**QUEEN'S
UNIVERSITY
BELFAST**

In silico analyses of neuropeptide-like protein (NLP) profiles in parasitic nematodes

McKay, F. M., McCoy, C. J., Crooks, B., Marks, N. J., Maule, A. G., Atkinson, L. E., & Mousley, A. (2022). In silico analyses of neuropeptide-like protein (NLP) profiles in parasitic nematodes. *International Journal for Parasitology*, 52(1), 77-85. <https://doi.org/10.1016/j.ijpara.2021.07.002>

Published in:
International Journal for Parasitology

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights
Copyright 2021 the authors.
This is an open access article published under a Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

In silico analyses of neuropeptide-like protein (NLP) profiles in parasitic nematodes

Fiona M. McKay, Ciaran J. McCoy, Bethany Crooks, Nikki J. Marks, Aaron G. Maule, 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

ARTICLE INFO

Article history:

Received 4 May 2021

Received in revised form 1 July 2021

Accepted 5 July 2021

Available online 24 August 2021

Keywords:

Neuropeptide

Neuropeptide-like protein

NLP

Nematode

Parasite

Ascaris

ABSTRACT

Nematode parasite infections cause disease in humans and animals and threaten global food security by reducing productivity in livestock and crop farming. The escalation of anthelmintic resistance in economically important nematode parasites underscores the need for the identification of novel drug targets in these worms. Nematode neuropeptide signalling is an attractive system for chemotherapeutic exploitation, with neuropeptide G-protein coupled receptors (NP-GPCRs) representing the lead targets. In order to successfully validate NP-GPCRs for parasite control it is necessary to characterise their function and importance to nematode biology. This can be aided through identification of receptor activating ligand (s) via deorphanisation. Such efforts require the identification of all neuropeptide ligands within parasites. Here we mined the genomes of nine therapeutically relevant pathogenic nematodes to characterise the neuropeptide-like protein complements and demonstrate that: (i) parasitic nematodes possess a reduced complement of neuropeptide-like protein-encoding genes relative to *Caenorhabditis elegans*; (ii) parasite neuropeptide-like protein profiles are broadly conserved between nematode clades; (iii) five *Ce-nlps* are completely conserved across the nematode species examined; (iv) the extent and position of neuropeptide-like protein-motif conservation is variable; (v) novel RPamide-encoding genes are present in parasitic nematodes; (vi) novel Allatostatin-C-like peptide encoding genes are present in both *C. elegans* and parasitic nematodes; (vii) novel neuropeptide-like protein families are absent in *C. elegans*; and (viii) highly conserved nematode neuropeptide-like proteins are bioactive. These data highlight the complexity of nematode neuropeptide-like proteins and reveal the need for nomenclature revision in this diverse neuropeptide family. The identification of neuropeptide-like protein ligands, and characterisation of those with functional relevance, advance our understanding of neuropeptide signalling to support exploitation of the neuropeptidergic system as an anthelmintic target.

© 2021 The Authors. Published by Elsevier Ltd on behalf of Australian Society for Parasitology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nematode parasites continue to threaten human health and global food security (Hotez et al., 2008; Payne and Fitchett, 2010; Elling, 2013; Roeber et al., 2013; Charlier et al., 2014; Mavrot et al., 2015; Coyne et al., 2018; Lopes-Caitar et al., 2019; Hotez and Lo, 2020). The impact of nematode parasites is exacerbated by the limited number of available anthelmintic drugs and the escalation of anthelmintic resistance in key nematode pathogens (Kaplan and Vidyashankar, 2012; Sangster et al., 2018). The disruption of nematode neuromuscular signalling is a well-established route to parasite control (McVeigh et al., 2006a, 2012; Li and

Kim, 2008), where the neuropeptide signalling component represents an appealing unexploited source of novel anthelmintic targets.

Nematode neuropeptide families encompass the FMRF-amide like peptides (FLPs), neuropeptide-like proteins (NLPs) and the insulin-like peptides (INSs). Whilst INSs peptides appear to signal through receptor tyrosine kinases (Kimura et al., 1997), the majority of FLPs and NLPs are believed to act as ligands for G-protein coupled receptors (GPCRs) (Li and Kim, 2008). Within the neuropeptide system, neuropeptide GPCRs (NP-GPCRs) emerge as the most promising novel anthelmintic targets, with a proven history of 'druggability' (for review see McVeigh et al., 2012). Ultimately, validation of NP-GPCRs as therapeutic targets requires functional characterisation to enable the selection of the most appealing candidates.

* Corresponding author.

E-mail address: A.Mousley@qub.ac.uk (A. Mousley).

While heterologous approaches to NP-GPCR deorphanisation in *Caenorhabditis elegans* have been successful, only a limited number of *C. elegans* NP-GPCRs have been functionally deorphanised, and only two reports of successful NP-GPCR deorphanisation have been described in parasites (Atkinson et al., 2013; Anderson et al., 2014). Whilst a major challenge of receptor deorphanisation is functional expression (McCoy et al., 2017), to aid deorphanisation efforts in therapeutically relevant nematode parasites we also require a complete profile of all potential interacting neuropeptide ligands and receptors. Recent studies have highlighted comprehensive analyses of FLP and NP-GPCR profiles in parasitic nematodes (McCoy et al., 2014), however our current understanding of NLP complements is significantly more limited (McVeigh et al., 2008; Buzy et al., 2021).

Here, we utilise an established neuropeptide Basic Local Alignment Search Tool (BLAST) approach to provide a comprehensive analyses of NLP-encoding gene profiles in nine key nematode parasites, and examine the bioactivity of the most highly conserved NLPs. These data advance understanding of nematode neuropeptide signalling and are critical to future deorphanisation efforts in parasitic nematodes that drive novel parasite control strategies.

