, H.; Fukushima, K.; Gengyo-Ando, K.; Mitani, S.; Dejima, K.; Nomura, K.; Yamashita, K. A *Caenorhabditis elegans* glycolipid-binding galectin functions in host defense against bacterial infection. *J. Biol. Chem.* **2009**, *284*, 26493–26501. [[CrossRef](#)] [[PubMed](#)]
69. Loukas, A.; Doedens, A.; Hintz, M.; Maizels, R.M. Identification of a new C-type lectin, TES-70, secreted by infective larvae of *Toxocara canis*, which binds to host ligands. *Parasitology* **2000**, *121*, 545–554. [[CrossRef](#)] [[PubMed](#)]
70. Ganji, S.; Jenkins, J.N.; Wubben, M.J. Molecular characterization of the reniform nematode C-type lectin gene family reveals a likely role in mitigating environmental stresses during plant parasitism. *Gene* **2014**, *537*, 269–278. [[CrossRef](#)] [[PubMed](#)]
71. Armstrong, S.D.; Babayan, S.A.; Lhermitte-Vallarino, N.; Gray, N.; Xia, D.; Martin, C.; Kumar, S.; Taylor, D.W.; Blaxter, M.L.; Wastling, J.M.; et al. Comparative analysis of the secretome from a model filarial nematode (*Litomosoides sigmodontis*) reveals maximal diversity in gravid female parasites. *Mol. Cell. Proteom.* **2014**, *13*, 2527–2544. [[CrossRef](#)] [[PubMed](#)]
72. Hewitson, J.P.; Harcus, Y.M.; Curwen, R.S.; Dowle, A.A.; Atmadja, A.K.; Ashton, P.D.; Wilson, A.; Maizels, R.M. The secretome of the filarial parasite, *Brugia malayi*: Proteomic profile of adult excretory-secretory products. *Mol. Biochem. Parasitol.* **2008**, *160*, 8–21. [[CrossRef](#)] [[PubMed](#)]

73. Geary, J.; Satti, M.; Moreno, Y.; Madrill, N.; Whitten, D.; Headley, S.A.; Agnew, D.; Geary, T.; Mackenzie, C. First analysis of the secretome of the canine heartworm, *Dirofilaria immitis*. *Parasit. Vectors*. **2012**, *5*, 140. [[CrossRef](#)] [[PubMed](#)]
74. Kim, J.Y.; Cho, M.K.; Choi, S.H.; Lee, K.H.; Ahn, S.C.; Kim, D.H.; Yu, H.S. Inhibition of dextran sulfate sodium (DSS)-induced intestinal inflammation via enhanced IL-10 and TGF- β production by galectin-9 homologues isolated from intestinal parasites. *Mol. Biochem. Parasitol.* **2010**, *174*, 53–61. [[CrossRef](#)] [[PubMed](#)]
75. Hewitson, J.P.; Grainger, J.R.; Maizels, R.M. Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. *Mol. Biochem. Parasitol.* **2009**, *167*, 1–11. [[CrossRef](#)] [[PubMed](#)]
76. Rzepecka, J.; Rausch, S.; Klotz, C.; Schnoller, C.; Kornprobst, T.; Hagen, J.; Ignatius, R.; Lucius, R.; Hartmann, S. Calreticulin from the intestinal nematode *Heligmosomoides polygyrus* is a Th2-skewing protein and interacts with murine scavenger receptor-A. *Mol. Immunol.* **2009**, *46*, 1109–1119. [[CrossRef](#)] [[PubMed](#)]
77. Kasper, G.; Brown, A.; Eberl, M.; Vallar, L.; Kieffer, N.; Berry, C.; Girdwood, K.; Eggleton, P.; Quinnell, R.; Pritchard, D.I. A calreticulin-like molecule from the human hookworm *Necator americanus* interacts with C1q and the cytoplasmic signalling domains of some integrins. *Parasite Immunol.* **2001**, *23*, 141–152. [[CrossRef](#)] [[PubMed](#)]
78. Li, Y.; Wang, K.; Xie, H.; Wang, Y.T.; Wang, D.W.; Xu, C.L.; Huang, X.; Wang, D.S. A nematode calreticulin, Rs-CRT, is a key effector in reproduction and pathogenicity of *Radopholus similis*. *PLoS ONE* **2015**, *10*, e0129351. [[CrossRef](#)] [[PubMed](#)]
79. Feng, H.; Wei, L.; Chen, H.; Zhou, Y. Calreticulin is required for responding to stress, foraging, and fertility in the white-tip nematode, *Aphelenchoides besseyi*. *Exp. Parasitol.* **2015**, *155*, 58–67. [[CrossRef](#)] [[PubMed](#)]
