

Genetic Characterisation of Vitellogenin Egg Proteins in *Heterodera schachtii*

Hannah Joy Wileman, BSc. (Hons), AFHEA

Submitted to the University of Hertfordshire in partial fulfilment of the
requirements of the degree of PhD with Industry Experience

January 2025

Acknowledgements

I would like to thank my funding bodies, Hertfordshire Local Enterprise Partnership, European Regional Development Fund, K. G. Davies Ltd. and the Norwegian Institute of Bioeconomy Research.

I am extremely grateful for the support and guidance provided to me by my primary supervisor Professor Roland Perry and my second supervisor/ Industry supervisor Professor Keith Davies, without their encouragement and expertise this research would not have been possible.

I would also like to express my gratitude towards the technical teams who have supported me throughout my research. I would especially like to thank Diana Francis for all the help provided by both her and her team. Also, I would like to give special thanks to the Geography, Environmental and Planning technical team, particularly Jamie Stone who was always willing to help. I am also thankful to Ramona Hirschfield for all her support through the final stages of my laboratory work.

I am also grateful for my supportive colleagues without whom this experience would have been very different. I would especially like to thank Evren Bingol and Valerija Parthala for your valued friendship throughout this journey. I would also like to thank Emily Brunt and Alexandra Jungeling for their unwavering support and willingness to discuss my research.

I would like to express my deepest gratitude to my family, especially my mum and dad, Judith Wileman and George Wileman, for their continual motivation and reassurance. Finally, I give my sincere thanks to Gage Holding, for his endless encouragement, belief in me and his patience for our life beyond this thesis submission.

Abstract

The beet cyst nematode, *Heterodera schachtii*, is an economically important crop pest. As part of the cyst nematode life cycle, the juvenile nematode develops inside the egg, where it is surrounded by perivitelline fluid. This fluid contains essential nutrients for embryogenesis; one of the most abundant proteins found is vitellin, formed from a precursor protein vitellogenin. In *Caenorhabditis elegans*, vitellogenin has been shown to have a role in nematode development by transporting lipids, providing amino acids, influencing post-embryonic phenotypes, and providing some environmental protection. Vitellogenin has not yet been fully identified and characterised in many species of plant-parasitic nematodes (PPN); some encoding genes have been identified within the genomes of some species, but the protein has only been partially characterised in *Heterodera glycines*. The objective of this work was to investigate and characterise vitellogenin in the closely related species *H. schachtii* and to elucidate its function in nematode development.

Initially using a bioinformatic approach, the presence of vitellogenin (*vit*) genes within species of cyst, root-knot (RKN) and root-lesion nematodes (RLN) was determined. The *vit* gene sequences from both *Globodera ellingtonae* and *Pratylenchus scribneri* were predicted, as these gene sequences had not previously been identified. The phylogenetic relationship between the PPN and *Caenorhabditis* spp. *vit* genes illustrated that the genes were clustered by their groups (cyst nematode, RKN, RLN or *Caenorhabditis* spp.) but were not clustered by species within these groups. It was shown that there was sequence conservation between the PPN species and both *C. elegans* and *C. briggsae*. All predicted genes contained the same protein domains as *C. elegans*: Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor domain type D. Also, the gene ontology annotation of many of the PPN *vit* genes corresponded to the *C. elegans vit* genes of lipid transport and nutrient reserve.

The gene expression of PPN *vit* genes was investigated by utilising expression data available in the public domain. It was shown that *vit* genes were most highly expressed at the adult female stage of the nematode's lifecycle and this expression was significantly greater than in earlier stages, such as the second-stage juvenile (J2) and the egg. In contrast to previous research, it was also shown that *vit* genes were also expressed at low levels in J2 and male stages of *H. schachtii*.

Proteins from *H. schachtii* were characterised using western blot analysis. An antibody developed to recognise a sequence within a vitellogenin gene was synthesised, and then validated by using dot blot analysis, demonstrating that the antibodies recognised proteins within cyst homogenates. The molecular weight of proteins likely to be vitellogenin were then obtained by western blot analysis. The molecular weight of these proteins in cyst homogenates

of *H. schachtii* were 146 kDa, 95 kDa, 89 kDa, 47 kDa, 34 kDa and 19 kDa. Also, the molecular weight of the *G. pallida* proteins obtained from cyst homogenates were 157 kDa, 107 kDa, 81 kDa, 55 kDa, 52 kDa and 46 kDa. Finally, the five proteins from *G. rostochiensis* cyst homogenates had molecular weights of 110 kDa, 82 kDa, 77 kDa, 47 kDa and 13 kDa. The antibody designed to recognise vitellogenin, detected proteins in the females, eggs, and J2 of *H. schachtii*, with proteins recognised that were specific to each stage. Furthermore, in preliminary investigations using the anti-vitellogenin antibody to immunolocalisation the proteins in *H. schachtii*, *G. pallida* and *G. rostochiensis*, proteins were detected at various stages of the nematode life cycle, including the female, egg and J2.

The genomic DNA sequence of the region of the *H. schachtii* containing the *vit* gene, Hsc_gene_6726 was determined using molecular methods of polymerase chain reaction, cloning and whole plasmid sequencing. Novel primers were developed to the region of the *H. schachtii* genome containing the *vit* gene. These primers were then used to amplify the sequence from DNA extracted from *H. schachtii* J2. This DNA sequence was inserted into a plasmid vector and sequenced by Oxford nanopore sequencing. The sequence was then analysed to obtain a partial vitellogenin genomic sequence for this *H. schachtii* population.

Antibodies have previously been developed with specificity to cyst nematode egg proteins in *G. pallida* and *G. rostochiensis*. Two antibodies tested, MAC 356 and PC320, were confirmed by western blot analysis to be recognising a 34 kDa protein, when using the same cyst homogenate samples. These proteins were isolated and identified using liquid chromatography–mass spectrometry, giving a putative identification of annexin in both species. Therefore, the target of both antibodies is the protein annexin. Immunofluorescent study was again used for the preliminary investigation of the target tissue within nematode samples of these antibodies. The PC320 antibody recognised protein within the eggshell of *G. pallida* and *G. rostochiensis*, whereas fluorescence was observed in the whole egg when using MAC 356. Greater fluorescence was observed in *G. pallida* samples than *G. rostochiensis* when using the MAC 356 antibody.

By determining the putative *vit* genes and their expression within PPN species, a greater understanding of vitellogenin and its role in PPN has been established. The development of an anti-vitellogenin antibody that successfully recognised proteins within cyst homogenates by western blotting, aided in the characterisation of the protein in cyst nematode species. Preliminary immunolocalisation studies have given an insight into the tissue location of vitellogenin in *H. schachtii*, *G. rostochiensis* and *G. pallida*. Furthermore, the determination of the vitellogenin DNA sequence from this population of *H. schachtii* can aid in further molecular study of the vitellogenin genes and proteins. Therefore, this work develops the understanding

of vitellogenin in cyst nematodes, indicating a role in lipid transport and nutrient provision for the development of the nematode within the egg, and forms a basis for further study of the function of vitellogenin in *H. schachtii* and other cyst nematodes.

Contents

Acknowledgements	2
Abstract.....	3
List of Figures	10
List of Tables	18
Abbreviations	21
1 General Introduction.....	23
1.1 Plant-Parasitic Nematodes	23
1.1.1 Morphology	24
1.1.2 Hatching	25
1.1.3 Host Plant Symptoms	25
1.2 Cyst Nematodes	26
1.2.1 Life Cycle	26
1.2.2 <i>Heterodera schachtii</i>	28
1.3 Root-Knot Nematodes	29
1.3.1 Life Cycle	29
1.4 Root-Lesion Nematodes	31
1.4.1 Life Cycle	31
1.5 Control and Detection of Plant-Parasitic Nematodes	32
1.5.1 Conventional Methods	32
1.5.2 Biological Control	34
1.5.3 Diagnosing Plant-Parasitic Nematodes	34
1.6 Plant-Parasitic Nematode Eggs	35
1.6.1 Egg Morphology	35
1.7 Vitellogenin	36
2 General Materials & Methods	40
2.1 Preparation of Biological Materials.....	40
2.1.1 Collection of Cysts	40
2.1.2 Collection of Second-Stage Juveniles	41
2.1.3 Collection of Eggs	42
2.1.4 Plant Inoculation	42
3 Bioinformatic Search of PPN <i>vit</i> Genes	44
3.1 Introduction.....	44
3.2 Materials and Methods	46
3.2.1 Data Acquisition	46

3.2.2	Gene Prediction	46
3.2.3	Phylogenetic Tree Construction	47
3.2.4	Protein Domain Prediction	47
3.3	Results	49
3.3.1	BLAST Search for PPN Vitellogenin.....	49
3.3.2	Genomic Location of <i>vit</i> Genes	58
3.3.3	Sequence and Pairwise Alignment.....	58
3.3.4	Phylogenetic Relationships of Vitellogenin	65
3.3.5	Protein Domains in Vitellogenin	67
3.3.6	Gene Ontology.....	72
3.3.7	Genome Availability	76
	Discussion.....	78
4	Gene Expression of Vitellogenin in PPN.....	83
4.1	Introduction.....	83
4.2	Materials and Methods	85
4.2.1	Data Acquisition	85
4.2.2	Data Analysis	89
4.3	Results	89
4.3.1	Vitellogenin Gene Expression in <i>Caenorhabditis briggsae</i>	89
4.3.2	Vitellogenin Gene Expression in <i>Caenorhabditis elegans</i>	91
4.3.3	Vitellogenin Gene Expression in <i>Globodera pallida</i>	94
4.3.4	Vitellogenin Gene Expression in <i>Globodera rostochiensis</i>	95
4.3.5	Vitellogenin Gene Expression in <i>Heterodera glycines</i>	97
4.3.6	Vitellogenin Gene Expression in <i>Heterodera schachtii</i>	98
4.3.7	Vitellogenin Gene Expression in <i>Meloidogyne incognita</i>	101
4.4	Discussion	103
5	Vitellogenin Protein in Cyst Nematodes	107
5.1	Introduction.....	107
5.2	Materials and Methods	109
5.2.1	Antibody Design and Synthesis.....	109
5.2.2	Protein Extraction.....	110
5.2.3	Protein Quantification.....	110
5.2.4	SDS-PAGE	111
5.2.5	Dot Blot Assay	111
5.2.6	Western Blot Assay	113

5.2.7	Analysis of Polyacrylamide Gels and Membranes.....	114
5.2.8	Protein Identification	114
5.2.9	Data Presentation and Statistical Analysis	115
5.3	Results	115
5.3.1	Comparison of Cyst Nematode Proteins using SDS-PAGE.....	115
5.3.2	Antibody Design and Synthesis.....	119
5.3.3	Dot Blot Assays.....	120
5.3.4	Western Blot Assays	135
5.3.5	Western Blot Analysis of Cyst Nematode Proteins Recognised by Anti-VIT A 138	
5.3.6	Sequencing of Vitellogenin Proteins.....	150
5.4	Discussion	150
6	Test of Antibodies with Specificity to Cyst Nematode Egg Proteins.....	155
6.1	Introduction.....	155
6.2	Materials and Methods	157
6.2.1	Immunoblotting of MAC 356 and PC320	157
6.2.2	Protein Identification	158
6.2.3	Preparation of Biological Material for Immunofluorescence	158
6.2.4	Immunofluorescence of Second-Stage Juveniles.....	159
6.2.5	Immunofluorescence of Cysts and Eggs	159
6.2.6	Visualisation of Samples	160
6.3	Results	161
6.3.1	Dot Blot Assays.....	161
6.3.2	Western Blot Analysis	161
6.3.3	Silver Stain of <i>G. pallida</i> Cysts	165
6.3.4	Protein Identification	166
6.3.5	Immunofluorescence.....	171
6.4	Discussion	187
7	Amplification of Genomic Region Containing <i>vit</i> Genes.....	191
7.1	Introduction.....	191
7.2	Materials & Methods.....	193
7.2.1	DNA Extraction	193
7.2.2	Primer Design	194
7.2.3	Polymerase Chain Reaction.....	195
7.2.4	Agarose Gel Electrophoresis.....	196
7.2.5	Ligation and Transformation.....	197

7.2.6	Colony PCR.....	199
7.2.7	Plasmid Purification	199
7.2.8	DNA Sequencing.....	200
7.3	Results	201
7.3.1	Comparison of DNA Extraction Methods	202
7.3.2	Amplification of Genomic DNA	203
7.3.3	Cloning of Genomic DNA	205
7.3.4	Sequencing of Genomic DNA	209
7.4	Discussion	211
8	General Discussion	214
8.1	PPN Genomes Contain Homologous <i>vit</i> Genes.....	214
8.2	Vitellogenin Genes are Expressed in PPN Species	215
8.3	Vitellogenin Proteins are Likely Present in Females and Eggs of Cyst Nematodes 216	
8.4	Preliminary Immunolocalisation of Vitellogenin Proteins in Cyst Nematodes.....	217
8.5	Amplification of <i>Heterodera schachtii</i> Genomic DNA containing the <i>vit</i> Gene	217
8.6	Antibodies MAC 356 and PC320 Target the Same Protein, Annexin	218
8.7	Future Study Recommendations.....	218
8.8	Conclusion.....	220
	References	221
	Appendices	242

List of Figures

Figure 1.1: Overview of methods of parasitism and feeding habits by different plant-parasitic nematodes; cyst nematodes, root-knot nematodes, migratory endoparasites and migratory ectoparasite (Siddique and Grundler, 2018).....	23
Figure 1.2: <i>Heterodera schachtii</i> second-stage juvenile (×100 magnification). [1] Stylet. [2] Digestive tract. (Photo: H. Wileman.).....	25
Figure 1.3: Diagram of a generalised life cycle of cyst nematodes, adapted from Tylka and Davis (2021).....	27
Figure 1.4: A: Female <i>Heterodera schachtii</i> cyst attached to root. B: Tanned cyst of <i>H. schachtii</i> attached to root. (Photo: H. Wileman.).....	28
Figure 1.5: Diagram of a generalised life cycle of root-knot nematodes, from Rehman <i>et al.</i> (2016).....	30
Figure 1.6: Root-lesion nematode within infected plant roots (Davis and MacGuidwin, 2000).	31
Figure 1.7: Diagram of layers of a nematode eggshell and perivitelline fluid, which surrounds the developing embryo. (From: Perry <i>et al.</i> , 2024, by courtesy of James Price.)..	36
Figure 2.1: <i>Heterodera schachtii</i> cyst nematode collection using Wye washer. Soil samples were placed in the Wye washer with a small amount of water (B), to separate cyst nematodes from larger soil debris. The Rotameter (A) was then set to 7.5 L/ min. The outflow ran through two sieves (850 µm and 250 µm mesh). The cyst nematodes were collected in the 250 µm sieve, filtrate was then processed as described in Section 2.1.1.....	41
Figure 2.2: Photograph of growth pouch containing four cabbage plants ‘Offenfam 2’, which have been inoculated with <i>Heterodera schachtii</i> . The black arrow indicates a cyst visible on the plant root.....	43
Figure 3.1: Phylogenetic tree constructed using Maximum-Likelihood method displaying the relationships between vitellogenin genes of some plant-parasitic nematode species of the genera <i>Globodera</i> , <i>Heterodera</i> , <i>Meloidogyne</i> and <i>Pratylenchus</i> and the free-living nematodes <i>Caenorhabditis elegans</i> and <i>Caenorhabditis briggsae</i> . RKN: Root-knot nematodes. RLN: Root-lesion nematodes. CN: Cyst nematodes. Bootstrap test: 100 replicates.....	66
Figure 3.2: Schematic summary of domain structures of vitellogenin proteins of species of <i>Heterodera</i> , <i>Globodera</i> , <i>Meloidogyne</i> , <i>Pratylenchus</i> and <i>Caenorhabditis</i> . Orange boxes represent the domains of Vitellogenin_N; blue boxes represent vitellogenin open beta-sheet domains; green boxes represent von Willebrand factor type D domains and, black arrows represent the end of the sequence. The diagram is not drawn to scale.....	68
Figure 4.1: Expression level in transcripts per million (logarithmic scale) of <i>Caenorhabditis briggsae vit</i> genes in both males and pseudofemales. Produced using data available from Thomas <i>et al.</i> (2012).....	90
Figure 4.2: Expression level in transcripts per million (logarithmic scale) of <i>Caenorhabditis briggsae vit</i> genes in mixed embryos, L1 larvae and young adults. Produced using data available from Grün <i>et al.</i> , (2014).....	91
Figure 4.3: Expression level in fragments per kilobase of transcript per million fragments mapped of <i>Caenorhabditis elegans vit</i> genes in embryonic development.	