2. Materials and methods

2.1. Species selection

Caenorhabditis elegans nlp gene sequeles were identified in the genomic datasets of nine parasitic nematode species (Supplementary Table S1). The species employed in this study were selected based on a number of criteria including: (i) quality and availability of genomic and transcriptomic data, (ii) species employed in previously published neuropeptide analyses (for comparative purposes; (McCoy et al., 2014)), (iii) species appeal from a parasite control perspective (key parasites of humans, livestock and plants), (iv) representatives across clades and lifestyles, (v) species that are experimentally tractable and/or have been studied with respect to nematode neurobiology, and (iv) species for which we have amassed data on the localisation and function of neuropeptides.

2.2. BLAST searches

A reciprocal BLAST method was used to identify *nlp* gene sequeles using the 'prepropeptide search string' approach (McCoy et al., 2014). Protein sequences for NLPs identified previously in *C. elegans* (Nathoo et al., 2001; Husson et al., 2005, 2007, 2014; McVeigh et al., 2008; Yamada et al., 2010; Howe et al., 2017; Van Bael et al., 2018; Supplementary Table S2) were obtained from WormBase (<https://www.wormbase.org>; WS276) and used as queries in translated nucleotide (tBLASTn) and predicted protein (BLASTp) searches of the genome datasets. tBLASTn searches were carried out on the WormBase ParaSite server (<https://parasite.wormbase.org/index.html>; Howe et al., 2017); BLASTp searches were carried out on the WormBase ParaSite Server or via a local Windows command line-based NCBI-BLAST. All BLAST searches were conducted between April 2017 and August 2018. Where there were multiple genomes available for a single species, all were mined (Supplementary Table S1). BLAST searches were also carried out in four additional *Caenorhabditis* spp. (*Caenorhabditis brenneri*, *Caenorhabditis briggsae*, *Caenorhabditis remanei* and *Caenorhabditis japonica*; as previously described (McCoy et al., 2014)). *nlp* genes (*nlp*-24–34), thought to have antimicrobial activity and distinguished by their glycine-rich sequences, were not included in this study due to their unsuitability for the BLAST search approach (McVeigh et al., 2008; Pujol et al., 2008; Dierking et al., 2016). Where *nlp* genes encoded multiple isoforms,

only the longest were used as query sequences unless isoforms were significantly different. Expect values were set to ≥ 1000 to avoid false negative returns. All returned sequences were examined manually for NLP motifs flanked by putative mono/dibasic cleavage sites, signal peptide cleavage sites, or C-terminal signatures to eliminate false positive hits (McVeigh et al., 2008; McCoy et al., 2014) and peptides were predicted accordingly (see Supplementary Table S7).

2.3. Post-BLAST analysis

Nematode parasite NLP-encoding gene sequeles were assigned primarily based on sequence homology to the predicted peptide region of *Ce-nlps* (Li and Kim, 2008; Van Bael et al., 2018). Sequences were aligned using Vector NTI Advance 11.5 AlignX® multiple sequence alignment tool (Lu and Moriyama, 2004). Signal peptide predictions were generated using SignalP4.1 (Nielsen, 2017), however lack of a predicted signal peptide did not exclude NLP designation.

2.4. Clustered analysis of sequences (CLANS) analysis

The Clustered Analysis of Sequences (CLANS) algorithm was used to perform all-against-all BLASTp comparisons between the identified nematode RPamide and Allatostatin-C like prepropeptide sequences, and generate a three-dimensional (3D) similarity matrix (Frickey and Lupas, 2004). Prepropeptide sequences were analysed using the CLANS tool (<https://toolkit.tuebingen.mpg.de/#/tools/clans>) and the most appropriate *E*-value limit was determined to facilitate sufficient cluster separation. All other parameters were set as default. The CLANS file output was examined and coloured after 10,000 clustering rounds using the Java-based desktop software.

2.5. RNAseq analysis

Publicly available life-stage-specific transcriptomic data were downloaded from the WormBase ParaSite Gene Expression server (Howe et al., 2017). Raw sequence reads for *Haemonchus contortus* and *Trichuris muris* were downloaded using the NCBI SRA Toolkit (Leinonen et al., 2011). Forward and reverse fastq files were trimmed to remove adapter, leading, trailing and low quality sequences using Trimmomatic (v0.36; parameter: LEADING:5 TRAILING:5 SLIDINGWINDOW:3:15 MINLEN:34 (Bolger et al., 2014)). Appropriate genome assemblies were accessed via WormBase ParaSite FTP server for previous versions of WormBase ParaSite (Howe et al., 2017) and HISAT2 (v2.1.0 (Kim et al., 2015)) was employed to map trimmed reads to genomic data. Raw gene counts were assigned to mapped reads using SubRead v 2.0.1 featureCounts (Liao et al., 2014). Fragments per Kilobase of Transcript per Million (FPKM) data were generated via transformation of raw counts of orthologous genes using countToFPKM in R Studio and median FPKMs calculated.

2.6. *Ascaris suum* ovijector physiology assays

Selected *Ascaris suum* NLP peptides were tested for activity on the *A. suum* ovijector (reproductive muscle) using an established physiology assay (Fellowes et al., 1998; Fellowes et al., 2000; Moffett et al., 2003). Adult female *A. suum* (>20 cm) were collected from the intestines of pigs at a local abattoir (Karro, Northern Ireland) and transported back to the laboratory in mammalian saline (0.9% NaCl, 37 °C). Worms were maintained in *Ascaris* Ringers Solution (ARS; 13.14 mM NaCl, 9.47 mM CaCl₂, 7.83 mM MgCl₂, 12.09 mM C₄H₁₁NO₃(Tris), 99.96 mM NaC₂H₃O₂, 19.64 mM KCl, pH 7.8) at 37 °C and 5 % CO₂ for up to 4 days with media changes

