

Original research paper

The Nematode Antimicrobial Peptidome: a novel opportunity for parasite control?

Allister Irvine, Sharon A. Huws, Louise E. Atkinson, Angela Mousley*

Microbes & Pathogen Biology, The Institute for Global Food Security, School of Biological Sciences,
Queen's University Belfast, 19 Chlorine Gardens, Belfast BT9 5DL, United Kingdom

*Corresponding author: Angela Mousley (a.mousley@qub.ac.uk). School of Biological Sciences,
Queen's University Belfast, 19 Chlorine Gardens, Belfast BT9 5DL, United Kingdom

Note: Supplementary data associated with this article

Abstract

Antimicrobial Peptides (AMPs) are key constituents of the invertebrate innate immune system where they provide critical protection against microbial threat. Knowledge of AMP complements within phylum Nematoda is limited however nematodes adopt diverse life strategies and frequently reside in microbe-rich environments such that they are likely possess broad AMP profiles with bioactivity against a range of microbiota. Indeed, parasitic nematode AMPs likely have roles in defence against invading pathogens and modulation of the host microbiome.

In this study the distribution and abundance of AMP-encoding genes were examined in 134 nematode genomes providing the most comprehensive profile of AMPs within phylum Nematoda. We reveal that phylum Nematoda is AMP-rich and -diverse, where 5887 genes encode AMPs. Genome and transcriptome analyses broadly reveal: (i) AMP family profiles that are influenced by nematode lifestyle where free-living nematodes appear to have an expansion of AMPs relative to parasitic species; (ii) major differences in the AMP profiles between nematode clades where Clade 9/V and 10/IV species possess expanded AMP repertoires; (iii) AMP families with highly restricted profiles (e.g. Cecropins and Diapausins) and others [e.g. Nemapores and Glycine Rich Secreted Peptides (GRSPs)] which are more widely distributed; (iv) complexity in the distribution and abundance of Defensin subfamily members; and (v) expression of AMPs in key nematode life stages.

These data indicate that phylum Nematoda has a diverse array of AMPs and underscores the need to interrogate AMP function to unravel their importance to nematode biology and host-worm-microbiome interactions. Enhanced understanding of the Nematode Antimicrobial Peptidome will inform drug discovery pipelines for pathogen control.

Keywords: Antimicrobial peptide, AMP, nematode, parasite, bioinformatics, *in-silico*

1. Introduction

Antimicrobial Peptides (AMPs) are natural immune effectors present in all classes of life (Wang, 2017). Invertebrate AMPs form the first line of defence against invading pathogens including bacteria, fungi and viruses by exerting broad-spectrum antimicrobial activities. AMPs also stimulate invertebrate immune regulation, enhancing the endogenous immune response to protect against microbial threat (van der Does et al., 2019). Whilst AMPs have gained significant attention as novel, resistance-breaking, antimicrobial agents in vertebrate medicine (Mwangi et al., 2019), their key roles in invertebrate innate immunity highlight an unexplored opportunity for the exploitation of AMPs as novel targets for the control of invertebrate pathogens, including nematode parasites. Current understanding of invertebrate AMP function is primarily derived from arthropods (Gomez et al., 2017) where they have been shown to be critical in defense against pathogens (Du et al., 2019; Zhang et al., 2021; Feng et al., 2022), however knowledge of the diversity, role and importance of AMPs, including in nematode parasites, is limited. Harnessing nematode-derived AMPs for parasite control demands a comprehensive understanding of AMP biology which is reliant on the characterisation of AMP diversity in key nematode pathogens.

Parasitic nematodes frequently reside in hazardous host environments such as the microbe-rich intestine where they survive in close association with a broad range of microorganisms such that gastrointestinal nematodes are likely to require a raft of endogenous AMPs for survival within the host. Knowledge of AMPs in nematodes has been derived from research using the model organism *Caenorhabditis elegans* and the gastrointestinal pig parasite *Ascaris suum*, and more recently broader *in silico* studies, where a number of AMP families have been identified (Cecropins, Defensins, Nemapores, Diapausins, Glycine Rich Secreted Peptides (GRSPs) (Lee et al., 1989; Kato and Komatsu, 1996; Banyai and Patthy, 1998; Andersson et al., 2003; Couillault et al., 2004; Dieterich et al., 2008; Tarr, 2012a; Ying et al., 2016; Gu et al., 2018). The ability to uncover AMPs in nematodes has been significantly enhanced by the availability of helminth genome and transcriptome datasets (Doyle, in press). Indeed, prior to this study, the most complete profile of AMPs in nematodes was conducted by Tarr (2012a), where draft genomes and Expressed Sequence Tag (EST) datasets were examined in 58 nematode species for a sub-set of

AMP families (Cecropin, Defensin, Nemapore). A constraint, noted by Tarr (2012a), was the lack of publicly available nematode 'omics data which prevented the comprehensive characterisation of nematode AMP families. Recent expansions and quality improvements in nematode 'omics data present a timely opportunity for the characterisation of pan-phylum nematode AMP diversity that will provide a springboard to functional biology and therapeutic exploitation.

This study examines 134 nematode genomes, representing 109 species, to profile five AMP-encoding gene families (Cecropin, Defensin, Diapausin, Nemapore, GRSP) using a homology-directed bioinformatics approach. To our knowledge, the data presented here represent the most comprehensive profile of nematode AMPs to date and signify a critical advance in our understanding of the nematode AMP repertoire. These data expose nematodes as a rich source of AMP diversity that displays both conservation and variation within and between nematode clades, lifestyles and life stages likely reflecting biologically relevant trends.

These data reveal AMP diversity within phylum Nematoda and provide a novel opportunity to probe AMP function in therapeutically relevant parasites to underpin novel nematode control approaches. In addition, this study has provided a novel library of nematode-derived natural antimicrobials which may exhibit desirable characteristics for future antimicrobial discovery.

2. Materials and Methods

2.1. Identification of putative AMP-encoding genes in phylum Nematoda

A summary of the homology-directed methods in this study have been provided graphically (see Figure 1A). Genes encoding putative AMPs representing five families (Cecropin, Diapausin, Defensin, Nemapore and GRSP) were identified based on homology to previously identified nematode AMPs using Hidden Markov Model (HMM) and Basic Local Alignment Search Tool (BLAST) searches. HMM profiles were built using HMMER v3.2.1 (www.hmmerr.org) for each AMP family and, where appropriate, for individual subfamilies. Previously identified AMPs (see Supplementary Table S1) were aligned using CLUSTAL Omega with default settings (Madeira et

104 al., 2019), followed by *hmmbuild* (with default settings) to construct specific AMP family HMM
105 profiles. Where required, alignments were manually adjusted to align AMP motifs.

106 A database of nematode predicted proteins was collated from the predicted protein datasets
107 derived from 134 nematode genomes (Wormbase Parasite V14; Supplementary Table S2; Bolt et
108 al., 2018; Harris et al., 2020). Constructed HMM models were subsequently searched against the
109 concatenated nematode predicted protein database using the *hmmsearch* command. The results
110 of each HMM search were cross-checked to ensure that the query sequences used to build the
111 model scored highly. This confirmed that the HMM profile successfully identified the query
112 sequences and was specific for each AMP family. For all AMP families, except the GRSPs, search
113 hits with an E value <0.01 were manually assessed for high peptide similarity and conserved AMP
114 family motifs. Where a sequence appeared incomplete, as a result of an assembly or annotation
115 error (i.e., missing an exon or incorrectly predicted as a longer protein), the output was retained to
116 provide a comprehensive profile of putative AMP-encoding genes. For the Cecropin family, new
117 hits were included based only on homology to the query sequences, whereas for the Defensin,
118 Diapausin and Nemapore families hits were included based on the presence of specific conserved
119 cysteine motifs. As the GRSPs are characterised by repetitive glycine-rich repeats and lack other
120 conserved motifs, individual subfamily HMMs were constructed based on the 17 GRSP subfamilies
121 identified in *C. elegans* and *Caenorhabditis briggsae* GRSPs (Ying et al., 2016). Due to the highly
122 repetitive glycine-rich nature of the GRSPs, all hits from these HMMs up to an E value of 10 were
123 assessed and were confirmed based on GRSP criteria established by Ying et al. (2016). As a
124 result, putative GRSP hits were only considered to be authentic if they were ≤ 200 amino acids,
125 contained a putative signal peptide sequence, and possessed mature peptide glycine content
126 >17%. Despite the construction of 17 subfamily HMM profiles for the GRSPs, results overlapped
127 making classification of putative GRSPs according to the pre-defined *Caenorhabditis* GRSP
128 subfamilies difficult. Consequently, GRSPs hits were not classified into discrete subfamilies in this
129 study.

130 Post HMM analyses, all HMM-derived putative AMPs were employed as BLASTp query sequences
131 in the NCBI command line BLAST application (version 2.9.0, April 2019) against the concatenated

132 nematode predicted protein database constructed previously. BLASTp returns were confirmed
133 using the same criteria as above for each AMP family and new confirmed hits were repeatedly
134 used as BLASTp queries until no new results were retrieved. Due to the high number of putative
135 GRSP hits returned by HMM and BLASTp searches and the potential for false-positive returns, all
136 putative GRSP hits were subjected to an additional BLASTp search against the *C. elegans* protein
137 database on the Wormbase Parasite BLAST server (with the low complexity filter off; Version 14);
138 where a GRSP identified previously by Ying et al. (2016) was not returned within the top five hits of
139 the additional BLASTp search (sorted by E value) the original putative GRSP hit was removed from
140 further analyses.

141 Putative AMP prepropeptides were analysed for the presence of signal peptide sequences and
142 propeptide cleavage sites using SignalP 5.0 (standard Sec/SPI settings, Eukarya organism;
143 Armenteros et al., 2019) and ProP 1.0 (default furin-specific prediction; Duckert et al., 2004)
144 respectively.

145 The Defensin family was subsequently classified into subfamilies (Mollusc/Nematode Defensin,
146 Traditional Insect Defensin, Macin and Drosomycin) based on number of cysteines and their
147 arrangement according to the cysteine reference array established previously (Tarr, 2016).

148 To reduce the possibility of false-negative returns post BLASTp, tBLASTn searches using query
149 sequences from closely related species with a positive hit were employed (via Wormbase Parasite
150 V14 BLAST server); this approach confirmed null returns and checked for the presence of
151 unannotated sequences. Where an unannotated putative AMP was identified, the nematode AMP
152 profile was updated to reflect this however unannotated genes could not be included in relative
153 abundance or transcriptome analyses. Due to high sequence variability and gene expansions in
154 specific genera it was not possible to accurately assign AMP genes as orthologs.

155 2.2. Transcriptome analyses of AMP-encoding genes

156 Publicly available life-stage specific RNAseq datasets for 15 nematode species, representing a
157 range of nematode clades and lifestyles, were retrieved from Wormbase Parasite v14 (see
158 Supplementary Table S2). A threshold of ≥ 2 Transcripts Per Million (TPM) was employed as a cut-

159 off for expression (Wagner et al., 2013). For heatmap construction TPM values were Log2
160 transformed and converted to Z values using Heatmapper (www.heatmapper.ca/; Babicki et al.,
161 2016). For Log transformation and heatmap visualisation TPM 0 was reported as TPM 0.01.
162 Heatmaps were scaled by row and genes were clustered by row using the Average Linkage
163 method with Euclidean Distance Measurement method.

