Nematology

Comparative phylogenetic analysis of vitellogenin in species of cyst and root-knot nematodes

--Manuscript Draft--

Manuscript Number:	NEMY-1659
Full Title:	Comparative phylogenetic analysis of vitellogenin in species of cyst and root-knot nematodes
Short Title:	Phylogenetic analysis of vitellogenin
Article Type:	Full Research Paper
Corresponding Author:	Hannah Wileman, BSc University of Hertfordshire Hatfield, Hertfordshire UNITED KINGDOM
Keywords:	Caenorhabditis; development; egg; Globodera; Heterodera; Meloidogyne; sedentary endoparasite; vitellin
Corresponding Author's Institution:	University of Hertfordshire
First Author:	Hannah Wileman, BSc
Order of Authors:	Hannah Wileman, BSc
	Keith G. Davies, PhD
	Roland N. Perry, PhD
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Summary Plant-parasitic nematodes (PPN) are an economically important group of crop pests and are oviparous animals; all nutrients required to develop and ensure the survival of their unhatched progeny need to be deposited within the egg, including proteins. The most abundant protein deposited is vitellin, formed of a precursor protein vitellogenin, which has roles in transporting lipids, providing amino acids and influencing post-embryonic development. The genes encoding vitellogenin have been well studied in Caenorhabditis elegans, but little is known about vitellogenin in PPN. Using the vitellogenin gene sequences from C. elegans, homologous sequences in the genomes of cyst and root-knot nematodes were identified and hypothetical vitellogenin genes were predicted. Protein domains were then determined. Sequences were aligned using MUSCLE and then used to construct phylogenetic trees using the maximum likelihood method. With the availability of genomic data and use of online local alignment tools, the vitellogenin encoding genes from C. elegans could be aligned to sequences from PPN genomes. All predicted genes contained the same protein domains as C. elegans; Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor domain type D. The constructed phylogenetic tree clustered the species into three characterised groups, root-knot nematodes, cyst nematodes and Caenorhabditis species. Vitellogenin genes in C. elegans were homologous to sequences within PPN genomes, allowing the hypothetical genes to be determined and the relationships between PPN vitellogenin genes to be inferred, forming a potential basis to understand further the role of vitellogenin in cyst and root-knot nematodes.
Funding Information:	K. G. Davies Ltd (HKEP University of Hertfordshire Project C004137) — Miss Hannah Wileman
	European Regional Development Fund (HKEP University of Hertfordshire Project C004137) — Miss Hannah Wileman
	Hertfordshire Local Enterprise Partnership (HKEP University of Hertfordshire Project C004137) — Miss Hannah Wileman

Comparative phylogenetic analysis of vitellogenin in species of cyst and root-knot nematodes

Hannah J. WILEMAN*, Roland N. PERRY and Keith G. DAVIES

Department of Clinical, Pharmaceutical and Biological Sciences, School of Life and Medical Sciences, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK

Running title: Phylogenetic analysis of vitellogenin

---

*Corresponding author, e-mail: h.wileman3@herts.ac.uk

Summary - Plant-parasitic nematodes (PPN) are an economically important group of crop pests and are oviparous animals; all nutrients required to develop and ensure the survival of their unhatched progeny need to be deposited within the egg, including proteins. The most abundant protein deposited is vitellin, formed of a precursor protein vitellogenin, which has roles in transporting lipids, providing amino acids and influencing post-embryonic development. The genes encoding vitellogenin have been well studied in *Caenorhabditis elegans*, but little is known about vitellogenin in PPN. Using the vitellogenin gene sequences from *C. elegans*, homologous sequences in the genomes of cyst and root-knot nematodes were identified and hypothetical vitellogenin genes were predicted. Protein domains were then determined. Sequences were aligned using MUSCLE and then used to construct phylogenetic trees using the maximum likelihood method. With the availability of genomic data and use of online local alignment tools, the vitellogenin encoding genes from *C. elegans* could be aligned to sequences from PPN genomes. All predicted genes contained the same protein domains as *C. elegans*; Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor domain type D. The constructed phylogenetic tree clustered the species into three characterised groups, root-knot nematodes, cyst nematodes and *Caenorhabditis* species. Vitellogenin genes in *C. elegans* were homologous to sequences within PPN genomes, allowing the hypothetical genes to be determined and the relationships between PPN vitellogenin genes to be inferred, forming a potential basis to understand further the role of vitellogenin in cyst and root-knot nematodes.

Keywords - *Caenorhabditis briggsae*, *Caenorhabditis elegans*, development, egg, *Globodera*, *Heterodera*, *Meloidogyne*, plant-parasitic nematode, sedentary endoparasite, vitellin.

Plant-parasitic nematodes (PPN) have been estimated to cause at least $US 80 billion of damage to crops per year (Nicol *et al.*, 2011). Infections by PPN cause non-specific symptoms such as stunted growth, wilting and yellowing of the leaves, caused by a decrease in function of the roots elicited by the nematode infection, which can lead to a reduced yield of the crop (Kumar & Yadav, 2020). Cyst nematodes (*Globodera* and *Heterodera*) and root-knot nematodes (*Meloidogyne*) are two of the most economically important groups of PPN (Jones *et al.*, 2013) and are found in the order Tylenchida.

Nematodes are oviparous animals; the developing nematode within the egg is surrounded by perivitelline fluid, which contains essential nutrients and protective antioxidants (Mkandawire *et al.*, 2021). For the successful development of the nematode, all nutrients for embryogenesis and further growth need to be deposited into the egg. Nutrients are deposited in the egg as proteins and other macromolecules; the primarily stored proteins are yolk proteins, also known as vitellin (Almenara *et al.*, 2013). Vitellin is formed of a precursor protein vitellogenin, which is a large glyco-lipoprotein complex (Winter *et al.*, 1996), a distant relative of the apoB protein in humans (Baker, 1988a). In *Caenorhabditis elegans*, synthesis of vitellogenin occurs in the hermaphrodite intestine, which is then secreted into the body cavity and taken up by the oocytes by receptor-mediated endocytosis (Perez & Lehner, 2019); the protein is then deposited in yolk granules of the egg (Winter *et al.*, 1996). Deposits of yolk protein remain within the egg after embryogenesis and prior to hatching (Bossinger & Schierenberg, 2003).

Vitellogenin is believed to have the primary function of the transport of lipids and micronutrients (Hayward *et al.*, 2010) and amino acid provision to developing progeny (Winter *et al.*, 1996). Additionally, vitellogenin has also been shown to influence post-embryonic phenotypes, whereby limited supply of the protein can result in smaller sized offspring and sterility in *C. elegans* (Perez & Lehner, 2019). Some studies report that

vitellogenin also serves a protective role against oxidative stress (Ishii *et al.*, 2002) and environmental protection from bacterial infections (Fischer *et al.*, 2013).

In *C. elegans*, vitellogenin proteins consist of four polypeptides: two large, 170 kDa yp170A and 170 kDa yp170B, and two small, 115 kDa yp115 and 88 kDa yp88 (Perez & Lehner, 2019). Six genes encoding vitellogenin have been characterised in *C. elegans* (*vit* 1-6) (Spieth *et al.*, 1985). The peptide encoded by *vit*-6 is proteolytically cleaved prior to entering the oocyte, to form yp115 and yp88 (Spieth *et al.*, 1991).

Vitellogenin proteins and the associated encoding genes have also been characterised in some species of both vertebrates (fish and birds) and invertebrates (insects). *Caenorhabditis elegans* has been the primary model for investigating vitellogenin in nematodes; however, orthologous genes have been found in other free-living nematodes: *Oscheius tipulae* (Almenara *et al.*, 2013) and *C. briggsase* (Zucker-Aprison & Blumenthal, 1989). Vitellogenin has also been identified in the animal-parasitic nematodes *Toxocara canis* (Zhu *et al*., 2017), *Haemonchus contortus* (Hartman *et al.*, 2001) and *Trichostongylus vitrinus* (Nisbet & Gasser, 2004). In *Heterodera glycines*, vitellogenin has been described and partially characterised as two major egg proteins (Masler, 1999). These proteins constitute over 50% of all proteins present within the egg with molecular weights of 190 kDa and 180 kDa, a comparative molecular weight to vitellogenin genes found in *C. elegans* (Masler, 1999). Additionally, when expressed sequence tag data from female *Globodera pallida* samples were analysed, vitellogenin-like transcripts were found to be most abundant (Jones *et al.*, 2009). Furthermore, the females of *G. pallida* develop from 21 to 35 days post infection (dpi) of the host plant; vitellogenin was shown to be highly expressed at 28 and 35 dpi, due to the deposition of yolk proteins within the oocytes (Cotton *et al.*, 2014). In the transcriptome of the root-lesion nematode, *Pratylenchus penetrans*, vitellogenin was again shown to be highly expressed (Vieira *et al.*, 2015).

Vitellogenin is an important protein in yolk provision for the developing nematode and possibly has a protective role; however, they have not been identified or characterised in many species of cyst and root-knot nematodes. With the increasing availability of genomic data for many of these economically important species, it is now possible to determine presence of vitellogenin genes within their genomes and infer their phylogenetic relationships.