twice daily. The ovijector was dissected from healthy, turgid worms and transferred to Hank's balanced salt solution (HBSS; Life technologies) at 37 °C in a 4 ml recording chamber, where the tissue was attached between a flexible and an inflexible pipette. Tissue activity was amplified and recorded via a photo-optic transducer system (Fetterer et al., 1977; Marks et al., 1996). Ovijectors which displayed regular, spontaneous contractility were equilibrated for at least 5 min before peptide addition. Inactive or erratically active ovijectors were discarded. Peptides were synthesised (Genosphere Biotechnologies Inc.) and 10 mM stock solutions prepared in double distilled (dd)H₂O, Dimethyl sulfoxide (DMSO), or Dimethyl formaldehyde (DMF) and stored in aliquots at –20 °C until use. Addition of 4 µl of ddH₂O, DMSO or DMF had no effect on ovijector activity. Peptide was added to the 4 ml water bath such that addition of 4 µl gave a final peptide concentration of 10 µM. Peptide effects were recorded for 10 min before media were replaced with fresh HBSS. The ovijector was subsequently recorded for a further 10 min to observe recovery. Contraction amplitude and frequency were measured 2 min prior to: time 0 (peptide addition), 2, 5, 10 and 20 min post-addition. In some cases, it was necessary to analyse additional time points to capture transient effects. The change in tension was also measured at time 0, 2, 5, 10 and 20 min post-addition, where time 0 = 0 mg/mm. Note that muscle relaxation caused an increase in tension (+mg/mm; shortening of the tissue) whereas muscle contraction caused a decrease in tension (–mg/mm; lengthening of tissue). Data were statistically analysed (Graphpad Prism 8) with repeated measures ANOVA followed by Dunnett's post-test to compare each time point with time 0.

3. Results and discussion

3.1. Parasitic nematodes possess a reduced complement of *C. elegans* NLP-encoding genes

nlp gene sequelogues (326) were identified in nine key nematode parasites (Table 1; Supplementary Fig. S1; Supplementary Table S3). These included 73 previously reported NLP-encoding genes identified through *in silico* and peptidomics studies (Nathoo et al., 2001; McVeigh et al., 2008; Jarecki et al., 2011; Koziol et al., 2016; Warnock et al., 2017; Van Bael et al., 2018; Buzy et al. 2021), and 253 novel sequences identified in this study. These data provide an insight into *nlp* conservation and diversity in nematode parasites that include multiple clades and lifestyles. As noted with parasite FLP and NP-GPCR datasets (McCoy et al., 2014) the parasitic nematode species investigated here displayed reduced complements of the *Ce*-NLP profile (Table 1). The human hookworm *Necator americanus* exhibited the largest share (67%) of *Ce*-*nlp*s; this is unsurprising considering they are both members of clade 9 (Holterman et al., 2006). In contrast, the clade 2 species *Trichinella spiralis* and *T. muris* displayed a significantly reduced *nlp* complement of 15% and 14% of the *Ce*-*nlp* profile, respectively. This is likely a result of both specific gene loss in the clade 2 species, as well as *nlp* gene duplication events within the lineages that led to the extant nematodes comprising the 'crown' clades (Holterman et al., 2006) that exhibit broader *nlp* complements. In addition, RNASeq data show that almost all of the predicted *nlp*s reported here are expressed in at least one of the parasitic nematodes examined and appear to be differentially expressed across lifecycle stages (see Supplementary Fig. S2). These data provide confidence that the *nlp*s identified by our *in silico* approach have roles in nematode biology.

3.2. Parasite *nlp* profiles are more conserved in closely related species within the same clade

There appeared to be less intra-clade than inter-clade *nlp* profile variation, mirroring trends observed for *flps* and NP-GPCRs

(Holterman et al., 2006; McCoy et al., 2014). The clade 2 nematodes (*T. spiralis*, *T. muris*) possess a near identical *nlp* profile, however notable differences in *nlp* profiles were observed within clade 8 species. Indeed, although the clade 8 filarids displayed near identical *nlp* profiles, they were different from that observed for *A. suum* which is also within clade 8. This is not surprising given that the filarid species are more closely related to each other than to *A. suum*, and likely reflects key differences in the lifestyles of the filarial species relative to *A. suum*.

3.3. Five *nlp*s are completely conserved across key nematode species

Five *Ce*-*nlp*s were completely conserved across all nine species interrogated in this study: *nlp*-12, -36, -47, -49 and *pdf*-1(*nlp*-74) (Table 1; Supplementary Fig. S1), highlighting their likely functional importance to nematode biology/behaviour, and supporting their presence in the last common ancestor of all nematodes. By contrast, 13 of the already annotated *Ce*-*nlp*s were absent from all parasites examined (*nlp*-4.1, -4.2, -16, -22, -23, -39, -41, -45, -53, -65, -78, -80, and *rgba*-1 (*nlp*-84)) and so have either, been lost independently in multiple parasitic nematode lineages or, have evolved relatively recently in the specific lineage that led to *C. elegans*. Indeed, although all other *Caenorhabditis* spp. examined here possess a similar NLP profile to that of *C. elegans* (Supplementary Table S4), the other *Caenorhabditids* examined lacked *nlp*-39 and *rgba*-1(*nlp*-84), suggesting that these genes arose following the split of the *C. elegans* lineage from *C. briggsae*, *C. brenneri* and *C. remanei* (~5–30 million years ago (MYA); Frézal and Félix, 2015). Similarly, the absence of *nlp*-4.1, -4.2, and -65 in *C. japonica* suggest that these genes likely arose following the separation of the 'Japonica' and 'Elegans' groups (~125–190 MYA; Lemos, 2007; Frézal and Félix, 2015). Please note that we designate the *C. elegans* genes F59C6.6 and F59C6.18 as *nlp*-4.1 and *nlp*-4.2, respectively, as both genes appear to have been independently annotated as '*nlp*-4' previously (see both WormBase (WS276) and Nathoo et al., 2001).

3.4. The extent and position of NLP motif conservation is variable

The extent of the amino acid residue conservation within the NLPs encoded on each gene varied (Supplementary Fig. S1). For example, in some instances primary sequence conservation was observed across the entire length of the encoded peptide (e.g. NLP-3, -43), however in other cases conservation was biased toward either the C- (e.g. NLP-17) or N-termini (e.g. NLP-9, -21). Some NLPs (e.g. NLP-6) displayed relatively low levels of sequence conservation across the entire length of the peptide, however this did not preclude their identification using the motif-based BLAST approach employed here. These observations suggest different levels of functional constraint between peptides, and are likely indicative of the motif that is important for receptor recognition or binding.