80. Van Damme, E.J.; Culerrier, R.; Barre, A.; Alvarez, R.; Rouge, P.; Peumans, W.J. A novel family of lectins evolutionarily related to class V chitinases: An example of neofunctionalization in legumes. *Plant Physiol.* **2007**, *144*, 662–672. [[CrossRef](#)] [[PubMed](#)]
81. Mast, S.W.; Diekman, K.; Karaveg, K.; Davis, A.; Sifers, R.N.; Moremen, K.W. Human EDEM2, a novel homolog of family 47 glycosidases, is involved in ER-associated degradation of glycoproteins. *Glycobiology* **2005**, *15*, 421–436. [[CrossRef](#)] [[PubMed](#)]
82. Williams, D.B. Beyond lectins: The calnexin/calreticulin chaperone system of the endoplasmic reticulum. *J. Cell Sci.* **2006**, *119*, 615–623. [[CrossRef](#)] [[PubMed](#)]
83. Ghenea, S. Endoplasmic reticulum associated with degradation of glycoproteins in *Caenorhabditis elegans*. *Rom. J. Biochem.* **2009**, *46*, 53–62.
84. Yoshida, Y.; Tanaka, K. Lectin-like ERAD players in ER and cytosol. *Biochim. Biophys. Acta* **2010**, *1800*, 172–180. [[CrossRef](#)] [[PubMed](#)]
85. Kang, M.J.; Ryoo, H.D. Suppression of retinal degeneration in *Drosophila* by stimulation of ER-associated degradation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17043–17048. [[CrossRef](#)] [[PubMed](#)]
86. Jakob, C.A.; Bodmer, D.; Spirig, U.; Battig, P.; Marcil, A.; Dignard, D.; Bergeron, J.J.; Thomas, D.Y.; Aebi, M. Htm1p, a mannosidase-like protein, is involved in glycoprotein degradation in yeast. *EMBO Rep.* **2001**, *2*, 423–430. [[CrossRef](#)] [[PubMed](#)]
87. Van Parijs, J.; Broekaert, W.F.; Goldstein, I.J.; Peumans, W.J. Hevein: An antifungal protein from rubber-tree (*Hevea brasiliensis*) latex. *Planta* **1991**, *183*, 258–264. [[CrossRef](#)] [[PubMed](#)]
88. Patel, H.V.; Vyas, A.A.; Vyas, K.A.; Liu, Y.S.; Chiang, C.M.; Chi, L.M.; Wu, W.G. Heparin and heparan sulfate bind to snake cardiotoxin. Sulfated oligosaccharides as a potential target for cardiotoxin action. *J. Biol. Chem.* **1997**, *272*, 1484–1492. [[CrossRef](#)] [[PubMed](#)]
89. Drenth, J.; Low, B.W.; Richardson, J.S.; Wright, C.S. The toxin-agglutinin fold. A new group of small protein structures organized around a four-disulfide core. *J. Biol. Chem.* **1980**, *255*, 2652–2655. [[PubMed](#)]
90. Asensio, J.L.; Siebert, H.C.; von der Lieth, C.W.; Laynez, J.; Bruix, M.; Soedjanaamadja, U.M.; Beintema, J.J.; Canada, F.J.; Gabius, H.J.; Jimenez-Barbero, J. NMR investigations of protein-carbohydrate interactions: Studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and *N,N',N''*-triacylchitotriose. *Proteins* **2000**, *40*, 218–236. [[PubMed](#)]
91. Shukurov, R.R.; Voblikova, V.D.; Nikonorova, A.K.; Komakhin, R.A.; Komakhina, V.V.; Egorov, T.A.; Grishin, E.V.; Babakov, A.V. Transformation of tobacco and *Arabidopsis* plants with *Stellaria media* genes encoding novel hevein-like peptides increases their resistance to fungal pathogens. *Transgenic Res.* **2012**, *21*, 313–325. [[CrossRef](#)] [[PubMed](#)]

92. Gao, J.; Yang, F.; Zhang, S.; Li, J.; Chen, H.; Liu, Q.; Zheng, J.; Xi, J.; Yi, K. Expression of a hevein-like gene in transgenic Agave hybrid No. 11648 enhances tolerance against zebra stripe disease. *Plant Cell Tissue Organ Cult.* **2014**, *119*, 579–585. [[CrossRef](#)]