Materials and methods

Data acquisition

Vitellogenin gene sequences from *C. elegans* (*vit*-1, *vit*-2, *vit*-3, *vit*-4, *vit*-5, and *vit*-6) were obtained using wormbase.org (Table S1). These sequences were then used as the query sequences for BLAST searches against the genomes of selected species of cyst nematodes (*Globodera* and *Heterodera* spp.), root-knot nematodes (*Meloidogyne* spp.) and *C. briggsae* on the parasite.wormbase.org server. BLAST searches were completed using tblastn, with a maximum of 100 target sequences, expect threshold of 0.01, and a word size of 2. The scoring parameters used were BLOSUM62 matrix and low complexity regions were filtered. The result gave homologous regions and a corresponding predicted 'overlapping gene' in each genome, these gene sequences were then selected from the database and collated for sequence alignment.

Gene prediction

The genome of *G. ellingtonae* was not available from the server of parasite.wormbase.org; instead NCBI BLAST was used to determine homologous regions within the species genome. Regions of the genome containing homologous sequences from the resulting BLAST searches were selected and uploaded into the Galaxy web platform (Afgan *et al.*, 2018); here the sequences were then inputted into the AUGUSTUS tool (Stanke & Morgenstern, 2005; Stanke *et al*., 2008). *Caenorhabditis elegans* was chosen as the training set model organism, the protein sequence and coding sequence were then predicted using both strands. The predicted vitellogenin genes were then selected by comparing the sequence to homologous regions identified in the BLAST search.

PHYLOGENETIC TREE CONSTRUCTION

All genes, predicted or database derived through BLAST searches, were then uploaded to Pfam (Bateman *et al.*, 2004) to determine theoretical protein domains for each protein sequence. Sequences which contained the same protein domains as *C. elegans* vitellogenin genes were included in the sequence alignment. Sequences identified to have homologous regions to the original queries but did not exhibit all domains shown in *C. elegans* vitellogenin were excluded from these data. Sequences were uploaded into Molecular Evolutionary Genetics Analysis version ten software (MEGA X) (Kumar *et al*., 2018) and pairwise alignment was performed using the MUSCLE (multiple sequence comparison by log-expectation) function in the software (Edgar, 2004). The cluster method used was UPGMA.

Phylogenetic trees were constructed using the maximum likelihood statistical method in the MEGA X software, using the Jones-Taylor-Thornton substitution model, all sites

including gaps were used. For each tree constructed the test of phylogeny used was the bootstrap method, this was replicated 100 times. Pairwise alignment scores were generated using Jalview (Waterhouse *et al.*, 2009).

Results

The result from the BLAST search indicated that there were homologous regions and genes to *C. elegans* vitellogenin within PPN genomes that have not previously been identified. Two genes from *G. ellingtonae* were predicted and were shown to be closely related to two *G. rostochiensis* vitellogenin genes (Fig. 1). Predicted vitellogenin gene *G. ellingtonae* 016 was shown to be most closely related to *G. rostochiensis* 2537, with a pairwise alignment percentage identity of 98.7%, but only showed 91.96% identity with *G. pallida* 0128. These three sequences differ in amino acid (aa) sequence length, 1963aa for *G. ellingtonae*, 2322aa for *G. rostochiensis* 2537 and 2406aa for *G. pallida* 0128. The predicted vitellogenin gene *G. ellingtonae* 060 was shown to be most closely related to *G. rostochiensis* 06238 with a percentage identity of 95.25% by pairwise alignment.

PROTEIN DOMAINS IN VITELLOGENIN

Predicted vitellogenin protein domains illustrated in Figure 2, determined by Pfam, show the approximate location of each protein domain within the genes. Three different protein domains are exhibited in each gene: Vitellogenin_N (a lipoprotein amino terminal region), vitellogenin open beta-sheet and von Willebrand factor type D domain. These domains consistently appear in the same order for all genes but have a varied approximate location. Figure 2 clearly illustrates that *C. elegans* and *C. briggsae* vitellogenin domains are

present at the similar estimated positions. Similarly, all *Heterodera* vitellogenin genes and *G. ellingtonae* 016 have protein domains in comparable positions to each other. Another set of genes with domains in similar positions include all *G. rostochiensis* vitellogenin genes and *G. pallida* 9459. Two *M. incognita* vitellogenin genes (28653 and 30865) have almost identical domain positions. Finally, many of the *M. javanica* predicted vitellogenin genes are comparable to each other excluding *M. javanica* 29487 and 43787.

PHYLOGENETIC RELATIONSHIPS OF VITELLOGENIN

A phylogenetic tree was constructed to infer the relationships between vitellogenin genes in cyst nematodes, root-knot nematodes and Rhabditida species, *C. elegans* and *C. briggsae* (Fig. 1). The tree was split into two major groups, the PPN species and the *Caenorhabditis* species; within the PPN group this was further divided into the root-knot and cyst nematode species with a small outgroup consisting of *G. pallida* 0564, *H. glycines* 13367 and *H. schachtii* 19700. The cyst nematodes species gave mixed locations for each gene and were not clustered based on their species. As shown in Figure 1, there were three smaller clusters of PPN each containing two, three and five more closely related genes. Two predicted genes from *H. schachtii* genome were included in the dataset, *H. schachtii* 007 was shown to be most closely related to *H. glycines* 23424, whereas *H. schachtii* 19700 was in the outgroup of the cyst nematode cluster.

The root-knot nematode cluster within Figure 1, comprising *M. hapla, M. incognita* and *M. javanica*, was split into two main groups, with a small outgroup including the only *M. hapla* hypothesised vitellogenin gene to possess all three protein domains. Like the cyst

nematodes, different genes from each species were in mixed locations and did not form clusters of individual species.

The divergence in genes between *C. elegans* and a closely related species, *C. briggsae*, was considered. Figure 1 gives *C. elegans vit*-6 and *C. briggsae* 006 as an outgroup of the *Caenorhabditis* cluster, where the remaining genes are split into two groups. Vitellogenin genes from *C. elegans, vit*-1, *vit*-2 and *vit*-3, were more closely related to *C. briggsae* 002 and 16767, whereas *C. elegans vit*-4 and *vit*-5 were more closely related to *C. briggsae* 14203 and 14234.

Discussion

This paper highlights the presence of vitellogenin genes within the genomes of cyst and root-knot nematodes that have not been previously extensively studied. The presence of vitellogenin genes in the genomes of PPN signifies a requirement for yolk provision to the eggs of the developing nematode. A significant proportion of the vitellogenin amino acid sequence is conserved between the cyst nematode, root-knot nematode and *Caenorhabditis* vitellogenin genes; however, areas of divergence remain. Conservation in parts of the amino acid sequence may be due to the functional domains of the proteins encoded, whereas the divergence of the genes may be a result of speciation events from a common ancestor.

Multiple paralogous vitellogenin genes are present within many species of both vertebrates and invertebrates, of differing amino acid composition (Smolenaars *et al.*, 2007). In *C. elegans* these genes encode for different subunits of a larger protein complex, which may allow for varying function of the encoded proteins (Perez & Lehner, 2019); it is possible that multiple genes in PPN are required to form similar complexes. Data presented here

illustrate the relationship between these paralogous genes, with each species having differing numbers of genes. Only one vitellogenin gene from the genome of *M. hapla* was identified that contained all three protein domains, whereas the genome of *M. incognita* has seven genes. This may be due to the relative sizes of their genomes, *M. hapla* genome contains 54 million base pairs (Opperman *et al.*, 2008), whilst the *M. incognita* genome contains 86 million base pairs (Abad *et al.*, 2008).

The genome of *G. ellingtonae* contains two not previously predicted vitellogenin genes, which are both most closely related to two genes from *G. rostochiensis. Globodera ellingtonae* has been shown to have similar behaviour and requirements for hatching and development to *G. rostochiensis*, but shows some molecular diversity (Hesse *et al.*, 2021). Like *G. rostochiensis, G. ellingtonae* is a pathogen of certain cultivars of potato but its pathogenicity has been inconsistent (Zasada *et al*., 2019). The inclusion of this recently discovered potato cyst nematode, *G. ellingtonae*, sought to further evaluate the nematode species against other members of the *Globodera* family.

As vitellogenin plays a key role in the development of the nematode, it could be anticipated that it would have a relatively conserved sequence, especially between closely related species. This has also been illustrated in the comparison of *C. elegans* with *C. briggsae* where sequences were shown to be 85% identical in coding regions (Zucker-Aprison & Blumenthal, 1989); therefore, vitellogenin genes in *C. elegans* and *C. briggsae* are highly conserved. The genome of *C. elegans* contains six vitellogenin genes whereas *C. briggsae* only contains five (Zucker-Aprison & Blumenthal, 1989). Perez & Lehner (2019) suggest that the *vit*-4 gene is missing from *C. briggsae* due to a duplication between *vit*-3 and *vit*-4. Contrary to this, data presented here indicate that there are two genes closely related to *C. elegans vit*-4 and *vit*-5 in *C. briggsae*, but only two genes closely related to *C. elegans vit*-

1, *vit*-2 and *vit*-3. This result suggests that either *vit*-1, *vit*-2 or *vit*-3 may be missing from *C. briggsae,* not *vit*-4 as originally predicted.