3.5. Pan-phylum *in silico* analyses aid NLP prediction

For the most part the conserved NLP motifs, identified through sequence alignments of putative prepropeptide nematode parasite NLP sequelogues, map to the previously predicted peptide regions identified via *C. elegans* *in silico* and mass spectrometry analyses (Husson et al., 2005; Van Bael et al., 2018). However, approximately one-third of the previously predicted NLP peptides, including multiple peptides identified by *C. elegans* mass spectrometry (Husson et al., 2005; Van Bael et al., 2018), are not conserved in any of the parasite species examined in this study (Supplementary Fig. S1; Supplementary Table S5). These peptides likely represent *C. elegans*-specific functionally important peptides or by-products of neuropeptide processing events. Indeed, sequence alignments of

Table 1
nlp-gene sequeleques in nine nematode parasite species. A black box indicates the presence of a sequeleque identified via BLAST search.

Clade	2		8			9			10	12
	Species									
Nlp-gene	Trichinella spiralis	Trichuris muris	Ascaris suum	Brugia malayi	Dirofilaria immitis	Caenorhabditis elegans	Necator americanus	Haemonchus contortus	Bursaphelenchus xylophilus	Globodera pallida
nlp-1										
nlp-2										
nlp-3										
nlp-4.1										
nlp-4.2										
nlp-5										
nlp-6										
nlp-7										
nlp-8										
nlp-9										
nlp-10										
nlp-11										
nlp-12										
nlp-13										
nlp-14										
nlp-15										
nlp-16										
nlp-17										
nlp-18										
nlp-19										
nlp-20										
nlp-21										
nlp-22										
nlp-23										
nlp-35										
nlp-36										
pdf-2/nlp-37										
nlp-38										
nlp-39										
nlp-40										
nlp-41										
nlp-42										
nlp-43										
capa-1/nlp-44										
nlp-45										
nlp-46										
nlp-47										
nlp-48										
crf-1/nlp-49										
nlp-50										
nlp-51										
nlp-52										
nlp-53										
trh-1/nlp-54										

Clade	2		8			9			10	12
	Species									
Nlp-gene	Trichinella spiralis	Trichuris muris	Ascaris suum	Brugia malayi	Dirofilaria immitis	Caenorhabditis elegans	Necator americanus	Haemonchus contortus	Bursaphelenchus xylophilus	Globodera pallida
nlp-55										
nlp-56										
nlp-57										
nlp-58										
nlp-59										
nlp-60										
nlp-61										
nlp-62										
nlp-63										
msrp-7/nlp-64										
nlp-65										
nlp-66										
nlp-67										
nlp-68										
nlp-69										
nlp-70										
nlp-71										
lury-1/nlp-72										
nlp-73										
pdf-1/nlp-74										
ntc-1/nlp-75										
nlp-76										
nlp-77										
nlp-78										
nlp-79										
nlp-80										
nlp-81										
nlp-82										
snet-1/nlp-83										
rgba-1/nlp-84										
nlp-85 ^a										
nlp-86 ^a										
nlp-87 ^a										
nlp-88 ^a										
nlp-89 ^a										
nlp-90 ^a										
nlp-91 ^a										
nlp-92 ^a										
nlp-93 ^a										
nlp-94 ^a										
nlp-95 ^a										
nlp-96 ^a										
nlp-97 ^a										
nlp-98 ^a										

^aNovel gene identified in this study.

some parasite- and Ce-NLP prepropeptides (e.g. NLP-56, NLP-79; Fig. 1) reveal alternative highly conserved motifs, that were not previously considered as putative peptides (Van Bael et al., 2018), but may represent functional NLPs. These data are fundamental to the construction of the peptide libraries that seed deorphanisation efforts and underscore the value of pan-phylum in silico analyses for peptide prediction.

3.6. Parasitic nematode in silico analyses reveal novel NLP-encoding genes and highlight the need for nomenclature revision

Phylum-spanning in silico analyses of NLP-encoding genes drives novel peptide discovery in parasites and provides an opportunity for identification of additional *nlp*s in *C. elegans*. This study discovered 14 novel *nlp*s (Table 1), which have been designated *nlp-85* to *nlp-98*. For continuity with previously identified Ce-*nlp*s (Nathoo et al., 2001; McVeigh et al., 2008; Van Bael et al., 2018), and before naming any further novel *nlp*-genes identified here, we have reassigned *snet-1* and *rgba-1* as *nlp-83* and -84, respectively. Although not classed as an *nlp* gene by the current version of WormBase (WS276), *snet-1* (*nlp-83*) was identified as a *C. elegans* neuropeptide gene in an olfactory plasticity study and previ-

ously considered as a member of the NLP family (Yamada et al., 2010; Hobert, 2013); *rgba-1* was identified as a novel *C. elegans* neuropeptide-encoding gene (Yin et al., 2017). The novel *nlp*s discovered in this study include additional members of RPamides and Allatostatins, and a number of entirely novel *nlp*s that had not previously been reported in *C. elegans*.

3.6.1. Novel RPamide-encoding genes were identified in parasitic nematodes

The *C. elegans* RPamide family consists of the *nlp-2*, -22 and -23 gene cluster first described by Nathoo et al. (2001), plus *nlp-46*, identified by McVeigh et al. (2008). The NLP-2, -22 and 23 peptides are highly similar, making sequeleque designation difficult for returned BLAST hits using the *C. elegans* directed BLAST approach. We therefore extended our BLAST searches using all of the RPamide encoding BLAST returns to query all available nematode genomes, and performed multiple sequence alignments to manually identify patterns of sequence conservation. Of the known RPamide encoding genes, *nlp-22* and -23 sequeleques were only identified in *Caenorhabditis* spp., whereas *nlp-2* and *nlp-46* were conserved in a number of parasites (Table 1, Supplementary Table S4).