Expression level was recorded at timed intervals post first-cleavage. Produced using data available from Hashimshony <i>et al.</i> (2014).....	92
Figure 4.4: Expression level in fragments per kilobase of transcript per million fragments mapped (logarithmic scale), of <i>Caenorhabditis elegans vit</i> genes in the embryo, L1, dauer and adults. Produced using data available from Dillman <i>et al.</i> (2015). 93	
Figure 4.5: Expression level in fragments per kilobase of transcript per million fragments mapped (logarithmic scale), of <i>Caenorhabditis elegans vit</i> genes in the embryo, L1-L4 larvae and young adult hermaphrodites. Produced using data available from Grün <i>et al.</i> , 2014.....	94
Figure 4.6: Expression level in transcripts per million (logarithmic scale) of <i>Globodera pallida vit</i> genes at 7, 14-, 21-, 28- and 35-days post infection, adult males, second-stage juveniles, and eggs. Produced using data available from Cotton <i>et al.</i> , 2014.....	95
Figure 4.7: Expression level in transcripts per million (logarithmic scale) of <i>Globodera rostochiensis vit</i> genes under various conditions: dry cysts, hatched second-stage juveniles (J2), hydrated cysts and hydrated cysts soaked in potato root diffusate (PRD) for 15 min, 1 h, 8 h, 24 h, 48 h or 7 days. Produced using data available from Duceppe <i>et al.</i> (2017).....	96
Figure 4.8: Normalised expression (logarithmic scale) of <i>Globodera rostochiensis vit</i> genes in dormant cysts, hydrated eggs, second-stage juveniles (J2) and 14 days post infection (DPI). Produced using data available from Eves-van den Akker <i>et al.</i> , (2016).....	97
Figure 4.9: Expression level in transcripts per million (logarithmic scale) of <i>Heterodera glycines vit</i> genes under different conditions: parasitic juveniles from a compatible or resistant host and pre-parasitic juveniles. Produced using data available from Gardner <i>et al.</i> (2018).....	98
Figure 4.10: Relative expression (logarithmic scale) of <i>Heterodera schachtii vit</i> genes at different stages of the nematode life cycle, including dormant cyst, pre-infective J2, 10- and 48-hours post infection (HPI), females and males 12 days post infection (DPI) and females 24 DPI. Produced using data available from Siddique <i>et al.</i> (2022).....	99
Figure 4.11: Relative expression (logarithmic scale) of <i>Heterodera schachtii</i> genes containing Vitellogenin_N domain, that showed homology to <i>Caenorhabditis elegans vit</i> genes. Relative expression calculated at different stages of the nematode life cycle, including dormant cyst, pre-infective J2, 10- and 48-hours post infection (HPI), females and males 12 days post infection (DPI) and females 24 DPI. Produced using data available from Siddique <i>et al.</i> (2022).....	100
Figure 4.12: Expression level in transcripts per million (logarithmic scale) of <i>Meloidogyne incognita vit</i> genes at different stages of the nematode life cycle, including the egg, pre-parasitic second-stage juveniles (J2), parasitic third and fourth-stage juveniles (J3-J4), adult females and adult males. Produced using data available from Blanc-Mathieu <i>et al.</i> (2017).....	101
Figure 4.13: Expression level in transcripts per million (logarithmic scale) of <i>Meloidogyne incognita vit</i> genes at different stages of the nematode life cycle, including the egg, second-stage juvenile (J2), third-stage juvenile (J3), fourth-stage juvenile (J4) and adult female. Produced using data available from Choi <i>et al.</i> (2017).....	102

- Figure 5.1: SDS-PAGE gel of *Globodera pallida* cyst homogenates, stained by silver staining. The first lane contained the molecular weight standard ladder (M). The lanes marked a, b, and c contained protein extracts from 1, 10 or 20 homogenised *G. pallida* cysts, respectively. Blue arrows and associated number indicate measured band, the molecular weights are given in Table 5.3.....116
- Figure 5.2: SDS-PAGE gel of cyst homogenates of potato cyst nematodes (PCN) and beet cyst nematodes (BCN), stained by silver staining. The first lane contains the molecular weight standard ladder (M). The lane marked PCN contains a protein extract from 10 homogenised *Globodera pallida* cysts and the lane marked BCN contains protein extracted from five homogenised *Heterodera schachtii* cysts. Blue arrows indicate bands in common between samples, whereas yellow arrows indicate bands only appearing in one sample.....117
- Figure 5.3: SDS-PAGE gel to compare concentration of cyst nematode protein extracts required for detectable protein by Coomassie staining. SDS-PAGE gel contains protein extracts from cysts of *Heterodera schachtii* (BCN) and *Globodera pallida* (PCN), stained by SimplyBlue™ safe stain, a Coomassie stain. The first lane contains the molecular weight standard ladder (M). Samples loaded into lanes A and D had a concentration of 800 µg/ mL, B and E at 400 µg/ mL and C and F at 200 µg/ mL.....118
- Figure 5.4: Sequence alignment of *vit* genes *Globodera pallida* 0128, *G. rostochiensis* 2537 and *Heterodera schachtii* 007 from *G. pallida*, *G. rostochiensis* and *H. schachtii*, with amino acid sequence of synthetic peptides highlighted. Synthetic peptide A was made from the sequence at amino acid positions 775-796; whereas synthetic peptide B was made from the sequence at amino acid positions 1688-1702. Black dotted boxes indicate the position of these sequences in the alignment.....119
- Figure 5.5: Dot blot assays of dilution series using primary antibodies produced against synthetic peptides from *vit* gene sequences. a: schematic drawing of dot blot assay presented in b, c and d, indicating sample position and dilution from 1:250 to 1:2000, from a 1 mg/ mL stock. Images b-d: images of dot blot assay results at different concentrations of goat anti-rabbit secondary antibody, at either 1:250 (b), 1:500 (c) or 1:1000 (d). A1 and A2: antibody A produced from two different rabbits. B1 and B2: antibody B produced from two different rabbits. A1-0, A2-0, B1-0 and B2-0: pre-immune serum from day zero rabbit bleeds. Each inoculation onto PVDF membrane was performed in duplicate and was visualised using DAB substrate.....120
- Figure 5.6: Area under the curve of mean dot intensity in dot blot assay, of dilution series using primary antibodies produced against synthetic peptides from *vit* gene sequences. A1 and A2: antibody one (A) produced from two different rabbits. B1 and B2: antibody two (B) produced from two different rabbits. A1-0, A2-0, B1-0 and B2-0: pre-immune serum from day zero rabbit bleeds. Incubated in goat anti-rabbit secondary antibody at 1:250 (a), 1:500 (b) or 1:1000 (c) (n=2).....123
- Figure 5.7: Schematic drawing of dot blot assays presented in Figure 5.5, indicating sample position of serial dilution using synthetic peptides A and B, from 1:125 to 1:2000.....124
- Figure 5.8: Dot blot assays of dilution series of synthetic peptides (A & B) from the *vit* gene sequences, on PVDF membrane. Incubated in anti-vitellogenin primary antibody A. a: A1 at 1:1000, b: A1 at 1:2000, c: A2 at 1:1000, and d: A2 at 1:2000, from

- rabbit day 35 serum. Incubated in goat anti-rabbit secondary antibody at 1:250, visualised using DAB substrate.....125
- Figure 5.9: Dot blot assays of dilution series of synthetic peptides (A and B) from *vit* gene sequences, on PVDF membrane. Incubated in anti-vitellogenin primary antibody B, a: B1 at 1:1000, b: B1 at 1:2000, c: B2 at 1:1000 and d: B2 at 1:2000, from rabbit day 35 serum. Incubated in goat anti-rabbit secondary antibody at 1:250, visualised using DAB substrate.....125
- Figure 5.10: Area under the curve of mean dot intensity from dot blot assay of dilution series using synthetic peptides (A + B) from *vit* gene sequences, from 1:125 to 1:2000. Membranes were incubated in either anti-vitellogenin A or anti-vitellogenin B, each produced in duplicate from day 35 bleeds from two rabbits per antibody, at 1:1000 or 1:2000; or incubated in pre-immune serum from bleeds at Day 0. Samples were then incubated in anti-rabbit secondary antibody at 1:250 (n=2).....128
- Figure 5.11: Schematic drawing of dot blot assays presented in Figure 5.12. Layout (a) was used in dot blots a, c and e; whereas layout (b) was used in dot blots b, d and f. Schematic drawings indicate sample position of serial dilution using synthetic peptide A (PA), protein extracts of *Heterodera schachtii* cysts (BCN) and *Globodera pallida* cysts (PCN), from 1:125 to 1:2000.....129
- Figure 5.12: Dot blot assays of dilution series of the synthetic peptide A (PA), protein homogenates of *Heterodera schachtii* cysts (BCN) or *Globodera pallida* cysts (PCN), protein concentration adjusted to 150 µg/ mL, as neat and diluted to 1:100. The nitrocellulose membrane was incubated in anti-vitellogenin primary antibodies, from rabbit day 70 serum. a) A1 at 1:250; b) A2 at 1:250; c) A1 at 1:500; d) A2 at 1:500; e) A1 at 1:1000; f) A2 at 1:1000. The membrane was also incubated in anti-rabbit secondary antibody at 1:250, visualised using 1-Step™ Ultra TMB-Blotting Solution.....132
- Figure 5.13: Area under the curve of mean dot intensity from dot blot assay of dilution series (neat to 1:100) using synthetic peptide A, protein homogenates adjusted to 150 µg/ mL of *Heterodera schachtii* cysts (BCN) or *Globodera pallida* cysts (PCN) (Figure 5.12), each dot was performed in duplicate. These samples were incubated in anti-vitellogenin primary antibodies from the day 70 serum from two rabbits, 'A1' and 'A2'. a) A1 at 1:250; b) A2 at 1:250; c) A1 at 1:500; d) A2 at 1:500; e) A1 at 1:1000; f) A2 at 1:1000. Samples were incubated in anti-rabbit secondary antibody at 1:250. Error bars represent standard error of the mean (n=2).....133
- Figure 5.14: Western blot analysis of dilution series of synthetic peptide A. Proteins were transferred onto a PVDF membrane and incubated in anti-VIT antibodies, from bleeds at day 35 (a and b) or from day 70 (c and d). Replicate A1: a and c; replicate A2: b and d. Membranes were then incubated in goat anti-rabbit secondary antibody at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; A: 1 mg/ mL; B: 0.5 mg/ mL; C: 0.25 mg/ mL; D: 0.15 mg/ mL; E: 0.0625 mg/ mL; F: 0.0313 mg/ mL; G: 0.0156 mg/ mL; H: 0.0078 mg/ mL.....136
- Figure 5.15: Area under the curve of band intensity in western blot analysis of dilution series of synthetic peptide A. Incubated in anti-vitellogenin antibodies (A1 or A2) at either day 35 or 70 of production, with a primary antibody concentration of 1:250 and incubated in goat anti-rabbit secondary antibody at 1:250 (n=1).....137

- Figure 5.16 & 5.17: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Protein concentrations were adjusted to 200 µg/ mL and transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts; GR C: *G. rostochiensis* cysts; HS C: *Heterodera schachtii* cysts; PA: synthetic peptide A (62.6 µg/ mL); *Caenorhabditis elegans* eggs used as negative control (200 µg/ mL). Blue arrows indicate measured bands.....139
- Figure 5.18: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Proteins were transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts (362 µg/ mL); HS C: *Heterodera schachtii* cysts (194 µg/ mL); HS J2: *H. schachtii* second-stage juveniles (30.43 µg/ mL); PA: synthetic peptide A (62.6 µg/ mL); *Caenorhabditis elegans* eggs used as negative control (150 µg/ mL). Blue arrows indicate measured bands.....140
- Figure 5.19: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Proteins were adjusted to 250 µg/ mL, transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts; GR C: *G. rostochiensis* cysts; HS C: *Heterodera schachtii* cysts; HS J2: *H. schachtii* second-stage juveniles; PA: synthetic peptide A (62.6 µg/ mL). Blue arrows indicate measured bands.....141
- Figure 5.20: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Proteins were transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts (79.63 µg/ mL); GP E: *G. pallida* eggs (155.25 µg/ mL); GR C: *G. rostochiensis* cysts (561.5 µg/ mL); GR E: *G. rostochiensis* eggs (214.63 µg/ mL); HS C- *Heterodera schachtii* cysts (219.63 µg/ mL); HS E: *H. schachtii* eggs (59.63 µg/ mL); HS J2- *H. schachtii* second-stage juveniles (concentration below detection limit); PA: synthetic peptide A (62.6 µg/ mL). Blue arrows indicate measured bands.....143
- Figure 5.21: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Proteins were adjusted to 70.5 µg/ mL, before being transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts; GP E: *G. pallida* eggs; GR C: *G. rostochiensis* cysts; GR E: *G. rostochiensis* eggs; HS C- *Heterodera schachtii* cysts; HS E: *H. schachtii* eggs; HS J2- *H. schachtii* second-stage juveniles; PA: synthetic peptide A at 280 µg/ mL. Blue arrows indicate measured bands.....144
- Figure 6.1: Photographs of seven western blot membranes using the MAC 356 monoclonal antibody (a-g). In each image, a protein band has been recognised in the *Globodera pallida* cyst homogenate samples. M: Molecular weight marker.