Vitellogenin genes have been shown to be upregulated in *G. pallida.* In supplementary data provided by Cotton *et al.* (2014), three *G. pallida* vitellogenin genes were shown to be upregulated in females (at 21, 28 and 35 dpi) compared to the early parasitic life-stage (7 and 14 dpi). The three genes shown to be upregulated are presented here and form part of the main cyst nematode cluster in the phylogenetic analyses. The gene *G. pallida* 0564, which forms part of an outgroup from the PPN cluster, is not shown to be expressed in that study; it is possible that this gene was not identified, or this gene is not expressed. Currently, no vitellogenin genes have been characterised in *G. rostochiensis.*

The genomes of species within the cyst nematode genus *Heterodera* contain vitellogenin genes; this study highlights their presence within the genomes of *H. schachtii* and *H. glycines*. Two *H. schachtii* genes are shown to be closely related to those in *H. glycines*; however, a third *H. glycines* gene appears to be more divergent. This result supports the findings of Singh *et al*. (2020), who generated a phylogenetic consensus tree containing 33 different populations of *Heterodera* spp., finding that *H. schachtii* and *H. glycines* were closely related. In a study that sequences the *H. schachtii* genome, data provided in the supplementary information listed two proteins that were annotated as containing a lipoprotein amino terminal region, von Willebrand factor type D domain and a domain of unknown function, whilst also indicating that they were highly expressed at 24 dpi (Siddique *et al.*, 2021). These two proteins have been incorporated into Figure 1 as HS_19700 and HS_007; establishing that they also have a vitellogenin open beta-sheet domain, which was not previously annotated.

Masler (1999) partially characterised two vitellogenin proteins in *H. glycines* with similar molecular weights (180 and 190 kDa) to the two larger protein subunits of *C. elegans*

(170 kDa yp170A and 170 kDa yp170B). Figure 1 presents three *H. glycines* hypothesised vitellogenin genes and it is possible that these genes encode for the 'major egg proteins' as described (Masler, 1999). The *H. glycines* transcriptome of effectors gave only one predicted vitellogenin gene sequence, provided in the supplementary information (Gardner *et al.*, 2018). Again, it may be possible that not all vitellogenin genes present in the genome are expressed to produce the vitellogenin proteins.

Root-knot nematodes are the most damaging nematode crop pests, so it was important to include some species of this genus. The relationship between 13 vitellogenin genes from root-knot nematode species has been illustrated, including one gene from the genome of *M. hapla,* five from *M. javanica* and seven from the genome of *M. incognita.* Currently, no studies have highlighted the presence of vitellogenin within the genomes of *M. hapla* or *M. javanica*. In previous analysis of the *M. incognita* genome, no vitellogenin encoding genes were identified (Abad *et al.*, 2008); however, in supplementary data provided by Bellafiore *et al*. (2008), five vitellogenin genes were found in the secreted effectors of female pharyngeal glands using mass spectrometry. A similar study of the effecters produced by female pharyngeal glands in *G. rostochiensis* resulted in no vitellogenin genes being detected (Maier *et al.*, 2012). The result from Bellafiore *et al.* (2008) is unexpected, as it is understood from *C. elegans* that vitellogenin genes are primarily expressed in the hermaphrodite intestine and then transported into the oocyte by receptor-mediated endocytosis (Perez & Lehner, 2019). However, it has now been shown that the *C. elegans* hermaphrodite vents vitellogenin as part of the yolk after egg laying ceases, for further nutritional provision for the developing nematode (Kern *et al.*, 2021). It may be possible that root-knot nematodes exhibit a similar behaviour when producing an egg mass, which primarily consists of glycoproteins (Sharon & Spiegel, 1993).

Protein domains present in *C. elegans* vitellogenin appear within the identified vitellogenin genes of cyst and root-knot nematodes. Using Pfam to predict the location of protein domains, *C. elegans, C. briggsae*, cyst and root-knot nematode vitellogenin genes were shown to contain three protein domains: Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor type D domain. In BLAST searches of PPN genomes, some genes contained only one or two of these domains, with a sequence length that was much shorter than other vitellogenin genes within the species; therefore, these results were not included in the analysis. Protein domain Vitellogenin_N is a lipoprotein amino terminal region that is predicted to be involved in lipid transport and is a domain present in the vitellogenin of certain insect, crustacean and nematode species (Smolenaars *et al.*, 2007). Nematode vitellogenin have been shown to contain amphipathic β-strands, which may also function in the transport of additional lipids (Smolenaars *et al.*, 2007). These strands may then form the vitellogenin open beta-sheet domain. Towards the C-terminus of the vitellogenin gene in nematodes, the protein domain von Willebrand type D is present. This domain has shown adhesive properties and has been hypothesised to be the binding site for membrane receptors on the oocyte (Baker, 1988b).

Whilst the presence of vitellogenin genes within certain PPN species has been determined, a complete understanding of their expression and function is yet to be fully elucidated. Each nematode species contains a different number of vitellogenin genes, even between closely related species. There is a possibility that the variance in the number of genes may be due to differences in expression, with some genes not being expressed. Further analysis of the stage- and tissue-specific expression of vitellogenin would help to understand further their presence and role in PPN. Vitellogenin genes in selected cyst and root-knot nematode species contain the same protein domains as those in *C. elegans* so could be hypothesised to exhibit a similar function; however, this requires further investigation.

Masler (1999) has partially characterised vitellogenin proteins in *H. glycines*; this work could be expanded to characterise the proteins in other PPN species. Vitellogenin plays a key role in the transport of lipids and nutritional provision to the developing nematode; if this function could be interrupted this may provide a novel route for control of PPN. Additionally, vitellogenin has been shown to have a role in protection against environmental stress in *C. elegans* (Fischer *et al.*, 2013). This has been demonstrated by the knockdown of *vit*-6 gene by RNA-interference, which increased the nematodes susceptibility to *Photorhabdus luminescens*, reducing its lifespan (Fischer *et al.*, 2013). If vitellogenin inhibited in PPN, this may provide a novel method of control by potentially increasing the nematode's susceptibility to environmental stress.

This study set out to highlight the presence of and predict vitellogenin genes within the genomes of some economically important PPN. The genes encoding the vitellogenin protein in *C. elegans* were found to be homologous to sequences in the genomes of cyst and root-knot nematodes and could be used to infer relationships between the PPN species. Additionally, the protein domains of vitellogenin genes that appear in *C. elegans* were also present in the PPN. This study contributes to understanding of vitellogenin in PPN; however, further investigation needs to be completed to fully characterise the proteins and understand their biological role. By determining the presence of vitellogenin genes within the genomes of some economically important PPN, the relationship between these genes could be inferred and this information could help form a basis for further investigation of vitellogenin in PPN.

Acknowledgements

This research was supported by funding from 1) Norwegian Institute of Bioeconomy Research, 2) K. G. Davies Ltd, 3) European Regional Development Fund, 4) Hertfordshire Local Enterprise Partnership.

References

ABAD, P., GOUZY, J., AURY, J.M., CASTAGNONE-SERENO, P., DANCHIN, E.G., DELEURY, E., PERFUS-BARBEOCH, L., ANTHOUARD, V., ARTIGUENAVE, F., BLOK, V.C. *et al.* (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology* 26, pp. 909-915. DOI: 10.1038/nbt.1482

AFGAN, E., BAKER, D., BATUT, B., VAN DEN BEEK, M., BOUVIER, D., ČECH, M., CHILTON, J., CLEMENTS, D., CORAOR, N., GRÜNING, B.A. *et al.* (2018). The Galaxy platform for accessible, reproducible, and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* 46, pp. 537-544. DOI:10.1093/nar/gky379

ALMENARA, D.P., DE MOURA, J.P., SCARABOTTO, C.P., ZINGALI, R.B. & WINTER, C.E. (2013). The molecular and structural characterization of two vitellogenins from the free-living nematode *Oscheius tipulae*. *PloS One* 8, e53460. DOI: 10.1371/journal.pone.0053460

BAKER, M.E., (1988a). Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase?. *Biochemical Journal* 255, pp. 1057-1060. DOI: 10.1042/bj2551057.

BAKER, M.E., (1988b). Invertebrate vitellogenin is homologous to human von Willebrand factor. *Biochemical Journal* 256, p. 1059. DOI: 10.1042/bj2561059

BATEMAN, A., COIN, L., DURBIN, R., FINN, R.D., HOLLICH, V., GRIFFITHS-JONES, S., KHANNA, A., MARSHALL, M., MOXON, S., SONNHAMMER, E.L. *et al.* (2004). The Pfam protein families database. *Nucleic Acids Research* 32, pp. 138-141. DOI: 10.1093/nar/gkh121

BELLAFIORE, S., SHEN, Z., ROSSO, M.N., ABAD, P., SHIH, P. AND BRIGGS, S.P. (2008). Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLoS Pathogens* 4, e1000192. DOI: 10.1371/journal.ppat.1000192

BOSSINGER, O. & SCHIERENBERG, E. (2003). The use of fluorescent marker dyes for studying intercellular communication in nematode embryos. *International Journal of Developmental Biology* 40, pp. 431-439.