A

B

Fig. 1. Novel predicted *Caenorhabditis elegans* neuropeptide-like proteins (Ce-NLPs) are conserved in parasitic nematodes. Multiple sequence alignments were performed using Vector NTI Advance 11.5 AlignX® (Lu and Moriyama, 2004). Text in italics highlights signal peptides as identified using Signal P 4.1 (Nielsen, 2017). Bold text indicates peptides predicted previously (see Van Bael et al., 2018). Bold underlined text indicates novel peptides predicted in this study based on their conservation in additional nematode species. Yellow and blue highlighted regions specify completely and partially conserved amino acid residues, respectively; green specifies similar residues. For all other novel NLP predictions made here (derived from *nlp-8*, -40, -48, -52, -56, -58, -66, -67, -79 and -81 prepropeptide alignments) see [Supplementary Table S5](#) and [Supplementary Fig. S1](#). Ce, *C. elegans*; As, *Ascaris suum*; Na, *Necator americanus*; Hc, *Haemonchus contortus*; Bx, *Bursaphelenchus xylophilus*; Gp, *Globodera pallida*.

Three novel RPamide-encoding genes were identified in parasitic nematodes: (i) a novel RPamide sequence A(A/V/T) MISGRGFRPG, initially identified here in *Brugia malayi* and *Dirofilaria immitis*, and sharing a common motif with RPamides encoded in 14 other filarial species, was classified as a novel filarid-specific RPamide-encoding gene (*nlp-85*); (ii) a second novel RPamide peptide-encoding gene (designated here as *nlp-86*; GRW(G/Q) LRPG), identified initially here in *A. suum*, *Bursaphelenchus xylophi-*

lius and *Globodera pallida* was also present in 11 additional nematode species representing both free-living and parasitic lifestyles; finally (iii) a third novel *nlp* (*nlp-87*; S(I/L)ALGR(F/L)(S/N)LRPG), identified initially here in *A. suum*, and subsequently in seven additional nematode species appears distinct from peptides encoded on *nlp-2* and *nlp-86*.

The presence of *nlp-2*, -86, and -87, that encode distinct RPamide peptide motifs, in the *A. suum* genome and subsequent CLANS

analysis provides evidence to support the novel *nlp* designations (Fig. 2A). However, several previous studies have been unable to accurately delineate the RPamide-encoding *nlp*s such that there are conflicting designations in the literature to what we describe here (McVeigh et al., 2008; Jarecki et al., 2011; Nelson et al., 2013; Koziol et al., 2016). This underscores the value of phylum-spanning in silico analyses to unravel the complexity of the NLP family and associated nomenclature. In addition, our approach offers a route to interrogate neuropeptide evolution; indeed, both motif and CLANS analyses reveal that *nlp-60* (a non-RPamide encoding gene) likely represents an RPamide family member that is closely related to *nlp-87* but has lost all RPamide peptides (Fig. 2A).

3.6.2. Novel Allatostatin-C-like peptide encoding genes were identified in *C. elegans* and parasitic nematodes

Our pan-phylum motif-based and reciprocal BLAST approach enabled us to identify nine novel Allatostatin-C-like neuropeptide encoding genes (*nlp-88*, -89, -90, 91, -92, -93, -94, -95 and -96), in addition to those previously reported (*nlp-55*, -62, -64, -66, -67, -71 and -79; (Mirabeau and Joly, 2013; Van Bael et al., 2018)).

Four of the novel Allatostatin-C-like neuropeptide encoding genes (*nlp-88*, -89, -90 and -91) were found in the *C. elegans* genome, share a common RNCFF(S/T)P(V/A)QC motif, possess a signal peptide, and appear to be enriched in neurons (see Worm-Base (WS276): WBGene00010848 (M04B2.6); WBGene00016436 (C35B1.7); WBGene00019292 (K02A6.1); WBGene00019293

Fig. 2. Clustered Analysis of Sequences (CLANS) clustering supports RPamide and Allatostatin-C like peptide encoding gene nomenclature. (A) Similarity matrix derived from all-against-all BLASTp comparisons between all identified nematode RPamide encoding prepropeptide sequences (*E*-value limit = 1). (B) Similarity matrix derived from all-against-all BLASTp comparisons between all identified nematode Allatostatin-C like prepropeptide sequences (*E*-value limit = 1E-5). *nlp*, neuropeptide-like protein.

Fig. 3. Selected neuropeptide-like protein predicted peptides (NLPs) are bioactive on the *Ascaris suum* ovijector muscle. (A) Representative muscle tension recordings showing the effect of As-NLP-36, As-NLP-47, As-NLP-49, and As-PDF-1 peptides. Arrows indicate peptide exposure. (B) The effects of NLPs on (i) *A. suum* ovijector contraction frequency, (ii) contraction amplitude, and (iii) baseline tension. Peptide exposure: 0–10 min. Scale: horizontal bar represents 1 min, vertical bar represents 1 mg.

(K02A6.1)). *nlp*-89, -90 and -91 appear to be restricted to Caenorhabditids, whereas *nlp*-88 is conserved across parasitic nematodes (Table 1; Supplementary Table S4). The additional five novel Allatostatin-C-like neuropeptide encoding genes identified here (*nlp*-92, -93, -94, -95 and -96) share a common (K/R)NC(F/Y) F motif, are conserved across key parasite species, but are absent from *C. elegans*. Again, CLANS analysis broadly supports the peptide motif based novel gene designations highlighted here (Fig. 2B).

3.6.3. Novel NLP families are absent from *Caenorhabditis elegans*

Two additional novel putative NLP-encoding genes, that display distinct peptide motifs (NPYSW, and S(L/V)AP(T/S)TSAX₃₋₄VS), were identified in this study and have been designated *nlp*-97 and -98, respectively; *nlp*-97 and -98 are absent from *C. elegans*. *nlp*-97 sequelogues were identified in a broad range of parasites, spanning clades 8, 9, 10 and 12 (Table 1), whilst *nlp*-98 appears to be *Globodera* spp. specific (see Supplementary Fig. S1). Note that *G. pallida* and *Globodera rostochiensis* each possess two highly similar *nlp*-98 sequelogues (*nlp*-98.1, -98.2); we also note recent *nlp* duplications in specific parasitic nematodes for *nlp*-1, -8, -10, -13,

-15, -19, -21, -36, -40, -47, -61, -69, -71, -76 and -79 and have designated these paralogues as *nlp*-X.1, *nlp*-X.2 etc (see Supplementary Fig. S1).