164 *2.3. Statistical analyses*

165 All graphs and statistical analyses were produced using GraphPad Prism 9 (www.graphpad.com/).
166 Data were tested for normality using the Kolmogorov-Smirnov test (Massey Jr, 1951). Non-
167 normally distributed data were tested for significance using Krustal Wallis Test with Dunn's multiple
168 comparisons test (Dunn, 1961).

169

170 3. Results and Discussion

171 3.1. *Phylum Nematoda is a rich source of antimicrobial peptides*

172 To investigate AMP abundance and diversity across phylum Nematoda, five known nematode
 173 AMP families (Cecropin, Diapausin, Defensin, Nemapore and GRSP (Tarr, 2012b; Zhu and Gao,
 174 2014; Ying et al., 2016; Gu et al., 2018) were profiled in 109 nematode species (134 genomes;
 175 representing 7 clades, 30 free-living nematode (FLN), 66 animal parasitic nematode (APN) and 13
 176 plant parasitic nematode (PPN) species; see Supplementary Table S2). Our analyses identified
 177 5887 putative AMP-encoding genes, of which >90% are novel, representing all known nematode
 178 AMP families [38 (0.65%) Cecropin-encoding genes, 116 (1.97%) Diapausin-encoding genes, 533
 179 (9.05%) Defensin-encoding genes, 2055 (34.91%) Nemapore-encoding genes and 3145 (53.42%)
 180 GRSP-encoding genes; see Figure 1B and Supplementary Table S3)]. Whilst the relative
 181 abundance and profile of AMP-encoding genes vary across the phylum (see Figure 2) the
 182 representation of AMPs in every nematode species examined underscores the importance of
 183 AMPs to nematode biology. These data provide the most comprehensive profile of AMPs across
 184 the phylum Nematoda and demonstrate that nematodes are AMP-rich.

185

186 3.2. *AMP profile, abundance, and family diversity vary across nematode lifestyles*

187 Our analyses reveal that AMP-encoding gene family profiles vary across nematode lifestyles (see
 188 Figure 1C). FLN and APN species display the greatest AMP family diversity (Diapausin, Defensin,
 189 Nemapore and GRSP families represented in FLNs; Cecropin, Defensin, Nemapore and GRSP
 190 families represented in APNs), however PPNs lack both Diapausin and Cecropin families.
 191 FLNs also possess the greatest number of AMP-encoding genes (98.31 ± 6.23 genes per genome)
 192 relative to both APNs (26.37 ± 3.53 genes per genome; $p < 0.0001$) and PPNs (29.33 ± 4.67 genes
 193 per genome; $p = 0.0001$) (see Figure 1D). The higher abundance of AMP-encoding genes in FLNs
 194 is not surprising given the diversity in the terrestrial and marine environments that they inhabit, and
 195 in the heterogeneity in microbes that they are likely to encounter. Parasitic nematodes, in contrast,
 196 live endogenously within animals and plants and are likely to benefit from host immunity which may

somewhat protect against microbial threat reducing the reliance on AMPs. Parasitic nematodes, however, also likely possess an AMP arsenal which is more specialised to host-dwelling microbes and pathogens.

3.3. Clade 9/V and 10/IV nematodes are AMP-rich

There was considerable variability in the distribution and abundance of AMP families across nematode clades (see Figure 2). For clade comparisons Clades 6/C and 11/IV were excluded as there is only a single species with a publicly available genome representing each of these clades [*Plectus sambesii* (Clade 6/C); *Acrobeloides nanus* (Clade 11/IV)]. Both of these species are free-living and display high AMP diversity (see Figure 2); expansion of genome datasets is required to determine whether the diverse AMP profiles noted here are conserved traits for Clade 6/C and Clade 11/IV species.

Clade 9/V species displayed the highest AMP diversity across the phylum (80.77 ± 5.736 AMP-encoding genes per genome). To account for potential differences in comparisons between genomes with different gene numbers, the number of AMP genes was calculated relative to the total number of protein coding genes (PCGs) for each genome (see Figure 2 and Supplementary Table S4). By this measure, Clade 9/V displayed the highest relative abundance of AMP genes (average $0.31 \pm 0.02\%$ per genome). Within Clade 9/V, there were differences between the relative abundance of AMP genes across species where *C. elegans* had the highest AMP complement (0.82% of total number of PCGs). In contrast, the lowest relative abundance of AMP genes was noted for the entomopathogenic nematode (EPN), *Heterohabditis bacteriophora*, where only 0.09% of PCGs encode AMP genes. The high abundance of AMP genes in *C. elegans* was unsurprising given that this species has provided the blueprint for nematode AMP research. Indeed, most of the query sequences used to identify AMP genes in this study originated from *C. elegans*. The reduced relative abundance of AMP genes in *H. bacteriophora* may be related to its symbiotic relationship with gram negative *Photorhabdus* bacteria (Parihar et al., 2022). *Photorhabdus* spp. are key to the lifecycle of *H. bacteriophora* and produce a range of toxins which kill the arthropod host (Clarke, 2020). Recently, an antibiotic with potent activity against gram negative bacteria was

225 discovered in *Photorhabdus* (Imai et al., 2019). It is therefore possible that *H. bacteriophora*
 226 benefits from AMPs produced by the bacterial symbiont reflecting the lower abundance of
 227 endogenous nematode AMPs in this species.

228 Beyond the AMP-rich *Caenorhabditis* genera, other genera within Clade 9/V also appear to display
 229 enriched AMP profiles. For example, free-living *Pristionchus* species have an enriched AMP
 230 profile with all but one species having >100 AMP genes ($0.33 \pm 0.02\%$ of all PCGs). *Pristionchus*
 231 feed on rotting plant matter and share mostly phoretic, and occasionally necromenic, relationships
 232 with arthropods (Félix et al., 2018). It may be the case that *Pristionchus* nematodes require an
 233 enriched arsenal of AMPs, to combat the microbe-rich environments they inhabit.

234 The Ancylostomatoidea and Trichostrongyloidea superfamilies, which include mammalian
 235 parasites, also have a higher AMP abundance than the other parasitic superfamilies, e.g. the
 236 Metastrongyloidea and Strongyloidea, in this clade. Many of these parasitic species that display
 237 higher AMP abundance, including *Necator americanus*, *Teladorsagia circumcincta* and
 238 *Haemonchus contortus*, are known to alter the host gut microbiome during infection (El-Ashram et
 239 al., 2017; Cortés et al., 2020; Jenkins et al., 2021). Whilst it is currently unclear whether AMPs
 240 from these species are secreted and/or excreted into the host environment to contribute to
 241 changes in the microbiome, nematode-derived AMPs, such as *A. suum* Cecropin P1, have been
 242 isolated from the host gut (Lee et al., 1989; Andersson et al., 2003). Future research should aim to
 243 establish whether excreted/secreted nematode AMPs are active against host microbiota and
 244 contribute to changes in the host microbiome during infection.

245 Clade 10/IV is also AMP-rich (64.79 ± 9.51 AMP-encoding genes per genome; $0.28 \pm 0.02\%$ of
 246 PCGs in the Clade 10/IV species examined in this study encode AMPs). Despite the diversity in
 247 lifestyles, most Clade 10/IV species share similar numbers of AMPs. *Steinernema* spp. appear to
 248 have an elevated AMP abundance relative to other Strongyloidoidea superfamily members such as
 249 *Strongyloides* spp. Interestingly, like Clade 9/V *H. bacteriophora*, *Steinernema* spp. are also EPNs
 250 that display endosymbiotic relationships with gram negative bacteria. *Steinernema* spp. harbour
 251 *Xenorhabdus* bacteria which are critical for worm establishment and ultimately kill the insect host
 252 (Goodrich-Blair, 2007). Consistent with the relationship between *Photorhabdus* and *H.*

bacteriophora, *Xenorhabdus* also produces antimicrobials which aid establishment and maintenance of the nematode parasites in the host (Dreyer et al., 2019). *Steinernema* spp. however do not appear to have the same reduction in AMP gene abundance as was observed in *H. bacteriophora*. At present, little is known about the role that worm-derived AMPs play in the lifecycle of EPNs. A recent excretory/secretory protein (ESP) analysis of *Steinernema feltiae* activated infective juveniles identified 266 proteins which were designated as ‘venom’ proteins involved in host infection (Chang et al., 2019); here we reveal that seven of these ESPs are Nemapores (L889_g29545, L889_g13905, L889_g31036, L889_g31037, L889_g17551, L889_g20297 and L889_g6015). Notably, these Nemapores were absent from the ESP of non-activated infective juveniles demonstrating that they likely play a role in *S. feltiae* infection and are actively secreted/excreted from the worm upon exposure to host tissue. The original study (Chang et al., 2019) did not identify these genes as Nemapores based on automatic annotation, demonstrating the value of the manual homology-directed approach used here. The presence of Nemapores in *S. feltiae* ESP also encourage more focussed peptidomic analyses of parasitic nematode ESP products to determine host/environment facing AMP profiles.

3.4. Clade 2/I, 8/III and 12/IV nematodes possess less diverse AMP profiles

Multiple nematode clades possess a limited AMP profile in comparison to Clade 9/V and 10/IV in terms of the number of AMP genes and the distribution of AMP families. Clade 2/I nematodes possess the most limited AMP profile of all the nematode clades profiled in this study (3.44 ± 0.50 AMP-encoding genes per genome; $0.02 \pm 0.01\%$ of PCGs encoding AMP genes). All species from this clade possess genes encoding Nemapores and GRSPs with only *Romanomermis culicivorax* possessing Defensin-encoding genes (see Figure 2). Clade 2/I nematodes are generally considered to be more basal and have more limited gene complements for other protein families including neuropeptides, antioxidant enzymes, G-protein-coupled receptors (GPCRs) and ATP-binding cassette transporters (ABC transporters) (McCoy et al., 2014; IHGC 2019; Xu et al., 2020; McKay et al., 2021). The reduced AMP arsenal observed in Clade 2/I may indicate that elevated AMP diversity in the Crown Clades (Clade 8-12) originated as a result of gene duplications and/or

selective pressures that are not observed in the basal clades. It is interesting that many Clade 2/I nematodes share similar host niches with some of the parasitic species that have more diverse AMP profiles. Indeed, both *Trichinella* and *Trichuris* species spend much of their lifecycles in the microbe-rich gut environment of their hosts. The reduced number of AMP-encoding genes in Clade 2/I nematodes poses an intriguing question about how these nematodes combat microbial threats without the range of AMPs that other intestinal parasitic nematodes possess. It is possible that Clade 2/I species may possess divergent AMPs or larger antimicrobial proteins that cannot be identified using the homology-based pipeline employed in this study.