COTTON, J.A., LILLEY, C.J., JONES, L.M., KIKUCHI, T., REID, A.J., THORPE, P., TSAI, I.J., BEASLEY, H., BLOK, V., COCK, P.J. *ET AL.* (2014). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology* 15, pp. 1-17. DOI: 10.1186/gb-2014-15-3-r43

EDGAR, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, pp. 1-19. DOI: 10.1186/1471-2105-5-113

FISCHER, M., REGITZ, C., KULL, R., BOLL, M. & WENZEL, U. (2013). Vitellogenins increase stress resistance of *Caenorhabditis elegans* after *Photorhabdus luminescens* infection depending on the steroid-signaling pathway. *Microbes and Infection* 15, pp. 569-578. DOI: 10.1016/j.micinf.2013.05.002

GARDNER, M., DHROSO, A., JOHNSON, N., DAVIS, E.L., BAUM, T.J., KORKIN, D. & MITCHUM, M.G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode *Heterodera glycines*. *Scientific Reports* 8, pp. 1-15. DOI: 10.1038/s41598-018-20536-5

HAYWARD, A., TAKAHASHI, T., BENDENA, W.G., TOBE, S.S. & HUI, J.H. (2010). Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer

proteins in metazoans. *FEBS Letters* 584, pp. 1273-1278. DOI: 10.1016/j.febslet.2010.02.056

HESSE, C.N., MORENO, I., PARDO, O.A., FUENTES, H.P., GRENIER, E., DANDURAND, L.M. & ZASADA, I.A. (2021). Characterization of Populations from Chile Utilizing Whole Genome Sequencing. *Journal of Nematology* 53, pp. 1-9. DOI: 10.21307/jofnem-2021-088

ISHII, N., GOTO, S. & HARTMAN, P.S. (2002). Protein oxidation during aging of the nematode *Caenorhabditis elegans*. *Free Radical Biology and Medicine* 33, pp. 1021-1025. DOI: 10.1016/S0891-5849(02)00857-2

JONES, J.T., KUMAR, A., PYLYPENKO, L.A., THIRUGNANASAMBANDAM, A., CASTELLI, L., CHAPMAN, S., COCK, P.J., GRENIER, E., LILLEY, C.J., PHILLIPS, M.S. *et al.* (2009). Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology* 10, pp. 815-828. DOI: 10.1111/j.1364-3703.2009.00585.x

JONES, J.T., HAEGEMAN, A., DANCHIN, E.G., GAUR, H.S., HELDER, J., JONES, M.G., KIKUCHI, T., MANZANILLA-LÓPEZ, R., PALOMARES-RIUS, J.E., WESEMAEL, W.M. *et al.* (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology* 14, pp. 946-961. DOI: 10.1111/mpp.12057

KERN, C.C., TOWNSEND, S., SALZMANN, A., RENDELL, N.B., TAYLOR, G.W., COMISEL, R.M., FOUKAS, L.C., BÄHLER, J. & GEMS, D. (2021). *C. elegans* feed yolk to their young in a form of primitive lactation. *Nature Communications* 12, pp. 1-11. DOI: 10.1038/s41467-021-25821-y

KUMAR, S., STECHER, G., LI, M., KNYAZ, C., & TAMURA, K. (2018). MEGAX: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35, p. 1547. DOI: 10.1093/molbev/msy096

KUMAR, Y. & YADAV, B.C. (2020). Plant-parasitic nematodes: Nature's most successful plant parasite. *International Journal of Research and Review* 7, pp. 379-386.

MAIER, T.R., HEWEZI, T., PENG, J. & BAUM, T.J. (2013). Isolation of whole esophageal gland cells from plant-parasitic nematodes for transcriptome analyses and effector identification. *Molecular Plant-Microbe Interactions* 26, pp. 31-35. DOI: 10.1094/MPMI-05-12-0121-FI

MASLER, E.P. (1999). Detection and partial characterization of egg polypeptides from *Heterodera glycines*. *Journal of Nematology* 31, p. 305-311.

MKANDAWIRE, T.T., GRENCIS, R.K., BERRIMAN, M. & DUQUE-CORREA, M.A. (2021). Hatching of parasitic nematode eggs: a crucial step determining infection. *Trends in Parasitology* 38, pp. 174-187. DOI: 10.1016/j.pt.2021.08.008

NICOL, J.M., TURNER, S.J., COYNE, D.L., NIJS, L.D., HOCKLAND, S. & MAAFI, Z.T. (2011). Current nematode threats to world agriculture. In: Jones, J., Gheyson, G. & Fenoll, C. (Eds). *Genomics and Molecular Genetics of Plant-Nematode Interactions.* Dordrecht, Netherlands, Springer, pp. 21-43. DOI: 10.1007/978-94-007-0434-3_2

OPPERMAN, C.H., BIRD, D.M., WILLIAMSON, V.M., ROKHSAR, D.S., BURKE, M., COHN, J., CROMER, J., DIENER, S., GAJAN, J., GRAHAM, S. *et al.* (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences* 105, pp. 14802-14807. DOI: 10.1073/pnas.0805946105

PEREZ, M.F. & LEHNER, B. (2019). Vitellogenins- yolk gene function and regulation in *Caenorhabditis elegans*. *Frontiers in Physiology* 10, p. 1067. DOI: 10.3389/fphys.2019.01067

SHARON, E. & SPIEGEL, Y. (1993). Glycoprotein characterization of the gelatinous matrix in the root-knot nematode *Meloidogyne javanica*. *Journal of Nematology* 25, p.585-589.

SIDDIQUE, S., RADAKOVIC, Z.S., HILTL, C., PELLEGRIN, C., BAUM, T.J., BEASLEY, H., CHITAMBO, O., CHOPRA, D., DANCHIN, E.G., GRENIER, E. *et al.* (2021). The genome and life stage-specific transcriptomes of a plant-parasitic nematode and its host reveal susceptibility genes involved in trans-kingdom synthesis of vitamin B. *bioRxiv*. DOI: 10.1101/2021.10.01.462558

SINGH, P.R., KARSSEN, G., COUVREUR, M. & BERT, W. (2020). Morphological and molecular characterization of n. sp. (Nematoda: Heteroderidae) from Gran Canaria, Canary Islands. *Journal of Nematology* 52, pp. 1-14. DOI: 10.21307/jofnem-2020-098

SMOLENAARS, M.M., MADSEN, O., RODENBURG, K.W. & VAN DER HORST, D.J. (2007). Molecular diversity and evolution of the large lipid transfer protein superfamilys. *Journal of Lipid Research* 48, pp. 489-502. DOI: 10.1194/jlr.R600028-JLR200

SPIETH, J., DENISON, K., ZUCKER, E. & BLUMENTHAL, T., (1985). The nucleotide sequence of a nematode vitellogenin gene. *Nucleic Acids Research* 13, pp. 7129-7138. DOI: 10.1093/nar/13.19.7129

SPIETH, J., NETTLETON, M., ZUCKER-APRISON, E., LEA, K. & BLUMENTHAL, T. (1991). Vitellogenin motifs conserved in nematodes and vertebrates. *Journal of Molecular Evolution* 32, pp. 429-438. DOI: 10.1007/BF02101283

STANKE, M., DIEKHANS, M., BAERTSCH, R. & HAUSSLER, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, pp. 637-644. DOI: 10.1093/bioinformatics/btn013

STANKE, M. & MORGENSTERN, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* 33, pp. 465-467. DOI: 10.1093/nar/gki458

VIEIRA, P., EVES-VAN DEN AKKER, S., VERMA, R., WANTOCH, S., EISENBACK, J.D. & KAMO, K. (2015). The *Pratylenchus penetrans* transcriptome as a source for the development of alternative control strategies: mining for putative genes involved in parasitism and evaluation of in planta RNAi. *PloS One* 10, e0144674. DOI: 10.1371/journal.pone.0144674

WATERHOUSE, A.M., PROCTER, J.B., MARTIN, D.M., CLAMP, M. & BARTON, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, pp. 1189-1191. DOI: 10.1093/bioinformatics/btp033

WINTER, C.E., PENHA, C. & BLUMENTHAL, T. (1996). Comparison of a vitellogenin gene between two distantly related rhabditid nematode species. *Molecular Biology and Evolution* 13, pp. 674-684. DOI: 10.1093/oxfordjournals.molbev.a025628

ZASADA, I. A., INGHAM, R. E., BAKER, H., & PHILLIPS, W. S. (2019). Impact of *Globodera ellingtonae* on yield of potato (*Solanum tuberosum*). *Journal of Nematology* 51. DOI: 10.21307/JOFNEM-2019-073

ZUCKER-APRISON, E. & BLUMENTHAL, T. (1989). Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. *Journal of Molecular Evolution* 28, pp. 487-496. DOI: 10.1007/BF02602929

Comparative phylogenetic analysis of vitellogenin in species of cyst and root-knot nematodes

Running title: Phylogenetic analysis of vitellogenin

Summary - Plant-parasitic nematodes (PPN) are an economically important group of crop pests and are oviparous animals; all nutrients required to develop and ensure the survival of their unhatched progeny need to be deposited within the egg, including proteins. The most abundant protein deposited is vitellin, formed of a precursor protein vitellogenin, which has roles in transporting lipids, providing amino acids and influencing post-embryonic development. The genes encoding vitellogenin have been well studied in *Caenorhabditis elegans*, but little is known about vitellogenin in PPN. Using the vitellogenin gene sequences from *C. elegans*, homologous sequences in the genomes of cyst and root-knot nematodes were identified and hypothetical vitellogenin genes were predicted. Protein domains were then determined. Sequences were aligned using MUSCLE and then used to construct phylogenetic trees using the maximum likelihood method. With the availability of genomic data and use of online local alignment tools, the vitellogenin encoding genes from *C. elegans* could be aligned to sequences from PPN genomes. All predicted genes contained the same protein domains as *C. elegans*; Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor domain type D. The constructed phylogenetic tree clustered the species into three characterised groups, root-knot nematodes, cyst nematodes and *Caenorhabditis* species. Vitellogenin genes in *C. elegans* were homologous to sequences within PPN genomes, allowing the hypothetical genes to be determined and the relationships between PPN vitellogenin genes to be inferred, forming a potential basis to understand further the role of vitellogenin in cyst and root-knot nematodes.