Measured size of band is indicated by the blue arrow and the molecular weight value above it. In (g), three samples of increasing concentration of protein were used and gave bands in the same position, of the same molecular weight (33 kDa).....	162
Figure 6.2: Photograph of western blot membrane incubated in MAC 356 monoclonal antibody against <i>Globodera pallida</i> (GP), <i>G. rostochiensis</i> (GR) and <i>Heterodera schachtii</i> (HS) cyst homogenate samples. M: Molecular weight ladder. Blue arrows indicate measured bands, each with a molecular weight of 34 kDa.....	163
Figure 6.3: Photograph of western blot analysis using PC320 to recognise proteins in cyst homogenates of <i>Globodera pallida</i> (GP) and <i>Heterodera schachtii</i> (HS). M: Molecular weight marker. Blue arrow indicates measured band.....	164
Figure 6.4: Western blot analysis of <i>Globodera pallida</i> (GP), <i>G. rostochiensis</i> (GR) and <i>Heterodera schachtii</i> (HS) incubated in monospecific polyclonal antibody, PC320. M: Molecular weight marker. Measured bands indicated by blue arrows.....	165
Figure 6.5: Silver stain image of cyst homogenates from <i>Globodera pallida</i> (GP). M: Molecular weight marker. Bands measured with molecular weights between 25 kDa and 35 kDa are indicated with blue arrows and molecular weight value.....	166
Figure 6.6: Immunofluorescent images of <i>Heterodera schachtii</i> female cysts, incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). A brightfield image of cyst shown in (b) and (d), is given in (e).....	172
Figure 6.7: Immunofluorescent images of <i>Heterodera schachtii</i> eggs incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield image of egg shown in (a) and (c), given in (e).....	173
Figure 6.8: Immunofluorescent images of <i>Globodera pallida</i> cysts incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies, at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield image of cysts shown in (b) and (d) is given in (e).....	174
Figure 6.9: Immunofluorescent images of <i>Globodera rostochiensis</i> second-stage juveniles extracted from the egg, incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of the J2 are given in (e) and (f).....	175
Figure 6.10: Immunofluorescent images of <i>Globodera rostochiensis</i> eggs and second-stage juveniles contained within the egg incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of egg are given in (e) and (f).....	176

Figure 6.11: Immunofluorescent images of <i>Globodera pallida</i> (a) and <i>Globodera rostochiensis</i> (b) cysts incubated in MAC 356 antibodies at 1:250 concentration. Samples were then incubated in anti-rat secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of cysts shown are given in (e) and (f).....	178
Figure 6.12: Immunofluorescent images of <i>Globodera pallida</i> (a) and <i>Globodera rostochiensis</i> (b) eggs incubated in MAC 356 antibodies at 1:500 concentration. Samples were then incubated in anti-rat secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of eggs are shown in (e) and (f).....	179
Figure 6.13: Immunofluorescent images of <i>Heterodera schachtii</i> cysts incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-VIT antibodies at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of cysts shown are given in (e) and (f).....	181
Figure 6.14: Immunofluorescent images of <i>Heterodera schachtii</i> eggs incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-VIT antibodies, at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of eggs shown are given in (e) and (f).....	182
Figure 6.15: Immunofluorescent images of <i>Heterodera schachtii</i> second-stage juvenile heads incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-VIT antibodies, at 1:250 concentration, then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d).....	183
Figure 6.16: Immunofluorescent images of <i>Globodera pallida</i> cysts incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-vitellogenin antibodies at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield image of cyst shown in (a) and (c), given in (e).....	184
Figure 6.17: Immunofluorescent images of <i>Globodera rostochiensis</i> cysts incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-vitellogenin antibodies at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of cysts are given in (e) and (f).....	185
Figure 6.18: Immunofluorescent images of <i>Globodera rostochiensis</i> second-stage juvenile manually released from an egg; incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-vitellogenin antibodies at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of second-stage juveniles shown in (e) and (f).....	186

Figure 7.1: Line diagram of <i>Heterodera schachtii</i> genome sequence (accession number: JAHGVF010000007.1) utilised for primer design, containing a putative <i>vit</i> gene, Hsc_gene_6726. Orange line indicates gene location. Primer pairs: A (blue arrow), B (purple arrow) and C (green arrow).....	194
Figure 7.2: Diagram of plasmid construct: pJET 1.2 blunt plasmid DNA (green) with genomic DNA sequence of <i>Heterodera schachtii</i> gene Hsc_gene_6726 (blue). Origin of replication (yellow), ampicillin resistance site (pale green) and sequencing primer sites (purple) are labelled.....	201
Figure 7.3: Comparison of two primers (A and B) for the amplification of <i>Heterodera schachtii</i> vitellogenin containing genomic DNA. M: DNA ladder (kilobase pairs). S1 – S6 represent different lysis conditions during sample preparation, described in section 7.2. Bands excised for further DNA amplification are highlighted (A-D).....	203
Figure 7.4: Amplification of <i>Heterodera schachtii</i> genomic DNA containing the vitellogenin gene encoding region, primers used were specific to this region. Samples 1-6 were obtained from DNA extracted from agarose gel slice of band A in Figure 7.3. M: DNA ladder (kilobase pairs). Bands excised for DNA sequencing are highlighted (E, F and G).....	204
Figure 7.5: Amplification of <i>Heterodera schachtii</i> genomic DNA using primers specific to sequences surrounding the vitellogenin gene encoding region. The DNA sample used in the PCR reaction was obtained from agarose gel slice of band B in Figure 7.3. Samples S1-S3 represent different volumes of PCR product used in gel electrophoresis; 0.5 µL, 1 µL and 2 µL. M: DNA ladder (kilobase pairs). Bands excised for DNA sequencing are highlighted (H and I).....	204
Figure 7.6: Amplification of <i>Heterodera schachtii</i> DNA isolated from agarose gel. Primers used were specific for the <i>H. schachtii</i> genomic DNA containing the vitellogenin gene sequence. M: DNA ladder (kilobase pairs). S1 and S2 were samples of DNA extracted from gel extracts of bands C and D highlighted in Figure 7.3. Bands excised for DNA sequencing are highlighted (J and K).....	205
Figure 7.7: Colony PCR to determine the successful ligation and transformation of <i>Heterodera schachtii</i> genomic DNA into the pJET 1.2 plasmid construct. Primers used were provided with the CloneJET PCR cloning kit. M: DNA ladder (kilobase pairs). Samples J1- J2 represent amplification from two colonies, whereas K1- K5 were five colonies selected from the bacterial transformation.....	206
Figure 7.8: Colony PCR to determine the successful ligation and transformation of <i>Heterodera schachtii</i> genomic DNA into the pJET 1.2 plasmid construct. Primers used were provided with the CloneJET PCR cloning kit. M: DNA ladder (kilobase pairs). Samples H1- H5 and I1- I5 were ten selected colonies from bacterial transformation using the purified DNA product of bands H and I in Figure 7.5, whereas E1- E2 was obtained from a purified DNA product of band E, in Figure 7.4. Sample C: Control DNA product ligated and transformed into plasmid construct.....	209

List of Tables

Table 3.1: Accession numbers of genes with a putative identification of vitellogenin, nematode species and gene descriptions as given in Section 3.3.....	47
Table 3.2: Summary of genes obtained from BLAST searches using <i>Caenorhabditis elegans</i> vitellogenin (<i>vit</i>) genes 1-6 as query sequences against plant-parasitic nematode genomes. Red highlight indicates genes that contain the same protein domains found in <i>C. elegans vit</i> genes and used in subsequent analysis.....	51
Table 3.3: Pairwise alignment of vitellogenin genes from <i>Caenorhabditis briggsae</i> , <i>C. elegans</i> , <i>Globodera pallida</i> , <i>G. rostochiensis</i> , <i>G. ellingtonae</i> , <i>Heterodera glycines</i> , <i>H. schachtii</i> , <i>Meloidogyne hapla</i> , <i>M. incognita</i> , <i>M. javanica</i> and <i>Pratylenchus scribneri</i> . Percentage identity (%) for each pairwise alignment is given; values are highlighted in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%). (1/5).....	60
Table 3.4: Amino acid position of protein domains in putative vitellogenin proteins in species of <i>Caenorhabditis</i> , <i>Globodera</i> , <i>Heterodera</i> , <i>Meloidogyne</i> and <i>Pratylenchus</i> . Domain position indicated by start and end point derived from Pfam (Bateman <i>et al.</i> , 2004), some genes were predicted to have the same domain in multiple positions, indicated by values. Full length of amino acid sequence given by total residues.....	69
Table 3.5: Putative <i>vit</i> genes with associated GO terms and descriptions, retrieved from InterPro using gene sequences.....	73
Table 3.6: Pairwise alignment of vitellogenin genes from three <i>G. pallida</i> genomes. Percentage identity (%) for each pairwise alignment is given; values given in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%).....	77
Table 3.7: Pairwise alignment of vitellogenin genes from three <i>G. rostochiensis</i> genomes. Percentage identity (%) for each pairwise alignment is given; values given in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%).....	78
Table 4.1: Summary table of gene expression studies of <i>Caenorhabditis</i> species and plant-parasitic nematodes that have included several stages of the nematode life cycle. Data source for analysis and data type from normalisation methods is given.....	85
Table 4.2: Summary of key methodology used in gene expression studies of <i>Caenorhabditis elegans</i> , <i>C. briggsae</i> and plant-parasitic nematode species; including the strain or population used, number of replications, RNA extraction method, sequencing method and data normalisation method used.....	87
Table 5.1: Position of peptide sequences of <i>Heterodera schachtii</i> , <i>Globodera rostochiensis</i> and <i>G. pallida</i> from sequence alignment for synthetic peptide design. Peptide synthesised indicated (*).....	109
Table 5.2: Reagents required for casting two gels and 1X sample buffer.....	111
Table 5.3: Approximate molecular weight (kDa) of protein bands measured in Figure 5.1.	116
Table 5.4: Mean dot blot intensity of dilution series using primary antibodies produced against synthetic peptides, designed from <i>vit</i> gene sequences. Each primary antibody (A1, A2, B1, B2) or pre-immune serum (A1-0, A2-0, B1-0 and B2-0) had a concentration of 1:250, 1:500, 1:1000 or 1:2000, diluted from 1 mg/ mL. Each	

primary antibody dilution was then incubated in goat anti-rabbit HRP conjugated secondary antibody at 1:250, 1:500 or 1:1000. Area under the curve of dot intensity of the dot blot assay was then measured using imageJ, area determined in arbitrary units.....	122
Table 5.5: Area under the curve of mean dot intensity of dot blot assays using synthetic peptides (A and B) from <i>vit</i> gene sequences. Samples were incubated in anti-vitellogenin A or anti-vitellogenin B, from Day 35 bleeds of two rabbits per antibody or incubated in pre-immune serum, from day zero bleeds, at 1:1000 or 1:2000. Samples were then incubated in anti-rabbit secondary antibody at 1:250. Area under the curve of dot intensity of the dot blot assays were then measured using imageJ, measured in arbitrary units.....	127
Table 5.6: Area under the curve of mean dot intensity of dot blot assays of dilution series using synthetic peptide A (pA), protein homogenates of <i>Heterodera schachtii</i> cysts and <i>Globodera pallida</i> cysts, adjusted to 150 µg/ mL (Figures 5.12). Samples were incubated in anti-vitellogenin primary antibodies (A1 or A2), from rabbit day 70 serum, at 1:250, 1:500 or 1:1000. Samples were then incubated in anti-rabbit secondary antibody at 1:250. Area under the curve of dot intensity of the dot blot assay was then measured using imageJ.....	134
Table 5.7: Area under the curve of band intensity in western blot analysis of dilution series of synthetic peptide A. Incubated in anti-vitellogenin antibodies (A1 or A2) at either day 35 or 70 of production, at 1:250 primary antibody concentration and 1:250 secondary antibody concentration.....	137
Table 5.8: Summary table of western blots presented in Figure 5.16 - 5.21. Molecular weight (MW) values (kDa) for each band measured from each figure are given. The molecular weight of visualised bands seen in both technical replicates Figures 5.16 and 5.17 are given. Figure 5.18 - Figure 5.21 represent different biological replicates for cyst samples. Molecular weights were calculated for <i>H. schachtii</i> second-stage juvenile samples, different biological replicates are presented in Figure 5.18 and 5.19, the same sample was used in Figures 5.20 and 5.21, the molecular weight of bands visualised in both samples are also given. The molecular weight of proteins from egg samples were derived from one biological replicate. The mean values from samples deriving from different biological replicates were then calculated.....	146
Table 6.1: Peptide sequences identified by MALDI-TOF spectrometry and corresponding protein identified through mascot searches with UniProt accession number.....	166
Table 6.2: Results of UniProt database search of <i>Globodera pallida</i> and <i>G. rostochiensis</i> proteins, restricted to molecular weights of 30-35 kDa. Sequences with significant alignments and associated UniProt accession numbers are recorded for alignments with an E-value less than 10. Alignments with E-values less than 100 are given in supplementary material 6.1.....	167
Table 6.3: Protein identification of 34 kDa protein band from <i>Globodera rostochiensis</i> using liquid chromatography mass spectrometry. The UniProt accession number, description and mass (Da) of the identified protein are given. Significant peptide spectral matches and the number of sequences are shown, the higher value of significant matches indicated likelihood of protein identification. Identifications with the highest significant matches are highlighted in red.....	169
Table 6.4: Mean fluorescence intensity measured by imageJ of cyst nematode samples: <i>Globodera pallida</i> , <i>G. rostochiensis</i> and <i>Heterodera schachtii</i> as shown in Figure 6.6- 6.10. The mean fluorescent value of pre-immune serum or PC320 antibody at a concentration of 1:500, is given in arbitrary units (a.u).....	177

Table 6.5: Mean fluorescence intensity measured by imageJ of cyst nematode samples, <i>Globodera pallida</i> , <i>G. rostochiensis</i> and <i>Heterodera schachtii</i> as given in Figure 6.13- 6.18. The mean fluorescent value of pre-immune serum or anti-VIT A antibody at a concentration of 1:500, is given in arbitrary units (a.u).....	187
Table 7.1: Lysis conditions of samples in DNA extraction method optimisation.....	193
Table 7.2: Primer design for use in PCR reactions to amplify the genomic region of the <i>H. schachtii</i> gene HSC_gene_6726.....	195
Table 7.3: Reaction mixture for ligation reaction of DNA fragments from second-stage juveniles of <i>Heterodera schachtii</i> using the pJET 1.2/blunt cloning vector. Sample E2 is a replicate of sample E.....	198
Table 7.4: pJET1.2 forward and reverse sequencing primers.....	199
Table 7.5: <i>Heterodera schachtii</i> genomic DNA was ligated into a plasmid vector (pJET 1.2). Transformed DH10B Escherichia coli competent cells were inoculated onto lysogeny broth agar supplemented with ampicillin, using different volumes (200 µL, 20 µL and 20 µL of a 1:10 dilution). Total number of bacterial colony forming units were counted. DNA samples were obtained by purified DNA from gel bands in Figure 7.4 and Figure 7.5.....	206
Table 7.6: <i>Heterodera schachtii</i> genomic DNA was ligated into a plasmid vector (pJET 1.2 blunt). Transformed DH10B Escherichia coli competent cells were inoculated onto lysogeny broth agar supplemented with ampicillin, using different volumes (200 µL, 20 µL and 20 µL of a 1:10 dilution). Total number of bacterial colony forming units were counted. DNA samples were obtained by purifying DNA from gel bands in Figure 7.3 and Figure 7.4.....	208
Table 7.7: Pairwise alignment and percentage identity (%) of samples H1, H2, H3, I2 and I5 compared to each other and the reference sequence of the expected cloned region from <i>Heterodera schachtii</i>	211

Abbreviations

AU	Arbitrary units
AHDB	Agriculture and Horticulture Development Board
ANOVA	Analysis of variance
APS	Ammonium persulfate
BCA	Bicinchoninic acid
BCN	Beet cyst nematode
BLAST	Basic Local Alignment Search Tool
CN	Cyst nematode
COI	Cytochrome oxidase I
DAB	3,3'-diaminobenzidine
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DPI	Days post infection
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMBL-EBI	European Bioinformatics Institute- European Molecular Biology Laboratory
FITC	Fluorescein Isothiocyanate
FPKM	Fragments per kilobase of transcript per million mapped reads
GO	Gene ontology
GP	<i>Globodera pallida</i>
GR	<i>Globodera rostochiensis</i>
GST	Glutathione S-transferase
HPI	Hours post infection
HRP	Horseradish peroxidase
HS	<i>Heterodera schachtii</i>
ITS	Internal transcribed spacer
J2	Second-stage juvenile
J3	Third-stage juvenile
J4	Fourth-stage juvenile
LAMP	Loop-mediated isothermal amplification
LB-A	Lysogeny broth agar with ampicillin
LCMS	Liquid chromatography–mass spectrometry
MALDI-TOF	Matrix-assisted laser desorption ionisation–time of flight
MEGA	Molecular Evolutionary Genetics Analysis
MEP	Major egg protein
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NFW	Nuclease free water
PA	Peptide A
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline with Tween 20
PCN	Potato cyst nematode
PCR	Polymerase chain reaction
PPN	Plant-parasitic nematode
PRD	Potato root diffusate
PVDF	Polyvinylidene fluoride

RKN	Root-knot nematode
RLN	Root-lesion nematode
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPKM	Reads per kilobase of transcript per million reads mapped
rRNA	Ribosomal RNA
SCAR	Sequence Characterized Amplified Region
SCN	Soybean cyst nematode
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween 20
TMB	3,3',5,5'-tetramethylbenzidine
TMM	Trimmed mean of M-values
TPM	Transcripts per million
UPGMA	Unweighted pair group method with arithmetic mean
VIT	Vitellogenin
WLB	Worm lysis buffer

1 General Introduction

1.1 Plant-Parasitic Nematodes

Nematodes are one of the most abundant groups of animals in the world and can be found in the soil of many habitats (Blaxter and Koutsovoulos, 2015); they are often grouped based on feeding habits, whether they feed on bacteria, fungi, algae, plants, or animals (Khan and Kim, 2007). Nematodes that feed on living plant matter are known as plant-parasitic nematodes (PPN). They can be found in association with the plants, usually in the top 15-20 centimetres of the soil (Jones and Jones, 1964). It is thought that PPN were first discovered in 1743; however, the symptoms caused would have been noticed by growers well before this date (Maggenti, 1981). PPN penetrate the cell wall of plants using a stylet or odontostyle to feed from the plant. The majority of PPN can be characterised as either ectoparasites or endoparasites; ectoparasites remain in the soil and feed on the surface of the root, whereas endoparasites will invade the root and feed from the tissue within (Sikder and Vestergård, 2020). Once the nematode has invaded the root, a sedentary endoparasite will establish a feeding site and complete its life cycle within the root, whereas a migratory endoparasite will continue to move through the roots (Siddique and Grundler, 2018) (Figure 1.1).