3.7. Highly conserved NLPs are bioactive on the *Ascaris suum* ovijector

To determine the potential role of the most highly conserved NLPs in regulating muscle function, we examined the effects of the following predicted *A. suum* NLPs: As-PDF-1 (A and B), As-PDF-2 (C), As-NLP-12 (A1, B, C, D), NLP-36 (A1, A2, B1, B2, C1, C2), As-NLP-47 (A) and As-NLP-49 (B) (see Supplementary Table S6; Supplementary Table S7) on the reproductive musculature (ovijector) of *A. suum*, using established methods (Fellowes et al., 1998, 2000; Moffett et al., 2003). Note that NLP-12 (A1, B, C and D), and NLP-36 (A1, A2, C1 and C2) peptides were inactive (see Supplementary Table S6; Supplementary Table S7).

NLP-36B1, -36B2, -47, PDF-1A, -1B peptides exhibited distinct excitatory effects on ovijector muscle (Fig. 3, Supplementary Table S6; Supplementary Table S7) as follows; (i) As-NLP-36B1 and As-NLP-36B2 induced an increase in contraction amplitude

and frequency that aligned with the previously described ovijector response type (RT) 5 (Moffett et al., 2003); (ii) As-NLP-47 induced a transient lengthening of the ovijector with an increase in contraction frequency, similar to RT2 type responses, whilst (iii) As-PDF-1A and As-PDF-1B caused a decrease in contraction amplitude alongside an increase in contraction frequency that aligned to RT5. By contrast, As-NLP-49 was the only peptide tested shown to induce relaxation of the ovijector muscle (RT1; Fig. 3, Supplementary Table S6; Supplementary Table S7).

The muscle physiology described here represents the first known reports of NLP activity on the *A. suum* ovijector (Moffett et al., 2003; Mousley et al., 2004; Mousley et al., 2005; McVeigh et al., 2006b; McVeigh et al., 2008). Intriguingly, *pdf-1*, *nlp-47* and *nlp-49* are all linked to the control of egg-laying in *C. elegans* (Lindemans et al., 2009; Meelkop et al., 2012; Chew et al., 2018), a role which may be conserved in parasitic nematodes.

4. Conclusions

This study provides a comprehensive library of NLPs in nine key parasitic nematodes, and highlights that parasites possess a reduced and variable complement of the *C. elegans* NLP profile. We identify 14 novel *nlp*s, 10 of which are not found in *C. elegans*, bringing the total number of *nlp*s in nematodes to 99. Five *nlp*s display complete conservation across the nine phylogenetically dispersed parasitic nematodes examined here; four of these encode peptides that modulate nematode muscle function.

Whilst this study has been successful in the identification of >300 *nlp* genes across key parasitic nematode species, there are caveats associated with BLAST-based identification of neuropeptide sequelogues. Indeed, the use of *C. elegans* NLP sequences as BLAST queries may not uncover all parasite sequelogues, particularly in the case of those NLPs which display divergent sequences. Enhancements in genome quality for any of these species may reveal additional neuropeptide sequelogues. Note that the non-identification of a specific gene using this BLAST-based approach does not definitively prove its absence from any given parasite. The value of peptidomics approaches, particularly de novo sequencing, to neuropeptide discovery is clear when we consider that many of the more recently identified *C. elegans* neuropeptides are not traditionally 'nlp-/flp-like' (Nathoo et al., 2001; Van Bael et al., 2018). *Caenorhabditis elegans* data show that highly conserved neuropeptides tend to be discovered first by in silico studies, however de novo peptidomics approaches have obvious advantages when sequences are divergent from search string motifs. Nevertheless, in silico studies are highly valuable as they support exploitation of the huge amount of available in silico data to expand our knowledge of invertebrate neuropeptide signalling.

The data presented here advance our understanding of neuropeptide signalling in parasitic nematodes, support the possibility for conservation of neuropeptide function across multiple nematode species, and inform neuropeptide-receptor deorphanisation studies in therapeutically relevant parasite species.

Acknowledgements

The authors are grateful to Karro, Cookstown, NI, UK, for assistance in the collection of nematodes, and wish to thank Dr Isabel Beets and Professor Timothy G. Geary for helpful discussions. The authors acknowledge support for this work from: Biotechnology and Biological Sciences Research Council, UK, grant (BB/H019472/1), Biotechnology and Biological Sciences Research Council, UK/Boehringer Ingelheim, Germany, grants (BB/M010392/1, BB/T016396/1), and Department for the Economy Northern Ireland (DfE).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijpara.2021.07.002>.