Keywords - *Caenorhabditis briggsae*, *Caenorhabditis elegans*, development, egg, *Globodera*, *Heterodera*, *Meloidogyne*, plant-parasitic nematode, sedentary endoparasite, vitellin.

Plant-parasitic nematodes (PPN) have been estimated to cause at least $US 80 billion of damage to crops per year (Nicol *et al.*, 2011). Infections by PPN cause non-specific symptoms such as stunted growth, wilting and yellowing of the leaves, caused by a decrease in function of the roots elicited by the nematode infection, which can lead to a reduced yield of the crop (Kumar & Yadav, 2020). Cyst nematodes (*Globodera* and *Heterodera*) and root-knot nematodes (*Meloidogyne*) are two of the most economically important groups of PPN (Jones *et al.*, 2013) and are found in the order Tylenchida.

Nematodes are oviparous animals; the developing nematode within the egg is surrounded by perivitelline fluid, which contains essential nutrients and protective antioxidants (Mkandawire *et al.*, 2021). For the successful development of the nematode, all nutrients for embryogenesis and further growth need to be deposited into the egg. Nutrients are deposited in the egg as proteins and other macromolecules; the primarily stored proteins are yolk proteins, also known as vitellin (Almenara *et al.*, 2013). Vitellin is formed of a precursor protein vitellogenin, which is a large glyco-lipoprotein complex (Winter *et al.*, 1996), a distant relative of the apoB protein in humans (Baker, 1988a). In *Caenorhabditis elegans*, synthesis of vitellogenin occurs in the hermaphrodite intestine, which is then secreted into the body cavity and taken up by the oocytes by receptor-mediated endocytosis (Perez & Lehner, 2019); the protein is then deposited in yolk granules of the egg (Winter *et al.*, 1996). Deposits of yolk protein remain within the egg after embryogenesis and prior to hatching (Bossinger & Schierenberg, 2003).

Vitellogenin is believed to have the primary function of the transport of lipids and micronutrients (Hayward *et al.*, 2010) and amino acid provision to developing progeny (Winter *et al.*, 1996). Additionally, vitellogenin has also been shown to influence post-embryonic phenotypes, whereby limited supply of the protein can result in smaller sized offspring and sterility in *C. elegans* (Perez & Lehner, 2019). Some studies report that

vitellogenin also serves a protective role against oxidative stress (Ishii *et al.*, 2002) and environmental protection from bacterial infections (Fischer *et al.*, 2013).

In *C. elegans*, vitellogenin proteins consist of four polypeptides: two large, 170 kDa yp170A and 170 kDa yp170B, and two small, 115 kDa yp115 and 88 kDa yp88 (Perez & Lehner, 2019). Six genes encoding vitellogenin have been characterised in *C. elegans* (*vit* 1-6) (Spieth *et al.*, 1985). The peptide encoded by *vit*-6 is proteolytically cleaved prior to entering the oocyte, to form yp115 and yp88 (Spieth *et al.*, 1991).

Vitellogenin proteins and the associated encoding genes have also been characterised in some species of both vertebrates (fish and birds) and invertebrates (insects). *Caenorhabditis elegans* has been the primary model for investigating vitellogenin in nematodes; however, orthologous genes have been found in other free-living nematodes: *Oscheius tipulae* (Almenara *et al.*, 2013) and *C. briggsase* (Zucker-Aprison & Blumenthal, 1989). Vitellogenin has also been identified in the animal-parasitic nematodes *Toxocara canis* (Zhu *et al*., 2017), *Haemonchus contortus* (Hartman *et al.*, 2001) and *Trichostongylus vitrinus* (Nisbet & Gasser, 2004). In *Heterodera glycines*, vitellogenin has been described and partially characterised as two major egg proteins (Masler, 1999). These proteins constitute over 50% of all proteins present within the egg with molecular weights of 190 kDa and 180 kDa, a comparative molecular weight to vitellogenin genes found in *C. elegans* (Masler, 1999). Additionally, when expressed sequence tag data from female *Globodera pallida* samples were analysed, vitellogenin-like transcripts were found to be most abundant (Jones *et al.*, 2009). Furthermore, the females of *G. pallida* develop from 21 to 35 days post infection (dpi) of the host plant; vitellogenin was shown to be highly expressed at 28 and 35 dpi, due to the deposition of yolk proteins within the oocytes (Cotton *et al.*, 2014). In the transcriptome of the root-lesion nematode, *Pratylenchus penetrans*, vitellogenin was again shown to be highly expressed (Vieira *et al.*, 2015).

Vitellogenin is an important protein in yolk provision for the developing nematode and possibly has a protective role; however, they have not been identified or characterised in many species of cyst and root-knot nematodes. With the increasing availability of genomic data for many of these economically important species, it is now possible to determine presence of vitellogenin genes within their genomes and infer their phylogenetic relationships.

Materials and methods

Data acquisition

Vitellogenin gene sequences from *C. elegans* (*vit*-1, *vit*-2, *vit*-3, *vit*-4, *vit*-5, and *vit*-6) were obtained using wormbase.org (Table S1). These sequences were then used as the query sequences for BLAST searches against the genomes of selected species of cyst nematodes (*Globodera* and *Heterodera* spp.), root-knot nematodes (*Meloidogyne* spp.) and *C. briggsae* on the parasite.wormbase.org server. BLAST searches were completed using tblastn, with a maximum of 100 target sequences, expect threshold of 0.01, and a word size of 2. The scoring parameters used were BLOSUM62 matrix and low complexity regions were filtered. The result gave homologous regions and a corresponding predicted 'overlapping gene' in each genome, these gene sequences were then selected from the database and collated for sequence alignment.

Gene prediction

The genome of *G. ellingtonae* was not available from the server of parasite.wormbase.org; instead NCBI BLAST was used to determine homologous regions within the species genome. Regions of the genome containing homologous sequences from the resulting BLAST searches were selected and uploaded into the Galaxy web platform (Afgan *et al.*, 2018); here the sequences were then inputted into the AUGUSTUS tool (Stanke & Morgenstern, 2005; Stanke *et al*., 2008). *Caenorhabditis elegans* was chosen as the training set model organism, the protein sequence and coding sequence were then predicted using both strands. The predicted vitellogenin genes were then selected by comparing the sequence to homologous regions identified in the BLAST search.

PHYLOGENETIC TREE CONSTRUCTION

All genes, predicted or database derived through BLAST searches, were then uploaded to Pfam (Bateman *et al.*, 2004) to determine theoretical protein domains for each protein sequence. Sequences which contained the same protein domains as *C. elegans* vitellogenin genes were included in the sequence alignment. Sequences identified to have homologous regions to the original queries but did not exhibit all domains shown in *C. elegans* vitellogenin were excluded from these data. Sequences were uploaded into Molecular Evolutionary Genetics Analysis version ten software (MEGA X) (Kumar *et al*., 2018) and pairwise alignment was performed using the MUSCLE (multiple sequence comparison by log-expectation) function in the software (Edgar, 2004). The cluster method used was UPGMA.

Phylogenetic trees were constructed using the maximum likelihood statistical method in the MEGA X software, using the Jones-Taylor-Thornton substitution model, all sites

including gaps were used. For each tree constructed the test of phylogeny used was the bootstrap method, this was replicated 100 times. Pairwise alignment scores were generated using Jalview (Waterhouse *et al.*, 2009).

Results

The result from the BLAST search indicated that there were homologous regions and genes to *C. elegans* vitellogenin within PPN genomes that have not previously been identified. Two genes from *G. ellingtonae* were predicted and were shown to be closely related to two *G. rostochiensis* vitellogenin genes (Fig. 1). Predicted vitellogenin gene *G. ellingtonae* 016 was shown to be most closely related to *G. rostochiensis* 2537, with a pairwise alignment percentage identity of 98.7%, but only showed 91.96% identity with *G. pallida* 0128. These three sequences differ in amino acid (aa) sequence length, 1963aa for *G. ellingtonae*, 2322aa for *G. rostochiensis* 2537 and 2406aa for *G. pallida* 0128. The predicted vitellogenin gene *G. ellingtonae* 060 was shown to be most closely related to *G. rostochiensis* 06238 with a percentage identity of 95.25% by pairwise alignment.