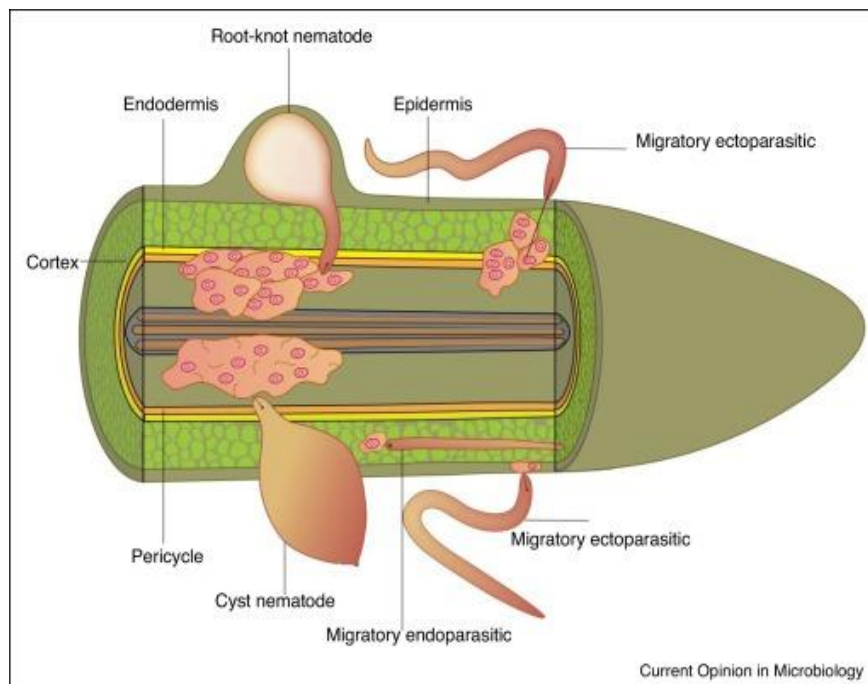


Figure 1.1: Overview of methods of parasitism and feeding habits by different plant-parasitic nematodes; cyst nematodes, root-knot nematodes, migratory endoparasites and migratory ectoparasite (Siddique and Grundler, 2018).

Many of the nematode species existing in soil are free-living nematodes (bacterial feeders), which are involved in the degradation of organic matter. A well-studied free-living nematode is *Caenorhabditis elegans* which has become a model organism for medical studies and has some application in comparative studies of PPN (Bürglin *et al.*, 1998).

PPN cause significant economic costs globally, with crop losses estimated at \$US 80 billion dollars. The most damaging genera contributing to these economic losses include *Meloidogyne* spp., *Heterodera* spp., *Globodera* spp., *Pratylenchus* spp., *Radopholus similis* and *Ditylenchus dipsaci* (Jones *et al.*, 2013). The economic losses caused by PPN may be much higher as damage may not be correctly attributed to them, but instead to insect pests or abiotic factors (De Waele and Elsen, 2008). In the UK, potato cyst nematodes (PCN) are estimated to cause crop losses of £26 million per year (Norshie *et al.*, 2016). In Europe, in the late 1990s, the estimated yield losses caused by the beet-cyst nematode (BCN) were €90 million (Amiri *et al.*, 2002).

The phylum Nematoda is part of the kingdom Animalia, in the 'super'-phylum Ecdysozoa, which includes other moulting animals. Nematodes are highly diverse and are grouped into five clades, with PPN in clades I and IV, as part of the orders Dorylaimida and Tylenchida (Blaxter, 2011). The genera *Globodera* spp., *Heterodera* spp., *Meloidogyne* spp. and *Pratylenchus* spp., are PPN, within the order Tylenchida (Elling, 2006).

1.1.1 Morphology

PPN vary in size dependent on species, ranging from 250 µm to 12 mm, with an average length of 1 mm (Lambert and Bekal, 2002). Nematodes have unsegmented cylindrical worm-like body shape that consists of an external cylinder encasing an internal cylinder, which is the nematode's digestive tract (Figure 1.2). The fluid in-between, described as the pseudo-coelom, contains the nematode's internal organs, which helps maintain pressure and has a circulatory function (Decraemer and Hunt, 2024). PPN have no skeleton or circular muscles, their high internal hydrostatic pressure acts as an antagonist to the longitudinal muscles which enables movement (Decraemer and Hunt, 2024). To allow nematodes to move through moisture in between soil particles, the somatic muscles contract alternately to allow sinusoidal movement through the soil (Niebur and Erdős, 1991). Juvenile nematodes have nervous, secretory-excretory, and digestive systems; and develop a fully functioning reproductive system once matured into adults (Decraemer and Hunt, 2024).

The cuticle of the nematode is highly variable between species but can also vary dependent on the sex or developmental stage (Decraemer and Hunt, 2024). Without a skeleton, the cuticle helps maintain their body shape and provides protection against the environment (Page and Johnstone, 2007). The cuticle is secreted from the epidermis in layers: the epicuticle

(outermost layer) contains glycoproteins and other surface proteins, the further three layers are the cortical, median, and basal zones. Generally, nematodes will moult four times in their life cycle, the cuticle will either be shed or reabsorbed (Decraemer and Hunt, 2024).

A key identifying feature of PPN is the stylet, a hollow mouth spear (Figure 1.2), which at times protrudes out through the mouth region. The stylet is used to penetrate roots and establish feeding. This process can be aided by the secretion of proteins and metabolites through the stylet (Abad and Williamson, 2010). The stylet is connected to the pharynx, which is connected to the intestine and then to the rectum (females) or cloaca (males) (Lambert and Bekal, 2002). The amphidal glands, which are located either side of the nematodes mouth (Perry, 1996) produce secretions which may aid in the parasitism of the plant roots (Curtis, 2007).

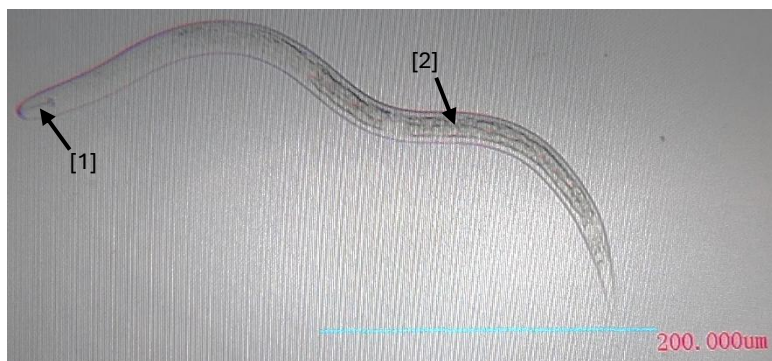


Figure 1.2: *Heterodera schachtii* second-stage juvenile ($\times 100$ magnification). [1] Stylet. [2] Digestive tract. (Photo: H. Wileman.)

1.1.2 Hatching

The juveniles within the eggs of different species of PPN will hatch either in response to a stimulus or under favourable conditions; for example, PCN will hatch in the presence of potato root diffusate, whereas *Meloidogyne javanica*, a root-knot nematode (RKN), will hatch when the temperature is 25-30°C in the presence of oxygen and water (Bird and Bird, 2012). The J2 of *H. schachtii* hatch readily in both water and root diffusate (Moens *et al.*, 2018). Additionally, it has been shown that young soybean crops have an increased stimulatory effect for nematode hatching than older plants (Koenning, 2004).

1.1.3 Host Plant Symptoms

A typical diagnostic feature of cyst nematodes is the presence of cysts on the roots. Generally, plants infected with cyst nematodes are stunted, have yellowing leaves, or wilt (Lilley *et al.*, 2005). RKN can similarly cause stunted growth and the presence of smaller and yellowing leaves. Infected roots can have galls which are typical of this infection, but they may also have smaller bushy roots. Plants that produce tubers, may have small swellings on the surface (Agrios, 2005). Furthermore, root-lesion nematodes (RLN) can cause non-specific above ground symptoms such as stunted growth and yellowing leaves but have typical lesions on

the root surface (Nicol *et al.*, 2011). These lesions make the roots more susceptible to damage from bacteria and fungi (Agrios, 2005).

1.2 Cyst Nematodes

Cyst nematodes are sedentary endoparasites and are obligate parasites of the host they infect (Jones *et al.*, 2013). They tend to be found in temperate regions and are widely distributed in Europe, parts of Asia and the Americas (Agrios, 2005). As a collective, PCN have a broad range of hosts; however, certain species will only have very specific hosts. For example, *Heterodera carotae* will only infect carrots (Nicol *et al.*, 2011), whereas *H. schachtii* can parasitise over 200 different plant species (Amiri *et al.*, 2002). Cyst nematodes are known to cause yield losses in crops including potatoes, sugar beets, cereals and brassicas (Evans and Rowe, 1998). In Europe, PCN have been shown to cause losses of up to 9% of total potato yield and without effective management can result in total yield loss (Subbotin and Hallmann, 2024). The most economically important species of cyst nematodes are in the genera *Heterodera* and *Globodera*; including *G. rostochiensis* and *G. pallida* (PCN), *H. glycines* (soybean cyst nematode), *H. avenae* and *H. filipjevi* (cereal cyst nematodes) and *H. schachtii* (beet cyst nematode) (Subbotin and Hallmann, 2024).

1.2.1 Life Cycle

The life cycle of cyst nematodes can vary from 3-6 weeks (Abad and Williamson, 2010). The life cycle may be paused if there are no host plants present or suitable hatching conditions; unhatched PCN juveniles can survive in a dormant state encased within the cysts in the soil for up to 20 years without a suitable host (Palomares-Rius *et al.*, 2016). The optimal hatching conditions vary for different species, including between closely related species such as *G. rostochiensis* and *G. pallida* whose hatching optima are 20°C and 16°C, respectively (Subbotin and Hallmann, 2024). Temperature can also affect the time to complete a lifecycle and influences the number of generations per year; for example, in temperate regions, cyst nematodes will complete one or two generations per year (Subbotin and Hallmann, 2024). With increased soil temperature the number of generations for some species is greater; for example, *H. schachtii* typically completes two or three generations per year, but in higher temperatures can achieve four (Reuther *et al.*, 2017).

Cyst nematodes have four juvenile stages and a male or female adult stage, between each developmental stage the nematode moults its cuticle (Bird and Bird, 2012). The infective stage of the cyst nematode is the second-stage juvenile (J2), which has a fully formed cuticle and morphology as described in section 1.1.1, excluding a formed reproductive system. The J2 migrates through the soil; this movement can be affected by the pore size, oxygen and water content of the soil (Koenning, 2004). During migration to locate a host root, aided by a concentration gradient of plant diffusates, the J2 relies on its lipid reserves and can only survive approximately 6-11 days prior to finding a host (Robinson *et al.*, 1987). The J2 invades near the end of the root tip (labelled penetrating juvenile in Figure 1.3) and migrates through the cells, which causes necrosis to root tissue from where it entered to the vascular cylinder (Abad and Williamson, 2010). When the nematode inserts its stylet into cells to establish a feeding cell, the protoplast may collapse or be coated in a callose-like material; in this case the nematode will instead find another cell (Jones *et al.*, 2013). Once a suitable feeding cell has been found, the J2 injects secretions through the stylet which break down cell walls between the initial cell and neighbouring cells. Genes encoding for cell wall degrading enzymes and those involved in metabolic pathways are upregulated, to increase nutrients and water to this site (Williamson and Gleason, 2003). These cells form a multinucleate feeding site known as the syncytium; in resistant hosts it will not be formed (Koenning, 2004). During syncytium formation plant defence responses are suppressed (Holbein *et al.*, 2016).

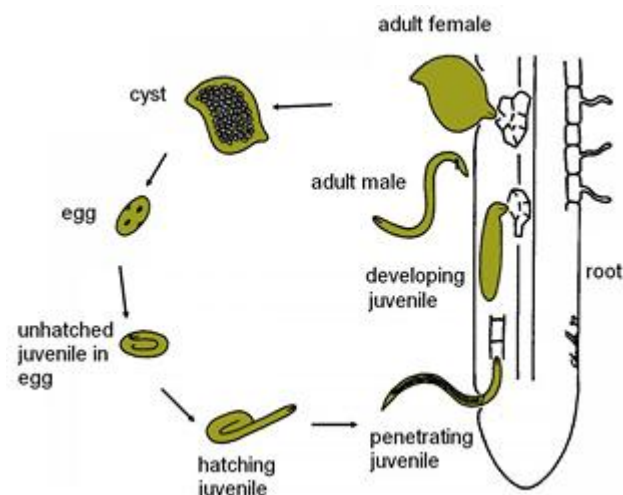


Figure 1.3: Diagram of a generalised life cycle of cyst nematodes, adapted from Tylka and Davis (2021).

The nematode, which is now sedentary, will go through a further two moults of its cuticle. If the feeding site contains abundant nutrients, the nematode will swell and become female (Williamson and Gleason, 2003). Alternatively, if the nutrients are limited due to water stress or overcrowding, the juvenile will become male, as they require fewer nutrients to mature (Koenning, 2004). The male remains vermiform and exits the root to find a female, whereas

as the female increases in size, it ruptures the epidermis of the root and is exposed to the rhizosphere. Cyst nematodes reproduce by sexual reproduction, and after fertilisation the female produces eggs. As the female dies, the cyst tans and the eggs remain encased within the cyst (female body) (Figure 1.4), this protects the eggs from the environment (Subbotin and Hallmann, 2024).

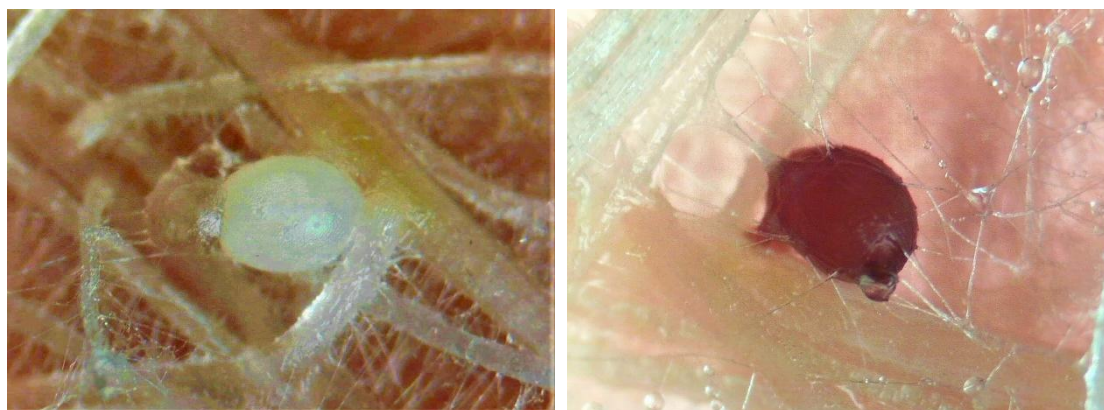


Figure 1.4: A: Female *Heterodera schachtii* cyst attached to root. B: Tanned cyst of *H. schachtii* attached to root. (Photo: H. Wileman.)

