RESEARCH ARTICLE

Exploring the antimicrobial peptidome of nematodes through phylum-spanning *in silico* analyses highlights novel opportunities for pathogen 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, Belfast, United Kingdom

* a.mousley@qub.ac.uk

Abstract

Antimicrobial Peptides (AMPs) are key constituents of the invertebrate innate immune system and provide critical protection against microbial threat. Nematodes display diverse life strategies where they are exposed to heterogenous, microbe rich, environments highlighting their need for an innate immune system. Within the Ecdysozoa, arthropod AMPs have been well characterised, however nematode-derived AMP knowledge is limited. In this study the distribution and abundance of putative AMP-encoding genes was examined in 134 nematode genomes providing the most comprehensive profile of AMP candidates within phylum Nematoda. Through genome and transcriptome analyses we reveal that phylum Nematoda is a rich source of putative AMP diversity and demonstrate (i) putative AMP group profiles that are influenced by nematode lifestyle where free-living nematodes appear to display enriched putative AMP profiles relative to parasitic species; (ii) major differences in the putative AMP profiles between nematode clades where Clade 9/V and 10/IV species possess expanded putative AMP repertoires; (iii) AMP groups 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 CSαβ subgroup members; and (v) that putative AMPs are expressed in host-facing life stages and biofluids of key nematode parasites. These data indicate that phylum Nematoda displays diversity in putative AMPs and underscores the need for functional characterisation to reveal their role and importance to nematode biology and host-nematode-microbiome interactions.

Author summary

Nematodes include some of the most abundant animals on Earth, found in almost every natural habitat. Globally, parasitic nematodes cause major impacts to human, animal and plant health. Little is known about how parasitic nematodes interact with microbes despite the microbe-rich environments that many of them inhabit. In other invertebrate animals,

Antimicrobial Peptides (AMPs) are critical in defense against pathogenic microbes. By characterising nematode AMP diversity and exploring the role of nematode-derived AMPs we could reveal novel targets for antimicrobial and anthelmintic development.

In this study we profiled the distribution of five known AMP groups across the nematode phylum. This homology-directed approach demonstrated that, nematodes appear rich in AMP diversity with major differences in the abundance and distribution of putative AMP-encoding genes across nematode lifestyles and clades. In addition, transcriptomic and proteomic analyses show that many of the identified AMP-encoding genes appear biologically relevant in nematode life stages that interact with the host. This study provides a database of potential nematode-derived AMPs which can drive functional analyses to explore the role of nematode AMPs.

Introduction

Antimicrobial Peptides (AMPs) are a group of host defence peptides that act as natural immune effectors and are present in all classes of life [1]. Invertebrates, which lack an adaptive immune system, rely on facets of innate immunity, including AMPs, to protect against microbial threat [2]. Invertebrate AMPs form the first line of defence against invading pathogens including bacteria, fungi and viruses by exerting broad-spectrum antimicrobial activities. Membranolytic AMPs are generally cationic amphipathic peptides which, through various interaction models [carpet, barrel-stave and toroidal pore; [3], bind to microbial cell membranes leading to cell lysis [4]. AMPs also stimulate invertebrate immune regulation, enhancing the endogenous immune response to protect against microbial threat [5]. Whilst AMPs have gained significant attention as novel, resistance-breaking, antimicrobial agents [6], 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. Phylum Nematoda belong to the clade Ecdysozoa which also includes arthropods and tardigrades [7]. Current understanding of ecdysozoan AMP function is primarily derived from arthropods [8] where they have been shown to be critical in defense against pathogens [9–11], however knowledge of the diversity, role and importance of AMPs, including in nematode parasites, is limited [12]. 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. Consequently, gastrointestinal nematodes are likely to require a raft of endogenous AMPs for survival within the host environment. 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 five distinct AMP groups have been identified: Cecropins, Cysteine stabilised α-helix and β-sheet fold (CSαβ) peptides, Nemapores, Diapausins, and Glycine Rich Secreted Peptides (GRSPs) [13–21]. Each of these AMP groups are characterised by distinct sequence features, where the CSαβ peptides, Nemapores and Diapausins are defined by cysteine motifs [22–24] and the Cecropins and GRSPs are linear peptides defined by a specific signature sequence [23] and an enrichment of glycine residues respectively [20]. The ability to uncover AMPs in nematodes has been significantly enhanced by the availability of helminth genome and transcriptome datasets [25]. Indeed, prior to this study, the most complete profile of putative AMPs in nematodes was

conducted by Tarr [19], where draft genomes and Expressed Sequence Tag (EST) datasets were examined in 58 nematode species for a sub-set of AMP groups (Cecropin, CSαβ [previously named Defensin], Nemapore). A constraint, noted by Tarr [19], was the lack of publicly available nematode 'omics data which prevented the comprehensive characterisation of nematode AMP groups. Recent expansions and quality improvements in nematode 'omics data present a timely opportunity for the *in-silico* 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 groups (Cecropin, CSαβ, Diapausin, Nemapore, GRSP) using a BLAST/HMM-directed bioinformatics approach. To our knowledge, the data presented here represent the most comprehensive profile of putative nematode AMPs to date and signify a critical advance in our understanding of the potential nematode AMP repertoire. These data expose nematodes as a rich source of putative AMP diversity that displays both conservation and variation within and between nematode clades, lifestyles and life stages likely reflecting biologically relevant trends.
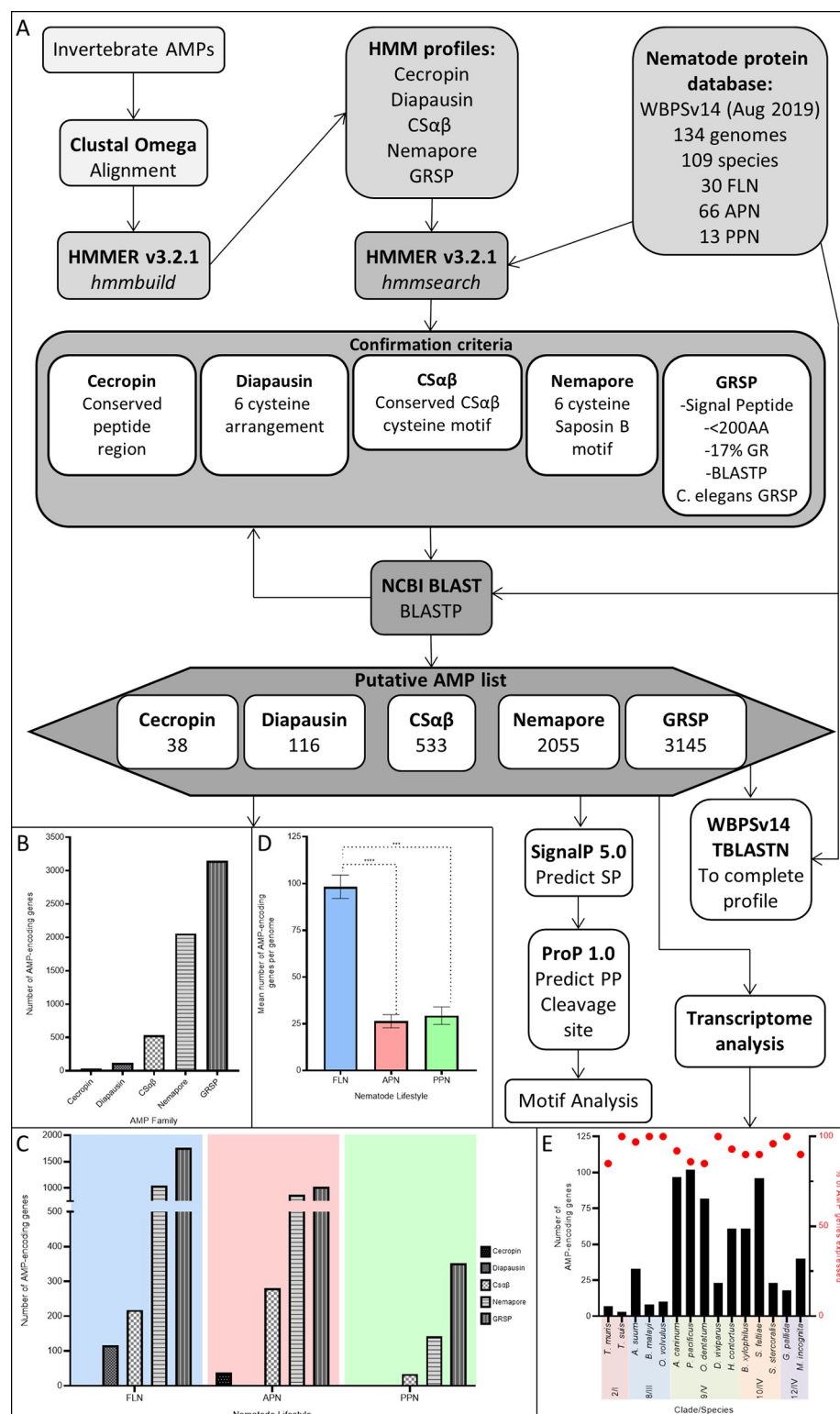
These data reveal putative AMP diversity within phylum Nematoda and provide a database of potential AMPs that can be used to explore AMP biology in therapeutically relevant parasites to underpin novel nematode control approaches. In addition, this study has provided a novel library of potential nematode-derived natural antimicrobials which may exhibit desirable characteristics for future antimicrobial discovery.

Materials and methods

Identification of putative AMP-encoding genes in phylum Nematoda

A summary of the methods in this study have been provided graphically (see Fig 1A). Genes encoding putative AMPs representing five groups (Cecropin, Diapausin, CSαβ, Nemapore and GRSP-see Table 1) were identified based on sequence similarity to the five previously identified nematode AMP groups using Hidden Markov Model (HMM) and Basic Local Alignment Search Tool (BLAST) searches. Each of these AMP groups are represented by distinct sequence features (see Table 1). HMM profiles were built using HMMER v3.2.1 (www.hmmer.org) for each AMP group and, where appropriate, for individual subgroups. Previously identified AMPs (see S1 File) were aligned using CLUSTAL Omega with default settings [26], followed by *hmmbuild* (with default settings) to construct specific AMP group HMM profiles. Where required, alignments were manually adjusted to align AMP motifs.