93. Slavokhotova, A.A.; Naumann, T.A.; Price, N.P.; Rogozhin, E.A.; Andreev, Y.A.; Vassilevski, A.A.; Odintsova, T.I. Novel mode of action of plant defense peptides—Hevein-like antimicrobial peptides from wheat inhibit fungal metalloproteases. *FEBS J.* **2014**, *281*, 4754–4764. [[CrossRef](#)] [[PubMed](#)]
94. Engelmann, I.; Griffon, A.; Tichit, L.; Montanana-Sanchis, F.; Wang, G.; Reinke, V.; Waterston, R.H.; Hillier, L.W.; Ewbank, J.J. A comprehensive analysis of gene expression changes provoked by bacterial and fungal infection in *C. elegans*. *PLoS ONE* **2011**, *6*, e19055. [[CrossRef](#)] [[PubMed](#)]
95. Wong, D.; Bazopoulou, D.; Pujol, N.; Tavernarakis, N.; Ewbank, J.J. Genome-wide investigation reveals pathogen-specific and shared signatures in the response of *Caenorhabditis elegans* to infection. *Genome Biol.* **2007**, *8*, R194. [[CrossRef](#)] [[PubMed](#)]
96. Treitz, C.; Cassidy, L.; Hockendorf, A.; Leippe, M.; Tholey, A. Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematocidal *Bacillus thuringiensis*. *J. Proteom.* **2015**, *113*, 337–350. [[CrossRef](#)] [[PubMed](#)]
97. Twumasi-Boateng, K.; Shapira, M. Dissociation of immune responses from pathogen colonization supports pattern recognition in *C. elegans*. *PLoS ONE* **2012**, *7*, e35400. [[CrossRef](#)] [[PubMed](#)]
98. Sinha, A.; Rae, R.; Iatsenko, I.; Sommer, R.J. System wide analysis of the evolution of innate immunity in the nematode model species *Caenorhabditis elegans* and *Pristionchus pacificus*. *PLoS ONE* **2012**, *7*, e44255. [[CrossRef](#)] [[PubMed](#)]
99. Pujol, N.; Zugasti, O.; Wong, D.; Couillault, C.; Kurz, C.L.; Schulenburg, H.; Ewbank, J.J. Anti-fungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides. *PLoS Pathog.* **2008**, *4*, e1000105. [[CrossRef](#)] [[PubMed](#)]
100. Dorris, M.; Viney, M.E.; Blaxter, M.L. Molecular phylogenetic analysis of the genus *Strongyloides* and related nematodes. *Int. J. Parasitol.* **2002**, *32*, 1507–1517. [[CrossRef](#)]
101. Burnell, A.M.; Stock, S.P. *Heterorhabditis*, *Steinernema* and their bacterial symbionts—Lethal pathogens of insects. *Nematology* **2000**, *2*, 31–42. [[CrossRef](#)]
102. Natarajan, K.; Dimasi, N.; Wang, J.; Mariuzza, R.A.; Marquies, D.H. Structure and function of natural killer cell receptors: Multiple molecular solutions to self, nonself discrimination. *Annu. Rev. Immunol.* **2002**, *20*, 853–885. [[CrossRef](#)] [[PubMed](#)]
103. Sano, H.; Kuroki, Y.; Honma, T.; Ogasawara, Y.; Sohma, H.; Voelker, D.R.; Akino, T. Analysis of chimeric proteins identifies the regions in the carbohydrate recognition domains of rat lung collectins that are essential for interactions with phospholipids, glycolipids, and alveolar type II cells. *J. Biol. Chem.* **1998**, *273*, 4783–4789. [[CrossRef](#)] [[PubMed](#)]
104. Weiss, I.M.; Kaufmann, S.; Mann, K.; Fritz, M. Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc *Haliothis laevegata*. *Biochem. Biophys. Res. Commun.* **2000**, *267*, 17–21. [[CrossRef](#)] [[PubMed](#)]
105. Ghedin, E.; Wang, S.; Spiro, D.; Caler, E.; Zhao, Q.; Crabtree, J.; Allen, J.E.; Delcher, A.L.; Guiliano, D.B.; Miranda-Saavedra, D.; et al. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **2007**, *317*, 1756–1760. [[CrossRef](#)] [[PubMed](#)]