1.2.2 *Heterodera schachtii*

Beet cyst nematodes (*Heterodera schachtii*) are a damaging and economically important crop pest, with a wide host range of 218 species, including brassica crops and sugar beet (*Beta vulgaris*) (Daub, 2021). It has been estimated that the total area in European sugar beet production infested by *H. schachtii* is 164,000 ha (Daub, 2021). These nematodes are generally found in soils that are sandy, loamy and organic, and are estimated to be present in greater than 6% of the UK sugar beet growing area (Wright *et al.*, 2022). In the UK, losses to BCN infestation could potentially cause up to £3.5 million losses per year (Wright *et al.*, 2017).

Methods of control for BCN rely on crop rotation with tolerant or resistant crops, as many chemical strategies are now banned, due to associated health and environmental concerns (Reuther *et al.*, 2017). The use of tolerant and resistant varieties of sugar beet can be used to reduce nematode populations, with several tolerant varieties now commercially available with reported good tolerance to BCN. Several resistance genes have been investigated, including *HS1^{pro1}* which was identified from *Patellifolia procumbens*, a wild relative of sugar beet (Jung *et al.*, 1998; Kleine *et al.*, 1995); however, it is not widely used due to poor yields (Wright *et al.*, 2022). Another wild relative to sugar beet, *Beta vulgaris* spp. *maritima*, has been a genetic resource for developing tolerant varieties, utilising the *HsBvm-1* gene (Stevanato *et al.*, 2015).

Mature females of the beet cyst nematode, *H. schachtii*, are typically lemon-shaped and demonstrate polyphenol oxidase tanning of the cuticle producing a light to dark brown colour. Tanning of the mature female cuticle is characteristic of cyst nematodes. The cuticle of *H.*

schachtii also demonstrates folds and ridges that are identifiable to the species (Subbotin and Hallmann, 2024).

1.3 Root-Knot Nematodes

It is believed that the sedentary life cycle of the RKN evolved separately to that of the cyst nematode; however, it is also a sedentary endoparasite (Williamson and Gleason, 2003). There are now nearly 100 different species of RKN described (Wesemael *et al.*, 2024), which are part of the genus *Meloidogyne*, Göldi (1892). These are distributed worldwide in both temperate and tropical regions, predominately found in sandy soils, with a 40- 60% moisture level (Wesemael *et al.*, 2024). The species deemed most important are *M. arenaria*, *M. incognita* and *M. javanica* (tropical), and *M. hapla* (temperate) (Jones *et al.*, 2013). *Meloidogyne hapla* can withstand soil temperatures of less than 10°C, whereas *M. incognita* and *M. javanica* are less adapted to low temperatures (Wesemael *et al.*, 2024). RKN can infect a multitude of plants, infecting nearly all known higher species (Wesemael *et al.*, 2024). These nematodes influence yield by impacting on the plant's ability to obtain water and nutrients, as well as translocating minerals around the plant; the root systems become damaged and the crop becomes less marketable (Kayani *et al.*, 2017). The quality of the crop, especially tubers and tap roots is diminished and causes economic losses (Wesemael *et al.*, 2024).

1.3.1 Life Cycle

Similar to cyst nematodes, the infective stage of the RKN is the J2; the nematode's cuticle will moult four times before becoming an adult nematode. The hatching of J2 can occur without the presence of root diffusate and is instead linked to the soil temperature; however, root diffusates can still provide a hatching stimulus (Wesemael *et al.*, 2024). Hatched RKN migrate through the soil and enter the root near the root cap with the use of their stylets and enzymes (cellulytic and pectolytic) secreted by the pharyngeal glands. Once entering the root, the J2 migrates intercellularly (Wesemael *et al.*, 2024). The nematode migrates to the meristem and goes back to the vascular cylinder and the zone of differentiation, where a feeding cell is established (Williamson and Gleason, 2003) (Figure 1.5). The head of the nematode is then embedded in the vascular tissue feeding on the protoxylem and protophloem (Wesemael *et al.*, 2024). The interaction between the nematode and the plant causes upregulation of auxins and cytokinins, which promote cell growth and cell division (Wesemael *et al.*, 2024). This causes giant cells to be formed in the phloem. These feeding sites initially have five to seven cells combined, which may eventually become 100× larger than normal cells and multinucleate (Williamson and Gleason, 2003), without the development of new cell walls. If the giant cells

are not successfully formed, the J2 will either migrate to an alternative root or starve (Wesemael *et al.*, 2024).

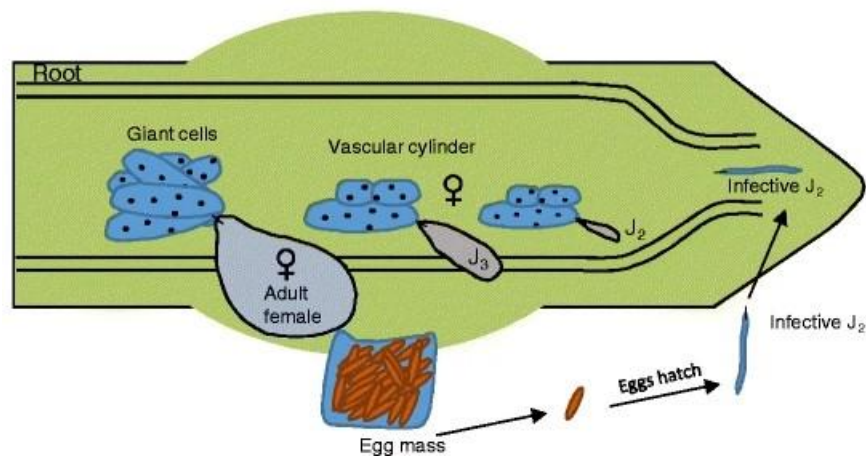


Figure 1.5: Diagram of a generalised life cycle of root-knot nematodes, from Rehman *et al.* (2016).

During the formation of giant cells, root tissue can become hypertrophic, which leads to gall formation around the giant cells (Williamson and Gleason, 2003). Roots become distorted and their function is reduced (Bird *et al.*, 2003). These root galls appear one or two days after the nematode has penetrated the roots; the size of this gall will be dependent on the size of the host plant roots, the number of nematodes infecting the plant and the infecting species (Wesemael *et al.*, 2024).

It is crucial for the J2 to feed prior to moulting as the stylet is not functional through the third and fourth juvenile stages. After the fourth moult both the digestive and reproductive organs are fully developed. If present, the vermiform males appear from the cuticle of the fourth moult and will leave the root, utilising its stylet to exit the root through the root tissue. By contrast, the female continues to feed and becomes more pyriform (Wesemael *et al.*, 2024). RKN males are more abundant under poorer conditions (Wesemael *et al.*, 2024). Some RKN reproduce by sexual reproduction; however, some species such as *M. incognita*, *M. javanica* and *M. arenaria* reproduce by mitotic parthenogenesis (Williamson and Gleason, 2003). Mitotic parthenogenesis is a form of asexual reproduction, where mitosis occurs in the oocytes which maintain a diploid chromosomal number (Wright and Perry, 2006).

The female will lay up to a thousand eggs into a gelatinous matrix on the root surface or within the gall, producing an egg mass (Jones *et al.*, 2013). At this time, the metabolic activity of the giant cells is greatly increased, with increased feeding (Wesemael *et al.*, 2024). If it is a young host plant, the eggs will hatch immediately, whereas if the host plant is an older plant, these eggs will remain dormant; however, these eggs cannot survive as long as eggs of cyst

nematodes. The egg mass will provide some protection to the eggs, dependent on the environmental conditions (Wesemael *et al.*, 2024).

1.4 Root-Lesion Nematodes

Another important group of PPN are the migratory endoparasites, root-lesion nematodes (RLN), which is the common name given to *Pratylenchus* species. Important species of this genus include *Pratylenchus penetrans*, *P. coffeae* and *P. thornei* (Jones *et al.*, 2013). These nematodes can cause significant damage resulting in poor growth and yield losses in a wide host range. Hosts include coarse grains, rice, potatoes, coffee, wheat, fruits and vegetables (Nicol *et al.*, 2011). RLN are distributed worldwide; however, 75% of species have been detected in central and South America, Africa, Asia and Australia (De Waele and Elsen, 2008). RLN are predominately migratory endoparasites (Figure 1.6); however, they have been observed demonstrating ectoparasitic behaviour, feeding on root hairs (Zunke, 1990). Control of RLN through crop rotation is difficult due to the large host range, including crops and weeds (Bélair *et al.*, 2007).

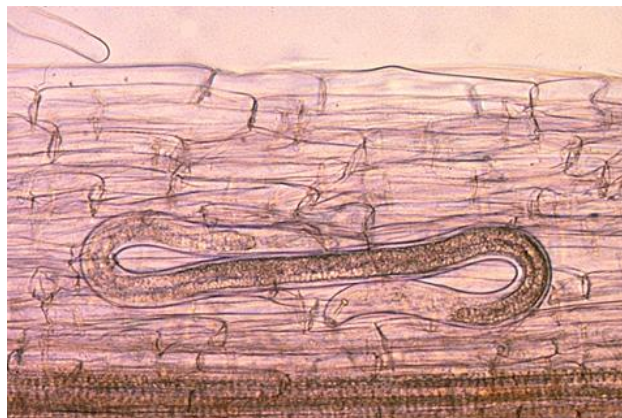


Figure 1.6: Root-lesion nematode within infected plant roots (Davis and MacGuidwin, 2000).

1.4.1 Life Cycle

The length of RLN life cycle is dependent on the *Pratylenchus* species and the environmental conditions; therefore, it can vary from 45 to 65 days (Nicol *et al.*, 2011). In all stages of the RLN life cycle (J2, J3, J4 and adult), the nematode remains vermiform and mobile, excluding the developing egg and the first-stage juvenile, which remains within the egg (Vieira *et al.*, 2018). After hatching the J2 migrates to detect a suitable penetration site and are mostly attracted to the zone of elongation (Archidona-Yuste *et al.*, 2024). Juveniles may perform ectoparasitic behaviours by feeding on the surface of the root prior to entering the root; then use their stylet to penetrate the root epidermis (Linsell *et al.*, 2014).

Once within the root, RLN feed as endoparasites; they do not feed on the endodermis or central stele, instead on root cortex cells (Archidona-Yuste *et al.*, 2024). RLN do not induce a

specialised feeding structure like cyst nematodes or RKN (Vieira *et al.*, 2018). Using its stylet, the nematode punctures the cell walls of the cortical cells and feeds. The nematode secretes enzymes and ingests the cytoplasm of the cell; cell death will occur if this continues for an extended period and the nematode may then enter a resting phase (Archidona-Yuste *et al.*, 2024). Other nematodes become attracted to this damaged cell, which causes further destruction to the root (Zunke, 1990). This attraction to the same point of entry and cells is beneficial for mating in amphimictic species; however, some species reproduce by parthenogenesis. *Pratylenchus penetrans* and *P. coffeae* both have males present; however, no males are present in *P. thornei* or *P. scribneri* (Archidona-Yuste *et al.*, 2024). The female nematode will lay eggs in cavities within the root caused by migration or nearby in the soil (Linsell *et al.*, 2014). RLN can overwinter in roots or in soil; eggs have the most protection against unfavourable environmental conditions (Agrios, 2005).

1.5 Control and Detection of Plant-Parasitic Nematodes

1.5.1 Conventional Methods

Due to the extensive damage certain species of PPN can cause, growers must utilise methods of crop protection to ensure yield and a marketable crop. These methods have previously included or currently include, chemical nematicides, solarisation, crop rotation, trap cropping, resistance genes, biofumigation, or biological control (Poveda *et al.*, 2020). Some of these methods may be used in conjunction with one another as part of an Integrated Pest Management (IPM).

Chemical nematicides were first introduced in 1881 when carbon disulphide was used against *H. schachtii* (Back *et al.*, 2018); by 1982, over \$US 1 billion was being spent on nematicides in the United States. Many nematicides have now been banned and removed from use due to the toxicity associated with humans and to the environment (Bird *et al.*, 2003). Nematicides are broadly classed as fumigants, non-fumigants and seed treatments (Back *et al.*, 2018). Some broad-spectrum nematicides are still used today, such as metam-sodium, a soil fumigant that affects the oxygen availability to cells of PPN (Lewis *et al.*, 2016). This compound has shown to be successful in reducing PPN present in the soil but also has a negative impact on soil microbial populations (Collins *et al.*, 2006).

Solarisation is a method of increasing the temperature of the soil by using clear polythene over wet ground to control PPN, by creating unfavourable conditions (Gill *et al.*, 2017). This method also increases nitrogen in the soil, which is beneficial to plant growth (Back *et al.*, 2018). Solarisation can also be used in weed management; the control of weeds is important as some weeds are hosts of certain species of PPN (Back *et al.*, 2018).

Crop rotation is the practice of growing non-host crops for nematode populations present for a season or several seasons. For example, in cotton growing areas of the USA, peanut is grown to control *M. incognita*, as it is not a host (Starr *et al.*, 2007). Unfortunately, the use of crop rotation is limited for the control of some RLN, as they can be highly polyphagous, with large host ranges (Wesemael *et al.*, 2024). Also, PPN species may have long survival rates within the soil, which can't be controlled using crop rotation (Li *et al.*, 2015). Plant volunteers can also be problematic for the control of PPN, as they are plants that have not been intentionally grown by the grower, such as potatoes or wheat emerging from a previous crop cycle or neighbouring field. Without management of volunteers, the presence of these plants within the field may provide a host for existing PPN (AHDB, n.d.) Trap cropping is used to encourage PPN to hatch without the ability to complete its life cycle. In the control of PCN, this is achieved using plants that are hosts for the nematode to invade the root but do not allow a syncytium to be formed (Back *et al.*, 2018).

Another method in nematode control is the use of biofumigation, which utilises the hydrolysis of glucosinolates into isothiocyanates from certain crops to have nematocidal effects (Brennan *et al.*, 2020). This method of control has been demonstrated using crops within the Brassicaceae family, such as mustard and radish. These crops are grown as cover crops and are incorporated into the soil once fully developed, to release the isothiocyanates (Brennan *et al.*, 2020). Furthermore, *Brassica juncea*, Indian mustard, has shown to be effective against PCN in glasshouse experiments (AHDB, 2019).

Host plants can be described as susceptible, tolerant or resistant to PPN (Blok *et al.*, 2018). Certain crops can be bred to contain resistance genes against PPN, where the plant is able to inhibit production of the feeding site and reproduction (Williamson and Kumar, 2006). By contrast, tolerant species will be parasitised by the nematode, but yield output is not greatly affected (Blok *et al.*, 2018).

Some PPN species, such as *Meloidogyne chitwoodi*, have been described as a quarantined pest, by having the potential to cause economic impacts in areas it is not yet present or being present in areas but has small distribution and is officially controlled (Hockland *et al.*, 2024). In Great Britain, legislation has been introduced to prevent the spread of PPN species and quarantined pests. The Plant Health Act (1967) was introduced for the control of pests and diseases damaging to agricultural or horticultural crops, subsequent amendments have referred to PPNs. This legislation makes provisions for the control of introduction of pests into Great Britain and the control of spread of pests. Phytosanitary certificates are also used to ensure plants or plant material are pest and disease free when imported into Great Britain. If plants or plant material are infected with *G. pallida* or *G. rostochiensis*, this is notifiable to the

Animal and Plant Health Agency (DEFRA and APHA, 2024). Furthermore, certification schemes such as the seed potato classification scheme ensures that these crops have not been grown in areas containing PCN. Prospective fields for seed potato production are sampled to determine the presence of PCN prior to approval (APHA, 2025).

1.5.2 Biological Control

Biological control is the use of living organisms or their metabolites to control a pest. The control of PPN using biological control has been investigated using bacteria, fungi, viruses and predatory nematodes (Poveda *et al.*, 2020). Biological control agents may either parasitise the nematode or compete with it for nutrients (Tian *et al.*, 2007).