A database of nematode predicted proteins was collated from the predicted protein datasets derived from 134 nematode genomes [WormBase ParaSite V14; see S2 File;] [34–35]. Constructed HMM models were subsequently searched against the concatenated nematode predicted protein database using the *hmmsearch* command. The results of each HMM search were cross-checked to ensure that the query sequences used to build the model scored highly. This confirmed that the HMM profile successfully identified the query sequences and was specific for each AMP group. For all AMP groups, except the GRSPs, search hits with an E value <0.01 were manually assessed for high peptide similarity and conserved AMP group motifs. Where a sequence appeared incomplete, as a result of an assembly or annotation error (i.e., missing an exon or incorrectly predicted as a longer protein), the output was retained to provide a comprehensive profile of putative AMP-encoding genes. For the Cecropin group, new hits were included based only on similarity to the query sequences, whereas for the CSαβ, Diapausin and Nemapore groups hits were included based on the presence of specific conserved cysteine motifs. Note that for the Nemapores, any sequences with additional protein domains (checked using InterPro; https://www.ebi.ac.uk/interpro/) were excluded from the study to ensure that Saposin-like proteins (SAPLIPs) with likely other functions [36] were not included.

Fig 1. An *in silico* AMP discovery pipeline reveals that phylum Nematoda is a rich source of putative AMPs. A) Summary of BLAST/HMM-based methods for putative AMP identification. B) Number of putative AMP-encoding genes identified for each nematode AMP group. C) Total number of putative AMP-encoding genes for each AMP group across nematode lifestyles. D) Mean number of putative AMP-encoding genes per genome across each nematode lifestyle. E) Summary of putative AMP-gene expression (>2 TPM) relative to the number of putative AMP-encoding genes for key nematode species. Error bars indicate standard error of the mean. *** P<0.001; **** P<0.0001.

https://doi.org/10.1371/journal.pntd.0011618.g001

As the GRSPs are characterised by repetitive glycine-rich repeats and lack other conserved motifs, individual HMMs were constructed based on the 17 GRSP subgroups identified in *C. elegans* and *Caenorhabditis briggsae* GRSPs [20]. Due to the highly repetitive glycine-rich nature of the GRSPs, all hits from these HMMs up to an E value of 10 were assessed and were confirmed based on GRSP criteria established previously [20]. As a result, putative GRSP hits were only considered to be authentic if they were ≤200 amino acids, contained a putative signal peptide sequence, and possessed mature peptide glycine content >17%. Despite the construction of 17 subgroup HMM profiles for the GRSPs, results overlapped making classification of putative GRSPs according to the pre-defined *Caenorhabditis* GRSP subgroups difficult. Consequently, GRSPs hits were not classified into discrete subgroups in this study.

Post HMM analyses, all HMM-derived putative AMPs were employed as BLASTp query sequences in the NCBI command line BLAST application (version 2.9.0, April 2019) against the concatenated nematode predicted protein database constructed previously. BLASTp returns were confirmed using the same criteria as above for each AMP group and new confirmed hits were repeatedly used as BLASTp queries until no new results were retrieved. Due to the high number of putative GRSP hits returned by HMM and BLASTp searches and the potential for false-positive returns, all putative GRSP hits were subjected to an additional BLASTp search against the *C. elegans* protein database on the WormBase ParaSite BLAST server (with the low complexity filter off; Version 14); where a GRSP identified previously [20] was not returned within the top five hits of the additional BLASTp search (sorted by E value) the original putative GRSP hit was removed from further analyses.

Putative AMP prepropeptides were analysed for the presence of signal peptide sequences and propeptide cleavage sites using SignalP 5.0 [standard Sec/SPI settings, Eukarya organism; [37] and ProP 1.0 [default furin-specific prediction; [38] respectively. For complete sequences, putative mature peptides (based on the removal of predicted signal peptide and propeptide regions) were submitted to the Antimicrobial Peptide Scanner [APS; model- vr.2 Feb2020; [39] to assess antimicrobial potential.

The CSαβ group was subsequently classified into subgroups (Mollusc/Nematode Defensin, Traditional Insect Defensin, Macin and Drosomycin) based on number of cysteines and their arrangement according to the cysteine reference array established previously [24].

Table 1. Summary of five known nematode AMP group characteristics.

Nematode AMP Group	Peptide characteristics	Example member
Cecropin	Amphipathic α-helical structure [27]	*Ascaris suum* Cecropin P1 [28]
Diapausin	Two α-helices and a triple-stranded β-sheet structure stabilised by six cysteine motif distinct to that of CSαβ peptides [29]	*Manduca sexta* Diapausin-1 [30]
CSαβ	α-helix and two β sheet structure stabilised by at least six cysteines which form three disulfide bonds [24]	*Ascaris suum* ASABF-α [31]
Nemapore	Saposin-like fold consisting of amphipathic α-helices, strengthened by three disulfide bonds [32]	*Caenorhabditis elegans* SPP-3 [33]
GRSP	Secreted peptides with a mature peptide glycine content between 17%-74% [20]	*Caenorhabditis elegans* NLP-31 [17]

https://doi.org/10.1371/journal.pntd.0011618.t001

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

Clustered analysis of sequences (CLANS) analysis

Sequence similarity between identified AMP candidate groups was visualised using the Clustered Analysis of Sequences (CLANS) Java application [40]. For all putative AMP groups except the GRSPs, a FASTA file of full-length complete sequences were submitted to the CLANS web-utility (https://toolkit.tuebingen.mpg.de/tools/clans) in the MPI Bioinformatics Toolkit [41–42] to generate all-against-all pairwise BLASTP searches using the BLOSUM62 scoring matrix. Sequence similarity networks were visualised in two dimensions using the CLANS Java application at the most appropriate E-value limit to facilitate cluster separation after 10,000 clustering rounds. The CLANS network was manually coloured and annotated to reflect group and or subgroup classification.

Transcriptome analyses of putative AMP-encoding genes

Publicly available life-stage specific RNAseq datasets for 15 nematode species, representing a range of nematode clades and lifestyles, were retrieved from WormBase ParaSite v14 (see S2 File). A threshold of ≥ 2 Transcripts Per Million (TPM) was employed as a cut-off for expression [43]. For heatmap construction TPM values (TPM +1) were Log2 transformed and converted to Z values using Heatmapper [www.heatmapper.ca/; [44]. Heatmaps were scaled by row and genes were clustered by row using the Average Linkage method with Euclidean Distance Measurement method.

Statistical analyses

All graphs and statistical analyses were produced using GraphPad Prism 9 (www.graphpad.com/). Data were tested for normality using the Kolmogorov-Smirnov test [45]. Non-normally distributed data were tested for significance using Krustal Wallis Test with Dunn's multiple comparisons test [46].

Results and discussion

Phylum Nematoda is a rich source of putative antimicrobial peptides

To investigate putative AMP abundance and diversity across phylum Nematoda, five known nematode AMP groups [Cecropin, Diapausin, CSαβ, Nemapore and GRSP] [19–21] were profiled in 109 nematode species (134 genomes; representing 7 clades, 30 free-living nematode (FLN), 66 animal parasitic nematode (APN) and 13 plant parasitic nematode (PPN) species; see S2 File). Our analyses identified 5887 putative AMP-encoding genes, of which >90% have not been previously reported, representing all known nematode AMP groups [38 (0.65%) Cecropin-encoding genes, 116 (1.97%) Diapausin-encoding genes, 533 (9.05%) CSαβ-encoding genes, 2055 (34.91%) Nemapore-encoding genes and 3145 (53.42%) GRSP-encoding genes; see Fig 1B and S3 File]. CLANS analyses of all nematode AMP groups except the GRSPs largely supported the designation of these sequences within each group and/or subgroup

Fig 2. Relative abundance of putative AMP genes reveal AMP profile diversity across nematode clades and lifestyles. Nematode species (109 nematode species; 134 genomes) are organised by clade [47–48] and superfamily designation (NCBI taxonomy browser; https://www.ncbi.nlm.nih.gov/taxonomy). Relative abundance was calculated as a percentage of putative 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] [34]. ^ indicates a genome with <50 BUSCO complete score. Asterisk indicates multiple genome assemblies were used.

https://doi.org/10.1371/journal.pntd.0011618.g002

(see S1 Fig). Beyond the identified genes displaying sequence homology to existing nematode AMPs, a high proportion of the predicted mature peptides from complete AMP candidates were also predicted to have antimicrobial activity by the APS. Indeed, 90% of putative Cecropin peptides, 96% of putative Diapausin peptides, 99% of putative CSαβ peptides, 81% of putative Nemapore peptides and 93% of putative GRSPs achieved a prediction probability >0.5 (see S3 File) supporting their designation as AMP candidates. Whilst the relative abundance and profile of putative AMP-encoding genes vary across the phylum (see Fig 2) the representation of putative AMPs in every nematode species examined underscores the potential importance of AMPs to nematode biology. These data provide the most comprehensive profile of AMP candidates across the phylum Nematoda and demonstrate that nematodes appear AMP-rich. As these putative AMPs were identified through *in-silico* methods and share conserved sequences with experimentally validated AMPs it is important to note that these genes require functional characterisation to identify those with conserved AMP activities.

AMP profile, abundance, and group diversity vary across nematode lifestyles

Our analyses reveal that putative AMP group profiles vary across nematode lifestyles (see Fig 1C). FLN and APN species display the greatest AMP group diversity (Diapausin, CSαβ, Nemapore and GRSP groups represented in FLNs; Cecropin, CSαβ, Nemapore and GRSP groups represented in APNs), however PPNs lack both Diapausin and Cecropin groups.