PROTEIN DOMAINS IN VITELLOGENIN

Predicted vitellogenin protein domains illustrated in Figure 2, determined by Pfam, show the approximate location of each protein domain within the genes. Three different protein domains are exhibited in each gene: Vitellogenin_N (a lipoprotein amino terminal region), vitellogenin open beta-sheet and von Willebrand factor type D domain. These domains consistently appear in the same order for all genes but have a varied approximate location. Figure 2 clearly illustrates that *C. elegans* and *C. briggsae* vitellogenin domains are

present at the similar estimated positions. Similarly, all *Heterodera* vitellogenin genes and *G. ellingtonae* 016 have protein domains in comparable positions to each other. Another set of genes with domains in similar positions include all *G. rostochiensis* vitellogenin genes and *G. pallida* 9459. Two *M. incognita* vitellogenin genes (28653 and 30865) have almost identical domain positions. Finally, many of the *M. javanica* predicted vitellogenin genes are comparable to each other excluding *M. javanica* 29487 and 43787.

PHYLOGENETIC RELATIONSHIPS OF VITELLOGENIN

A phylogenetic tree was constructed to infer the relationships between vitellogenin genes in cyst nematodes, root-knot nematodes and Rhabditida species, *C. elegans* and *C. briggsae* (Fig. 1). The tree was split into two major groups, the PPN species and the *Caenorhabditis* species; within the PPN group this was further divided into the root-knot and cyst nematode species with a small outgroup consisting of *G. pallida* 0564, *H. glycines* 13367 and *H. schachtii* 19700. The cyst nematodes species gave mixed locations for each gene and were not clustered based on their species. As shown in Figure 1, there were three smaller clusters of PPN each containing two, three and five more closely related genes. Two predicted genes from *H. schachtii* genome were included in the dataset, *H. schachtii* 007 was shown to be most closely related to *H. glycines* 23424, whereas *H. schachtii* 19700 was in the outgroup of the cyst nematode cluster.

The root-knot nematode cluster within Figure 1, comprising *M. hapla, M. incognita* and *M. javanica*, was split into two main groups, with a small outgroup including the only *M. hapla* hypothesised vitellogenin gene to possess all three protein domains. Like the cyst

nematodes, different genes from each species were in mixed locations and did not form clusters of individual species.

The divergence in genes between *C. elegans* and a closely related species, *C. briggsae*, was considered. Figure 1 gives *C. elegans vit*-6 and *C. briggsae* 006 as an outgroup of the *Caenorhabditis* cluster, where the remaining genes are split into two groups. Vitellogenin genes from *C. elegans, vit*-1, *vit*-2 and *vit*-3, were more closely related to *C. briggsae* 002 and 16767, whereas *C. elegans vit*-4 and *vit*-5 were more closely related to *C. briggsae* 14203 and 14234.

Discussion

This paper highlights the presence of vitellogenin genes within the genomes of cyst and root-knot nematodes that have not been previously extensively studied. The presence of vitellogenin genes in the genomes of PPN signifies a requirement for yolk provision to the eggs of the developing nematode. A significant proportion of the vitellogenin amino acid sequence is conserved between the cyst nematode, root-knot nematode and *Caenorhabditis* vitellogenin genes; however, areas of divergence remain. Conservation in parts of the amino acid sequence may be due to the functional domains of the proteins encoded, whereas the divergence of the genes may be a result of speciation events from a common ancestor.

Multiple paralogous vitellogenin genes are present within many species of both vertebrates and invertebrates, of differing amino acid composition (Smolenaars *et al.*, 2007). In *C. elegans* these genes encode for different subunits of a larger protein complex, which may allow for varying function of the encoded proteins (Perez & Lehner, 2019); it is possible that multiple genes in PPN are required to form similar complexes. Data presented here

illustrate the relationship between these paralogous genes, with each species having differing numbers of genes. Only one vitellogenin gene from the genome of *M. hapla* was identified that contained all three protein domains, whereas the genome of *M. incognita* has seven genes. This may be due to the relative sizes of their genomes, *M. hapla* genome contains 54 million base pairs (Opperman *et al.*, 2008), whilst the *M. incognita* genome contains 86 million base pairs (Abad *et al.*, 2008).

The genome of *G. ellingtonae* contains two not previously predicted vitellogenin genes, which are both most closely related to two genes from *G. rostochiensis. Globodera ellingtonae* has been shown to have similar behaviour and requirements for hatching and development to *G. rostochiensis*, but shows some molecular diversity (Hesse *et al.*, 2021). Like *G. rostochiensis, G. ellingtonae* is a pathogen of certain cultivars of potato but its pathogenicity has been inconsistent (Zasada *et al*., 2019). The inclusion of this recently discovered potato cyst nematode, *G. ellingtonae*, sought to further evaluate the nematode species against other members of the *Globodera* family.

As vitellogenin plays a key role in the development of the nematode, it could be anticipated that it would have a relatively conserved sequence, especially between closely related species. This has also been illustrated in the comparison of *C. elegans* with *C. briggsae* where sequences were shown to be 85% identical in coding regions (Zucker-Aprison & Blumenthal, 1989); therefore, vitellogenin genes in *C. elegans* and *C. briggsae* are highly conserved. The genome of *C. elegans* contains six vitellogenin genes whereas *C. briggsae* only contains five (Zucker-Aprison & Blumenthal, 1989). Perez & Lehner (2019) suggest that the *vit*-4 gene is missing from *C. briggsae* due to a duplication between *vit*-3 and *vit*-4. Contrary to this, data presented here indicate that there are two genes closely related to *C. elegans vit*-4 and *vit*-5 in *C. briggsae*, but only two genes closely related to *C. elegans vit*-

1, *vit*-2 and *vit*-3. This result suggests that either *vit*-1, *vit*-2 or *vit*-3 may be missing from *C. briggsae,* not *vit*-4 as originally predicted.

Vitellogenin genes have been shown to be upregulated in *G. pallida.* In supplementary data provided by Cotton *et al.* (2014), three *G. pallida* vitellogenin genes were shown to be upregulated in females (at 21, 28 and 35 dpi) compared to the early parasitic life-stage (7 and 14 dpi). The three genes shown to be upregulated are presented here and form part of the main cyst nematode cluster in the phylogenetic analyses. The gene *G. pallida* 0564, which forms part of an outgroup from the PPN cluster, is not shown to be expressed in that study; it is possible that this gene was not identified, or this gene is not expressed. Currently, no vitellogenin genes have been characterised in *G. rostochiensis.*

The genomes of species within the cyst nematode genus *Heterodera* contain vitellogenin genes; this study highlights their presence within the genomes of *H. schachtii* and *H. glycines*. Two *H. schachtii* genes are shown to be closely related to those in *H. glycines*; however, a third *H. glycines* gene appears to be more divergent. This result supports the findings of Singh *et al*. (2020), who generated a phylogenetic consensus tree containing 33 different populations of *Heterodera* spp., finding that *H. schachtii* and *H. glycines* were closely related. In a study that sequences the *H. schachtii* genome, data provided in the supplementary information listed two proteins that were annotated as containing a lipoprotein amino terminal region, von Willebrand factor type D domain and a domain of unknown function, whilst also indicating that they were highly expressed at 24 dpi (Siddique *et al.*, 2021). These two proteins have been incorporated into Figure 1 as HS_19700 and HS_007; establishing that they also have a vitellogenin open beta-sheet domain, which was not previously annotated.

Masler (1999) partially characterised two vitellogenin proteins in *H. glycines* with similar molecular weights (180 and 190 kDa) to the two larger protein subunits of *C. elegans*

(170 kDa yp170A and 170 kDa yp170B). Figure 1 presents three *H. glycines* hypothesised vitellogenin genes and it is possible that these genes encode for the 'major egg proteins' as described (Masler, 1999). The *H. glycines* transcriptome of effectors gave only one predicted vitellogenin gene sequence, provided in the supplementary information (Gardner *et al.*, 2018). Again, it may be possible that not all vitellogenin genes present in the genome are expressed to produce the vitellogenin proteins.

Root-knot nematodes are the most damaging nematode crop pests, so it was important to include some species of this genus. The relationship between 13 vitellogenin genes from root-knot nematode species has been illustrated, including one gene from the genome of *M. hapla,* five from *M. javanica* and seven from the genome of *M. incognita.* Currently, no studies have highlighted the presence of vitellogenin within the genomes of *M. hapla* or *M. javanica*. In previous analysis of the *M. incognita* genome, no vitellogenin encoding genes were identified (Abad *et al.*, 2008); however, in supplementary data provided by Bellafiore *et al*. (2008), five vitellogenin genes were found in the secreted effectors of female pharyngeal glands using mass spectrometry. A similar study of the effecters produced by female pharyngeal glands in *G. rostochiensis* resulted in no vitellogenin genes being detected (Maier *et al.*, 2012). The result from Bellafiore *et al.* (2008) is unexpected, as it is understood from *C. elegans* that vitellogenin genes are primarily expressed in the hermaphrodite intestine and then transported into the oocyte by receptor-mediated endocytosis (Perez & Lehner, 2019). However, it has now been shown that the *C. elegans* hermaphrodite vents vitellogenin as part of the yolk after egg laying ceases, for further nutritional provision for the developing nematode (Kern *et al.*, 2021). It may be possible that root-knot nematodes exhibit a similar behaviour when producing an egg mass, which primarily consists of glycoproteins (Sharon & Spiegel, 1993).