Bacteria have the potential as a biocontrol agent for PPN, as certain species of bacteria can parasitise nematodes, producing toxins or enzymes that have nematicidal effects. For example, the bacterial genus *Pasteuria* are Gram positive endospore-forming bacteria that are obligate pathogens of approximately 300 different nematode species (Tian *et al.*, 2007).

Fungi can trap, be endoparasitic, produce toxins against, or compete with, the nematode for nutrients (Li *et al.*, 2015). Nematode-trapping fungi utilise fungal hyphae or larger looping structures to form a trapping mechanism; populations of *M. chitwoodi* were reduced by this method (DiLegge *et al.*, 2019). Endoparasitic fungi can attach and penetrate the cuticle of PPN, as in *Catenaria auxilaris* infections of *H. schachtii* (Kerry, 1980). Finally, fungi can compete with nematodes for nutrients, therefore protecting the plant in a symbiotic relationship.

Predatory nematodes have been investigated for their use as a biocontrol agent since the early 1900s (Poveda *et al.*, 2020). Predatory nematodes reside in the orders Mononchida, Dorylaimida, Diplogasterida and Aphelenchida. Mononchid predators possess a buccal cavity containing teeth or one large tooth, whereas Dorylaimid predators have an odontostyle that punctures their prey. Diplogasterid predators also have teeth, feeding on both nematodes and bacteria. Finally, Aphelenchid predators have a stylet to penetrate the cuticle of the nematode prey (Khan and Kim, 2007).

1.5.3 Diagnosing Plant-Parasitic Nematodes

Accurately and quickly identifying the population of nematodes in the soil can be important for quarantine programmes and determining a suitable strategy for crop protection (Alemu, 2014). The causative agent of a PPN infection can be difficult to determine due to many of the symptoms caused being non-specific and their small size. Many PPN only have very subtle differences in their morphology; therefore, trained personnel are required to identify nematodes accurately by light microscopy (Carneiro *et al.*, 2017). There is now a decreasing number of people trained in this field; subsequently, more reliance on integrating

morphological with molecular techniques or the use of molecular techniques alone is needed (de Oliveira *et al.*, 2011). PPN can be identified by features of the migratory J2, females or males (Ye *et al.*, 2015); for example, perineal patterns of females are used to identify female *Meloidogyne* spp. (Carneiro *et al.*, 2017).

Many molecular techniques for the diagnosis and identification of PPN have now been investigated and some have been adopted for routine use in diagnostic laboratories (Berry *et al.*, 2007). These methods may be qualitative, determining presence or absence, or quantitative, through use of DNA or antibody-based methods (Davies *et al.*, 1996). These techniques are generally less time consuming and more accurate than morphological techniques (Carneiro *et al.*, 2017).

1.6 Plant-Parasitic Nematode Eggs

1.6.1 Egg Morphology

Most PPN eggs have a similar size across different species, they have an ellipsoidal shape with a transparent shell, the average length and width shows inter- and intra-specific variation from 53 μm to 133 μm in length and 17 to 79 μm in width. Eggs are formed of several layers: vitelline, chitinous and lipid or chondroitin proteoglycan layer, the extra-embryonic matrix and the permeability membrane (Figure 1.7); however, some species will have an additional outermost layer, the uterine layer. The vitelline layer is formed from the oolemma and is comprised of lipoproteins (Bird and Bird, 2012); this layer prevents polyspermy (Perry *et al.*, 2024). The chitinous layer contains chitin which is a polymer of repeating units of β -(1,4)-linked N-acetylglucosamine and is often the thickest layer of the eggshell and provides structure to the egg (Wharton, 1980). The thickness of this layer varies for different species and can vary between 50-400 nm (Perry *et al.*, 2024). It was believed that nematodes contained a lipid layer that was impermeable to small molecules and protected the juvenile against chemicals (Perry and Moens, 2011). In some species, this lipid layer was described to be hydrolysed prior to hatching to increase its permeability (Bird and Bird, 2012). It is now suggested that this lipid layer is instead a layer formed of chondroitin proteoglycans, these have been identified in the eggshells of both *C. elegans* (Olson *et al.*, 2012) and *G. rostochiensis* (Price *et al.*, 2023). The extra-embryonic matrix is positioned between the chondroitin proteoglycan layer and the permeability barrier; this layer acts to separate the plasma membrane of the embryo from the

preceding layers of the eggshell. The final layer is a lipid-rich permeability membrane, which protects the developing embryo from the external environment (Perry *et al.*, 2024).

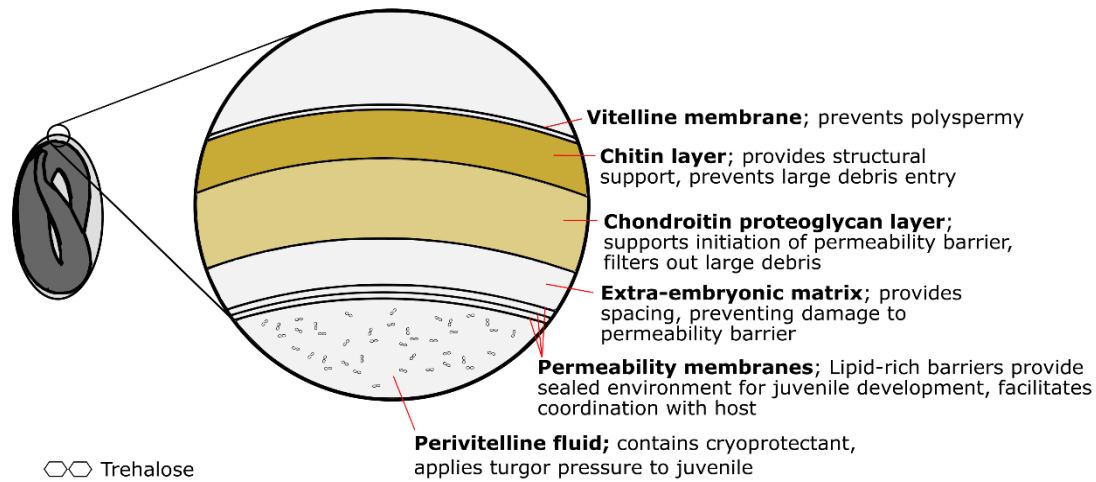


Figure 1.7: Diagram of layers of a nematode eggshell and perivitelline fluid, which surrounds the developing embryo. (From: Perry *et al.*, 2024, by courtesy of James Price.)

Beneath these eggshell layers, the developing embryo and eventually the juvenile is surrounded by the perivitelline fluid; this fluid protects the juvenile (Perry and Moens, 2011). Perivitelline fluid provides a nutritional source, protection and 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, 1988b). Vitellogenin is one of the most abundant proteins within the egg (Sharrock, 1983) and is associated with lipids to form yolk (Grant and Hirsh, 1999). In a proteomic study of lipid droplets in *C. elegans*, an association between vitellogenin and lipid droplets was determined (Vrablik *et al.*, 2015). Yolk is important to the developing embryo for the provision of proteins and lipids (Grant and Hirsh, 1999).

1.7 Vitellogenin

The location of vitellogenin synthesis varies between vertebrate and non-vertebrates. Chicken (*Gallus gallus*) vitellogenin is synthesised in the animal's liver and is regulated by oestrogen prior to being deposited into the bloodstream and taken up by developing oocytes. Within the oocyte, vitellogenin is cleaved into phosvitin and lipovitellin (Wallace, 1985). In *Xenopus laevis*, vitellogenin is also synthesised in the liver and secreted into the bloodstream prior to uptake by oocytes. Whilst in the liver the vitellogenin undergo lipidation, phosphorylation and

glycosylation (Wahli *et al.*, 1981). Conversely, in insects, vitellogenin is abundantly expressed in the fat body and secreted into the haemolymph before being transported by receptor-mediated endocytosis into the developing oocytes (Sappington and Raikhel, 1998).

In *C. elegans*, synthesis of vitellogenin occurs in the hermaphrodite intestine and is then secreted into the body cavity by the classic exocytotic pathway (Zhai *et al.*, 2024), passed through the gonadal basal lamina and the somatic gonad, then finally taken up by the oocytes by receptor-mediated endocytosis (Perez and Lehner, 2019), mediated by RME-2 (Grant and Hirsh, 1999). The protein is then deposited in yolk granules/ yolk organelles of the egg (Winter *et al.*, 1996), in late-stage oocytes and early embryos (Zhai *et al.*, 2024). Whilst proteins are utilised during embryogenesis, deposits of yolk protein remain within the egg after embryogenesis and prior to hatching (Bossinger and Schierenberg, 2003). It has also been shown that yolk proteins are present within the nematode's intestine prior to hatch (Grant and Hirsh, 1999).

Caenorhabditis elegans contains two lipoprotein complexes formed from vitellogenin (A and B) and are made up of four polypeptides: two large, 170 kDa YP170A and 170 kDa YP170B, and two small, 115 kDa YP115 and 88 kDa YP88. Complex A is comprised of YP170A, YP115 and YP88, whereas complex B is a dimer of YP170B (Sharrock *et al.*, 1990). Six genes encoding vitellogenin have been characterised in *C. elegans* (*vit* 1-6) (Spieth *et al.*, 1985). The YP170B subunit is encoded by *vit*-3, *vit*-4 and *vit*-5, whereas YP170A is encoded by *vit*-1 and *vit*-2. The peptide encoded by *vit*-6 is proteolytically cleaved prior to entering the oocyte, to form yp115 and yp88 (Spieth and Blumenthal, 1985).

Vitellogenin is transported into cells by the receptor RME-2. It has been shown when *C. elegans* mutants lack the encoding gene, *rme-2*, vitellogenin accumulates within the body cavity and is not taken up into the oocyte. Additionally, cholesterol remains within the body cavity (Matyash *et al.*, 2001). In further studies of *rme-2* knockdown mutants, it was shown that no yolk was present within the egg and greatly reduced viability (Grant and Hirsh, 1999). The embryonic vitellogenin produced increases with the ageing of the hermaphrodite and influences post-embryonic phenotypes (Perez *et al.*, 2017). The progeny of older mothers are larger, develop more quickly and are more resistant to starvation, than progeny of younger mothers (Perez and Lehner, 2019). Conversely, the greater production of vitellogenin in ageing nematodes can shorten their lifespan (Sornda *et al.*, 2019).

The expression of *vit* genes in *C. elegans* is regulated by several methods. Regulators bind directly to *vit* gene promoters: VPE1, VPE2 and an unnamed promoter. The regulator ELT-2 binds directly to VPE2, whereas the transcription factor MAB-3 binds to the unnamed promoter (Perez and Lehner, 2019). The transcription factor UNC-62 is vital for vitellogenin expression

and is greatly decreased on its inhibition (van Nostrand *et al.*, 2013). The transcription factor MAB-3 confers sex specificity of *vit* gene expression (Yi and Zarkower, 1999). Vitellogenin is also regulated by several signalling pathways: insulin-like growth factor, TGF- β and target of rapamycin signalling. Studies have also suggested that *vit* gene expression is regulated by SKN-1, *mafr-1*, *kri-1*, *lrp-2* and *ceh-60* (Perez and Lehner, 2019).

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 and 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).

Whilst research of yolk proteins in nematodes has primarily been conducted using *C. elegans*, there have been few studies that have focused on vitellogenin in PPN. Masler (1999) investigated proteins that had been deposited into the eggs of *H. glycines*; two proteins were identified to be the most abundant, constituting over 50% of all proteins present within the egg. These proteins were labelled MEP-I and MEP-II, with molecular weights of 190 kDa and 180 kDa, respectively. These proteins may not have been from eggshell origin, as most proteins previously attributed to the *H. glycines* eggshell were of 70 kDa or less (Kennedy *et al.*, 1997). Masler (1999) suggests that these proteins could potentially be vitellin, due to similar molecular weights identified to vitellin proteins found in *C. elegans*. Antibodies were later generated against MEP-I and MEP-II, which only recognised proteins in samples containing females and eggs, but not J2 (Masler, 2002).

Additionally, when expressed sequence tag data from female *G. 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, *P. penetrans*, vitellogenin was again shown to be highly expressed (Vieira *et al.*, 2015). Furthermore, in a proteomic study of *M. hapla*, four vitellogenin proteins were identified in mass spectrometry analysis (Mbeunkui *et al.*, 2010). These proteins were identified in the J2 stage and were hypothesised to be localised to the intestine, if the expression was consistent with *C. elegans*.

Vitellogenin is an important protein in yolk provision and lipid transportation for the developing nematode and possibly has a protective role; however, limited research has focused on the

identification and characterisation in many species of cyst, root-knot and root-lesion nematodes. Therefore, the aims of this work are:

1. Determine the presence of *vit* genes within species of cyst, root-knot and root-lesion nematodes and examine the phylogenetic relationship between the genes.
2. Examine PPN *vit* gene expression and the life-stage specificity of these genes.
3. Determine the molecular weight of vitellogenin proteins in *H. schachtii*.
4. Investigate the use of antibodies raised against egg proteins of cyst nematodes.
5. Perform preliminary investigations to determine the localisation of vitellogenin in cyst nematodes using immunofluorescence.
6. Compare the immunolocalisation of antibodies raised against cyst nematode egg proteins.
7. Determine the vitellogenin DNA sequence of populations of *H. schachtii*.

2 General Materials & Methods

2.1 Preparation of Biological Materials

2.1.1 Collection of Cysts

Heterodera schachtii cysts were retrieved from infected soil, kindly donated by Dr Alistair Wright of British Beet Research Organisation. *Heterodera schachtii* cysts were also obtained from growth pouches inoculated with second-stage juveniles (J2), as described in section 2.1.4. In SDS-PAGE and western blot analysis, *Globodera pallida* and *G. rostochiensis* cyst homogenates were used as a comparison to the *H. schachtii* proteins.

Cysts of *H. schachtii* were collected either using a Wye washer at Rothamsted Research (Figure 2.1) or by using a modified Cobb's (1918) sieving method. Prior to commencing any washing of soil, water was run for ten minutes to remove any residual copper from the pipes. When using the Wye washer, 250 g of a soil sample was placed in the Wye washer, which contained a small amount of water. The washer was then filled so there was a steady stream of water being flushed from the spout through two sieves (850 µm and 250 µm mesh size). The rate of flow of water was then increased to 7.5 L/ min at 20°C. The Wye washer was run until the water appeared clear or after 10 min.

When using the modified Cobb's (1918) sieving method, 50 g increments of soil sample were used. The soil was placed into a 250 mL cylinder, filled with tap water, and inverted five times to mix the soil. The solution was then carefully poured through two sieves (850 µm and 250 µm mesh size). The cylinder was refilled, mixed, and decanted through the sieves twice more, the remaining soil was discarded. This was repeated until 500 g to 1 kg of soil had been processed. For both the Wye washer method and the sieving method, the filtrate collected in the 250 µm sieve was washed into a 50 mL Falcon tube. To separate the cysts from the remaining soil filtrate, the 50 mL filtrate was poured through double lined cheese cloth and observed at 10× magnification under a stereomicroscope. Individual cysts were picked using forceps, counted, and placed in a damp Petri dish.



Figure 2.1: *Heterodera schachtii* cyst nematode collection using Wye washer. Soil samples were placed in the Wye washer with a small amount of water (B), to separate cyst nematodes from larger soil debris. The Rotameter (A) was then set to 7.5 L/ min. The outflow ran through two sieves (850 μ m and 250 μ m mesh). The cyst nematodes were collected in the 250 μ m sieve (C), filtrate was then processed as described in Section 2.1.1.