FLNs also possess the greatest number of putative AMP-encoding genes (98.31 ± 6.23 genes per genome) relative to both APNs (26.37 ±3.53 genes per genome; p<0.0001) and PPNs

(29.33±4.67 genes per genome; p = 0.0001) (see Fig 1D). The higher abundance of putative AMP-encoding genes in FLNs may reflect the diversity in the terrestrial and marine environments that FLNs inhabit, and the potential heterogeneity in microbes that they are likely to encounter. Parasitic nematodes, in contrast, live endogenously within animals and plants and could 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. This may also suggest that parasitic nematodes possess distinct AMP groups which have yet to be identified and require novel approaches to AMP discovery.

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

There was considerable variability in the distribution and abundance of putative AMP groups across nematode clades (see Fig 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 putative AMP diversity (see Fig 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 putative AMP diversity across the phylum (80.77 ±5.74 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 Fig 2 and S4 File). By this measure, Clade 9/V displayed the highest relative abundance of putative AMP genes (average 0.31 ±0.02% per genome). Within Clade 9/V, there were differences between the relative abundance of putative 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 putative AMP genes was noted for the entomopathogenic nematode (EPN), *Heterohabditis bacteriophora*, where only 0.09% of PCGs encode AMP genes. The high abundance of putative 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 putative AMP genes in *H. bacteriophora* may be related to its symbiotic relationship with gram negative *Photorhabdus* bacteria [49]. *Photorhabdus* spp. are key to the lifecycle of *H. bacteriophora* and produce a range of toxins which kill the arthropod host [50]. Recently, an antibiotic with potent activity against gram negative bacteria was discovered in *Photorhabdus* [51]. It is therefore possible that *H. bacteriophora* benefits from AMPs produced by the bacterial symbiont reflecting the lower abundance of endogenous nematode AMPs in this species.

Beyond the AMP-rich *Caenorhabditis* genera, other genera within Clade 9/V also appear to display enriched putative AMP profiles. For example, free-living *Pristionchus* species have an enriched AMP profile with all but one species having >100 putative AMP genes (0.33 ±0.02% of all PCGs). *Pristionchus* feed on rotting plant matter and share mostly phoretic, and occasionally necromenic, relationships with arthropods [52]. It may be the case that *Pristionchus* nematodes require an enriched arsenal of AMPs, to combat the microbe-rich environments they inhabit.

The Ancylostomatoidea and Trichostrongyloidea superfamilies, which include mammalian parasites, also have a higher putative AMP abundance than the other parasitic superfamilies, e.g. the Metastrongyloidea and Strongyloidea, in this clade. Many of these parasitic species that display higher putative AMP abundance, including *Necator americanus*, *Teladorsagia*

circumcincta and *Haemonchus contortus*, are known to alter the host gut microbial composition during infection [53–55]. Whilst it is currently unclear whether the putative AMPs identified here are secreted and/or excreted into the host environment and may contribute to changes in the microbial composition, nematode-derived AMPs, such as *A. suum* Cecropin P1, have been isolated from the host gut [13,16]. Future research should aim to establish whether excreted/secreted nematode AMPs are active against host microbiota and contribute to changes in the host microbiome during infection.

Clade 10/IV also appear to be AMP-rich (64.79 ±9.51 AMP-encoding genes per genome; 0.28 ±0.02% of PCGs in the Clade 10/IV species examined in this study encode putative AMPs). Despite the diversity in lifestyles, most Clade 10/IV species share similar numbers of putative AMPs. *Steinernema* spp. appear to have an elevated putative AMP abundance relative to other Strongyloidoidea superfamily members such as *Strongyloides* spp. Interestingly, like Clade 9/V *H. bacteriophora*, *Steinernema* spp. are also EPNs that display endosymbiotic relationships with gram negative bacteria. *Steinernema* spp. harbour *Xenorhabdus* bacteria which are critical for nematode establishment and ultimately kill the insect host [56]. 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 [57]. *Steinernema* spp. however do not appear to have the same reduction in putative AMP gene abundance as was observed in *H. bacteriophora*. At present, little is known about the role that nematode-derived AMPs play in the lifecycle of EPNs.

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

Multiple nematode clades possess a limited putative AMP profile in comparison to Clade 9/V and 10/IV in terms of the number of AMP genes and the distribution of AMP groups. These more limited putative AMP profiles do not appear to be a result of poorer genome quality as many of the genomes with less diverse putative AMP profiles display genome BUSCO quality scores >90. Clade 2/I nematodes possess the most limited putative AMP profile of all the nematode clades profiled in this study (3.44 ±0.50 AMP-encoding genes per genome; 0.02 ±0.01% of PCGs encoding AMP genes). All species from this clade possess genes encoding Nemapores and GRSPs with only *Romanomermis culicivorax* possessing CSαβ-encoding genes (see Fig 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) [58–61]. 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 putative 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 BLAST/HMM-based pipeline employed in this study.

With the exception of the Ascaridoidea (31.38 ±4.14 AMP-encoding genes per genome; 0.17 ±0.02% of PCGs encoding AMPs) and Oxyuroidea (20.50± 0.50 AMP-encoding genes per genome; 0.17 ± 0.02% of PCGs encoding AMPs) superfamilies, Clade 8/III nematodes also broadly displayed a reduced putative AMP complement (6.95 ±0.01 AMP-encoding genes per
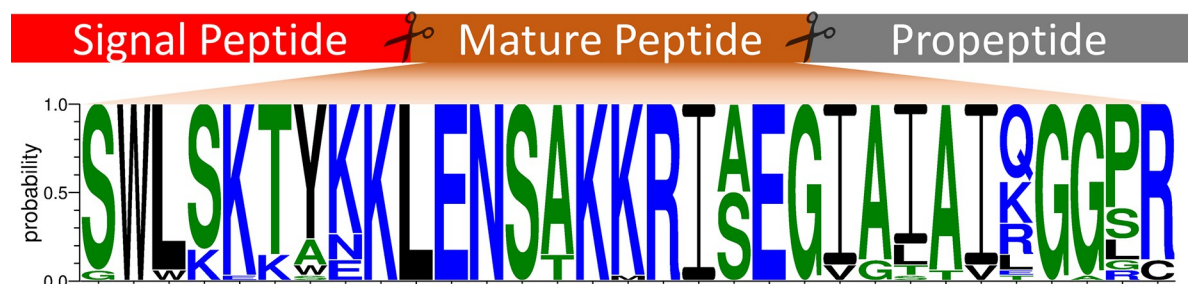
genome; 0.06 ± 0.01% of PCGs encoding AMPs) in comparison to Clade 9/V species. This reduced AMP complement was particularly evident within the Filarioidea species which may suggest that there has been a loss of putative AMP genes from filarial nematodes. In line with the more basal nematodes in Clade 2/I, reduced profiles of metabolic and neuropeptide pathway components are also observed in filarial nematodes [59,61]. 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 [62]. It is possible that *Wolbachia* provides a level of protection for filarial parasites such that the nematodes require fewer endogenous AMPs. Alternatively, the reduced putative 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 putative AMP profile (27.47± 4.54 AMP-encoding genes per genome; 0.09± 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 putative AMP-encoding genes (0.34% relative abundance), exhibiting a profile much more in line with Clade 10/IV species. As noted for Clade 2/I nematodes there is also potential that there are currently undiscovered AMP groups 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 distinct AMP groups.

Cecropin genes are entirely restricted to Ascaridoidea species

The pan-phylum data presented here provides evidence to support the restriction of putative 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 [63]. 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 [13]. Despite this, the effect of these peptides on host microbiota has yet to be characterised.

Putative Cecropins have not undergone the same gene expansion observed in other AMP groups (4.75 ± 0.88 genes per genome). The Ascarid Cecropins appear to be highly conserved suggesting that these peptides have similar antimicrobial activities (Fig 3). Indeed, chemically

Fig 3. Ascaridoidea-restricted Cecropins display significant amino acid conservation. Weblogo generated using 34 putative cecropin peptide sequences (4 peptides excluded due to gene mispredictions; GS_20806, TCNE_0000065801, Tcan_02070, GS_06967) using WebLogo3 [https://weblogo.threeplusone.com/] [65–66], coloured according to amino acid hydrophobicity (hydrophilic = blue, neutral = green, hydrophobic = black). Amino acid letter height reflects the probability of conservation at a specific location. Signal peptide, mature peptide and propeptide indicated.

https://doi.org/10.1371/journal.pntd.0011618.g003

synthesised *A. suum* Cecropin P1-P4 have similar activity profiles against bacteria and fungi [28]. 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 [64]. Recently, Cecropin peptides from *Anisakis* species were shown to have varying antibacterial activities against five bacterial species as well as differences in the transcript expression of Cecropin genes according to the host animal infected [63]. These studies demonstrate the need to characterise AMPs from the same groups across different species.