With the exception of the Ascaridoidea (31.38 \pm 4.14 AMP-encoding genes per genome; 0.17 \pm 0.02% of PCGs encoding AMPs) and Oxyuroidea (20.50 \pm 0.50 AMP-encoding genes per genome; 0.17 \pm 0.02% of PCGs encoding AMPs) superfamilies, Clade 8/III nematodes also broadly displayed a reduced AMP complement (6.95 \pm 0.01 AMP-encoding genes per genome; 0.06 \pm 0.01% of PCGs encoding AMPs) in comparison to Clade 9/V species. This reduced AMP complement was particularly evident within the Filarioidea species which may suggest that there has been a loss of AMP genes from filarial nematodes. In line with the more basal nematodes in Clade 2/I, reduced profiles of metabolic and neuropeptide pathway components are also observed in filarial nematodes (IHGC 2019; McKay et al., 2021). Filarial nematodes typically reside in the circulatory system and subcutaneous tissues of the definitive host and may, therefore, be less reliant on AMPs due to the more sterile host niches they occupy. Additionally, many filarial nematodes including *Brugia malayi*, *Onchocerca volvulus* and *Wuchereria bancrofti* share well characterised endosymbiotic relationships with *Wolbachia* which are critical for key aspects of nematode biology (Taylor et al., 2005). It is possible that *Wolbachia* provides a level of protection for filarial parasites such that the nematodes require fewer endogenous AMPs. Alternatively, the reduced AMP profiles in Filarids could help shape a favourable environment for *Wolbachia* growth.

Clade 12/IV species, which include the PPNs, also possess a reduced AMP profile (27.47 \pm 4.54 AMP-encoding genes per genome; 0.09 \pm 0.01% of PCGs encoding AMPs) in comparison to Clade 9/V nematodes. Notably, the only PPN in this study that is not a Clade 12/IV species, *Bursaphelenchus xylophilus* (Clade 10/IV), possesses 61 AMP-encoding genes (0.34% relative

abundance), exhibiting a profile much more in line with Clade 10/IV species. As noted for Clade 2/1 nematodes there is also potential that there are currently undiscovered AMP families within Clade 12/IV species with diverse motifs that are not identifiable via the methods employed here. This highlights the need to expand AMP studies beyond *C. elegans* and *A. suum* to discover more evolutionarily distinct AMP families.

3.5 *Cecropin* genes are entirely restricted to *Ascaridoidea* species

The pan-phylum data presented here provides evidence to support the restriction of Cecropin-encoding genes to Ascarid species; these data also report the first identification of Cecropin-encoding genes from *Parascaris* spp and confirm the recent identification of Cecropins from *Anisakis* species (Rončević et al., 2022). The restriction of Cecropins to Ascaridoidea suggests Ascarid-specific roles that require further characterisation. As noted above, Cecropins were originally isolated from the porcine intestinal environment (Lee et al., 1989). Despite this, the effect of these peptides on host microbiota has yet to be characterised.

Cecropins have not undergone the same gene expansion observed in other AMP families (4.75 ± 0.88 genes per genome). The Ascarid Cecropins appear to be highly conserved suggesting that these peptides have similar antimicrobial activities (Figure 3). Indeed, chemically synthesised *A. suum* Cecropin P1-P4 have similar activity profiles against bacteria and fungi (Pillai et al., 2005). However, minor differences in amino acids can significantly change the antimicrobial activity of these peptides; for example, a single amino acid substitution of a Proline residue for Isoleucine residue at position 22 of *A. suum* Cecropin P1 significantly reduces antibacterial and membrane permeabilising activity (Gazit et al., 1996). Antimicrobial activity should, therefore, be investigated for Cecropin peptides from other Ascarid species.

3.6 *Diapausin* genes are restricted to specific free-living soil nematodes

Nematode Diapausin-encoding genes were originally identified from *Pristionchus pacificus* (Dieterich et al., 2008). The pan-phylum analyses performed here reveals that Diapausins are

restricted to Clades 9/V and 11/IV in soil-dwelling FLNs, primarily the Clade 9/V Diplogasteroidea which includes *Pristionchus* spp. Beyond the Diplogasteroidea, two Rhabditoidea, *Mesorhabditis belari* and *Oscheius tipulae*, also encode Diapausins while the closely related *Caenorhabditis* genera do not. The absence of Diapausin-encoding genes in *C. elegans* is interesting given the similar environmental niche occupied by these species, highlighting the importance of AMP research beyond *C. elegans*. Diapausin-encoding genes were also identified in *Acrobeloides nanus*, the only Clade 11/IV species in this study; this is an intriguing observation from an evolutionary perspective.

It has been suggested that *P. pacificus* Diapausin-encoding genes have been acquired through horizontal gene transfer from beetles with which they share phoretic/necromenic associations (Rodelsperger and Sommer, 2011). This theory is supported by the presence of a Diapausin-like sequences in an insect iridovirus implicating the virus as the vector for gene transfer (Tanaka et al., 2003). Phylogenetic analyses did not provide any further insight into these relationships or gene evolution (data not shown) primarily due to poor resolution as a consequence of the rapid evolution of AMP-encoding genes.

In total 116 Diapausin-encoding genes were identified across phylum Nematoda with an average of 8.92 ± 1.62 genes per genome. Diapausins possess a cysteine motif that consists of six cysteine residues which form three disulfide bonds (see Figure 4A). The position of these cysteines is conserved compared to arthropod Diapausins, suggesting conservation in Diapausin tertiary structure. Interestingly, some of the Diapausin-encoding genes identified in this study possessed two complete cysteine Diapausin motifs (e.g. ACRNAN_scaffold769.g15805; 12 cysteine residues). So far, Diapausin-encoding genes have only been reported to possess a single Diapausin cysteine motif and thus it is unclear whether the genes identified here contain multiple domains from errors in gene annotation or whether they represent a novel type of Diapausin-encoding gene, where multiple peptides could be cleaved from a single gene precursor. Beyond the nematode Diapausin cysteine motif there is significant amino acid variation that may result in different antimicrobial activity profiles between individual peptides. While Diapausins in beetles are

potent antifungal agents which play a role in diapause (Tanaka et al., 2003), the role of Diapausins in nematodes has yet to be established.

3.7. Classification of nematode Defensins reveals complexity in subfamily distribution

Defensins are a superfamily of cysteine-stabilised alpha-beta (CS $\alpha\beta$) peptides which share a tertiary structure comprising an α -helix and two antiparallel β -sheets stabilised by three disulfide bonds (Zhu et al., 2005). Defensins can be classified into multiple groups of AMPs displaying differences in cysteine organisation. 533 Defensin-encoding genes (3.98 ± 0.50 genes per genome) were identified in this study and were broadly distributed across phylum Nematoda, with representatives in every nematode clade (see Figure 2). Defensin-encoding genes are more restricted in Clades 2/I (only *R. culicivoxax*) and 8/III (only Ascaridoidea and Oxyuroidea spp.), which is consistent with a previous study (Tarr, 2012a). In contrast, Defensin-encoding genes are more widespread in Clades 9/V, 10/IV and 12/IV (see Figure 2); note that two Defensin-encoding genes, previously identified from *P. trichosuri* EST datasets (Tarr, 2012a), were not identified here.

Defensin-encoding genes identified in this study were classified into four subfamilies based on the cysteine reference array reported previously (Tarr, 2016): Mollusc/Nematode Defensins (MN), Traditional Insect Defensins (TIDs), Macins and Drosomycins. The distribution of these subfamilies was variable across phylum Nematoda (see Figure 5A). The MN subfamily, previously known as Antibacterial Factors (ABFs; Tarr, 2012b), was the most widely distributed, identified in almost all clades examined (see Figure 5A). MNs are characterised by a cysteine motif consisting of eight cysteine residues which form four disulfide bonds (see Figure 4B). In line with the Diapausins, high amino acid variability exists between the key MN cysteine residues, which may contribute to diversity in antimicrobial activity. MN antimicrobial activity has only been investigated in *A. suum* ASABF- α and *C. elegans* ABF-1 and -2 which possess potent antibacterial and antifungal activity (Kato and Komatsu, 1996; Kato et al., 2002). MNs were identified in every nematode clade except for Clade 12/IV, suggesting that they may have been acquired by an ancestral nematode and subsequently lost in the genera now lacking MN-encoding genes (Clade 2/I *Trichinella* and *Trichuris* spp., Clade 8/III filarial species, and all Clade 12/IV species). The

391 reduced distribution of the MN superfamily is consistent with the broadly reduced AMP
392 complement in Clade 2/I and 8/III suggesting a negative selective pressure for AMP-encoding
393 genes in these clades, whereas Clade 12/IV displays other Defensin subfamilies such that the lack
394 of MNs could be due to functional redundancy.

395 A MN subfamily member with only six cysteine residues and a rearranged disulfide bonding
396 pattern, known as ASABF-6Cys- α , was previously identified in *A. suum* (Minaba et al., 2009).
397 Previous analysis of EST and draft genomic datasets identified homologs of ASABF-6Cys- α
398 beyond Ascaridida in *Ancylostoma ceylanicum*, *Caenorhabditis briggsae*, *Caenorhabditis remanei*
399 and *Meloidogyne hapla* (Tarr, 2012a). However, this study has revealed that the *A. ceylanicum*, *C.*
400 *briggsae* and *C. remanei* sequences are, in fact, not homologs of ASABF-6Cys- α but are traditional
401 MN Defensins which were previously incomplete. The *M. hapla* sequences were not included in
402 this analysis as they lack key MN motifs. In this study, a single putative ortholog of *A. suum*
403 ASABF-6Cys- α was identified in *Parascaris univalens* supporting the original hypothesis that the 6-
404 cysteine variation in the MN Defensin sequence is only present in the Ascaridida (Minaba et al.,
405 2009).

406 Nematode Macin-encoding genes have been previously reported by Gerdol et al. (2012) however
407 this study provides the first profile of these genes across phylum Nematoda. Macin-encoding
408 genes were identified in Clades 6/C and 9/V. Within Clade 9/V, Macin-encoding genes are found
409 in all species except all *Caenorhabditis* spp., *Dictyocaulus viviparus* and both *Angiostrongylus*
410 spp.. Notably, Macin-encoding genes are found in parasites of health and economic importance
411 including *H. contortus*, *N. americanus* and *T. circumcincta*. The absence of Macin-encoding genes
412 from the *Caenorhabditis* genera suggests that they have been lost in these species while they
413 have been retained in other free-living Clade 9/V nematodes. The reason for this loss is unclear
414 but reemphasises the need to conduct AMP research beyond *C. elegans*.

415 Macins from other invertebrates possess antibacterial and antifungal activities and can promote
416 regeneration of neuronal tissue in annelids (Jung et al., 2012), however nematode Macins have not
417 been functionally characterised. Macins typically display eight cysteine residues that form four
418 disulfide bonds, however nematode Macins have an additional cysteine pair forming another

419 putative disulfide bond (see Figure 4B). Theromacin from the annelid *Hirudo medicinalis*, which
 420 possesses activity against gram positive bacteria and promotes neuron regeneration, also has an
 421 additional cysteine pair suggesting that nematode Macins may possess similar functional profiles.

422 Drosomycin Defensins possess eight cysteine residues which form four disulfide bonds to produce
 423 a tertiary structure that is distinct from the MN Defensins (see Figure 4B). Drosomycin-encoding
 424 genes were previously identified in free-living *Panagrolaimus superbus* (Clade 10/IV) EST datasets
 425 (Tarr, 2012a). In this study, there were no *Panagrolaimus* genomes included in the nematode
 426 database however, the addition of four *Panagrolaimus* genomes in a subsequent Wormbase
 427 Parasite update (Version 15, October 2020) allowed the confirmation of Drosomycin genes in the
 428 *Panagrolaimus* genera (data not shown).