2.1.2 Collection of Second-Stage Juveniles

In order to collect *H. schachtii* J2 for molecular study, freshly extracted cysts were placed on a hatching tray. The tray consisted of two plastic rings securing a single layered mesh above a glass evaporating dish, which was filled with approximately 50 mL tap water. Cysts were placed on top of the mesh to allow hatched J2 to fall into the water below. The water was collected every 48 h to determine the J2 count. The total number of J2 hatched was determined by decanting the water from the hatching tray into a measuring cylinder and mixing; 2 mL of the hatched nematode solution was placed on a Doncaster plate and counted. This was completed in triplicate and an average count was determined.

If the total number of J2 were to be collected, the nematode solution was allowed to settle in the measuring cylinder for 1 h. Excess water was removed from the cylinder until 10 mL of the nematode solution remained. The total volume was then reduced over a series of

centrifugation steps (16,100 g, 15°C, 2.5 min). The final volume of the J2 suspension obtained was calculated based on the concentration required for the study.

2.1.3 Collection of Eggs

Heterodera schachtii eggs were collected by placing a single cyst in 10 µL of sterile deionised water on a microscope slide and observed at 30× magnification under a stereomicroscope. The cyst was held in place using forceps and cut open using a scalpel. Debris of the cyst was removed, and eggs were washed into a 1.5 mL Eppendorf tube. This process was repeated until sufficient eggs were collected. The water was removed, and the eggs were washed several times in sterile deionised water. The total number of eggs required depended on the quantity needed for the experiment conducted.

2.1.4 Plant Inoculation

The cysts of *H. schachtii* were also collected using growth pouches (Figure 2.2) containing cabbage 'Offenham 2' (Mr Fothergill's vegetable seeds). Four cabbage seeds were germinated at the top and left to grow until the root tips were approximately 5 cm from the bottom of the pouch. Second-stage juveniles were collected as described in Section 2.1.2, to give a final concentration of approximately 20-30 J2 per 5 µL of water. Multiple root tips were inoculated with 5 µL of the solution, filter paper was placed over each point of inoculation to assist the nematodes in entering the roots and was removed after 1 week. The pouches were then stood in a tray containing a 2 cm depth of deionised water within a growth cabinet. The cabinet had a constant temperature of 20°C and 40% relative humidity, with a day/night cycle of 12 h. Cabbage roots were checked every 2 weeks for the emergence of the females. Once the cyst cuticle tanned, the cyst was removed using forceps and used for further study.



Figure 2.2: Photograph of growth pouch containing four cabbage plants 'Offenfam 2', which have been inoculated with *Heterodera schachtii*. The black arrow indicates a cyst visible on the plant root.

3 Bioinformatic Search of PPN *vit* Genes

3.1 Introduction

Vitellogenin (*vit*) genes have primarily been researched in the free-living nematode *Caenorhabditis elegans*. In this species 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 and Lehner, 2019). Six genes encoding vitellogenin have been characterised in *C. elegans* (Spieth *et al.*, 1985, 1991). Five *vit* genes (*vit* 1-5) are located on the X chromosome, whereas *vit*-6 is autosomal and found on chromosome IV. Although five of these genes are on the X chromosome, they are not clustered in the chromosome (Heine and Blumenthal, 1986). The peptide encoded by *vit*-6 is proteolytically cleaved prior to entering the oocyte, to form yp115 and yp88 (Spieth *et al.*, 1991). Polypeptide yp170A is encoded by *vit*-2 and yp170B is encoded by *vit*-3, *vit*-4 and *vit*-5 (Spieth and Blumenthal, 1985).

The *C. elegans* gene *vit*-1 is 82% identical to *vit*-2 and was originally suggested to be a pseudogene; however, it has now been shown that it is transcribed and translated. The genes *vit*-3 and *vit*-4 are next to each other on the X chromosome and are 99% identical. It has also been discovered that *vit*-5 is only 67% similar to *vit*-2, but 96% identical to *vit*-3 and *vit*-4 (Perez and Lehner, 2019). The gene *vit*-6 has diverged the most from the other *vit* genes, by being only 22% identical to *vit*-2 and 50% to *vit*-5. The amino acid composition of all *vit* genes is similar, with many cysteine residues conserved (Spieth *et al.*, 1991). The *C. elegans vit* genes contain four introns, these have been found to be in similar positions to vertebrate *vit* genes and may be a result of deriving from a common ancestor (Nardelli *et al.*, 1987).

The *vit* genes of *C. elegans* contain three protein domains, Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor type D domain. 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, 1988a). Another common protein domain associated with vitellogenin in oviparous animals is phosvitin. This domain has been recognised in chicken (*Gallus gallus*) and the African claw frog (*Xenopus laevis*) but was

shown to be the least conserved domain between the two species and is not present in *C. elegans* vitellogenin (Nardelli *et al.*, 1987).

Vitellogenin proteins and the associated encoding genes have also been characterised in some species of both vertebrates (fish and birds) and invertebrates (insects). Nematode vitellogenin can be aligned against both vertebrate and insect vitellogenin but is highly divergent, and nematode and vertebrate are more closely related than they are to insect vitellogenin (Winter *et al.*, 1996). *Caenorhabditis elegans* has been the primary model for investigating vitellogenin in nematodes; however, orthologous genes have been found in the free-living nematodes *Oscheius tipulae* (Almenara *et al.*, 2013) and *C. briggsae* (Zucker-Aprison and Blumenthal, 1989). Vitellogenin has been also identified in the animal-parasitic nematodes *Toxocara canis* (Zhu *et al.*, 2017), *Haemonchus contortus* (Hartman *et al.*, 2001) and *Trichostongylus vitrinus* (Nisbet and Gasser, 2004).

Vitellogenin is an important protein in yolk provision for the developing nematode; however, little research has focused on the encoding genes in cyst, root-knot (RKN) and root-lesion nematodes (RLN). The presence of *vit* genes within the genomes of *Globodera pallida* has been reported in a transcriptome study, which described the high expression of *vit* genes during female life stages (Cotton *et al.*, 2014).

With the increasing availability of genomic data for many economically important species of plant-parasitic nematodes (PPN), it is now possible to determine presence of *vit* genes within their genomes. The first multicellular organism to have its DNA sequenced was the nematode *C. elegans* in 1998 (The *C. elegans* Sequencing Consortium, 1998). Subsequently, many PPN genomes were sequenced; however, until 2017 there were fewer than 20 genomes published (Kikuchi *et al.*, 2017). Many sequencing projects have focused on species in the order Tylenchida (Montarry *et al.*, 2021), including the cyst nematodes: *G. pallida* (Cotton *et al.*, 2014; van Steenbrugge *et al.*, 2023), *G. rostochiensis* (Eves-van den Akker *et al.*, 2016), *G. ellingtonae* (Phillips *et al.*, 2017), *Heterodera glycines* (Masonbrink *et al.*, 2021) and *H. schachtii* (Siddique *et al.*, 2022). The genomes of selected root-knot nematode species have also been sequenced, including: *Meloidogyne incognita* (Abad *et al.*, 2008), *M. javanica* (Blanc-Mathieu *et al.*, 2017), *M. enterolobii* (Koutsovoulos *et al.*, 2020) and *M. hapla* (Opperman *et al.*, 2008). The root-lesion nematode species, *Pratylenchus coffeae* (Burke *et al.*, 2015) and *P. scribneri* (Arora *et al.*, 2023) have also been sequenced. Wormbase parasite (Howe *et al.*, 2017) is a key repository for PPN genomic data, with published and unpublished genomes, allowing a wider comparison of genomes of a single species. Furthermore, genomic sequences can be used to predict the biological function of the encoded protein by determining

the biological process, molecular function and cellular component, this is described as a gene ontology (Ashburner *et al.*, 2000).

Vitellogenin genes have not yet been identified or characterised in many species of cyst, root-knot or root-lesion nematodes. Genomic data for many of these economically important species has now been sequenced, therefore it is now possible to determine the presence and sequence of *vit* genes within their genomes and infer their phylogenetic relationships. Therefore, this chapter aims to highlight the presence of and predict *vit* genes within the genomes of some economically important PPN, use phylogenetic analysis to compare these sequences and to infer biological function from the use of bioinformatic resources.

3.2 Materials and Methods

3.2.1 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 3.1). 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. NCBI BLAST was used for searches against the genomes of *Globodera ellingtonae* and the root lesion nematode, *Pratylenchus scribneri*, as these data were not available from the parasite.wormbase.org server. Multiple genome assemblies are available for *G. pallida*, *G. rostochiensis* and *H. schachtii*, which allowed comparison between these data to determine if there were differences in vitellogenin gene prediction across sequencing studies.

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 of the genome and a corresponding predicted 'overlapping gene' in each genome; these gene sequences were then selected from the database and collated for sequence alignment.

3.2.2 Gene Prediction

NCBI BLAST was used to determine homologous regions to the *vit* genes within the genomes of *G. ellingtonae* and *P. scribneri*. 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 and 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.

3.2.3 Phylogenetic Tree Construction

Sequences were uploaded into Molecular Evolutionary Genetics Analysis version ten software (MEGA X) (Kumar *et al.*, 2018) and multiple alignments were 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, and all sites including gaps were used. For each tree constructed the test of phylogeny used was the bootstrap method, this was replicated 100 times. The divergence in sequence of the putative *vit* genes was assessed by calculating pairwise alignments to give percentage identity scores. The scores were generated by inserting sequences into the Jalview software (Waterhouse *et al.*, 2009). Sequences with greater than 35-40% similarity in the percentage identity matrix were likely to have structural similarity (Krissinel, 2007).

3.2.4 Protein Domain Prediction

All genes, predicted or database derived through BLAST searches, were then uploaded to Pfam (Bateman *et al.*, 2004) and later InterPro (Paysan-Lafosse *et al.*, 2023) 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. GO terms were identified by inserting putative *vit* gene sequences into the InterPro platform (Paysan-Lafosse *et al.*, 2023).

Table 3.1: Accession numbers of genes with a putative identification of vitellogenin, nematode species and gene descriptions as given in Section 3.3.

Nematode species	Gene description	Accession number
<i>Caenorhabditis briggsae</i>	<i>C. briggsae</i> 002	WBGene00035047
	<i>C. briggsae</i> 006	WBGene00024796
	<i>C. briggsae</i> 14203	WBGene00034777
	<i>C. briggsae</i> 14234	WBGene00034800
	<i>C. briggsae</i> 16767	WBGene00000365
<i>Caenorhabditis elegans</i>	<i>C. elegans vit-1</i>	WBGene00006925
	<i>C. elegans vit-2</i>	WBGene00006926
	<i>C. elegans vit-3</i>	WBGene00006927

Nematode species	Gene description	Accession number
	<i>C. elegans vit-4</i>	WBGene00006928
	<i>C. elegans vit-5</i>	WBGene00006929
	<i>C. elegans vit-6</i>	WBGene00006930
<i>Globodera ellingtonae</i>	<i>G. ellingtonae</i> 016	<i>G. ellingtonae</i> isolate Ge_1 G_ellingtonae_scaf_0016 (GenBank: MEIZ01000016.1)
	<i>G. ellingtonae</i> 060	<i>G. ellingtonae</i> isolate Ge_1 G_ellingtonae_scaf_0060 (GenBank: MEIZ01000060.1)
<i>Globodera pallida</i>	<i>G. pallida</i> 0128	GPLIN_000012800
	<i>G. pallida</i> 0564	GPLIN_000564000
	<i>G. pallida</i> 7024	GPLIN_000702400
	<i>G. pallida</i> 9459	GPLIN_000945900
	<i>G. pallida</i> 4789	GPAL_D383_g04789
	<i>G. pallida</i> 5968	GPAL_D383_g05968
	<i>G. pallida</i> 7045	Gpal_D383_g07045
	<i>G. pallida</i> 5538	GPALN_005538
	Predicted GP_90	<i>Globodera pallida</i> isolate Newton Gpscaff_6 (GenBank: JAGDFN010000090.1.g39)
<i>Globodera rostochiensis</i>	<i>G. rostochiensis</i> 2537	GROS_g02537
	<i>G. rostochiensis</i> 2714	GROS_g02714
	<i>G. rostochiensis</i> 6238	GROS_g06238
	<i>G. rostochiensis</i> 10898	Gr22_v10_g10898
	<i>G. rostochiensis</i> 18657	Gr22_v10_g18657
	<i>G. rostochiensis</i> 2260	Gr22_v10_g2260
	<i>G. rostochiensis</i> 2396	Gr22_v10_g2396
	<i>G. rostochiensis</i> 16066	Gr19_v10_g16066
	<i>G. rostochiensis</i> 6962	Gr19_v10_g6962
	<i>G. rostochiensis</i> 8204	Gr19_v10_g8204
	<i>G. rostochiensis</i> 8560	Gr19_v10_g8560
<i>Heterodera glycines</i>	<i>H. glycines</i> 13367	Hetgly.G000013367
	<i>H. glycines</i> 23424	Hetgly.G000023424
	<i>H. glycines</i> 6510	Hetgly.G000006510
<i>Heterodera schachtii</i>	<i>H. schachtii</i> 007	Hsc_gene_6726
	<i>H. schachtii</i> 19700	Hsc_gene_19700
<i>Meloidogyne hapla</i>	<i>M. hapla</i> 3631	MhA1_Contig363.frz3.fgene1
<i>Meloidogyne incognita</i>	<i>M. incognita</i> 11129	Minc3s00369g11129
	<i>M. incognita</i> 17578	Minc3s00805g17578
	<i>M. incognita</i> 24409	Minc3s01514g24409

Nematode species	Gene description	Accession number
	<i>M. incognita</i> 28653	Minc3s02166g28653
	<i>M. incognita</i> 30865	Minc3s02607g30865
	<i>M. incognita</i> 6902	Minc3s00180g06902
	<i>M. incognita</i> 9435	Minc3s00286g09435
<i>Meloidogyne javanica</i>	<i>M. javanica</i> 15377	M.Javanica_Scaff1399g015377
	<i>M. javanica</i> 21515	M.Javanica_Scaff2226g021515
	<i>M. javanica</i> 29487	M.Javanica_Scaff3500g029487
	<i>M. javanica</i> 3127	M.Javanica_Scaff3804g031270
	<i>M. javanica</i> 3894	M.Javanica_Scaff242g003894
	<i>M. javanica</i> 43787	M.Javanica_Scaff6405g043787
	<i>M. javanica</i> 5192	M.Javanica_Scaff349g005192
<i>Pratylenchus scribneri</i>	<i>P. scribneri</i> 001	<i>Pratylenchus scribneri</i> isolate ND ptg000001l_1 (Genbank: JAQSEB010000001.1)
	<i>P. scribneri</i> 017	<i>Pratylenchus scribneri</i> isolate ND ptg000029l_1 (Genbank: JAQSEB010000017.1)
	<i>P. scribneri</i> 019	<i>Pratylenchus scribneri</i> isolate ND ptg000033l_1 (Genbank: JAQSEB010000019.1)
	<i>P. scribneri</i> 064	<i>Pratylenchus scribneri</i> isolate ND ptg000106l_1 (Genbank: JAQSEB010000064.1)
	<i>P. scribneri</i> 111	<i>Pratylenchus scribneri</i> isolate ND ptg000177l_1 (Genbank: JAQSEB010000111.1)
	<i>P. scribneri</i> 147	<i>Pratylenchus scribneri</i> isolate ND ptg000228l_1 (Genbank: JAQSEB010000147.1)
	<i>P. scribneri</i> 032	<i>Pratylenchus scribneri</i> isolate ND ptg000057l_1 (GenBank: JAQSEB010000032.1)

3.3 Results

3.3.1 BLAST Search for PPN Vitellogenin

BLAST searches were conducted to determine putative vitellogenin genes in species of plant-parasitic nematodes (PPN). The full BLAST search results are given in Supplementary File 3.1, Tables 3.1.1-3.1.10; a summary of all genes obtained in this search is given in Table 3.2. The result from the BLAST search indicated that there were homologous regions and genes to *Caenorhabditis elegans* vitellogenin within PPN genomes that have not previously been identified.