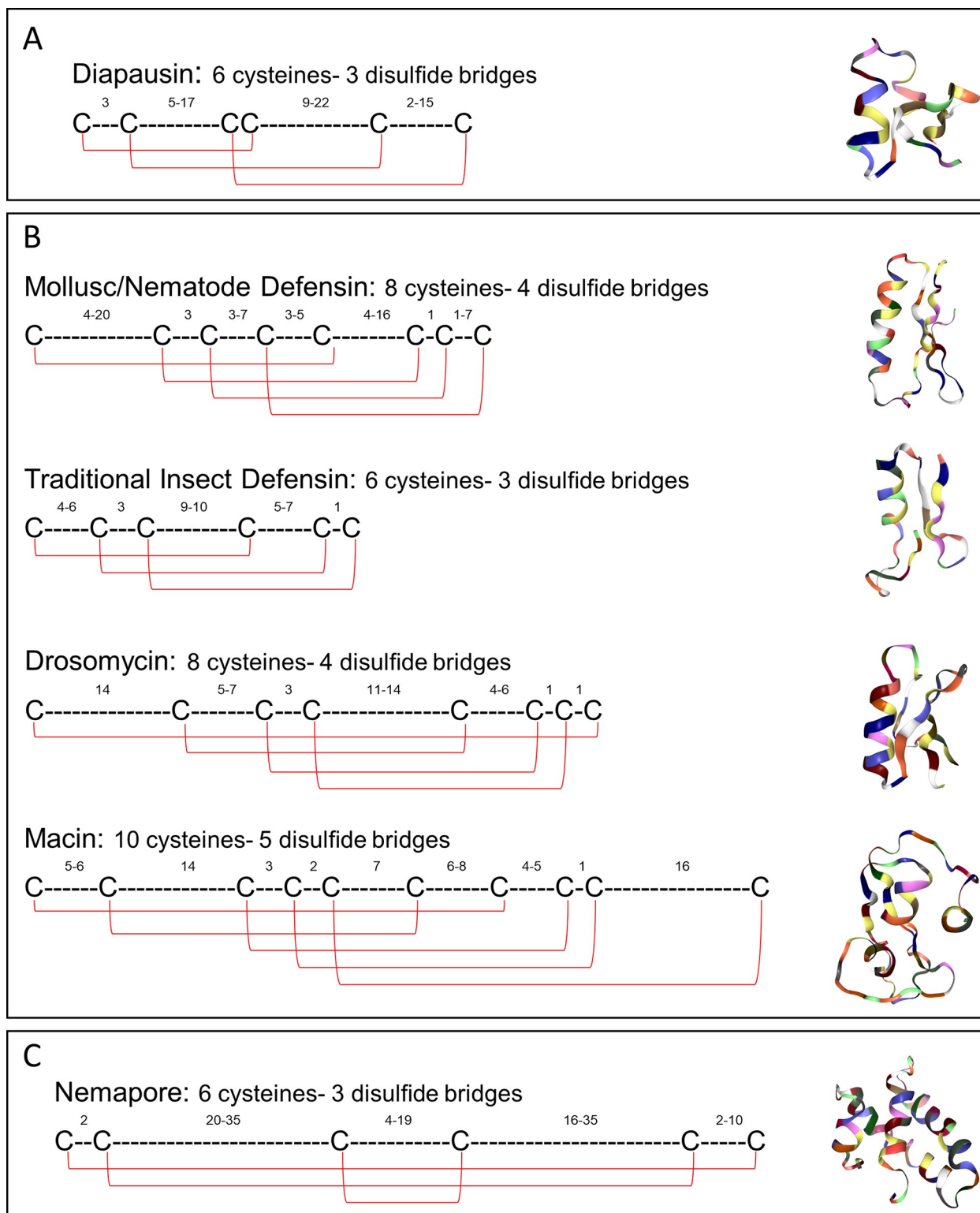## Diapausin genes are restricted to specific free-living soil nematodes

Nematode Diapausin-encoding genes were originally identified from *Pristionchus pacificus* [18]. 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 [67]. 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 [22]. Phylogenetic analyses did not provide any further insight into these relationships or gene evolution primarily due to poor resolution as a consequence of the rapid evolution of AMP-encoding genes. More work is needed to explore the distribution of Diapausin-like genes across diverse animal taxa in order to confirm this hypothesis.

In total 116 putative 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 Fig 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 [22], the role of Diapausins in nematodes has yet to be established.

Classification of nematode CSαβ peptides reveals complexity in subgroup distribution

CSαβ peptides are a complex group of peptides which share a tertiary structure comprising an α-helix and two antiparallel β-sheets stabilised by three disulfide bonds [69]. CSαβ peptides

Fig 4. Cysteine organisation of the nematode cysteine-stabilised AMP groups enables delineation of AMP subgroups. A: The putative Diapausin cysteine organisation. Peptide group 3D structure example: *Gastrophysa atrocyanea* Diapause-specific peptide (RCSB: 2E2F). B: Organisation of cysteine residues in the putative CSαβ members. Peptide group 3D structure examples: Mollusc/Nematode Defensin (*Ascaris suum* ASABF-α; RCSB:2D56), Traditional Insect Defensin (*Sarcophaga peregrina* Sapecin; RCSB: 1L4V), Drosomycin (*Drosophila melanogaster* Drosomycin; RCSB: 1MYN), Macin (*Hirudo medicinalis* Theromacin; RCSB:2LN8). C: The putative Nemapore cysteine organisation. Peptide

group 3D structure example: *Caenorhabditis elegans* Caenopore-5 (RCSB:2JS9). 'C' indicates a Cysteine residue. '-' indicates any other amino acid. Subscript numbers indicate the range in number of amino acids between each cysteine residue. Red lines indicate putative disulphide bonds between cysteine residues. Peptide group 3D structures obtained from Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) [https://www.rcsb.org/] [68].
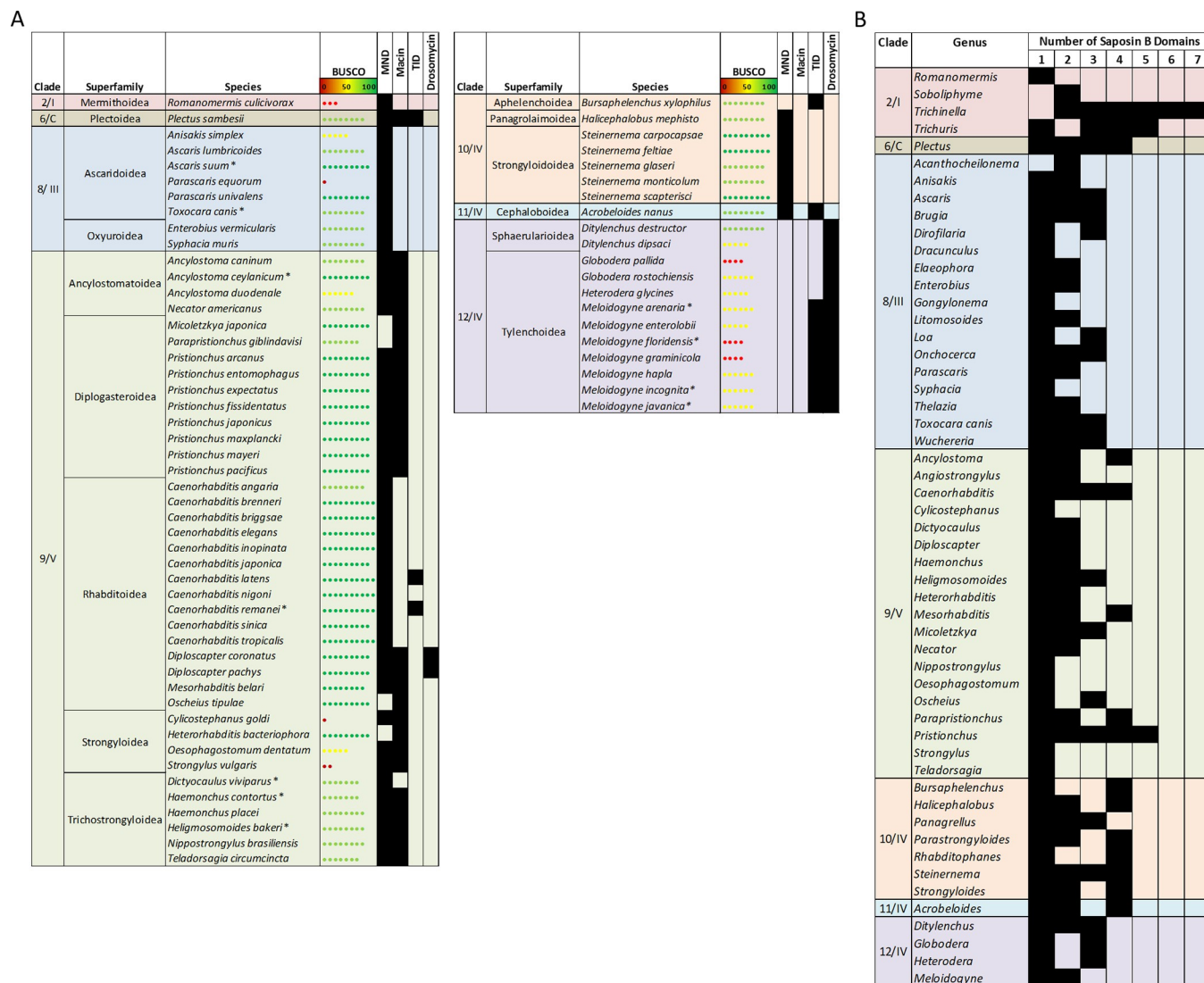
can be classified into multiple subgroups of AMPs displaying differences in cysteine organisation. The putative CSαβ peptides identified in this study have been classified based on the presence of a shared cysteine organisation but this does not indicate that they share the same evolutionary origin. 533 putative CSαβ-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 Fig 2). Putative CSαβ-encoding genes are more restricted in Clades 2/I (only *R. culicivorax*) and 8/III (only Ascaridoidea and Oxyuroidea spp.), which is consistent with a previous study [19]. In contrast, CSαβ-encoding genes are more widespread in Clades 9/V, 10/IV and 12/IV (see Fig 2); note that two CSαβ-encoding genes, previously identified from *P. trichosuri* EST datasets [19], were not identified here.

CSαβ-encoding genes identified in this study were classified into four subgroups based on the cysteine reference array reported previously [24]: Mollusc/Nematode Defensins (MND), Traditional Insect Defensins (TIDs), Macins and Drosomycins. The distribution of these subgroups was variable across phylum Nematoda (see Fig 5A). The MND subgroup, previously known as Antibacterial Factors [ABFs; 23], was the most widely distributed, identified in almost all clades examined (see Fig 5A). MNDs are characterised by a cysteine motif consisting of eight cysteine residues which form four putative disulfide bonds (see Fig 4B). In line with the Diapausins, high amino acid variability exists between the key MND cysteine residues, which may contribute to diversity in antimicrobial activity. MND 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 [14,70]. Putative MNDs 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 MND-encoding genes (Clade 2/I *Trichinella* and *Trichuris* spp., Clade 8/III filarial species, and all Clade 12/IV species). The reduced distribution of the MND subgroup is consistent with the broadly reduced AMP complement in Clade 2/I and 8/III suggesting a negative selective pressure for AMP-encoding genes in these clades, whereas Clade 12/IV displays other CSαβ subgroups such that the lack of MNDs could be due to functional redundancy.