Protein domains present in *C. elegans* vitellogenin appear within the identified vitellogenin genes of cyst and root-knot nematodes. Using Pfam to predict the location of protein domains, *C. elegans, C. briggsae,* cyst and root-knot nematode vitellogenin genes were shown to contain three protein domains: Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor type D domain. In BLAST searches of PPN genomes, some genes contained only one or two of these domains, with a sequence length that was much shorter than other vitellogenin genes within the species; therefore, these results were not included in the analysis. Protein domain Vitellogenin_N is a lipoprotein amino terminal region that is predicted to be involved in lipid transport and is a domain present in the vitellogenin of certain insect, crustacean and nematode species (Smolenaars *et al.*, 2007). Nematode vitellogenin have been shown to contain amphipathic β-strands, which may also function in the transport of additional lipids (Smolenaars *et al.*, 2007). These strands may then form the vitellogenin open beta-sheet domain. Towards the C-terminus of the vitellogenin gene in nematodes, the protein domain von Willebrand type D is present. This domain has shown adhesive properties and has been hypothesised to be the binding site for membrane receptors on the oocyte (Baker, 1988b).

Whilst the presence of vitellogenin genes within certain PPN species has been determined, a complete understanding of their expression and function is yet to be fully elucidated. Each nematode species contains a different number of vitellogenin genes, even between closely related species. There is a possibility that the variance in the number of genes may be due to differences in expression, with some genes not being expressed. Further analysis of the stage- and tissue-specific expression of vitellogenin would help to understand further their presence and role in PPN. Vitellogenin genes in selected cyst and root-knot nematode species contain the same protein domains as those in *C. elegans* so could be hypothesised to exhibit a similar function; however, this requires further investigation.

Masler (1999) has partially characterised vitellogenin proteins in *H. glycines*; this work could be expanded to characterise the proteins in other PPN species. Vitellogenin plays a key role in the transport of lipids and nutritional provision to the developing nematode; if this function could be interrupted this may provide a novel route for control of PPN. Additionally, vitellogenin has been shown to have a role in protection against environmental stress in *C. elegans* (Fischer *et al.*, 2013). This has been demonstrated by the knockdown of *vit*-6 gene by RNA-interference, which increased the nematodes susceptibility to *Photorhabdus luminescens*, reducing its lifespan (Fischer *et al.*, 2013). If vitellogenin inhibited in PPN, this may provide a novel method of control by potentially increasing the nematode's susceptibility to environmental stress.

This study set out to highlight the presence of and predict vitellogenin genes within the genomes of some economically important PPN. The genes encoding the vitellogenin protein in *C. elegans* were found to be homologous to sequences in the genomes of cyst and root-knot nematodes and could be used to infer relationships between the PPN species. Additionally, the protein domains of vitellogenin genes that appear in *C. elegans* were also present in the PPN. This study contributes to understanding of vitellogenin in PPN; however, further investigation needs to be completed to fully characterise the proteins and understand their biological role. By determining the presence of vitellogenin genes within the genomes of some economically important PPN, the relationship between these genes could be inferred and this information could help form a basis for further investigation of vitellogenin in PPN.

References

ABAD, P., GOUZY, J., AURY, J.M., CASTAGNONE-SERENO, P., DANCHIN, E.G., DELEURY, E., PERFUS-BARBEOCH, L., ANTHOUARD, V., ARTIGUENAVE, F., BLOK, V.C. *et al.* (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology* 26, pp. 909-915. DOI: 10.1038/nbt.1482

AFGAN, E., BAKER, D., BATUT, B., VAN DEN BEEK, M., BOUVIER, D., ČECH, M., CHILTON, J., CLEMENTS, D., CORAOR, N., GRÜNING, B.A. *et al.* (2018). The Galaxy platform for accessible, reproducible, and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* 46, pp. 537-544. DOI:10.1093/nar/gky379

ALMENARA, D.P., DE MOURA, J.P., SCARABOTTO, C.P., ZINGALI, R.B. & WINTER, C.E. (2013). The molecular and structural characterization of two vitellogenins from the free-living nematode *Oscheius tipulae*. *PloS One* 8, e53460. DOI: 10.1371/journal.pone.0053460

BAKER, M.E., (1988a). Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase?. *Biochemical Journal* 255, pp. 1057-1060. DOI: 10.1042/bj2551057.

BAKER, M.E., (1988b). Invertebrate vitellogenin is homologous to human von Willebrand factor. *Biochemical Journal* 256, p. 1059. DOI: 10.1042/bj2561059

BATEMAN, A., COIN, L., DURBIN, R., FINN, R.D., HOLLICH, V., GRIFFITHS-JONES, S., KHANNA, A., MARSHALL, M., MOXON, S., SONNHAMMER, E.L. *et al.* (2004). The Pfam protein families database. *Nucleic Acids Research* 32, pp. 138-141. DOI: 10.1093/nar/gkh121

BELLAFIORE, S., SHEN, Z., ROSSO, M.N., ABAD, P., SHIH, P. AND BRIGGS, S.P. (2008). Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLoS Pathogens* 4, e1000192. DOI: 10.1371/journal.ppat.1000192

BOSSINGER, O. & SCHIERENBERG, E. (2003). The use of fluorescent marker dyes for studying intercellular communication in nematode embryos. *International Journal of Developmental Biology* 40, pp. 431-439.

COTTON, J.A., LILLEY, C.J., JONES, L.M., KIKUCHI, T., REID, A.J., THORPE, P., TSAI, I.J., BEASLEY, H., BLOK, V., COCK, P.J. *ET AL.* (2014). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology* 15, pp. 1-17. DOI: 10.1186/gb-2014-15-3-r43

EDGAR, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, pp. 1-19. DOI: 10.1186/1471-2105-5-113

FISCHER, M., REGITZ, C., KULL, R., BOLL, M. & WENZEL, U. (2013). Vitellogenins increase stress resistance of *Caenorhabditis elegans* after *Photorhabdus luminescens* infection depending on the steroid-signaling pathway. *Microbes and Infection* 15, pp. 569-578. DOI: 10.1016/j.micinf.2013.05.002

GARDNER, M., DHROSO, A., JOHNSON, N., DAVIS, E.L., BAUM, T.J., KORKIN, D. & MITCHUM, M.G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode *Heterodera glycines*. *Scientific Reports* 8, pp. 1-15. DOI: 10.1038/s41598-018-20536-5

HAYWARD, A., TAKAHASHI, T., BENDENA, W.G., TOBE, S.S. & HUI, J.H. (2010). Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer proteins in metazoans. *FEBS Letters* 584, pp. 1273-1278. DOI: 10.1016/j.febslet.2010.02.056

HESSE, C.N., MORENO, I., PARDO, O.A., FUENTES, H.P., GRENIER, E., DANDURAND, L.M. & ZASADA, I.A. (2021). Characterization of Populations from Chile Utilizing Whole

Genome Sequencing. *Journal of Nematology* 53, pp. 1-9. DOI: 10.21307/jofnem-2021-088

ISHII, N., GOTO, S. & HARTMAN, P.S. (2002). Protein oxidation during aging of the nematode *Caenorhabditis elegans*. *Free Radical Biology and Medicine* 33, pp. 1021-1025. DOI: 10.1016/S0891-5849(02)00857-2

JONES, J.T., KUMAR, A., PYLYPENKO, L.A., THIRUGNANASAMBANDAM, A., CASTELLI, L., CHAPMAN, S., COCK, P.J., GRENIER, E., LILLEY, C.J., PHILLIPS, M.S. *et al.* (2009). Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology* 10, pp. 815-828. DOI: 10.1111/j.1364-3703.2009.00585.x

JONES, J.T., HAEGEMAN, A., DANCHIN, E.G., GAUR, H.S., HELDER, J., JONES, M.G., KIKUCHI, T., MANZANILLA-LÓPEZ, R., PALOMARES-RIUS, J.E., WESEMAEL, W.M. *et al.* (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology* 14, pp. 946-961. DOI: 10.1111/mpp.12057

KERN, C.C., TOWNSEND, S., SALZMANN, A., RENDELL, N.B., TAYLOR, G.W., COMISEL, R.M., FOUKAS, L.C., BÄHLER, J. & GEMS, D. (2021). *C. elegans* feed yolk to their young in a form of primitive lactation. *Nature Communications* 12, pp. 1-11. DOI: 10.1038/s41467-021-25821-y

KUMAR, S., STECHER, G., LI, M., KNYAZ, C., & TAMURA, K. (2018). MEGAX: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35, p. 1547. DOI: 10.1093/molbev/msy096

KUMAR, Y. & YADAV, B.C. (2020). Plant-parasitic nematodes: Nature's most successful plant parasite. *International Journal of Research and Review* 7, pp. 379-386.