429 Drosomycin-encoding genes were also identified in Clade 12/IV species and two Clade 9/V
 430 species (*Diploscapter coronatus* and *Diploscapter pachys*). It is unclear whether other Clade 10/IV
 431 species have lost these genes or whether the *Panagrolaimus* species acquired them
 432 independently. Nematode Drosomycin peptides have yet to be functionally characterised however
 433 insect Drosomycin peptides display potent antifungal activity against filamentous fungi (Zhang and
 434 Zhu, 2009). Note that cysteine-rich nematode peptides have previously been designated as
 435 Drosomycin peptides but have since been reclassified as TID subfamily members based on their
 436 cysteine motif positioning (Zhu and Gao, 2014; Tarr, 2016).

437 Traditional Insect Defensin peptides are characterised by the typical insect Defensin cysteine array
 438 where six cysteines form three disulfide bonds (see Figure 4B). Traditional Insect Defensins
 439 display a more limited distribution across phylum Nematoda (28 TID-encoding genes in 12
 440 nematode species). Intriguingly, the 12 nematode species that possess TID-encoding genes are
 441 phylogenetically distinct FLN and PPN species that span five nematode clades; this raises
 442 questions about the evolutionary history of these genes. The TID subfamily members appear to
 443 have functionally diversified such that each peptide has a specific activity range. For example,
 444 *Caenorhabditis remanei* Cremycin-5 possesses antifungal activity whereas Cremycin-15 lacks
 445 antifungal activity but is bioactive against gram positive bacteria (Zhu and Gao, 2014).

3.8. Nemapores are the most widely distributed AMP family, and can be delineated based on Saposin domain characteristics

Nemapores are cysteine-stabilised AMPs found in a wide range of animal species. They are members of the SAPosin-Like Protein (SAPLIP) superfamily which includes other membrane-interacting proteins such as metallophosphoesterases and cell proliferator regulators (Bruhn, 2005). Nemapores are characterised by a signal peptide sequence followed by at least one Saposin domain. Saposin domains comprise of six cysteine residues in a specific arrangement (see Figure 4C). This study updates previous analyses which indicate that Nemapores are the most widely distributed AMP family in phylum Nematoda (see Figure 2; Tarr, 2012a). Indeed, at least one Nemapore-encoding gene ($\sim 15.34 \pm 1.38$ genes per genome) was identified from every species analysed, suggesting that Nemapores are a key component of nematode innate immunity. Nemapore-encoding genes, which were previously reported as absent in *N. americanus*, *Strongyloides stercoralis* and *Meloidogyne incognita* (Tarr, 2012a), were identified in this study reemphasising the value in utilising updated genomic datasets.

Despite broad representation across the phylum, the numbers of Nemapore-encoding genes varies considerably across clades and species. Clades 6/C, 9/V and 10/IV appear to be enriched for Nemapore-encoding genes; for example, *Plectus sambesii* has >100 Nemapore-encoding genes. Nemapore-encoding gene expansions imply that there may be diversification of Nemapore function in these species. As discussed previously Nemapores have already been identified in the ESP of *S. feltiae* activated juveniles and this, coupled with the expansion of Nemapore-encoding genes in Clade 10/IV, highlights that Nemapores may play a key role in Clade 10/IV nematode host infection. Additionally, *C. elegans* Nemapores have been implicated in nutrition, where knockdown of *Ce-spp-5* resulted in a diminished capability to kill ingested *Escherichia coli* resulting in impeded development (Roeder et al., 2010).

There is also variation in the number of Saposin domains encoded by Nemapore genes across nematode genera (see Figure 5B). Previous analyses indicate that Nemapores containing three and four Saposin domains were restricted to species from Clades 2/I and 9/V (Tarr, 2012a). Data generated in this study, however, show that multi-domain Nemapores are more widespread where

475 Nemapores with three or more Saposin domains are represented in all clades. Indeed, *Trichinella*
 476 spp. appear to be enriched for multi-domain Nemapores with some genes containing seven
 477 Saposin domains; multi-domain Nemapore genes have yet to be functionally characterised. In the
 478 amoeba *Naegleria fowleri* multiple bioactive Saposin peptides are processed from larger
 479 multi-peptide precursor sequences (Herbst et al., 2004). It is also possible that nematode multi-
 480 domain Nemapores are post translationally cleaved to produce multiple bioactive single domain
 481 Nemapores however previous attempts to unravel the processing of a four Saposin domain protein
 482 from *Trichinella spiralis* were unable to resolve protein cleavage (Selkirk et al., 2004).

483

484 3.9. Expansion of the GRSPs is not restricted to the *Caenorhabditis* genera

485 This study provides the first pan-phylum profile of the GRSP-encoding genes across phylum
 486 Nematoda; previously GRSPs had only been profiled in *C. elegans* and *C. briggsae* where they
 487 were found to be enriched relative to several other model organisms (Ying et al., 2016). The data
 488 presented here demonstrates that GRSPs are, in fact, the most abundant AMP family in phylum
 489 Nematoda (3145 genes; 23.47 ± 2.26 genes per genome) with Clades 9/V and 10/IV exhibiting
 490 enriched profiles. This evolutionary expansion of GRSPs extends beyond *Caenorhabditis* genera
 491 (see Figure 2), where GRSP-encoding genes are absent from only two nematode species
 492 [*Dracunculus medinensis* and *Litomosoides sigmodontis* (Clade 8/III)] which suggests importance
 493 to nematode biology. Functional characterisation of GRSPs has been restricted to *C. elegans*
 494 where their expression has been profiled post microbial infection (Ying et al., 2016), and
 495 antibacterial and antifungal assays reveal antimicrobial properties that do not involve membrane
 496 permeabilisation (Couillault et al., 2004; Lim et al., 2016). Further work will reveal if this
 497 characteristic is conserved across nematode GRSPs. In addition, it is interesting to note that a
 498 number of Neuropeptide-Like Proteins (NLPs) have been classified as GRSPs (NLP-10, NLP-12,
 499 NLP-21 and NLP-24-34), highlighting a need to unravel the role of these peptides.

500

3.10. Stage specific transcriptome data highlight that AMPs are expressed in key nematode life stages.

To assess the biological relevance of AMPs across nematode life stages, publicly available life stage specific RNAseq datasets were examined for 15 species (10 APN, 1 EPN, 3 PPN, 1 FLN). >85% of AMP genes were expressed (TPM >2) in at least one life stage (see Figure 1E). Notably, despite significant variation in the total number of AMP genes encoded within individual nematode species, this did not translate to major variability in the proportion of total AMP genes expressed. For example, in *Trichuris muris* six of the seven (86%) identified AMP genes were expressed whereas in *P. pacificus* 88 of the 102 (86%) identified AMP genes were expressed. There appear to be differences in the expression of AMP genes from the same AMP family such that no AMP family is consistently expressed in a specific life stage in the five parasitic species examined in more detail (see Supplementary Figure S1). This suggests that closely related AMPs may have functionally diversified and reinforces the need to further characterise these families in other species. Notably, in *A. suum* and *H. contortus* there is clear elevation in the expression of AMP genes in life stages which are more exposed to the host gastrointestinal system further supporting that these genes may play a role in nematode innate immunity.

3.11. Improved genome and transcriptome data will drive AMP discovery and functional characterisation in phylum Nematoda

Genome and transcriptome studies possess inherent caveats that make extrapolation of the data challenging. For example, in this study there are significant differences in the quality of the genome datasets where some chromosomal level assemblies (e.g. *Strongyloides ratti*) are available while others remain at contig level (e.g. *M. hapla*). The quality of genome datasets was considered through the integration of Benchmarking Universal Single-Copy Ortholog (BUSCO) complete scores for each genome (see Figure 2; Simão et al., 2015) such that there appears to be a correlation between the number of AMPs genes identified and genome quality as denoted by BUSCO score; for example, 7 AMP genes in *Parascaris equorum* (BUSCO complete score of 5) relative to 31 AMP genes in the closely related *Parascaris univalens* (BUSCO complete score of

90.9). As nematode genome assemblies and annotations are improved, more AMP genes will be uncovered to reveal a more comprehensive nematode antimicrobial peptidome.

The homology-derived approach used in this study to identify AMPs was driven by query sequences primarily derived from *C. elegans* and *A. suum*, such that the HMM models are likely to be more biased towards the identification of *Caenorhabditis*- or *Ascaris*-like AMPs. Consequently, it is possible that AMPs which have evolved rapidly in distantly related species may be overlooked using this approach, underscoring the need to develop alternative strategies for *in silico* AMP discovery. Indeed, it is difficult to assess the significance of the apparent reduced AMP profiles in Clades 2/I, 8/III and 12/IV until comprehensive AMP discovery approaches have been deployed across phylum Nematoda. Verification of AMPs in nematode tissues and biofluids using peptidomics technologies (Atkinson et al., 2021) will seed follow-on functional biology.

4. Conclusions:

This study provides a comprehensive genome-level pan-phylum profile of nematode AMP-encoding genes which will inform fundamental nematode biology and seed novel antimicrobial discovery. The data demonstrate that nematodes have a rich AMP complement that includes 5887 AMP-encoding genes in 109 species, ~90% of which are novel. The pan-phylum approach has revealed that APNs and PPNs possess fewer AMP-encoding genes than FLNs suggesting that parasitic nematodes may possess a more specific AMP arsenal tailored towards the host environments they inhabit. Major differences in the distribution and abundance of AMP families were evident across the phylum where Clade 2/I, 8/III and 12/IV species possess fewer AMP-encoding genes, while Clade 9/V and 10/IV species have expanded AMP repertoires. Two AMP families, Cecropins and Diapausins, appear to be restricted to Ascarids and FLNs respectively whereas Defensins, Nemapores and GRSPs were more widely distributed across the phylum. This study provides a springboard for functional biology to unravel the importance of AMPs to parasite and infection biology and will inform future drug discovery pipelines.

Acknowledgments

The authors acknowledge support for this work from: Biotechnology and Biological Sciences Research Council grant (BB/H019472/1), Biotechnology and Biological Sciences Research Council/Boehringer Ingelheim grants (BB/MO10392/1, BB/T016396/1), and Department of Agriculture Environment and Rural Affairs Northern Ireland (DAERA).

Figure Legends:

Figure 1: An *in silico* AMP discovery pipeline reveals that phylum Nematoda is a rich source

of AMPs. A) Summary of homology-based methods for putative AMP identification. B) Number of AMP-encoding genes identified for each nematode AMP family. C) Mean number of AMP genes per genome for each AMP family across nematode lifestyles. D) Total number of AMP-encoding genes across each nematode lifestyle. E) Summary of AMP-gene expression (> 2TPM) relative to the number of AMP-encoding genes for key nematode species. Error bars indicate standard error of the mean. *** P<0.001; **** P<0.0001. AA, Amino Acid; AMP, Antimicrobial Peptide; APN, Animal Parasitic Nematode; BLAST, Basic Local Alignment Search Tool; BLASTP, Protein BLAST; CSαβ, Cysteine-Stabilised Alpha Beta peptide superfamily; FLN, Free-Living Nematode; GR, Glycine Rich; GRSP, Glycine Rich Secreted Peptide; HMM, Hidden Markov Model; PP, Propeptide; PPN, Plant Parasitic Nematode; SP, Signal Peptide; TBLASTN, Translated nucleotide BLAST; WBPSv14, WormBase Parasite Version 14

Figure 2. Relative abundance of AMP genes reveal AMP profile diversity across nematode

clades and lifestyles. Nematode species (109 nematode species; 134 genomes) are organised by clade (Blaxter et al., 1998, van Megan et al 2009) and superfamily designation (NCBI taxonomy browser; <https://www.ncbi.nlm.nih.gov/taxonomy>). Relative abundance was calculated as a percentage of AMP genes relative to total protein coding genes per genome. BUSCO complete scores were obtained from Wormbase Parasite V14 (<https://parasite.wormbase.org/index.html>; Bolt et al 2018). Asterisk indicates multiple genome assemblies were used.