A MND member with only six cysteine residues and a rearranged disulfide bonding pattern, known as ASABF-6Cys-α, was previously identified in *A. suum* [71]. Previous analysis of EST and draft genomic datasets identified homologs of ASABF-6Cys-α beyond Ascaridida in *Ancylostoma ceylanicum*, *Caenorhabditis briggsae*, *Caenorhabditis remanei* and *Meloidogyne hapla* [19]. However, this study has revealed that the *A. ceylanicum*, *C. briggsae* and *C. remanei* sequences are, in fact, not homologs of ASABF-6Cys-α but are traditional MND peptides which were previously incomplete. The *M. hapla* sequences were not included in this analysis as they lack key MND motifs. In this study, a single putative ortholog of *A. suum* ASABF-6Cys-α was identified in *Parascaris univalens* supporting the original hypothesis that the 6-cysteine variation in the MND sequence is only present in nematodes in the Ascaridida [71].

Nematode Macin-encoding genes have been previously reported [72] however this study provides the most comprehensive profile of these genes across phylum Nematoda. Putative Macin-encoding genes were identified in Clades 6/C and 9/V. Within Clade 9/V, Macin-encoding genes are found in all species except all *Caenorhabditis* spp., *Dictyocaulus viviparus* and both *Angiostrongylus* spp. Notably, putative Macin-encoding genes are found in parasites

Fig 5. The distribution of putative CSαβ subgroups and putative multi-domain Nemapores reveal additional AMP group diversity across nematode clades. A: The distribution of four putative CSαβ subgroups [Mollusc/Nematode Defensin (MND), Macin, Traditional Insect Defensin (TID), Drosomycin] in CSαβ-containing nematode species. B: The distribution of putative multi-domain Nemapore-encoding genes across nematode genera. Nematode species/genera are organised by clade [47–48] and superfamily designation (NCBI taxonomy browser; https://www.ncbi.nlm.nih.gov/taxonomy). Black shading indicates the presence of at least one putative AMP gene identified via HMM, BLASTp and tBLASTn analyses. BUSCO complete scores were obtained from WormBase ParaSite V14 [https://parasite.wormbase.org/index.html] [34]. Note that species which do not possess CSαβ encoding genes are not included in A.

https://doi.org/10.1371/journal.pntd.0011618.g005

of health and economic importance including *H. contortus*, *N. americanus* and *T. circumcincta*. The absence of putative Macin-encoding genes from the *Caenorhabditis* genera could suggest that they have been lost in these species while they have been retained in other free-living Clade 9/V nematodes. The reason for this loss is unclear but reemphasises the need to conduct AMP research beyond *C. elegans*.

Macins from other invertebrates possess antibacterial and antifungal activities and can promote regeneration of neuronal tissue in annelids [73], however nematode Macins have not been functionally characterised. In other invertebrate phyla, Macin peptides possess eight to twelve cysteine residues [72] whereas nematodes appear to encode Macins with ten cysteine residues (see Fig 4B). Theromacin from the annelid *Hirudo medicinalis*, which exhibits activity

against gram positive bacteria and promotes neuron regeneration, also possesses ten cysteine residues suggesting that nematode Macins may possess similar functional profiles.

Drosomycin peptides possess eight cysteine residues which form four putative disulfide bonds to produce a tertiary structure that is distinct from the MNDs (see Fig 4B). Putative Drosomycin-encoding genes were previously identified in free-living *Panagrolaimus superbus* (Clade 10/IV) EST datasets [19]. In this study, there were no *Panagrolaimus* genomes included in the nematode database however, the addition of four *Panagrolaimus* genomes in a subsequent WormBase ParaSite update (Version 15, October 2020) allowed the confirmation of Drosomycin genes in the *Panagrolaimus* genera.

Putative Drosomycin-encoding genes were also identified in Clade 12/IV species and two Clade 9/V species (*Diploscapter coronatus* and *Diploscapter pachys*). It is unclear whether other Clade 10/IV species have lost these genes or whether the *Panagrolaimus* species acquired them independently. Nematode Drosomycin peptides have yet to be functionally characterised however insect Drosomycin peptides display potent antifungal activity against filamentous fungi [74]. Note that cysteine-rich nematode peptides have previously been designated as Drosomycin peptides but have since been reclassified as TID subgroup members based on their cysteine motif positioning [24].

Traditional Insect Defensin peptides are characterised by the typical insect Defensin cysteine array where six cysteines form three disulfide bonds (see Fig 4B). Putative TIDs display a more limited distribution across phylum Nematoda (28 TID-encoding genes in 12 nematode species). Intriguingly, the 12 nematode species that possess putative TID-encoding genes are phylogenetically distinct FLN and PPN species that span five nematode clades; this raises questions about the evolutionary history of these genes. The TID subgroup members appear to have functionally diversified such that each peptide has a specific activity range. For example, *Caenorhabditis remanei* Cremycin-5 possesses antifungal activity whereas Cremycin-15 lacks antifungal activity but is bioactive against gram positive bacteria [75].

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

Nemapores are cysteine-stabilised AMPs found in a wide range of nematode species. They are members of the SAPosin-LIke Protein (SAPLIP) superfamily which includes other membrane-interacting proteins such as metallophosphoesterases and cell proliferator regulators [36]. Nemapores are characterised by a signal peptide sequence followed by at least one Saposin B domain. Saposin B domains comprise of six cysteine residues in a specific arrangement (see Fig 4C). This study updates previous analyses which indicate that Nemapores are the most widely distributed AMP group in phylum Nematoda [see Fig 2; 19]. Indeed, at least one putative Nemapore-encoding gene (~15.34 ±1.38 genes per genome) was identified from every species analysed, suggesting that Nemapores may be a key component of nematode innate immunity. Putative Nemapore-encoding genes, which were previously reported as absent in *N. americanus*, *Strongyloides stercoralis* and *Meloidogyne incognita* [19], were identified in this study reemphasising the value in utilising updated genomic datasets.

Despite broad representation across the phylum, the numbers of putative 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 putative Nemapore-encoding genes. The elevated number of putative Nemapore-encoding genes in Clade 10/IV, highlights that Nemapores may play a key role in Clade 10/IV nematodes. 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 [76].
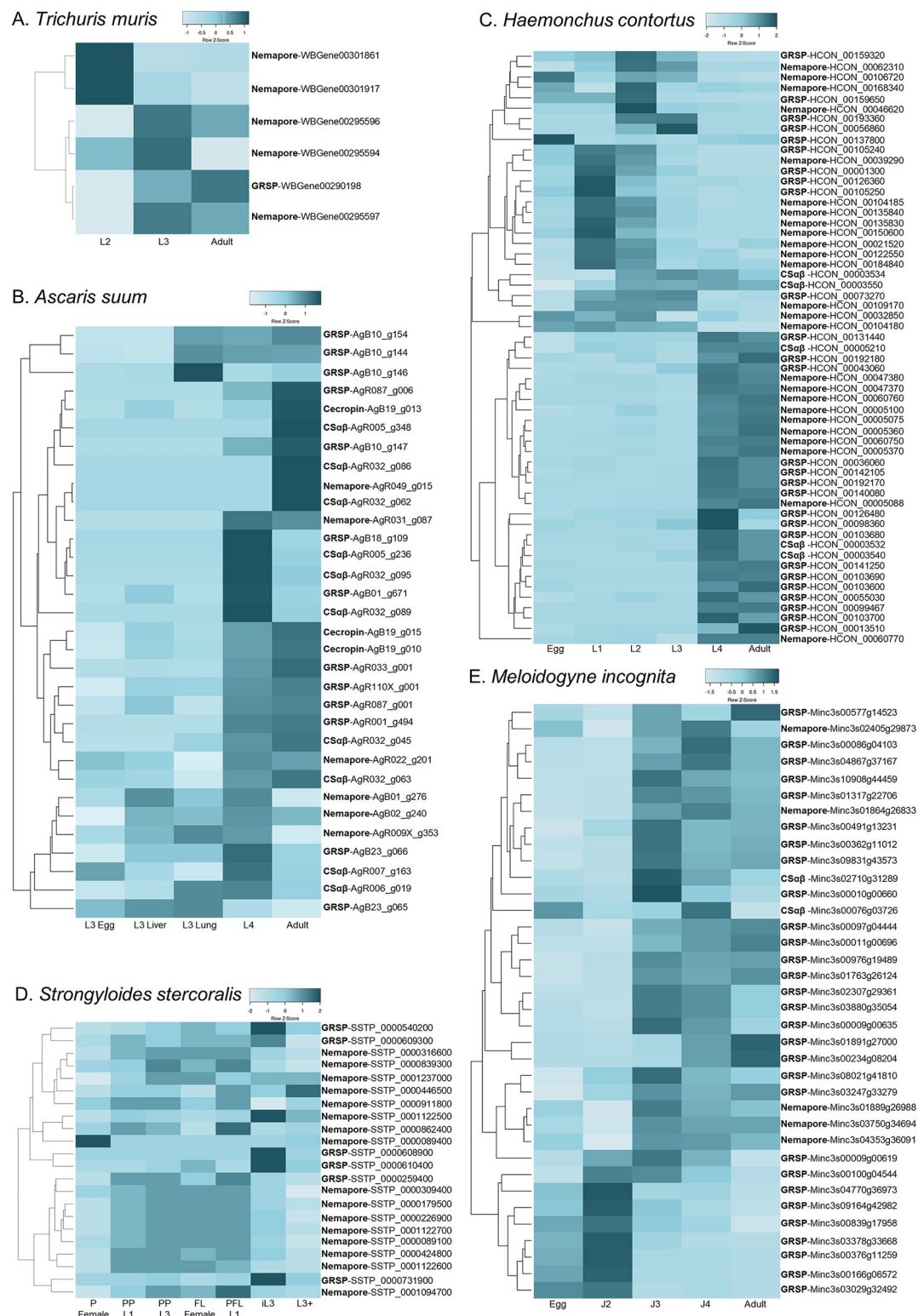
There is also variation in the number of Saposin B domains encoded by Nemapore genes across nematode genera (see Fig 5B). Previous analyses indicate that Nemapores containing three and four Saposin domains were restricted to species from Clades 2/I and 9/V [19]. Data generated in this study, however, show that putative multi-domain Nemapores are more widespread where Nemapores with three or more Saposin B domains are represented in all clades. Indeed, *Trichinella* spp. appear to be enriched for multi-domain Nemapores with some genes containing seven Saposin B domains; multi-domain Nemapore genes have yet to be functionally characterised. In the amoeba *Naegleria fowleri* multiple bioactive Saposin peptides are processed from larger multipeptide precursor sequences [77]. It is also possible that nematode multi-domain Nemapores are post translationally cleaved to produce multiple bioactive single domain Nemapores however previous attempts to unravel the processing of a four Saposin domain protein from *Trichinella spiralis* were unable to resolve protein cleavage [78].