References

- Anderson, R.C., Newton, C.L., Millar, R.P., Katz, A.A., 2014. The *Brugia malayi* neuropeptide receptor-4 is activated by FMRFamide-like peptides and signals via G α i. *Mol. Biochem. Parasitol.* 195, 54–58.
- Atkinson, L.E., Stevenson, M., McCoy, C.J., Marks, N.J., Fleming, C., Zamanian, M., Day, T.A., Kimber, M.J., Maule, A.G., Mousley, A., 2013. flp-32 Ligand/receptor silencing phenocopy faster plant pathogenic nematodes. *PLoS Pathog.* 9.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Buzy, A., Allain, C., Harrington, J., Lesuisse, D., Mikol, V., Bruhn, D.F., Maule, A.G., Guillemot, J.C., 2021. Peptidomics of *Haemonchus contortus*. *ACS Omega* 6, 10288–10305.
- Charlier, J., van der Voort, M., Kenyon, F., Skuce, P., Vercruysse, J., 2014. Chasing helminths and their economic impact on farmed ruminants. *Trends Parasitol.* 30, 361–367.
- Chew, Y.L., Grundy, L.J., Brown, A.E., Beets, I., Schafer, W.R., 2018. Neuropeptides encoded by *nlp-49* modulate locomotion, arousal and egg-laying behaviours in *Caenorhabditis elegans* via the receptor SEB-3. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373, 20170368.
- Coyne, D.L., Cortada, L., Dalzell, J.J., Claudius-Cole, A.O., Haukeland, S., Luambano, N., Talwana, H., 2018. Plant-parasitic nematodes and food security in Sub-Saharan Africa. *Annu. Rev. Phytopathol.* 56, 381–403.
- Dierking, K., Yang, W., Schulenburg, H., 2016. Antimicrobial effectors in the nematode *Caenorhabditis elegans*: an outgroup to the Arthropoda. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150299.
- Elling, A.A., 2013. Major emerging problems with minor *Meloidogyne* species. *Phytopathology* 103, 1092–1102.
- Fellowes, R., Maule, A., Marks, N., Geary, T., Thompson, D., Halton, D., 2000. Nematode neuropeptide modulation of the vagina vera of *Ascaris suum*: in vitro effects of PF1, PF2, PF4, AF3 and AF4. *Parasitology* 120, 79–89.
- Fellowes, R., Maule, A., Marks, N., Geary, T., Thompson, D., Shaw, C., Halton, D., 1998. Modulation of the motility of the vagina vera of *Ascaris suum* in vitro by FMRFamide-related peptides. *Parasitology* 116, 277–287.
- Fetterer, R.H., Pax, R.A., Bennett, J.L., 1977. *Schistosoma mansoni*: Direct method for simultaneous recording of electrical and motor activity. *Exp. Parasitol.* 43, 286–294.
- Frézal, L., Félix, M.A., 2015. *C. elegans* outside the Petri dish. *Elife* 4.
- Frickey, T., Lupas, A., 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20, 3702–3704.
- Hoebert, O., 2013. The neuronal genome of *Caenorhabditis elegans*. *WormBook*, ed. The *C. elegans* Research Community, *WormBook*, doi/10.1895/wormbook.1.161.1, <http://www.wormbook.org>.
- Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., Bongers, T., Holovachov, O., Bakker, J., Helder, J., 2006. Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. *Mol. Biol. Evol.* 23, 1792–1800.
- Hotez, P.J., Brindley, P.J., Bethony, J.M., King, C.H., Pearce, E.J., Jacobson, J., 2008. Helminth infections: the great neglected tropical diseases. *J. Clin. Invest.* 118, 1311–1321.
- Hotez, P.J., Lo, N.C., 2020. Neglected tropical diseases: public health control programs and mass drug administration. In: Edward T. Ryan, David R. Hill, Tom Solomon, Naomi E. Aronson, Timothy P. Endy, (Eds.), *Hunter's Tropical Medicine and Emerging Infectious Diseases* (Tenth Edition). Elsevier, online, pp. 209–213.
- Howe, K.L., Bolt, B.J., Shafie, M., Kersey, P., Berriman, M., 2017. WormBase ParaSite – a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10.
- Husson, S.J., Clynen, E., Baggerman, G., De Loof, A., Schoofs, L., 2005. Discovering neuropeptides in *Caenorhabditis elegans* by two dimensional liquid chromatography and mass spectrometry. *Biochem. Biophys. Res. Commun.* 335, 76–86.
- Husson, S.J., Mertens, I., Janssen, T., Lindemans, M., Schoofs, L., 2007. Neuropeptidergic signaling in the nematode *Caenorhabditis elegans*. *Prog. Neurobiol.* 82, 33–55.
- Husson, S.J., Reumer, A., Temmerman, L., De Haes, W., Schoofs, L., Mertens, I., Baggerman, G., 2014. Worm peptidomics. *EuPA Open Proteom.* 3, 280–290.
- Jarecki, J.L., Frey, B.L., Smith, L.M., Stretton, A.O., 2011. Discovery of neuropeptides in the nematode *Ascaris suum* by database mining and tandem mass spectrometry. *J. Proteome Res.* 10, 3098–3106.
- Kaplan, R.M., Vidyashankar, A.N., 2012. An inconvenient truth: global worming and anthelmintic resistance. *Vet. Parasitol.* 186, 70–78.
- Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360.
- Kimura, K.D., Tissenbaum, H.A., Liu, Y., Ruvkun, G., 1997. daf-2, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* 277, 942–946.