Figure 3: Ascaridoidea-restricted Cecropins display significant amino acid conservation.

Weblogo generated using 34 cecropin peptide sequences (4 peptides excluded due to gene mispredictions; GS_20806, TCNE_0000065801, Tcan_02070, GS_06967) using WebLogo3

585 (<https://weblogo.threeplusone.com/>), coloured according to amino acid hydrophobicity
 586 (hydrophilic=blue, neutral=green, hydrophobic=black). Amino acid letter height reflects the
 587 probability of conservation at a specific location. Signal peptide, mature peptide and propeptide
 588 indicated.

589 **Figure 4: Cysteine organisation of the nematode cysteine-stabilised AMP families enables**
 590 **delineation of AMP subfamilies.** A: The Diapausin cysteine organisation. Peptide family 3D
 591 structure example: *Gastrophysa atrocyanea* Diapause-specific peptide (RCSB: 2E2F). B:
 592 Organisation of cysteine residues in the CS α β Superfamily members. Peptide family 3D structure
 593 examples: Mollusc/Nematode Defensin (*Ascaris suum* ASABF- α ; RCSB:2D56), Traditional Insect
 594 Defensin (*Sarcophaga peregrina* Sapecin; RCSB: 1L4V), Drosomycin (*Drosophila melanogaster*
 595 Drosomycin; RCSB: 1MYN), Macin (*Hirudo medicinalis* Theromacin; RCSB:2LN8). C: The
 596 Nemapore cysteine organisation. Peptide family 3D structure example: *Caenorhabditis elegans*
 597 Caenopore-5 (RCSB:2JS9). 'C' indicates a Cysteine residue. '-' indicates any other amino acid.
 598 Subscript numbers indicate the range in number of amino acids between each cysteine residue.
 599 Red lines indicate putative disulphide bonds between cysteine residues. Peptide family 3D
 600 structures obtained from Research Collaboratory for Structural Bioinformatics Protein Data Bank
 601 (RCSB PDB) (<https://www.rcsb.org/>; Berman et al., 2000).

602 **Figure 5: The distribution of Defensin subfamilies and multi-domain Nemapores reveal**
 603 **additional AMP family diversity across nematode clades.** A: The distribution of four Defensin
 604 subfamilies [Mollusc/Nematode Defensin (MN), Macin, Traditional Insect Defensin (TID),
 605 Drosomycin] in Defensin-containing nematode species. B: The distribution of multi-domain
 606 Nemapore-encoding genes across nematode genera. Nematode species/genera are organised by
 607 clade (Blaxter et al., 1998, van Megan et al 2009) and superfamily designation (NCBI taxonomy
 608 browser; <https://www.ncbi.nlm.nih.gov/taxonomy>). Black shading indicates the presence of at least
 609 one AMP gene identified via HMM, BLASTp and tBLASTn analyses. BUSCO complete scores
 610 were obtained from Wormbase Parasite V14 (<https://parasite.wormbase.org/index.html>; Bolt et al
 611 2018). Asterisk indicates multiple genome assemblies were used. Note that species which do not
 612 possess Defensin genes are not included in A.

613

614 **Supplementary Data:**

615 **Supplementary Table S1: AMP query sequences used for construction of AMP family**

616 **HMMs.**

617 **Supplementary Table S2: Genomic and transcriptomic datasets used in the study.**

618 **Supplementary Table S3: AMP gene IDs identified from 134 nematode predicted protein**

619 **datasets.**

620 **Supplementary Table S4: Number and relative abundance of AMP genes identified in this**

621 **study.**

622 **Supplementary Figure S1: Summary of transcriptomic data.** Life stage specific heatmaps of
 623 AMP expression for AMP genes from B: *Trichuris muris* (6 genes), C: *Ascaris suum* (32 genes), D:
 624 *Haemonchus contortus* (58 genes), E: *Strongyloides stercoralis* (22 genes), F: *Meloidogyne*
 625 *incognita* (36 genes). Heatmaps generated using Heatmapper (<http://www.heatmapper.ca/>;
 626 Babicki et al., 2016). Expression is shown in Z scores and scaled by row. Genes are clustered
 627 using Average Linkage and Euclidean distance method. Darker shades represent higher
 628 expression and lighter shades represent lower expression.

629

630 **References:**

631 Andersson, M., Boman, A., Boman, H.G., 2003. *Ascaris* nematodes from pig and human make
 632 three antibacterial peptides: isolation of cecropin P1 and two ASABF peptides. Cell. Mol. Life. Sci.
 633 60, 599-606.

634 Armenteros, J.J.A., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von
 635 Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural
 636 networks. Nat. Biotechnol. 37, 420-423.

637 Atkinson, L.E., Liu, Y., McKay, F., Vandewyler, E., Viau, C., Irvine, A., Rosa, B.A., Li, Z., Liang, Q.,
638 Marks, N.J., Maule, A.G., Mitreva, M., Beets, I., Li, L., Mousley, A. 2021. *Ascaris suum* Informs
639 Extrasynaptic Volume Transmission in Nematodes. ACS Chem. Neurosci. 12, 3176-3188.

640 Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., Wishart, D.S., 2016.
641 Heatmapper: web-enabled heat mapping for all. Nucleic Acids. Res. 44, W147-W153.

642 Bányai, L., Patthy, L., 1998. Amoebapore homologs of *Caenorhabditis elegans*. Biochim. Biophys.
643 Acta. 1429, 259-264.

644 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N.,
645 Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids. Res. 28, 235-242.

646 Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R.,
647 Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T., Thomas, W.K., 1998. A molecular evolutionary
648 framework for the phylum Nematoda. Nature 392, 71-75.

649 Bolt, B.J., Rodgers, F.H., Shafie, M., Kersey, P.J., Berriman, M., Howe, K.L., 2018. Using
650 Wormbase Parasite: An Integrated Platform for Exploring Helminth Genomic Data. In: Kollmar, M.
651 (Ed.), Eukaryotic Genomic Databases. Springer, pp. 471-491.

652 Bruhn, H., 2005. A short guided tour through functional and structural features of saposin-like
653 proteins. Biochem. J. 389, 249-257.

654 Chang, D.Z., Serra, L., Lu, D., Mortazavi, A., Dillman, A.R., 2019. A core set of venom proteins is
655 released by entomopathogenic nematodes in the genus *Steinernema*. PLoS Pathog. 15,
656 e1007626.

657 Clarke, D.J., 2020. *Photorhabdus*: a tale of contrasting interactions. Microbiology (Reading). 166.
658 International Helminth Genomes Consortium. 2019. Comparative genomics of the major parasitic
659 worms. Nat. Genet. 51, 163-174.

660 Cortés, A., Wills, J., Su, X., Hewitt, R.E., Robertson, J., Scotti, R., Price, D.R., Bartley, Y.,
661 McNeilly, T.N., Krause, L., Powell, J. J., Nisbet, A.J., Cantacessi, C., 2020. Infection with the sheep

gastrointestinal nematode *Teladorsagia circumcincta* increases luminal pathobionts. Microbiome 8, 1-15.

Couillault, C., Pujol, N., Reboul, J., Sabatier, L., Guichou, J.F., Kohara, Y., Ewbank, J.J., 2004. TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. Nat. Immunol. 5, 488-494.

Dieterich, C., Clifton, S.W., Schuster, L.N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R., Godfrey, J., Minx, P., Mitreva, M., Roeseler, W., Tian, H., Witte, H., Yang, S.P., Wilson, R.K., Sommer, R.J., 2008. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. Nat. Genet. 40, 1193-1198.

Doyle, S.R., in press. Improving helminth genome resources in the post-genomic era. Trends Parasitol.

Dreyer, J., Rautenbach, M., Booyesen, E., van Staden, A.D., Deane, S.M., Dicks, L.M.T., 2019. *Xenorhabdus khoisanus* SB10 produces Lys-rich PAX lipopeptides and a Xenocoumacin in its antimicrobial complex. BMC Microbiol. 19, 132.

Du, Z.-q., Li, B., Shen, X.-l., Wang, K., Du, J., Yu, X.-d., Yuan, J.-j., 2019. A new antimicrobial peptide isoform, Pc-crustin 4 involved in antibacterial innate immune response in fresh water crayfish, *Procambarus clarkii*. Fish Shellfish Immunol. 94, 861-870.

Duckert, P., Brunak, S., Blom, N., 2004. Prediction of proprotein convertase cleavage sites. Protein Eng. Des. Sel. 17, 107-112.

Dunn, O.J., 1961. Multiple comparisons among means. J. Am. Stat. Assoc. 56, 52-64.

El-Ashram, S., Al Nasr, I., Abouhajer, F., El-Kemary, M., Huang, G., Dinçel, G., Mehmood, R., Hu, M., Suo, X., 2017. Microbial community and ovine host response varies with early and late stages of *Haemonchus contortus* infection. Vet. Res. Commun. 41, 263-277.

Félix, M.-A., Ailion, M., Hsu, J.-C., Richaud, A., Wang, J., 2018. *Pristionchus* nematodes occur frequently in diverse rotting vegetal substrates and are not exclusively necromenic, while *Panagrellus redivivoides* is found specifically in rotting fruits. PloS One 13, e0200851.

688 Feng, K., Li, W., Tang, X., Luo, J., Tang, F., 2022. Termicin silencing enhances the toxicity of
689 *Serratia marcescens* Bizio (SM1) to *Odontotermes formosanus* (Shiraki). Pestic. Biochem. Physiol.
690 185, 105120.

691 Gazit, E., Miller, I.R., Biggin, P.C., Sansom, M.S.P., Shai, Y., 1996. Structure and Orientation of
692 the Mammalian Antibacterial Peptide Cecropin P1 within Phospholipid Membranes. J. Mol. Biol.
693 258, 860-870.

694 Gerdol, M., De Moro, G., Manfrin, C., Venier, P., Pallavicini, A., 2012. Big defensins and
695 mytimacins, new AMP families of the Mediterranean mussel *Mytilus galloprovincialis*. Dev. Comp.
696 Immunol. 36, 390-399.

697 Gomez, E.A., Giraldo, P., Orduz, S., 2017. InverPep: A database of invertebrate antimicrobial
698 peptides. J. Glob. Antimicrob. Resist. 8, 13-17.

699 Goodrich-Blair, H., 2007. They've got a ticket to ride: *Xenorhabdus nematophila*-*Steinernema*
700 *carpocapsae* symbiosis. Curr. Opin. Microbiol. 10, 225-230.

701 Gu, J., Gao, B., Zhu, S., 2018. Characterization of bi-domain drosomycin-type antifungal peptides
702 in nematodes: An example of convergent evolution. Dev. Comp. Immunol. 87, 90-97.