Enriched numbers of putative GRSPs are not restricted to the *Caenorhabditis* genera

This study provides the first pan-phylum profile of putative GRSP-encoding genes across phylum Nematoda; previously GRSPs had only been profiled in *C. elegans* and *C. briggsae* where they were found to be enriched relative to several other model organisms [20]. The data presented here demonstrates that GRSPs appear to be the most abundant AMP group in phylum Nematoda (3145 genes; 23.47 ±2.26 genes per genome) with Clades 9/V and 10/IV exhibiting enriched profiles. The enriched number of putative GRSPs extends beyond *Caenorhabditis* genera (see Fig 2), where GRSP-encoding genes are absent from only two nematode species [*Dracunculus medinensis* and *Litomosoides sigmodontis* (Clade 8/III)] which suggests importance to nematode biology. Despite sharing sequence similarity, it is not clear if the putative GRSPs identified here share an evolutionary origin. It is important to note that the highly repetitive glycine rich nature of the GRSP group makes establishing evolutionary relationships challenging and further research is required to understand the evolution of GRSPs across phylum Nematoda. Functional characterisation of GRSPs has been restricted to *C. elegans* where their expression has been profiled post microbial infection [20], and antibacterial and antifungal assays reveal antimicrobial properties that do not involve membrane permeabilisation [17,79]. Further work is crucial to confirm that the genes identified here do encode bioactive AMPs. In addition, it is interesting to note that a number of Neuropeptide-Like Proteins (NLPs) have been classified as GRSPs (NLP-10, NLP-12, NLP-21 and NLP-24-34), highlighting a need to unravel the role of these peptides.

Publicly available transcriptome and proteomic datasets support the prioritisation of nematode putative AMPs

To assess the biological relevance of putative 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). Of the putative AMP genes identified >85% were expressed (TPM >2) in at least one life stage (see Fig 1E). Notably, despite significant variation in the total number of putative 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 putative AMP genes from the same AMP group such that no AMP group is consistently expressed in a specific life stage in the five parasitic species examined in more detail (see Fig 6). This suggests that closely related putative AMPs may have functionally diversified and

A. *Trichuris muris*

B. *Ascaris suum*

C. *Haemonchus contortus*

D. *Strongyloides stercoralis*

E. *Meloidogyne incognita*

Fig 6. Summary of transcriptomic data. Life stage specific heatmaps of AMP expression for AMP genes from A: *Trichuris muris* (6 genes), B: *Ascaris suum* (32 genes), C: *Haemonchus contortus* (58 genes), D: *Strongyloides stercoralis* (22 genes), E: *Meloidogyne incognita* (36 genes). Heatmaps generated using Heatmapper [http://www.heatmapper.ca/] [44]. Expression is shown in Z scores and scaled by row. Genes are clustered using Average Linkage and Euclidean distance method. Darker shades represent higher expression and lighter shades represent lower expression.

https://doi.org/10.1371/journal.pntd.0011618.g006

reinforces the need to further characterise these putative AMPs in other species. Notably, in *A. suum* and *H. contortus* there is clear elevation in the expression of putative AMP genes in life stages which are exposed to the host gastrointestinal system further supporting that these genes may play a role in nematode innate immunity.

To validate the prediction of the putative AMP-encoding genes identified through *in-silico* analyses in this study, previously published proteomic/peptidomic datasets (see Table 2) were analysed for the presence of AMP candidates in the excretory/secretory products (ESP) and biofluids of *A. suum* (ESP, uterine fluid and pseudocoelomic cavity fluid [80,81]), *H. contortus* (L3, L4 and adult ESP [82]), *Steinernema feltiae* (infective juvenile ESP [83]) and *Trichuris suis* (adult ESP [84]). Twenty-six AMP candidates were detected (see Table 2) supporting their potential biological relevance in nematodes. Interestingly, no putative AMPs were detected in the pseudocoelomic cavity fluid using proteomics approaches [80] in contrast to five putative AMPs that were detected using peptidomics approaches [81]. This demonstrates that a peptidomic workflow may be more appropriate for AMP discovery. In addition, the peptides identified here (see Table 2), through analysis of existing proteomics datasets, were not originally identified as putative AMPs as the peptides were uncharacterised. The comprehensive BLAST/HMM approach employed here enabled the retrospective interrogation of all putative AMP peptides in these datasets underscoring the value of this study.

Table 2. Overview of AMP candidates detected in proteomic [80,82–83,84] or peptidomic [81] studies.

Species	Source	Gene ID	AMP Group
Ascaris suum	ESP [80]	GS_10369 (ASABF-β)	CSαβ
		GS_05608 (ASABF-α)	CSαβ
		GS_08337 (ASABF-γ)	CSαβ
	UF [80]	GS_19916	CSαβ
	PCF [81]	AgB19_g015 (Cecropin P1)	Cecropin
		GS_14502 (Cecropin P2)	Cecropin
		AgB19_g010 (Cecropin P3)	Cecropin
		AgR032_g063 (ASABF-α)	CSαβ
		AgR032_g045 (ASABF-6Cys-α)	CSαβ
Haemonchus contortus	L3 ESP [82]	HCON_00062310	Nemapore
	L3, L4 & Adult ESP [82]	HCON_00047370	Nemapore
	L4 & Adult ESP [82]	HCON_00003550	CSαβ
		HCON_00104180	Nemapore
		HCON_00060770	Nemapore
		HCON_00060760	Nemapore
	Adult ESP [82]	HCON_00005370	Nemapore
		HCON_00005360	Nemapore
		HCON_00005100	Nemapore
Steinernema feltiae	Activated infective juvenile ESP [83]	L889_g29545	Nemapore
		L889_g13905	Nemapore
		L889_g31036	Nemapore
		L889_g31037	Nemapore
		L889_g17551	Nemapore
		L889_g20297	Nemapore
		L889_g6015	Nemapore
Trichuris suis	Adult ESP [84]	D918_09288	Nemapore

ESP, excretory/secretory products; PCF, pseudocoelomic fluid; UF, uterine fluid.

https://doi.org/10.1371/journal.pntd.0011618.t002

Limitations of BLAST/HMM-driven AMP discovery in nematodes

Antimicrobial peptide identification through *in silico* studies is useful for providing peptide libraries which can enable prioritisation of putative AMP targets for functional analyses. However, there are limitations associated with *in-silico* AMP discovery. The putative AMPs identified in this study were discovered through a sequence similarity pipeline and require functional characterisation in order to confirm their putative antimicrobial activities and determine their putative role(s) in nematode biology.

Genome and transcriptome studies also 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 BUSCO complete scores for each genome [see Fig 2 and S2 File] [85]. It is important to note that BUSCO datasets are primarily derived from free-living organisms and as a result, gene losses as a result of parasitism may incorrectly give the impression of poorer genome quality based on BUSCO score. Despite this, there appears to be an association 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). For many lower quality genomes, there were fewer than expected AMPs identified and many of those identified appeared incomplete, either predicted as part of longer proteins or missing key sequence motifs. As nematode genome assemblies and annotations are improved, more AMP genes will be uncovered to reveal a more comprehensive nematode antimicrobial peptidome.

The BLAST/HMM-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 [81] will seed follow-on functional biology.

Conclusions

This study provides an *in silico* comprehensive genome-level pan-phylum profile of five groups of putative nematode AMP-encoding genes which will inform future analyses into the function of these putative AMPs. These data demonstrate that nematodes appear to have a rich AMP complement that includes 5887 putative AMP-encoding genes in 109 species, ~90% of which are novel. The pan-phylum approach has revealed that APNs and PPNs possess fewer putative 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 putative AMP groups were evident across the phylum where Clade 2/I, 8/III and 12/IV species possess fewer putative AMP-encoding genes, while Clade 9/V and 10/IV species have expanded AMP repertoires. Two AMP groups, Cecropins and Diapausins, appear to be restricted to Ascarids and FLNs respectively whereas CSαβ, Nemapores and GRSPs were more widely distributed across the phylum. This study provides a database of putative AMPs in a range of economically relevant nematode species that should be functionally characterised to unravel the importance of AMPs to parasite biology.

Supporting information

S1 Fig. Coloured CLANS sequence similarity network for nematode Cecropins, Diapausins, CSαβs and Nemapores. Network was generated after 10,000 rounds of clustering from all-against all protein BLAST searches between all complete sequences (E-value limit = 1e-4). Sequences are coloured according to group or subgroup designation.
(TIF)

S1 File. AMP query sequences used for construction of AMP group HMMs.
(XLSX)

S2 File. Genomic and transcriptomic datasets used in this study.
(XLSX)

S3 File. AMP gene IDs identified from 134 nematode predicted protein datasets.
(XLSX)

S4 File. Number and relative abundance of putative AMP genes identified in this study.
(XLSX)

Author Contributions

Conceptualization: Louise E. Atkinson, Angela Mousley.