MAIER, T.R., HEWEZI, T., PENG, J. & BAUM, T.J. (2013). Isolation of whole esophageal gland cells from plant-parasitic nematodes for transcriptome analyses and effector identification. *Molecular Plant-Microbe Interactions* 26, pp. 31-35. DOI: 10.1094/MPMI-05-12-0121-FI

MASLER, E.P. (1999). Detection and partial characterization of egg polypeptides from *Heterodera glycines*. *Journal of Nematology* 31, p. 305-311.

MKANDAWIRE, T.T., GRENCIS, R.K., BERRIMAN, M. & DUQUE-CORREA, M.A. (2021). Hatching of parasitic nematode eggs: a crucial step determining infection. *Trends in Parasitology* 38, pp. 174-187. DOI: 10.1016/j.pt.2021.08.008

NICOL, J.M., TURNER, S.J., COYNE, D.L., NIJS, L.D., HOCKLAND, S. & MAAFI, Z.T. (2011). Current nematode threats to world agriculture. In: Jones, J., Gheyson, G. & Fenoll, C. (Eds). *Genomics and Molecular Genetics of Plant-Nematode Interactions.* Dordrecht, Netherlands, Springer, pp. 21-43. DOI: 10.1007/978-94-007-0434-3_2

OPPERMAN, C.H., BIRD, D.M., WILLIAMSON, V.M., ROKHSAR, D.S., BURKE, M., COHN, J., CROMER, J., DIENER, S., GAJAN, J., GRAHAM, S. *et al.* (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences* 105, pp. 14802-14807. DOI: 10.1073/pnas.0805946105

PEREZ, M.F. & LEHNER, B. (2019). Vitellogenins- yolk gene function and regulation in *Caenorhabditis elegans*. *Frontiers in Physiology* 10, p. 1067. DOI: 10.3389/fphys.2019.01067

SHARON, E. & SPIEGEL, Y. (1993). Glycoprotein characterization of the gelatinous matrix in the root-knot nematode *Meloidogyne javanica*. *Journal of Nematology* 25, p.585-589.

SIDDIQUE, S., RADAKOVIC, Z.S., HILTL, C., PELLEGRIN, C., BAUM, T.J., BEASLEY, H., CHITAMBO, O., CHOPRA, D., DANCHIN, E.G., GRENIER, E. *et al.* (2021). The genome

and life stage-specific transcriptomes of a plant-parasitic nematode and its host reveal susceptibility genes involved in trans-kingdom synthesis of vitamin B. *bioRxiv*. DOI: 10.1101/2021.10.01.462558

SINGH, P.R., KARSSEN, G., COUVREUR, M. & BERT, W. (2020). Morphological and molecular characterization of n. sp. (Nematoda: Heteroderidae) from Gran Canaria, Canary Islands. *Journal of Nematology* 52, pp. 1-14. DOI: 10.21307/jofnem-2020-098

SMOLENAARS, M.M., MADSEN, O., RODENBURG, K.W. & VAN DER HORST, D.J. (2007). Molecular diversity and evolution of the large lipid transfer protein superfamilys. *Journal of Lipid Research* 48, pp. 489-502. DOI: 10.1194/jlr.R600028-JLR200

SPIETH, J., DENISON, K., ZUCKER, E. & BLUMENTHAL, T., (1985). The nucleotide sequence of a nematode vitellogenin gene. *Nucleic Acids Research* 13, pp. 7129-7138. DOI: 10.1093/nar/13.19.7129

SPIETH, J., NETTLETON, M., ZUCKER-APRISON, E., LEA, K. & BLUMENTHAL, T. (1991). Vitellogenin motifs conserved in nematodes and vertebrates. *Journal of Molecular Evolution* 32, pp. 429-438. DOI: 10.1007/BF02101283

STANKE, M., DIEKHANS, M., BAERTSCH, R. & HAUSSLER, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, pp. 637-644. DOI: 10.1093/bioinformatics/btn013

STANKE, M. & MORGENSTERN, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* 33, pp. 465-467. DOI: 10.1093/nar/gki458

VIEIRA, P., EVES-VAN DEN AKKER, S., VERMA, R., WANTOCH, S., EISENBACK, J.D. & KAMO, K. (2015). The *Pratylenchus penetrans* transcriptome as a source for the development of alternative control strategies: mining for putative genes involved in parasitism and

evaluation of in planta RNAi. *PloS One* 10, e0144674. DOI: 10.1371/journal.pone.0144674

WATERHOUSE, A.M., PROCTER, J.B., MARTIN, D.M., CLAMP, M. & BARTON, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, pp. 1189-1191. DOI: 10.1093/bioinformatics/btp033

WINTER, C.E., PENHA, C. & BLUMENTHAL, T. (1996). Comparison of a vitellogenin gene between two distantly related rhabditid nematode species. *Molecular Biology and Evolution* 13, pp. 674-684. DOI: 10.1093/oxfordjournals.molbev.a025628

ZASADA, I. A., INGHAM, R. E., BAKER, H., & PHILLIPS, W. S. (2019). Impact of *Globodera ellingtonae* on yield of potato (*Solanum tuberosum*). *Journal of Nematology* 51. DOI: 10.21307/JOFNEM-2019-073

ZUCKER-APRISON, E. & BLUMENTHAL, T. (1989). Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. *Journal of Molecular Evolution* 28, pp. 487-496. DOI: 10.1007/BF02602929

Figure 1. Phylogenetic tree

Click here to access/download;Figure;Figure1. Phylogenetic tree.jpg

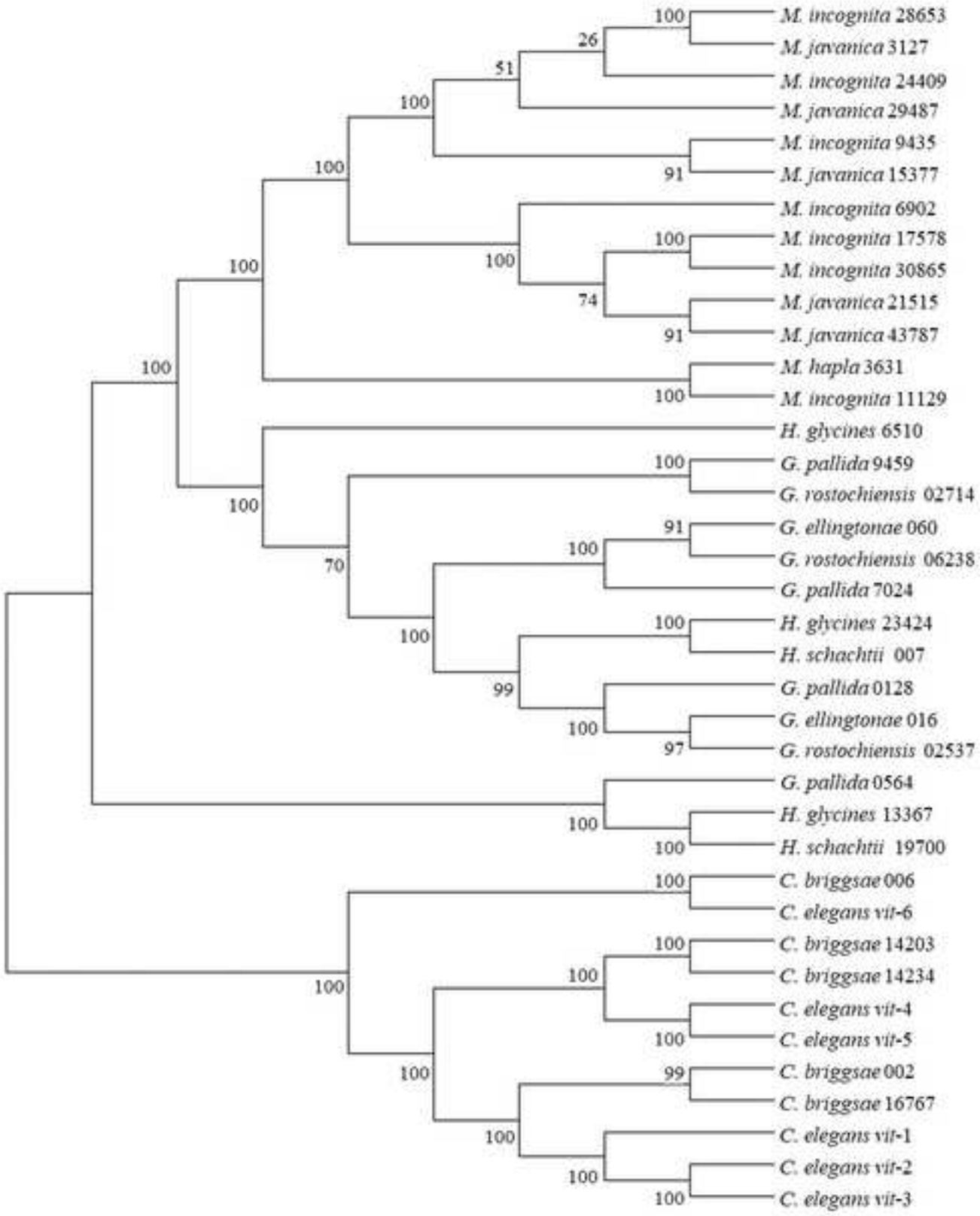

Figure 2. Protein Domain Schematic

Click here to access/download;Figure;Figure2. Protein domans.jpg

Figure legend information

Supplementary material Accession Table

Cover Letter