703 Harris, T.W., Arnaboldi, V., Cain, S., Chan, J., Chen, W.J., Cho, J., Davis, P., Gao, S., Grove, C.A.,
704 Kishore, R., Lee, R.Y.N., Muller H-S., Nakamura, C., Nuin, P., Paulini, M., Raciti, D., Rodgers,
705 F.H., Russell, M., Schindelman, G., Auken, K.V., Wang, Q., Williams, G., Wright, A.J., Yook, K.,
706 Howe, K.L., Schedl, T., Stein, L., Sternberg, P.W., 2020. WormBase: a modern Model Organism
707 Information Resource. Nucleic Acids. Res. 48, D762-D767.

708 Herbst, R., Marciano-Cabral, F., Leippe, M., 2004. Antimicrobial and pore-forming peptides of free-
709 living and potentially highly pathogenic *Naegleria fowleri* are released from the same precursor
710 molecule. J. Biol. Chem. 279, 25955-25958.

711 Imai, Y., Meyer, K.J., Iinishi, A., Favre-Godal, Q., Green, R., Manuse, S., Caboni, M., Mori, M.,
712 Niles, S., Ghiglieri, M., Honrao, C., Ma, X., Guo, J.J., Makriyannis, A., Linares-Otoya, L., Bohringer,
713 N., Wuisan, Z.G., Kaur, H., Wu, R., Mateus, A., Typas, A., Savitski, M.M., Espinoza, J.L.,

714 O'Rourke, A., Nelson, K.E., Hiller, S., Noinaj, N., Schaberle, T.F., D'Onofrio, A., Lewis, K., 2019. A
715 new antibiotic selectively kills Gram-negative pathogens. *Nature* 576, 459-464.

716 Jenkins, T.P., Pritchard, D.I., Tanasescu, R., Telford, G., Papaiakevou, M., Scotti, R., Cortés, A.,
717 Constantinescu, C.S., Cantacessi, C., 2021. Experimental infection with the hookworm, *Necator*
718 *americanus*, is associated with stable gut microbial diversity in human volunteers with relapsing
719 multiple sclerosis. *BMC Biol* 19, 1-17.

720 Jung, S., Sonnichsen, F.D., Hung, C.W., Tholey, A., Boidin-Wichlacz, C., Haeusgen, W., Gelhaus,
721 C., Desel, C., Podschun, R., Waetzig, V., Tasiemski, A., Leippe, M., Grotzinger, J., 2012. Macin
722 family of antimicrobial proteins combines antimicrobial and nerve repair activities. *J. Biol. Chem.*
723 287, 14246-14258.

724 Kato, Y., Aizawa, T., Hoshino, H., Kawano, K., Nitta, K., Zhang, H., 2002. abf-1 and abf-2, ASABF-
725 type antimicrobial peptide genes in *Caenorhabditis elegans*. *Biochem. J.* 361, 221-230.

726 Kato, Y., Komatsu, S., 1996. ASABF, a Novel Cysteine-rich Antibacterial Peptide Isolated from the
727 Nematode *Ascaris suum* - Purification, primary structure, and molecular cloning of cDNA. *J. Biol.*
728 *Chem.* 271, 30493-30498.

729 Lee, J.Y., Boman, A., Sun, C.X., Andersson, M., Jörnvall, H., Mutt, V., Boman, H.G., 1989.
730 Antibacterial peptides from pig intestine: isolation of a mammalian cecropin. *Proc. Natl. Acad. Sci.*
731 *U. S. A.* 86, 9159-9162.

732 Lim, M.P., Firdaus-Raih, M., Nathan, S., 2016. Nematode Peptides with Host-Directed Anti-
733 inflammatory Activity Rescue *Caenorhabditis elegans* from a *Burkholderia pseudomallei* Infection.
734 *Front. Microbiol.* 7, 1436.

735 Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey,
736 A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence analysis
737 tools APIs in 2019. *Nucleic Acids. Res.* 47, W636-W641.

738 Massey Jr, F.J., 1951. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* 46,
739 68-78.

740 McCoy, C.J., Atkinson, L.E., Zamanian, M., McVeigh, P., Day, T.A., Kimber, M.J., Marks, N.J.,
741 Maule, A.G., Mousley, A., 2014. New insights into the FLPerigic complements of parasitic
742 nematodes: Informing deorphanisation approaches. *EuPA Open. Proteom.* 3, 262-272.

743 McKay, F.M., McCoy, C.J., Crooks, B., Marks, N.J., Maule, A.G., Atkinson, L.E., Mousley, A.,
744 2021. In silico analyses of neuropeptide-like protein (NLP) profiles in parasitic nematodes. *Int. J.*
745 *Parasitol.* 52, 77-85.

746 Minaba, M., Ueno, S., Pillai, A., Kato, Y., 2009. Evolution of ASABF (*Ascaris suum* antibacterial
747 factor)-type antimicrobial peptides in nematodes: putative rearrangement of disulfide bonding
748 patterns. *Dev. Comp. Immunol.* 33, 1147-1150.

749 Mwangi, J., Hao, X., Lai, R., Zhang, Z.-Y., 2019. Antimicrobial peptides: new hope in the war
750 against multidrug resistance. *Zool. Res.* 40, 488.

751 Parihar, R.D., Dhiman, U., Bhushan, A., Gupta, P.K., Gupta, P., 2022. *Heterorhabditis* and
752 *Photorhabdus* Symbiosis: A Natural Mine of Bioactive Compounds. *Front. Microbiol.* 13, 790339.

753 Pillai, A., Ueno, S., Zhang, H., Lee, J.M., Kato, Y., 2005. Cecropin P1 and novel nematode
754 cecropins: a bacteria-inducible antimicrobial peptide family in the nematode *Ascaris suum*.
755 *Biochem. J.* 390, 207-214.

756 Rödelberger, C., Sommer, R.J., 2011. Computational archaeology of the *Pristionchus pacificus*
757 genome reveals evidence of horizontal gene transfers from insects. *BMC Evol. Biol.* 11, 239.

758 Roeder, T., Stanisak, M., Gelhaus, C., Bruchhaus, I., Grotzinger, J., Leippe, M., 2010. Caenopores
759 are antimicrobial peptides in the nematode *Caenorhabditis elegans* instrumental in nutrition and
760 immunity. *Dev. Comp. Immunol.* 34, 203-209.

761 Rončević, T., Gerdol, M., Mardirossian, M., Maleš, M., Cvjetan, S., Benincasa, M., Maravić, A.,
762 Gajski, G., Krce, L., Aviani, I., Hrabar, J., Trumbić, Ž., Derks, M., Pallavicini, A., Weingarth, M.,
763 Zoranić, L., Tossi, A., Mladineo, I. 2022. Anisaxins, helical antimicrobial peptides from marine
764 parasites, kill resistant bacteria by lipid extraction and membrane disruption. *Acta. Biomater.* 146,
765 131-144.

766 Selkirk, M.E., Hussein, A.S., Chambers, A.E., Goulding, D., Gares, M.P., Vásquez-Lopez, C.,
767 Gárate, T., Parkhouse, R.M.E., Gounaris, K., 2004. *Trichinella spiralis* secretes a homologue of
768 prosaposin. Mol. Biochem. Parasitol. 135, 49-56.

769 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO:
770 assessing genome assembly and annotation completeness with single-copy orthologs.
771 Bioinformatics 31, 3210-3212.

772 Tanaka, H., Sato, K., Saito, Y., Yamashita, T., Agoh, M., Okunishi, J., Tachikawa, E., Suzuki, K.,
773 2003. Insect diapause-specific peptide from the leaf beetle has consensus with a putative iridovirus
774 peptide. Peptides 24, 1327-1333.

775 Tarr, D.E.K., 2012a. Distribution and characteristics of ABFs, cecropins, nemapores, and
776 lysozymes in nematodes. Dev. Comp. Immunol. 36, 502-520.

777 Tarr, D.E.K., 2012b. Nematode antimicrobial peptides. ISJ-Invert. Surviv. J. 9, 122-133.

778 Tarr, D.E.K., 2016. Establishing a reference array for the CS- $\alpha\beta$ superfamily of defensive peptides.
779 BMC Res. Notes. 9, 490.

780 Taylor, M.J., Bandi, C., Hoerauf, A., 2005. *Wolbachia*. Bacterial endosymbionts of filarial
781 nematodes. Adv. Parasitol. 60, 245-284.

782 van der Does, A.M., Hiemstra, P.S., Mookherjee, N., 2019. Antimicrobial Host Defence Peptides:
783 Immunomodulatory Functions and Translational Prospects. In: Matsuzaki, K. (Ed.) Antimicrobial
784 Peptides: Basics for Clinical Application. Springer Signapore, pp. 149-171.

785 van Megen, H., van den Elsen, S., Holterman, M., Karssen, G., Mooyman, P., Bongers, T.,
786 Holovachov, O., Bakker, J., Helder, J., 2009. A phylogenetic tree of nematodes based on about
787 1200 full-length small subunit ribosomal DNA sequences. Nematology 11, 927-950.

788 Wagner, G.P., Kin, K., Lynch, V.J., 2013. A model based criterion for gene expression calls using
789 RNA-seq data. Theory Biosci. 132, 159-164.

790 Wang, G., 2017. Discovery, Classification and Functional Diversity of Antimicrobial Peptides. In:
791 Wang, G. (Ed.) Antimicrobial peptides: Discovery, Design and Novel Therapeutic Strategies. CABI
792 Publishing, pp 1-19.

793 Xu, L., Yang, J., Xu, M., Shan, D., Wu, Z., Yuan, D., 2020. Speciation and adaptive evolution
794 reshape antioxidant enzymatic system diversity across the phylum Nematoda. BMC Biol. 18, 1-18.