Formal analysis: Allister Irvine.

Funding acquisition: Louise E. Atkinson, Angela Mousley.

Investigation: Allister Irvine.

Methodology: Allister Irvine.

Project administration: Louise E. Atkinson, Angela Mousley.

Resources: Louise E. Atkinson, Angela Mousley.

Supervision: Sharon A. Huws, Louise E. Atkinson, Angela Mousley.

Writing – original draft: Allister Irvine, Louise E. Atkinson, Angela Mousley.

Writing – review & editing: Allister Irvine, Sharon A. Huws, Louise E. Atkinson, Angela Mousley.

References

1. Wang G. Discovery, classification and functional diversity of antimicrobial peptides. In: Wang G, editor. Antimicrobial peptides: Discovery, Design and Novel Therapeutic Strategies. CABI Publishing; 2017. pp 1–19.

2. Bosch TCG, Zasloff M. Antimicrobial peptides–or how our ancestors learned to control the microbiome. mBio. 2021; 12(5): e0184721. https://doi.org/10.1128/mBio.01847-21 PMID: 34579574

3. Matsuzaki K. Membrane permeabilization mechanisms. In: Matsuzaki K, editor. Antimicrobial Peptides: Basics for Clinical Application. Springer Singapore; 2019. pp 9–16.

4. Zasloff M. Antimicrobial peptides of multicellular organisms: my perspective. In: Matsuzaki K, editor. Antimicrobial Peptides: Basics for Clinical Application. Springer Singapore; 2019 pp 3–6.

5. van der Does AM, Hiemstra PS, Mookherjee N. Antimicrobial host defence peptides: immunomodulatory functions and translational prospects. In: Matsuzaki K, editor. Antimicrobial Peptides: Basics for Clinical Application. Springer Singapore; 2019 pp. 149–171.

6. Mwangi J, Hao X, Lai R, Zhang ZY. Antimicrobial peptides: new hope in the war against multidrug resistance. Zool. Res. 2019; 40(6): 488–505. https://doi.org/10.24272/j.issn.2095-8137.2019.062 PMID: 31592585

7. Giribet G, Edgecombe GD.Current understanding of Ecdysozoa and its internal phylogenetic relationships. Integr. Comp. Biol. 2017; 57(3): 455–466. https://doi.org/10.1093/icb/icx072 PMID: 28957525

8. Gomez EA, Giraldo P, Orduz S. InverPep: A database of invertebrate antimicrobial peptides. J. Glob. Antimicrob. Resist. 2017; 8: 13–17. https://doi.org/10.1016/j.jgar.2016.10.003 PMID: 27888793

9. Du ZQ, Li B, Shen XL, Wang K, Du J, Yu XD, et al. A new antimicrobial peptide isoform, Pc-crustin 4 involved in antibacterial innate immune response in fresh water crayfish, *Procambarus clarkii*. Fish Shellfish Immunol. 2019; 94: 861–870.

10. Zhang S, Hou C, Xiao B, Yao Y, Xiao W, Li C, et al. Identification and function of an Arasin-like peptide from *Litopenaeus vannamei*. Dev. Comp. Immunol. 2021; 125: 104174.

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

12. Mladineo I, Rončević T, Gerdol M, Tossi A. Helminthic host defense peptides: using the parasite to defend the host. Trends Parasitol. 2023, 39(5): 345–357. https://doi.org/10.1016/j.pt.2023.02.004 PMID: 36890022

13. Lee JY, Boman A, Sun CX, Andersson M, Jörnvall H, Mutt V, et al. Antibacterial peptides from pig intestine: isolation of a mammalian cecropin. Proc. Natl. Acad. Sci. U. S. A. 1989; 86(23): 9159–9162. https://doi.org/10.1073/pnas.86.23.9159 PMID: 2512577

14. Kato Y, Komatsu S. ASABF, a novel cysteine-rich antibacterial peptide isolated from the nematode *Ascaris suum*—purification, primary structure, and molecular cloning of cDNA. J. Biol. Chem. 1996; 271(48): 30493–30498.

15. Bányai L, Patthy L. Amoebapore homologs of *Caenorhabditis elegans*. Biochim. Biophys. Acta. 1998; 1429(1): 259–264.

16. Andersson M, Boman A, Boman HG. *Ascaris* nematodes from pig and human make three antibacterial peptides: isolation of cecropin P1 and two ASABF peptides. Cell. Mol. Life. Sci. 2003; 60(3): 599–606.

17. Couillault C, Pujol N, Reboul J, Sabatier L, Guichou JF, Kohara Y, et al. TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. Nat. Immunol. 2004; 5(5): 488–494.

18. Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, et al. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. Nat. Genet. 2008; 40(10): 1193–1198.

19. Tarr DEK. Distribution and characteristics of ABFs, cecropins, nemapores, and lysozymes in nematodes. Dev. Comp. Immunol. 2012; 36(3): 502–520. https://doi.org/10.1016/j.dci.2011.09.007 PMID: 21978453

20. Ying MY, Qiao YY, Yu LH. Evolutionary expansion of nematode-specific glycine-rich secreted peptides. Gene. 2016; 587(1): 76–82. https://doi.org/10.1016/j.gene.2016.04.049 PMID: 27129940

21. Gu J, Gao B, Zhu S. Characterization of bi-domain drosomycin-type antifungal peptides in nematodes: An example of convergent evolution. Dev. Comp. Immunol. 2018; 87, 90–97. https://doi.org/10.1016/j.dci.2018.06.005 PMID: 29894713

22. Tanaka H, Sato K, Saito Y, Yamashita T, Agoh M, Okunishi J, et al. Insect diapause-specific peptide from the leaf beetle has consensus with a putative iridovirus peptide. Peptides. 2003; 24(9): 1327–1333. https://doi.org/10.1016/j.peptides.2003.07.021 PMID: 14706547

23. Tarr DEK. Nematode antimicrobial peptides. Invert. Surviv. J. 2012; 9: 122–133.

24. Tarr DEK. Establishing a reference array for the CS-αβ superfamily of defensive peptides. BMC Res. Notes. 2016; 9(1): 490.

25. Doyle SR. Improving helminth genome resources in the post-genomic era. Trends Parasitol. 2022; 38: 831–840. https://doi.org/10.1016/j.pt.2022.06.002 PMID: 35810065

26. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2019; 47(W1): W636–641.

27. Sipos D, Andersson M, Ehrenberg A. The structure of the mammalian antibacterial peptide cecropin P1 in solution, determined by proton-NMR. Eur. J. Biochem. 1992; 209(1): 163–169. https://doi.org/10.1111/j.1432-1033.1992.tb17273.x PMID: 1396696

28. Pillai A, Ueno S, Zhang H, Lee JM, Kato Y. Cecropin P1 and novel nematode cecropins: a bacteria-inducible antimicrobial peptide family in the nematode *Ascaris suum*. Biochem. J. 2005; 390(Pt 1): 207–214.

29. Kouno T, Mizuguchi M, Tanaka H, Yang P, Mori Y, Shinoda H, et al. The structure of a novel insect peptide explains its Ca^{2+} channel blocking and antifungal activities. Biochemistry. 2007; 46(48): 13733–13741.

30. Al Souhail Q, Hiromasa Y. Rahnamaeian M, Giraldo MC, Takahashi D, Valent B, et al Characterization and regulation of expression of an antifungal peptide from hemolymph of an insect, *Manduca sexta*. Dev. Comp. Immunol. 2016; 61: 258–268.

31. Pillai A, Ueno S, Zhang H, Kato Y. Induction of ASABF (*Ascaris suum* antibacterial factor)-type antimicrobial peptides by bacterial injection: novel members of ASABF in the nematode *Ascaris suum*. Biochem. J. 2003; 371(Pt 3): 663–668.

32. Mysliwy J, Dingley AJ, Stanisak M, Jung S, Lorenzen I, Roeder T, et al. Caenopore-5: the three-dimensional structure of an antimicrobial protein from *Caenorhabditis elegans*. Dev. Comp. Immunol. 2010; 34(3): 323–330.

33. Hoeckendorf A, Leippe M. SPP-3, a saposin-like protein of *Caenorhabditis elegans*, displays antimicrobial and pore-forming activity and is located in the intestine and in one head neuron. Dev. Comp. Immunol. 2012; 38(1), 181–186.

34. Bolt BJ, Rodgers FH, Shafie M, Kersey PJ, Berriman M. Howe KL. Using WormBase ParaSite: An integrated platform for exploring helminth genomic data. In: Kollmar M, editor. Eukaryotic Genomic Databases. Springer; 2018. pp. 471–491.

35. Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, et al. WormBase: a modern Model Organism Information Resource. Nucleic Acids. Res. 2020; 48(D1): D762–D767. https://doi.org/10.1093/nar/gkz920 PMID: 31642470

36. Bruhn H. A short guided tour through functional and structural features of saposin-like proteins. Biochem. J. 2005; 389(Pt 2): 249–257. https://doi.org/10.1042/BJ20050051 PMID: 15992358

37. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 2019; 37(4): 420–423. https://doi.org/10.1038/s41587-019-0036-z PMID: 30778233

38. Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. Protein Eng. Des. Sel. 2004; 17(1): 107–112. https://doi.org/10.1093/protein/gzh013 PMID: 14985543

39. Veltri D, Kamath U, Shehu A. Deep learning improves antimicrobial peptide recognition. Bioinformatics. 2018; 34(16): 2740–2747. https://doi.org/10.1093/bioinformatics/bty179 PMID: 29590297

40. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics. 2004; 20(18): 3702–3704. https://doi.org/10.1093/bioinformatics/bth444 PMID: 15284097

41. Zimmerman L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. J. Mol. Bio. 2018; 430(15): 2237–2243.

42. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, et al. Protein sequence analysis using the MPI Bioinformatics Toolkit. Curr. Protoc. Bioinformatics. 2020; 72(1): e108. https://doi.org/10.1002/cpbi.108 PMID: 33315308

43. Wagner GP, Kin K, Lynch VJ. A model based criterion for gene expression calls using RNA-seq data. Theory Biosci. 2013; 132(3): 159–164. https://doi.org/10.1007/s12064-013-0178-3 PMID: 23615947

44. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, et al. Heatmapper: web-enabled heat mapping for all. Nucleic Acids. Res. 2016; 44(W1): W147–W153. https://doi.org/10.1093/nar/gkw419 PMID: 27190236

45. Massey FJ Jr. The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assoc. 1951; 46(253): 68–78.

46. Dunn OJ. Multiple comparisons among means. J. Am. Stat. Assoc. 1961; 56(293), 52–64.

47. Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, et al. A molecular evolutionary framework for the phylum Nematoda. Nature. 1998; 392(6671): 71–75. https://doi.org/10.1038/32160 PMID: 9510248

48. van Megen H, van den Elsen S, Holterman M, Karssen G, Mooyman P, Bongers T, et al. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology. 2009; 11(6): 927–950.

49. Parihar RD, Dhiman U, Bhushan A, Gupta PK, Gupta P. *Heterorhabditis* and *Photorhabdus* symbiosis: a natural mine of bioactive compounds. Front. Microbiol. 2022; 13: 790339.

50. Clarke DJ. *Photorhabdus*: a tale of contrasting interactions. Microbiology (Reading). 2020: 166(4): 335–348.

51. Imai Y, Meyer KJ, Iinishi A, Favre-Godal Q, Green R, Manuse S, et al. A new antibiotic selectively kills Gram-negative pathogens. Nature. 2019; 576(7787): 459–464. https://doi.org/10.1038/s41586-019-1791-1 PMID: 31747680

52. Félix M-A, Ailion M, Hsu J-C, Richaud A, Wang J. *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. 2018; 13(8): e0200851.

53. El-Ashram S, Al Nasr I, Abouhajer F, El-Kemary M, Huang G, Dinçel G, et al. Microbial community and ovine host response varies with early and late stages of *Haemonchus contortus* infection. Vet. Res. Commun. 2017; 41(4): 263–277.

54. Cortés A, Wills J, Su X, Hewitt RE, Robertson J, Scotti R, et al. Infection with the sheep gastrointestinal nematode *Teladorsagia circumcincta* increases luminal pathobionts. Microbiome. 2020; 8(1): 60.

55. Jenkins TP, Pritchard DI, Tanasescu R, Telford G, Papaiakovou M, Scotti R, et al. Experimental infection with the hookworm, *Necator americanus*, is associated with stable gut microbial diversity in human volunteers with relapsing multiple sclerosis. BMC Biol. 2021; 19(1): 74.

56. Goodrich-Blair H. They've got a ticket to ride: *Xenorhabdus nematophila-Steinernema carpocapsae* symbiosis. Curr. Opin. Microbiol. 2007; 10(3): 225–230.

57. Dreyer J, Rautenbach M, Booysen E, van Staden AD, Deane SM, Dicks LMT. *Xenorhabdus khoisanae* SB10 produces Lys-rich PAX lipopeptides and a Xenocoumacin in its antimicrobial complex. BMC Microbiol. 2019; 19(1): 132.

58. McCoy CJ, Atkinson LE, Zamanian M, McVeigh P, Day TA, Kimber MJ, et al. New insights into the FLPergic complements of parasitic nematodes: Informing deorphanisation approaches. EuPA Open. Proteom. 2014; 3: 262–272. https://doi.org/10.1016/j.euprot.2014.04.002 PMID: 26366373

59. International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms. Nat. Genet. 2019; 51(1): 163–174. https://doi.org/10.1038/s41588-018-0262-1 PMID: 30397333

60. Xu L, Yang J, Xu M, Shan D, Wu Z, Yuan D. Speciation and adaptive evolution reshape antioxidant enzymatic system diversity across the phylum Nematoda. BMC Biol. 2020; 18(1): 181. https://doi.org/10.1186/s12915-020-00896-z PMID: 33243226

61. McKay FM, McCoy CJ, Crooks B, Marks NJ, Maule AG, Atkinson LE, et al. In silico analyses of neuropeptide-like protein (NLP) profiles in parasitic nematodes. Int. J. Parasitol. 2021; 52(1): 77–85. https://doi.org/10.1016/j.ijpara.2021.07.002 PMID: 34450132

62. Taylor MJ, Bandi C, Hoerauf A. *Wolbachia*. Bacterial endosymbionts of filarial nematodes. Adv. Parasitol. 2005; 60: 245–284.

63. Rončević T, Gerdol M, Mardirossian M, Maleš M, Cvjetan S, Benincasa M, et al. Anisaxins, helical antimicrobial peptides from marine parasites, kill resistant bacteria by lipid extraction and membrane disruption. Acta. Biomater. 2022; 146: 131–144. https://doi.org/10.1016/j.actbio.2022.04.025 PMID: 35470073

64. Gazit E, Miller IR, Biggin PC, Sansom MSP, Shai Y. Structure and orientation of the mammalian antibacterial peptide Cecropin P1 within phospholipid membranes. J. Mol. Biol. 1996; 258(5): 860–870. https://doi.org/10.1006/jmbi.1996.0293 PMID: 8637016

65. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990; 18(20): 6097–6100. https://doi.org/10.1093/nar/18.20.6097 PMID: 2172928

66. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14(6): 1188–1190. https://doi.org/10.1101/gr.849004 PMID: 15173120

67. Rödelsperger C, Sommer RJ. Computational archaeology of the *Pristionchus pacificus* genome reveals evidence of horizontal gene transfers from insects. BMC Evol. Biol. 2011; 11: 239.

68. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1): 235–242. https://doi.org/10.1093/nar/28.1.235 PMID: 10592235

69. Zhu S, Gao B, Tytgat J. Phylogenetic distribution, functional epitopes and evolution of the CSαβ superfamily. Cell. Mol. Life. Sci. 2005; 62(19–20): 2257–2269.

70. Kato Y, Aizawa T, Hoshino H, Kawano K, Nitta K, Zhang H. abf-1 and abf-2, ASABF-type antimicrobial peptide genes in *Caenorhabditis elegans*. Biochem. J. 2002; 361(Pt 2): 221–230.

71. Minaba M, Ueno S, Pillai A, Kato Y. Evolution of ASABF (*Ascaris suum* antibacterial factor)-type antimicrobial peptides in nematodes: putative rearrangement of disulfide bonding patterns. Dev. Comp. Immunol. 2009; 33(11): 1147–1150.

72. Gerdol M, De Moro G, Manfrin C, Venier P, Pallavicini A. Big defensins and mytimacins, new AMP families of the Mediterranean mussel *Mytilus galloprovincialis*. Dev. Comp. Immunol. 2012; 36(2): 390–399.

73. Jung S, Sonnichsen FD, Hung CW, Tholey A, Boidin-Wichlacz C, Haeusgen W, et al. Macin family of antimicrobial proteins combines antimicrobial and nerve repair activities. J. Biol. Chem. 2012; 287(17): 14246–14258. https://doi.org/10.1074/jbc.M111.336495 PMID: 22396551

74. Zhang ZT, Zhu SY. Drosomycin, an essential component of antifungal defence in *Drosophila*. Insect Mol. Biol. 2009; 18(5): 549–556.

75. Zhu S, Gao B. Nematode-derived drosomycin-type antifungal peptides provide evidence for plant-to-ecdysozoan horizontal transfer of a disease resistance gene. Nat. Commun. 2014; 5: 3154. https://doi.org/10.1038/ncomms4154 PMID: 24434635

76. Roeder T, Stanisak M, Gelhaus C, Bruchhaus I, Grotzinger J, Leippe M. Caenopores are antimicrobial peptides in the nematode *Caenorhabditis elegans* instrumental in nutrition and immunity. Dev. Comp. Immunol. 2010; 34(2): 203–209.

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

78. Selkirk ME, Hussein AS, Chambers AE, Goulding D, Gares MP, Vásquez-Lopez C, et al. *Trichinella spiralis* secretes a homologue of prosaposin. Mol. Biochem. Parasitol. 2004; 135(1): 49–56.

79. Lim MP, Firdaus-Raih M, Nathan S. Nematode peptides with host-directed anti-inflammatory activity rescue *Caenorhabditis elegans* from a *Burkholderia pseudomallei* infection. Front. Microbiol. 2016; 7: 1436.

80. Chehayeb JF, Roberson AP, Martin RJ, Geary TG. Proteomic analysis of adult *Ascaris suum* fluid compartments and secretory products. PLoS Negl. Trop. Dis. 2014. 8(6): e2939.

81. Atkinson LE, Liu Y, McKay F, Vandewyer E, Viau C, Irvine A, et al. *Ascaris suum* informs extrasynaptic volume transmission in nematodes. ACS Chem. Neurosci. 2021; 12(17): 3176–3188.

82. Wang T, Ma G, Ang C-S, Korhoren PK, Koehler AV, Young ND, Nie S, Williamson NA, Gasser RB. High throughput LC-MS/MS-based proteomic analysis of excretory-secretory products from short-term in vitro culture of *Haemonchus contortus*. J. Proteomics. 2019; 204: 103375.

83. Chang DZ, Serra L, Lu D, Mortazavi A, Dillman AR. A core set of venom proteins is released by entomopathogenic nematodes in the genus *Steinernema*. PLoS Pathog. 2019; 15(5): e1007626.

84. Leroux L-P, Nasr M, Valanparambil R, Tam M, Rosa BA, Siciliani E, et al. Analysis of the *Trichuris suis* excretory/secretory proteins as a function of life cycle stage and their immunomodulatory properties. Sci. Rep. 2018; 8: 15921.

85. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015; 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351 PMID: 26059717