106. Jex, A.R.; Liu, S.; Li, B.; Young, N.D.; Hall, R.S.; Li, Y.; Yang, L.; Zeng, N.; Xu, X.; Xiong, Z.; et al. *Ascaris suum* draft genome. *Nature* **2011**, *479*, 529–533. [[CrossRef](#)] [[PubMed](#)]
107. Cotton, J.A.; Lilley, C.J.; Jones, L.M.; Kikuchi, T.; Reid, A.J.; Thorpe, P.; Tsai, I.J.; Beasley, H.; Blok, V.; Cock, P.J.; et al. The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol.* **2014**, *15*, R43. [[CrossRef](#)] [[PubMed](#)]
108. Kikuchi, T.; Cotton, J.A.; Dalzell, J.J.; Hasegawa, K.; Kanzaki, N.; McVeigh, P.; Takanashi, T.; Tsai, I.J.; Assefa, S.A.; Cock, P.J.; et al. Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog.* **2011**, *7*, e1002219. [[CrossRef](#)] [[PubMed](#)]
109. Opperman, C.H.; Bird, D.M.; Williamson, V.M.; Rokhsar, D.S.; Burke, M.; Cohn, J.; Cromer, J.; Diener, S.; Gajan, J.; Graham, S.; et al. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14802–14807. [[CrossRef](#)] [[PubMed](#)]

110. Abad, P.; Gouzy, J.; Aury, J.M.; Castagnone-Sereno, P.; Danchin, E.G.; Deleury, E.; Perfus-Barbeoch, L.; Anthouard, V.; Artiguenave, F.; Blok, V.C.; et al. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* **2008**, *26*, 909–915. [[CrossRef](#)] [[PubMed](#)]
111. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
112. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [[CrossRef](#)] [[PubMed](#)]
113. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [[CrossRef](#)] [[PubMed](#)]
114. Bauters, L.; Haegeman, A.; Kyndt, T.; Gheysen, G. Analysis of the transcriptome of *Hirschmanniella oryzae* to explore potential survival strategies and host-nematode interactions. *Mol. Plant Pathol.* **2013**, *15*, 352–363. [[CrossRef](#)] [[PubMed](#)]
115. Haegeman, A.; Bauters, L.; Kyndt, T.; Rahman, M.M.; Gheysen, G. Identification of candidate effector genes in the transcriptome of the rice root knot nematode *Meloidogyne graminicola*. *Mol. Plant Pathol.* **2012**, *14*, 379–390. [[CrossRef](#)] [[PubMed](#)]
116. Kumar, M.; Gantasala, N.P.; Roychowdhury, T.; Thakur, P.K.; Banakar, P.; Shukla, R.N.; Jones, M.G.; Rao, U. De novo transcriptome sequencing and analysis of the cereal cyst nematode, *Heterodera avenae*. *PLoS ONE* **2014**, *9*, e96311. [[CrossRef](#)] [[PubMed](#)]
117. Haegeman, A.; Joseph, S.; Gheysen, G. Analysis of the transcriptome of the root lesion nematode *Pratylenchus coffeae* generated by 454 sequencing technology. *Mol. Biochem. Parasitol.* **2011**, *178*, 7–14. [[CrossRef](#)] [[PubMed](#)]
118. Nicol, P.; Gill, R.; Fosu-Nyarko, J.; Jones, M.G. De novo analysis and functional classification of the transcriptome of the root lesion nematode, *Pratylenchus thornei*, after 454 GS FLX sequencing. *Int. J. Parasitol.* **2012**, *42*, 225–237. [[CrossRef](#)] [[PubMed](#)]
119. Howe, K.L.; Bolt, B.J.; Cain, S.; Chan, J.; Chen, W.J.; Davis, P.; Done, J.; Down, T.; Gao, S.; Grove, C.; et al. WormBase 2016: Expanding to enable helminth genomic research. *Nucleic Acids Res.* **2016**, *44*, D774–D780. [[CrossRef](#)] [[PubMed](#)]
120. Milne, I.; Lindner, D.; Bayer, M.; Husmeier, D.; McGuire, G.; Marshall, D.F.; Wright, F. TOPALi v2: A rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* **2009**, *25*, 126–127. [[CrossRef](#)] [[PubMed](#)]
121. Stamatakis, A.; Hoover, P.; Rougemont, J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* **2008**, *57*, 758–771. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).