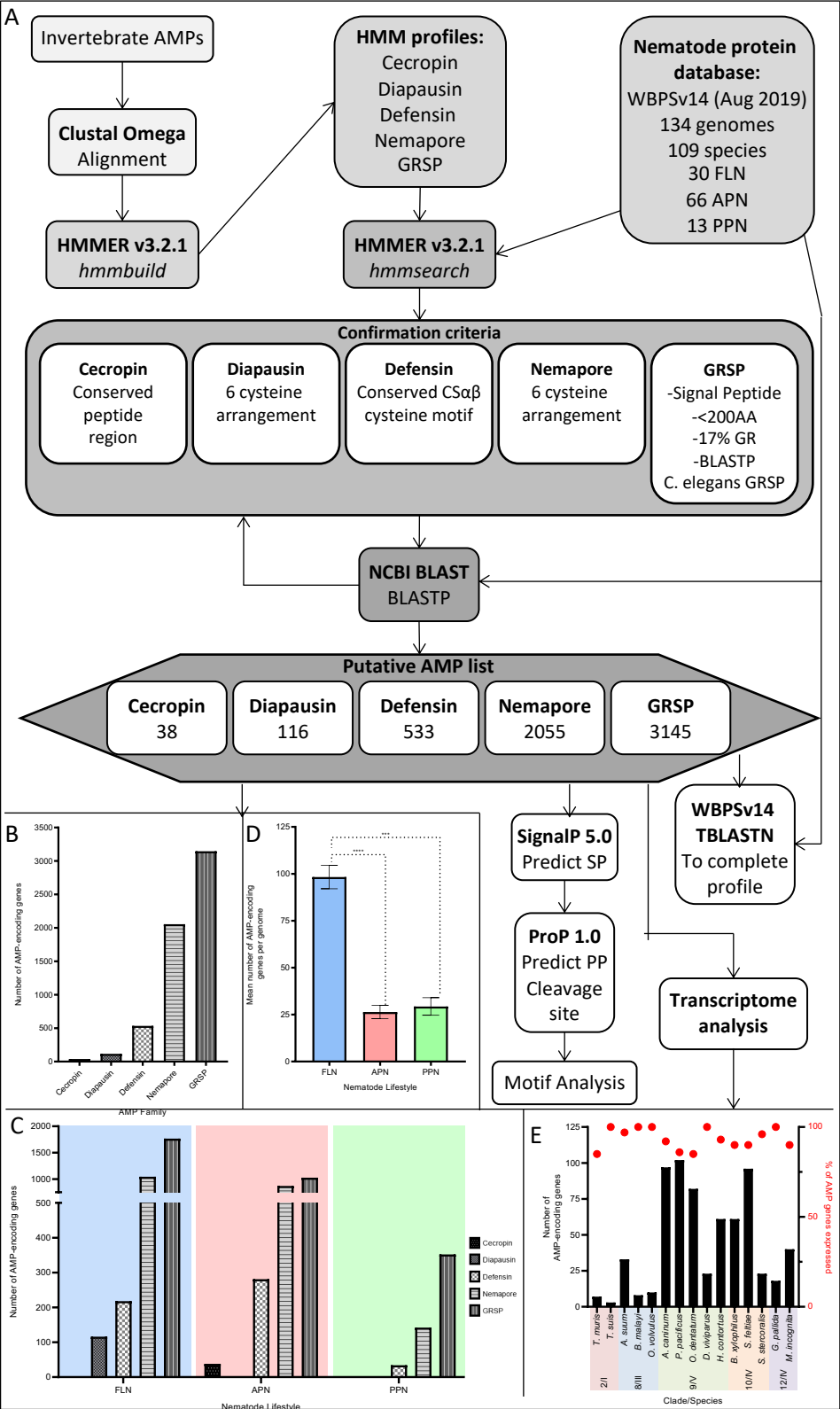
795 Ying, M.Y., Qiao, Y.Y., Yu, L.H., 2016. Evolutionary expansion of nematode-specific glycine-rich
796 secreted peptides. Gene 587, 76-82.










































797 Zhang, S., Hou, C., Xiao, B., Yao, Y., Xiao, W., Li, C., Shi, L., 2021. Identification and function of
798 an Arasin-like peptide from *Litopenaeus vannamei*. Dev. Comp. Immunol. 125, 104174.

799 Zhang, Z.T., Zhu, S.Y., 2009. Drosomycin, an essential component of antifungal defence in
800 *Drosophila*. Insect Mol. Biol. 18, 549-556.








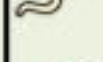



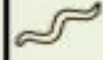










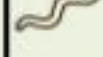


















801 Zhu, S., Gao, B., 2014. Nematode-derived drosomycin-type antifungal peptides provide evidence
802 for plant-to-ectoderm horizontal transfer of a disease resistance gene. Nat. Commun. 5, 3154.




























803 Zhu, S., Gao, B., Tytgat, J., 2005. Phylogenetic distribution, functional epitopes and evolution of
804 the CSαβ superfamily. Cell. Mol. Life. Sci. 62, 2257-2269.



Clade	Superfamily	Species	Lifestyle	AMP Rel % Abund								
				BUSCO			Total	Cecropin	Diapausin	Defensin	Nemapore	GRSP
				0	50	100						
2/I	Diectophymatoidea	<i>Soboliphyme baturini</i>		●●●●								
	Mermithoidea	<i>Romanomermis culicivorax</i>		●●●								
	Trichinelloidea	<i>Trichinella britovi</i>		●●●●●●●●								
		<i>Trichinella murrelli</i>		●●●●●●●●								
		<i>Trichinella nativa</i> *		●●●●●●●●								
		<i>Trichinella nelsoni</i>		●●●●●●●●								
		<i>Trichinella papuae</i>		●●●●●●●								
		<i>Trichinella patagoniensis</i>		●●●●●●●●								
		<i>Trichinella pseudospiralis</i> *		●●●●●●●●								
		<i>Trichinella spiralis</i> *		●●●●●●●●								
		<i>Trichinella sp. T6</i>		●●●●●●●●								
		<i>Trichinella sp. T8</i>		●●●●●●●●								
		<i>Trichinella sp. T9</i>		●●●●●●●●								
		<i>Trichinella zimbabwensis</i>		●●●●●●●●								
		<i>Trichuris muris</i>		●●●●●●●								
		<i>Trichuris suis</i> *		●●●●●●●●								
		<i>Trichuris trichiura</i>		●●●●●●●●								
6/C	Plectoidea	<i>Plectus sambesii</i>		●●●●●●●●								
8/III	Ascaridoidea	<i>Anisakis simplex</i>		●●●●●								
		<i>Ascaris lumbricoides</i>		●●●●●●●●								
		<i>Ascaris suum</i> *		●●●●●●●●								
		<i>Parascaris equorum</i>		●								
		<i>Parascaris univalens</i>		●●●●●●●●								
		<i>Toxocara canis</i> *		●●●●●●●●								
		Dracunculoidea	<i>Dracunculus medinensis</i>		●●●●●●●●							
	Filarioidea		<i>Acanthocheilonema viteae</i>		●●●●●●●●							
		<i>Brugia malayi</i>		●●●●●●●●								
		<i>Brugia pahangi</i>		●●●●●●●●								
		<i>Brugia timori</i>		●●●●●●●								
		<i>Dirofilaria immitis</i>		●●●●●●●●								
		<i>Elaeophora elaphi</i>		●●●●●●●●								
		<i>Litomosoides sigmodontis</i>		●●●●●●●●								
		<i>Loa loa</i> *		●●●●●●●●								
		<i>Onchocerca flexuosa</i> *		●●●●●●●								
		<i>Onchocerca ochengi</i> *		●●●●●●●●								
		<i>Onchocerca volvulus</i>		●●●●●●●●								
		<i>Wuchereria bancrofti</i> *		●●●●●●●●								
	Oxyuroidea	<i>Enterobius vermicularis</i>		●●●●●●●●								
		<i>Syphacia muris</i>		●●●●●●●●								
	Spiruroidea	<i>Gongylonema pulchrum</i>		●●●●●								
	Thelazioidea	<i>Thelazia callipaeda</i>		●●●●●●●●								

bioRxiv preprint doi: <https://doi.org/10.1101/2022.09.26.507570>; this version posted September 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Clade	Superfamily	Species	Lifestyle	BUSCO	AMP Rel % Abund						
					Total	Cecropin	Diapausin	Defensin	Nemapore	GRSP	
				0	50	100					
9/V	Ancylostomatoidea	<i>Ancylostoma caninum</i>		<div><div></div></div>							
		<i>Ancylostoma ceylanicum</i> *		<div><div></div></div>							
		<i>Ancylostoma duodenale</i>		<div><div></div></div>							
		<i>Necator americanus</i>		<div><div></div></div>							
	Diplogasteroidea	<i>Micoletzkyia japonica</i>		<div><div></div></div>							
		<i>Parapristionchus giblindavisi</i>		<div><div></div></div>							
		<i>Pristionchus arcanus</i>		<div><div></div></div>							
		<i>Pristionchus entomophagus</i>		<div><div></div></div>							
		<i>Pristionchus expectatus</i>		<div><div></div></div>							
		<i>Pristionchus fissidentatus</i>		<div><div></div></div>							
		<i>Pristionchus japonicus</i>		<div><div></div></div>							
		<i>Pristionchus maxplancki</i>		<div><div></div></div>							
		<i>Pristionchus mayeri</i>		<div><div></div></div>							
		<i>Pristionchus pacificus</i>		<div><div></div></div>							
		Metastrongyloidea	<i>Angiostrongylus cantonensis</i>		<div><div></div></div>						
			<i>Angiostrongylus costaricensis</i>		<div><div></div></div>						
	Rhabditoidea	<i>Caenorhabditis angaria</i>		<div><div></div></div>							
		<i>Caenorhabditis brenneri</i>		<div><div></div></div>							
		<i>Caenorhabditis briggsae</i>		<div><div></div></div>							
		<i>Caenorhabditis elegans</i>		<div><div></div></div>							
		<i>Caenorhabditis inopinata</i>		<div><div></div></div>							
		<i>Caenorhabditis japonica</i>		<div><div></div></div>							
		<i>Caenorhabditis latens</i>		<div><div></div></div>							
		<i>Caenorhabditis nigoni</i>		<div><div></div></div>							
		<i>Caenorhabditis remanei</i> *		<div><div></div></div>							
		<i>Caenorhabditis sinica</i>		<div><div></div></div>							
		<i>Caenorhabditis tropicalis</i>		<div><div></div></div>							
		<i>Diploscapter coronatus</i>		<div><div></div></div>							
		<i>Diploscapter pachys</i>		<div><div></div></div>							
		<i>Mesorhabditis belari</i>		<div><div></div></div>							
		<i>Oscheius tipulae</i>		<div><div></div></div>							
		Strongyloidea	<i>Cylicostephanus goldi</i>		<div><div></div></div>						
	<i>Heterorhabditis bacteriophora</i>			<div><div></div></div>							
<i>Oesophagostomum dentatum</i>			<div><div></div></div>								
<i>Strongylus vulgaris</i>			<div><div></div></div>								
Trichostrongyloidea	<i>Dictyocaulus viviparus</i> *		<div><div></div></div>								
	<i>Haemonchus contortus</i> *		<div><div></div></div>								
	<i>Haemonchus placei</i>		<div><div></div></div>								
	<i>Heligmosomoides bakeri</i> *		<div><div></div></div>								
	<i>Nippostrongylus brasiliensis</i>		<div><div></div></div>								
	<i>Teladorsagia circumcincta</i>		<div><div></div></div>								

Clade	Superfamily	Species	Lifestyle	AMP Rel % Abund								
				BUSCO			Total	Cecropin	Diapausin	Defensin	Nemapore	GRSP
				0	50	100						
10/IV	Aphelenchoidea	<i>Bursaphelenchus xylophilus</i>		●●●●●●●●								
	Panagrolaimoidea	<i>Panagrellus redivivus</i>		●●●●●●●●								
		<i>Halicephalobus mephisto</i>		●●●●●●●●								
	Strongyloidoidea	<i>Parastrongyloides trichosuri</i>		●●●●●●●●								
		<i>Rhabditophanes sp. KR3021</i>		●●●●●●●●								
		<i>Steinernema carpocapsae</i>		●●●●●●●●								
		<i>Steinernema feltiae</i>		●●●●●●●●								
		<i>Steinernema glaseri</i>		●●●●●●●●								
		<i>Steinernema monticolum</i>		●●●●●●●●								
		<i>Steinernema scapterisci</i>		●●●●●●●●								
		<i>Strongyloides papillosus</i>		●●●●●●●●								
		<i>Strongyloides ratti</i>		●●●●●●●●								
		<i>Strongyloides stercoralis</i>		●●●●●●●●								
		<i>Strongyloides venezuelensis</i>		●●●●●●●●								
		11/IV	Cephaloboidea	<i>Acrobeloides nanus</i>		●●●●●●●●						
12/IV	Sphaerularioidea	<i>Ditylenchus destructor</i>		●●●●●●●●								
	Tylenchoidea	<i>Ditylenchus dipsaci</i>		●●●●●●								
		<i>Globodera pallida</i>		●●●●●●								
		<i>Globodera rostochiensis</i>		●●●●●●								
		<i>Heterodera glycines</i>		●●●●●●								
		<i>Meloidogyne arenaria</i> *		●●●●●●								
		<i>Meloidogyne enterolobii</i>		●●●●●●								
		<i>Meloidogyne floridensis</i> *		●●●●●●								
		<i>Meloidogyne graminicola</i>		●●●●●●								
		<i>Meloidogyne hapla</i>		●●●●●●								
		<i>Meloidogyne incognita</i> *		●●●●●●								
		<i>Meloidogyne javanica</i> *		●●●●●●								