- Kozioł, U., Kozioł, M., Preza, M., Costabile, A., Brehm, K., Castillo, E., 2016. De novo discovery of neuropeptides in the genomes of parasitic flatworms using a novel comparative approach. *Int. J. Parasitol.* 46, 709–721.
- Leinonen, R., Sugawara, H., Shumway, M., International Nucleotide Sequence Database Collaboration, 2011. The sequence read archive. *Nucleic Acids Res.* 39, D19–D21.
- Lemos, B., 2007. The Opossum genome reveals further evidence for regulatory evolution in mammalian diversification. *Genome Biol.* 8, 223.
- Li, C., Kim, K., 2008. Neuropeptides. *WormBook*, ed. The C. elegans Research Community, WormBook, doi/10.1895/wormbook.1.142.1, <http://www.wormbook.org>.
- Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- Lindemans, M., Liu, F., Janssen, T., Husson, S.J., Mertens, I., Gäde, G., Schoofs, L., 2009. Adipokinetic hormone signaling through the gonadotropin-releasing hormone receptor modulates egg-laying in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U S A* 106, 1642–1647.
- Lopes-Caitar, V., Pinheiro, J., Marcelino-Guimarães, F., 2019. Nematodes in horticulture: an overview. *J. Hortic. Sci. Crop. Res.* 1, 105.
- Lu, G., Moriyama, E.N., 2004. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief Bioinform.* 5, 378–388.
- Marks, N., Johnson, S., Maule, A., Halton, D., Shaw, C., Geary, T., Moore, S., Thompson, D., 1996. Physiological effects of platyhelminth RFamide peptides on muscle-strip preparations of *Fasciola hepatica* (Trematoda: Digenea). *Parasitology* 113, 393–401.
- Mavrot, F., Hertzberg, H., Torgerson, P., 2015. Effect of gastro-intestinal nematode infection on sheep performance: a systematic review and meta-analysis. *Parasit. Vectors* 8, 557.
- McCoy, C.J., Atkinson, L.E., Robb, E., Marks, N.J., Maule, A.G., Mousley, A., 2017. Tool-driven advances in neuropeptide research from a nematode parasite perspective. *Trends Parasitol.* 33, 986–1002.
- McCoy, C.J., Atkinson, L.E., Zamanian, M., McVeigh, P., Day, T.A., Kimber, M.J., Marks, N.J., Maule, A.G., Mousley, A., 2014. New insights into the FLPergic complements of parasitic nematodes: Informing deorphanisation approaches. *EuPA Open Proteom.* 3, 262–272 <https://www.frontiersin.org/articles/10.3389/fendo.2021.718363/full>.
- McVeigh, P., Alexander-Bowman, S., Veal, E., Mousley, A., Marks, N.J., Maule, A.G., 2008. Neuropeptide-like protein diversity in phylum Nematoda. *Int. J. Parasitol.* 38, 1493–1503.
- McVeigh, P., Atkinson, L., Marks, N.J., Mousley, A., Dalzell, J.J., Sluder, A., Hammerland, L., Maule, A.G., 2012. Parasite neuropeptide biology: seeding rational drug target selection? *Int. J. Parasitol. Drugs Drug Resist.* 2, 76–91.
- McVeigh, P., Geary, T.G., Marks, N.J., Maule, A.G., 2006a. The FLP-side of nematodes. *Trends Parasitol.* 22, 385–396.
- McVeigh, P., Leech, S., Marks, N.J., Geary, T.G., Maule, A.G., 2006b. Gene expression and pharmacology of nematode NLP-12 neuropeptides. *Int. J. Parasitol.* 36, 633–640.
- Meelkop, E., Temmerman, L., Janssen, T., Suetens, N., Beets, I., Van Rompay, L., Shanmugam, N., Husson, S.J., Schoofs, L., 2012. PDF receptor signaling in *Caenorhabditis elegans* modulates locomotion and egg-laying. *Mol. Cell. Endocrinol.* 361, 232–240.
- Mirabeau, O., Joly, J.-S., 2013. Molecular evolution of peptidergic signaling systems in bilaterians. *Proc. Natl. Acad. Sci. USA* 110, E2028–E2037.
- Moffett, C., Beckett, A., Mousley, A., Geary, T., Marks, N., Halton, D., Thompson, D., Maule, A., 2003. The ovijector of *Ascaris suum*: multiple response types revealed by *Caenorhabditis elegans* FMRFamide-related peptides. *Int. J. Parasitol.* 33, 859–876.
- Mousley, A., Marks, N.J., Halton, D.W., Geary, T.G., Thompson, D.P., Maule, A.G., 2004. Arthropod FMRFamide-related peptides modulate muscle activity in helminths. *Int. J. Parasitol.* 34, 755–768.
- Mousley, A., Moffett, C.L., Duve, H., Thorpe, A., Halton, D.W., Geary, T.G., Thompson, D.P., Maule, A.G., Marks, N.J., 2005. Expression and bioactivity of allatostatin-like neuropeptides in helminths. *Int. J. Parasitol.* 35, 1557–1567.
- Nathoo, A.N., Moeller, R.A., Westlund, B.A., Hart, A.C., 2001. Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. *Proc. Natl. Acad. Sci. USA* 98, 14000–14005.
- Nelson, M.D., Trojanowski, N.F., George-Raizen, J.B., Smith, C.J., Yu, C.C., Fang-Yen, C., Raizen, D.M., 2013. The neuropeptide NLP-22 regulates a sleep-like state in *Caenorhabditis elegans*. *Nat. Commun.* 4, 2846.
- Nielsen, H., 2017. Predicting secretory proteins with SignalP. *Methods Mol. Biol.* 1611, 59–73.
- Payne, L., Fitchett, J.R., 2010. Bringing neglected tropical diseases into the spotlight. *Trends Parasitol.* 26, 421–423.
- Pujol, N., Zugasti, O., Wong, D., Couillault, C., Kurz, C.L., Schulenburg, H., Ewbank, J.J., 2008. Anti-fungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides. *PLoS Pathog.* 4.
- Roeber, F., Jex, A.R., Gasser, R.B., 2013. Impact of gastrointestinal parasitic nematodes of sheep, and the role of advanced molecular tools for exploring epidemiology and drug resistance—an Australian perspective. *Parasit. Vectors* 6, 153.
- Sangster, N.C., Cowling, A., Woodgate, R.G., 2018. Ten events that defined anthelmintic resistance research. *Trends Parasitol.* 34, 553–563.
- Van Bael, S., Zels, S., Boonen, K., Beets, I., Schoofs, L., Temmerman, L., 2018. A *Caenorhabditis elegans* mass spectrometric resource for neuropeptidomics. *J. Am. Soc. Mass Spectrom.* 29, 879–889.
- Warnock, N.D., Wilson, L., Patten, C., Fleming, C.C., Maule, A.G., Dalzell, J.J., 2017. Nematode neuropeptides as transgenic nematicides. *PLoS Pathog.* 13, e1006237.
- Yamada, K., Hirotsu, T., Matsuki, M., Butcher, R.A., Tomioka, M., Ishihara, T., Clardy, J., Kunitomo, H., Iino, Y., 2010. Olfactory plasticity is regulated by pheromonal signaling in *Caenorhabditis elegans*. *Science* 329, 1647–1650.
- Yin, J.A., Gao, G., Liu, X.J., Hao, Z.Q., Li, K., Kang, X.L., Li, H., Shan, Y.H., Hu, W.L., Li, H. P., Cai, S.Q., 2017. Genetic variation in glia-neuron signalling modulates ageing rate. *Nature* 551, 198–203.