0

Unannotated Gene

>0


0.25


0.5


>0

0.125

0.25

 =Animal Parasitic Nematode

 =Free-living Nematode

 =Plant Parasitic Nematode

Signal Peptide



Mature Peptide

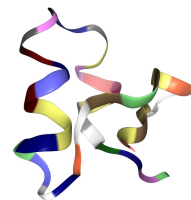
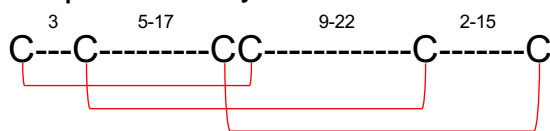


Propeptide



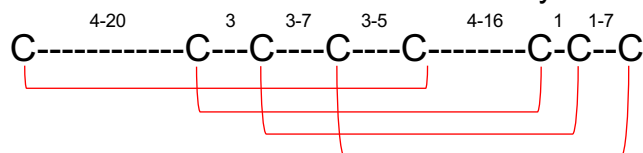
A

Diapausin: 6 cysteines- 3 disulfide bridges

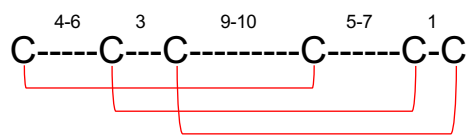


B

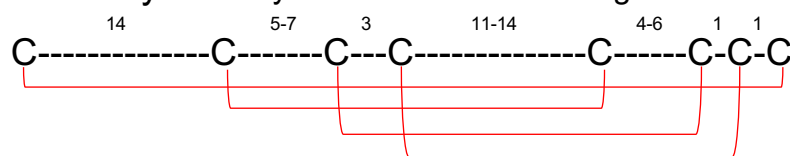
Mollusc/Nematode Defensin: 8 cysteines- 4 disulfide bridges



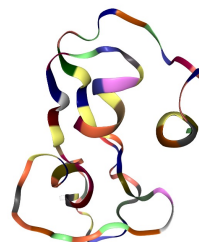
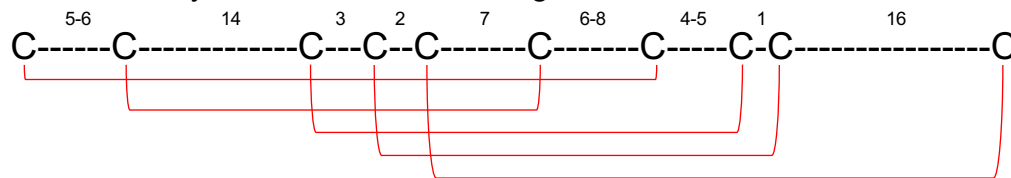
Traditional Insect Defensin: 6 cysteines- 3 disulfide bridges



Drosomycin: 8 cysteines- 4 disulfide bridges

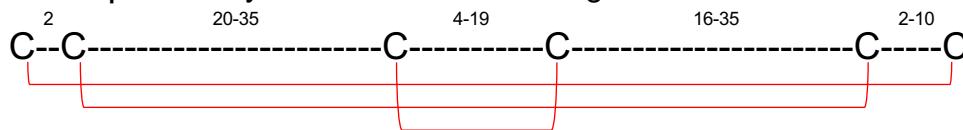


Macin: 10 cysteines- 5 disulfide bridges



C

Nemapore: 6 cysteines- 3 disulfide bridges



A

Clade	Superfamily	Species	BUSCO			MN	Macin	TID	Drosomycin
			0	50	100				
2/I	Mermithoidea	<i>Romanomermis culicivorax</i>	●●●						
6/C	Plectoidea	<i>Plectus sambesii</i>	●●●●●●●●						
8/III	Ascaridoidea	<i>Anisakis simplex</i>	●●●●●●						
		<i>Ascaris lumbricoides</i>	●●●●●●●●						
		<i>Ascaris suum</i> *	●●●●●●●●						
		<i>Parascaris equorum</i>	●●●●●●●●						
		<i>Parascaris univalens</i>	●●●●●●●●						
		<i>Toxocara canis</i> *	●●●●●●●●						
	Oxyuroidea	<i>Enterobius vermicularis</i>	●●●●●●●●						
		<i>Syphacia muris</i>	●●●●●●●●						
9/IV	Ancylostomatoidea	<i>Ancylostoma caninum</i>	●●●●●●●●						
		<i>Ancylostoma ceylanicum</i> *	●●●●●●●●						
		<i>Ancylostoma duodenale</i>	●●●●●●●●						
		<i>Necator americanus</i>	●●●●●●●●						
		<i>Micoletzky japonica</i>	●●●●●●●●						
	Diplogasteroidea	<i>Parapristionchus giblindavisi</i>	●●●●●●●●						
		<i>Pristionchus arcanus</i>	●●●●●●●●						
		<i>Pristionchus entomophagus</i>	●●●●●●●●						
		<i>Pristionchus expectatus</i>	●●●●●●●●						
		<i>Pristionchus fissidentatus</i>	●●●●●●●●						
		<i>Pristionchus japonicus</i>	●●●●●●●●						
		<i>Pristionchus maxplancki</i>	●●●●●●●●						
		<i>Pristionchus mayeri</i>	●●●●●●●●						
		<i>Pristionchus pacificus</i>	●●●●●●●●						
		<i>Caenorhabditis angaria</i>	●●●●●●●●						
	Rhabditoidea	<i>Caenorhabditis brenneri</i>	●●●●●●●●						
		<i>Caenorhabditis briggsae</i>	●●●●●●●●						
		<i>Caenorhabditis elegans</i>	●●●●●●●●						
		<i>Caenorhabditis inopinata</i>	●●●●●●●●						
		<i>Caenorhabditis japonica</i>	●●●●●●●●						
		<i>Caenorhabditis latens</i>	●●●●●●●●						
		<i>Caenorhabditis nigoni</i>	●●●●●●●●						
		<i>Caenorhabditis remanei</i> *	●●●●●●●●						
		<i>Caenorhabditis sinica</i>	●●●●●●●●						
		<i>Caenorhabditis tropicalis</i>	●●●●●●●●						
	Strongyloidea	<i>Diploscapter coronatus</i>	●●●●●●●●						
		<i>Diploscapter pachys</i>	●●●●●●●●						
		<i>Mesorhabditis belari</i>	●●●●●●●●						
		<i>Oscheius tipulae</i>	●●●●●●●●						
		<i>Cylicostephanus goldi</i>	●						
	Trichostrongyloidea	<i>Heterorhabditis bacteriophora</i>	●●●●●●●●						
		<i>Oesophagostomum dentatum</i>	●●●●●●●●						
		<i>Strongylus vulgaris</i>	●●						
		<i>Dictyocaulus viviparus</i> *	●●●●●●●●						
		<i>Haemonchus contortus</i> *	●●●●●●●●						
		<i>Haemonchus placei</i>	●●●●●●●●						
		<i>Heligmosomoides bakeri</i> *	●●●●●●●●						
		<i>Nippostrongylus brasiliensis</i>	●●●●●●●●						
		<i>Teladorsagia circumcincta</i>	●●●●●●●●						

B

Clade	Superfamily	Species	BUSCO			MN	Macin	TID	Drosomycin
			0	50	100				
10/IV	Aphelenchoidea	<i>Bursaphelenchus xylophilus</i>	●●●●●●●●						
	Panagrolaimoidea	<i>Halickephalobus mephisto</i>	●●●●●●●●						
	Strongyloidoidea	<i>Steinernema carpocapsae</i>	●●●●●●●●						
		<i>Steinernema feltiae</i>	●●●●●●●●						
		<i>Steinernema glaseri</i>	●●●●●●●●						
		<i>Steinernema monticolum</i>	●●●●●●●●						
11/IV	Cephaloboidea	<i>Steinernema scapterisci</i>	●●●●●●●●						
		<i>Acrobeloides nanus</i>	●●●●●●●●						
12/IV	Sphaerularioidea	<i>Ditylenchus destructor</i>	●●●●●●●●						
		<i>Ditylenchus dipsaci</i>	●●●●●●●●						
	Tylenchoidea	<i>Globodera pallida</i>	●●●●						
		<i>Globodera rostochiensis</i>	●●●●●●						
		<i>Heterodera glycines</i>	●●●●●●						
		<i>Meloidogyne arenaria</i> *	●●●●●●						
		<i>Meloidogyne enterolobii</i>	●●●●●●						
		<i>Meloidogyne floricola</i> *	●●●●						
		<i>Meloidogyne graminicola</i>	●●●●						
		<i>Meloidogyne hapla</i>	●●●●●●						
		<i>Meloidogyne incognita</i> *	●●●●●●						
<i>Meloidogyne javanica</i> *	●●●●●●								

Clade	Genus	Number of Saposin Domains						
		1	2	3	4	5	6	7
2/I	<i>Romanomermis</i>							
	<i>Soboliphyme</i>							
	<i>Trichinella</i>							
	<i>Trichuris</i>							
6/C	<i>Plectus</i>							
8/III	<i>Acanthocheilonema</i>							
	<i>Anisakis</i>							
	<i>Ascaris</i>							
	<i>Brugia</i>							
	<i>Dirofilaria</i>							
	<i>Dracunculus</i>							
	<i>Elaeophora</i>							
	<i>Enterobius</i>							
	<i>Gongylonema</i>							
	<i>Litomosoides</i>							
	<i>Loa</i>							
	<i>Onchocerca</i>							
	<i>Parascaris</i>							
	<i>Syphacia</i>							
	<i>Thelazia</i>							
	<i>Toxocara canis</i>							
	<i>Wuchereria</i>							
9/V	<i>Ancylostoma</i>							
	<i>Angiostrongylus</i>							
	<i>Caenorhabditis</i>							
	<i>Cylicostephanus</i>							
	<i>Dictyocaulus</i>							
	<i>Diploscapter</i>							
	<i>Haemonchus</i>							
	<i>Heligmosomoides</i>							
	<i>Heterorhabditis</i>							
	<i>Mesorhabditis</i>							
	<i>Micoletzky</i>							
	<i>Necator</i>							
	<i>Nippostrongylus</i>							
	<i>Oesophagostomum</i>							
	<i>Oscheius</i>							
	<i>Parapristionchus</i>							
	<i>Pristionchus</i>							
10/IV	<i>Strongylus</i>							
	<i>Teladorsagia</i>							
	<i>Bursaphelenchus</i>							
	<i>Halickephalobus</i>							
	<i>Panagrellus</i>							
	<i>Parastrongyloides</i>							
11/IV	<i>Rhabditophanes</i>							
	<i>Steinernema</i>							
	<i>Strongyloides</i>							
12/IV	<i>Acrobeloides</i>							
12/IV	<i>Ditylenchus</i>							
	<i>Globodera</i>							
	<i>Heterodera</i>							
	<i>Meloidogyne</i>							