As given in Table 3.2, five genes within the *C. briggsae* genome had sequence similarity to all six *C. elegans vit* genes. Four genes within the *Globodera pallida* genome were found to have some sequence homology to the six *C. elegans vit* genes. In searches of the *G. rostochiensis* genome, five genes were found using *C. elegans vit* 1-5 query sequences; however, only three were found using the *C. elegans vit*-6 gene sequence query. *Heterodera glycines* genome contained 17 different genes that had sequence similarity to the *C. elegans vit* 1-6 genes; however, only two genes were in common across all six query sequences. In BLAST searches using the *H. schachtii* genome ten genes were found to be similar to at least one of the six *C. elegans vit* gene sequences. In the *Meloidogyne hapla* genome, 13 genes were identified, ten were in common across all six *C. elegans* query sequences. Furthermore, in a search of the *M. incognita* genome, 29 genes were found using the *C. elegans* query sequences, 20 were in common across all six gene queries. Finally, the *M. javanica* genome contained 30 genes that had regions of similarity to the *C. elegans vit* genes.

As a result of BLAST searches multiple genes were identified that had similarity to the *C. elegans vit* genes. Once these gene sequences were obtained, they were inserted into Pfam as described in Section 3.2. Gene sequences that contained the same protein domains as *C. elegans* (Vitellogenin_N, open beta-sheet and von Willebrand factor type D domains) were included in phylogenetic tree analysis and subsequent comparison of protein domain position. All genes that were present in BLAST searches are given in Table 3.2; genes containing the same protein domains as present in *C. elegans* vitellogenin are highlighted. Final gene sequences obtained are given in Supplementary File 3.2.

The genomes of *G. ellingtonae* and *Pratylenchus scribneri* were not available from Wormbase Parasite, therefore data available from NCBI was used to predict genes as described in Section 3.2. BLAST searches resulted in four regions with considerable alignments to the *G. ellingtonae* genome (Supplementary File 3.1, Table 9). This enabled four *vit* genes to be predicted. Two of these genes were included in the phylogenetic tree and protein domain comparisons; however, the predicted gene from accession number MEIZ01000007.1, contained only the Vitellogenin_N and open-beta sheet domains. Furthermore, the predicted gene from accession number MEIZ01000069.1, contained the Vitellogenin_N domain only. Therefore, these two were excluded from this study. BLAST searches of the *P. scribneri* genome resulted in nine significant alignments to the genome (Supplementary File 3.1.10). Predicting genes from the resulting regions gave rise to nine putative *vit* genes. Only seven genes were included in the phylogenetic tree analysis and protein domain comparisons, as two genes were excluded from the analysis. The genes excluded were predicted from accession numbers JAQSEB010000053.1 and JAQSEB010000130.1, as they only contained the Vitellogenin_N and the open beta-sheet domains, or the Vitellogenin_N domain only.

Table 3.2: Summary of genes obtained from BLAST searches using *Caenorhabditis elegans* vitellogenin (*vit*) genes 1-6 as query sequences against plant-parasitic nematode genomes. Red highlight indicates genes that contain the same protein domains found in *C. elegans vit* genes and used in subsequent analysis.

Target genome/ Query sequence	<i>C. elegans vit</i> -1	<i>C. elegans vit</i> -2	<i>C. elegans vit</i> -3	<i>C. elegans vit</i> -4	<i>C. elegans vit</i> -5	<i>C. elegans vit</i> -6
<i>Caenorhabditis briggsae</i>	CBG14203	CBG14203	CBG14203	CBG14203	CBG14203	CBG14203
	CBG14234	CBG14234	CBG14234	CBG14234	CBG14234	CBG14234
	CBG16767	CBG16767	CBG16767	CBG16767	CBG16767	CBG16767
	Cbr-vit-2	Cbr-vit-2	Cbr-vit-2	Cbr-vit-2	Cbr-vit-2	Cbr-vit-2
	Cbr-vit-6	Cbr-vit-6	Cbr-vit-6	Cbr-vit-6	Cbr-vit-6	Cbr-vit-6
<i>Globodera pallida</i>	GPLIN_0000128 00	GPLIN_0000128 00	GPLIN_0000128 00	GPLIN_0000128 00	GPLIN_0000128 00	GPLIN_0000128 00
	GPLIN_0005640 00	GPLIN_0005640 00	GPLIN_0005640 00	GPLIN_0005640 00	GPLIN_0005640 00	
	GPLIN_0007024 00	GPLIN_0007024 00	GPLIN_0007024 00	GPLIN_0007024 00	GPLIN_0007024 00	GPLIN_0007024 00
	GPLIN_0009459 00	GPLIN_0009459 00	GPLIN_0009459 00	GPLIN_0009459 00	GPLIN_0009459 00	GPLIN_0009459 00
<i>Globodera rostochiensis</i>	GROS_g02537	GROS_g02537	GROS_g02537	GROS_g02537	GROS_g02537	GROS_g02537
	GROS_g02714	GROS_g02714	GROS_g02714	GROS_g02714	GROS_g02714	GROS_g02714
	GROS_g03015	GROS_g03015	GROS_g03015	GROS_g03015	GROS_g03015	
	GROS_g03016	GROS_g03016	GROS_g03016	GROS_g03016	GROS_g03016	
	GROS_g06238	GROS_g06238	GROS_g06238	GROS_g06238	GROS_g06238	GROS_g06238
<i>Heterodera glycines</i>			Hetgly00448			
						Hetgly01116
						Hetgly07363
						Hetgly15212 (Hetgly.G000006 510)
	Hetgly05118	Hetgly05118				Hetgly15218
	Hetgly05124	Hetgly05124				
	Hetgly05126	Hetgly05126				
	Hetgly05131	Hetgly05131				
	Hetgly17961 (Hetgly.G000013 367)					

Target genome/ Query sequence	<i>C. elegans vit-1</i>	<i>C. elegans vit-2</i>	<i>C. elegans vit-3</i>	<i>C. elegans vit-4</i>	<i>C. elegans vit-5</i>	<i>C. elegans vit-6</i>
	Hetgly21536	Hetgly21536	Hetgly21536	Hetgly21536	Hetgly21536	Hetgly21536
	Hetgly21542 (Hetgly.G000023 424)	Hetgly21542 (Hetgly.G000023 424)	Hetgly21542 (Hetgly.G000023 424)	Hetgly21542 (Hetgly.G000023 424)	Hetgly21542 (Hetgly.G000023 424)	Hetgly21542 (Hetgly.G000023 424)
	Hetgly22125					
			Hetgly21550	Hetgly21550	Hetgly21550	
						Hetgly21883
						Hetgly21884
			Hetgly22088	Hetgly22088	Hetgly22088	Hetgly22088
						Hetgly22100
<i>Heterodera schachtii</i>	Hsc_gene_1970 0	Hsc_gene_1970 0	Hsc_gene_1970 0	Hsc_gene_1970 0	Hsc_gene_1970 0	
						Hsc_gene_1583 7
						Hsc_gene_2236 8
			Hsc_gene_2569 8	Hsc_gene_2569 8	Hsc_gene_2569 8	Hsc_gene_2569 8
			Hsc_gene_2569 9	Hsc_gene_2569 9	Hsc_gene_2569 9	Hsc_gene_2569 9
	Hsc_gene_5564	Hsc_gene_5564	Hsc_gene_5564	Hsc_gene_5564	Hsc_gene_5564	Hsc_gene_5564
	Hsc_gene_5566	Hsc_gene_5566	Hsc_gene_5566	Hsc_gene_5566	Hsc_gene_5566	Hsc_gene_5566
						Hsc_gene_6329
	Hsc_gene_6726	Hsc_gene_6726	Hsc_gene_6726	Hsc_gene_6726	Hsc_gene_6726	Hsc_gene_6726
	Hsc_gene_6737	Hsc_gene_6737	Hsc_gene_6737	Hsc_gene_6737	Hsc_gene_6737	Hsc_gene_6737
<i>Meloidogyne hapla</i> (All gene names begin MhA1_Contig...)	Contig1717.frz3. gene3	Contig1717.frz3. gene3	Contig1717.frz3. gene3	Contig1717.frz3. gene3	Contig1717.frz3. gene3	Contig1717.frz3. gene3
	Contig1717.frz3. gene4	Contig1717.frz3. gene4	Contig1717.frz3. gene4	Contig1717.frz3. gene4	Contig1717.frz3. gene4	Contig1717.frz3. gene4
	Contig1717.frz3. gene5	Contig1717.frz3. gene5	Contig1717.frz3. gene5	Contig1717.frz3. gene5	Contig1717.frz3. gene5	Contig1717.frz3. gene5
	Contig1717.frz3. gene6	Contig1717.frz3. gene6	Contig1717.frz3. gene6	Contig1717.frz3. gene6	Contig1717.frz3. gene6	
	Contig2399.frz3.f gene1	Contig2399.frz3.f gene1	Contig2399.frz3.f gene1	Contig2399.frz3.f gene1	Contig2399.frz3.f gene1	Contig2399.frz3.f gene1

Target genome/ Query sequence	<i>C. elegans vit-1</i>	<i>C. elegans vit-2</i>	<i>C. elegans vit-3</i>	<i>C. elegans vit-4</i>	<i>C. elegans vit-5</i>	<i>C. elegans vit-6</i>
	Contig2399.frz3.f gene2	Contig2399.frz3.f gene2	Contig2399.frz3.f gene2	Contig2399.frz3.f gene2	Contig2399.frz3.f gene2	
	Contig2399.frz3. gene6	Contig2399.frz3. gene6	Contig2399.frz3. gene6	Contig2399.frz3. gene6	Contig2399.frz3. gene6	Contig2399.frz3. gene6
	Contig2399.frz3. gene7	Contig2399.frz3. gene7	Contig2399.frz3. gene7	Contig2399.frz3. gene7	Contig2399.frz3. gene7	Contig2399.frz3. gene7
	Contig2584.frz3. gene1	Contig2584.frz3. gene1	Contig2584.frz3. gene1	Contig2584.frz3. gene1	Contig2584.frz3. gene1	Contig2584.frz3. gene1
						Contig2800.frz3. gene1
	Contig363.frz3.fg ene1	Contig363.frz3.fg ene1	Contig363.frz3.fg ene1	Contig363.frz3.fg ene1	Contig363.frz3.fg ene1	Contig363.frz3.fg ene1
	Contig363.frz3.g ene10	Contig363.frz3.g ene10	Contig363.frz3.g ene10	Contig363.frz3.g ene10	Contig363.frz3.g ene10	Contig363.frz3.g ene10
	Contig363.frz3.g ene9	Contig363.frz3.g ene9	Contig363.frz3.g ene9	Contig363.frz3.g ene9	Contig363.frz3.g ene9	Contig363.frz3.g ene9
<i>Meloidogyne incognita</i>	Minc3s00180g06 901	Minc3s00180g06 901	Minc3s00180g06 901	Minc3s00180g06 901	Minc3s00180g06 901	Minc3s00180g06 901
		Minc3s00180g06 902	Minc3s00180g06 902	Minc3s00180g06 902	Minc3s00180g06 902	Minc3s00180g06 902
	Minc3s00286g09 435	Minc3s00286g09 435	Minc3s00286g09 435	Minc3s00286g09 435	Minc3s00286g09 435	Minc3s00286g09 435
	Minc3s00369g11 129	Minc3s00369g11 129	Minc3s00369g11 129	Minc3s00369g11 129	Minc3s00369g11 129	Minc3s00369g11 129
						Minc3s00369g11 130
	Minc3s00369g11 131	Minc3s00369g11 131				
	Minc3s00805g17 578	Minc3s00805g17 578	Minc3s00805g17 578	Minc3s00805g17 578	Minc3s00805g17 578	Minc3s00805g17 578
	Minc3s00805g17 579	Minc3s00805g17 579	Minc3s00805g17 579	Minc3s00805g17 579	Minc3s00805g17 579	Minc3s00805g17 579
	Minc3s01514g24 409	Minc3s01514g24 409	Minc3s01514g24 409	Minc3s01514g24 409	Minc3s01514g24 409	Minc3s01514g24 409
	Minc3s01575g24 833	Minc3s01575g24 833	Minc3s01575g24 833	Minc3s01575g24 833	Minc3s01575g24 833	Minc3s01575g24 833

Target genome/ Query sequence	<i>C. elegans vit-1</i>	<i>C. elegans vit-2</i>	<i>C. elegans vit-3</i>	<i>C. elegans vit-4</i>	<i>C. elegans vit-5</i>	<i>C. elegans vit-6</i>
	Minc3s01575g24 834	Minc3s01575g24 834	Minc3s01575g24 834	Minc3s01575g24 834	Minc3s01575g24 834	Minc3s01575g24 834
	Minc3s01861g26 817	Minc3s01861g26 817	Minc3s01861g26 817	Minc3s01861g26 817	Minc3s01861g26 817	Minc3s01861g26 817
	Minc3s01861g26 818	Minc3s01861g26 818				Minc3s01861g26 818
	Minc3s01975g27 537	Minc3s01975g27 537				Minc3s01975g27 537
	Minc3s01975g27 538	Minc3s01975g27 538				Minc3s01975g27 538
	Minc3s01975g27 539	Minc3s01975g27 539	Minc3s01975g27 539	Minc3s01975g27 539	Minc3s01975g27 539	Minc3s01975g27 539
	Minc3s02166g28 653	Minc3s02166g28 653	Minc3s02166g28 653	Minc3s02166g28 653	Minc3s02166g28 653	Minc3s02166g28 653
	Minc3s02166g28 654	Minc3s02166g28 654	Minc3s02166g28 654	Minc3s02166g28 654	Minc3s02166g28 654	Minc3s02166g28 654
	Minc3s02260g29 127	Minc3s02260g29 127	Minc3s02260g29 127	Minc3s02260g29 127	Minc3s02260g29 127	Minc3s02260g29 127
	Minc3s02607g30 865	Minc3s02607g30 865	Minc3s02607g30 865	Minc3s02607g30 865	Minc3s02607g30 865	Minc3s02607g30 865
	Minc3s02607g30 866	Minc3s02607g30 866	Minc3s02607g30 866	Minc3s02607g30 866	Minc3s02607g30 866	Minc3s02607g30 866
	Minc3s03289g33 421	Minc3s03289g33 421	Minc3s03289g33 421	Minc3s03289g33 421	Minc3s03289g33 421	Minc3s03289g33 421
			Minc3s06543g39 965	Minc3s06543g39 965	Minc3s06543g39 965	Minc3s06543g39 965
	Minc3s06972g40 563	Minc3s06972g40 563	Minc3s06972g40 563	Minc3s06972g40 563	Minc3s06972g40 563	
	Minc3s06972g40 564	Minc3s06972g40 564	Minc3s06972g40 564	Minc3s06972g40 564	Minc3s06972g40 564	Minc3s06972g40 564
	Minc3s07387g41 062	Minc3s07387g41 062	Minc3s07387g41 062	Minc3s07387g41 062	Minc3s07387g41 062	Minc3s07387g41 062
	Minc3s08451g42 267	Minc3s08451g42 267	Minc3s08451g42 267	Minc3s08451g42 267	Minc3s08451g42 267	Minc3s08451g42 267
	Minc3s09698g43 458	Minc3s09698g43 458	Minc3s09698g43 458	Minc3s09698g43 458	Minc3s09698g43 458	Minc3s09698g43 458

Target genome/ Query sequence	<i>C. elegans vit-1</i>	<i>C. elegans vit-2</i>	<i>C. elegans vit-3</i>	<i>C. elegans vit-4</i>	<i>C. elegans vit-5</i>	<i>C. elegans vit-6</i>
	Minc3s10886g44441					Minc3s10886g44441
<i>Meloidogyne javanica</i> (All gene names begin <i>M. javanica_Scaff...</i>)	Scaff10759g059768	Scaff10759g059768			Scaff10759g059768	Scaff10759g059768
	Scaff1399g015376		Scaff1399g015376	Scaff1399g015376	Scaff1399g015376	
	Scaff1399g015377	Scaff1399g015377	Scaff1399g015377	Scaff1399g015377	Scaff1399g015377	Scaff1399g015377
	Scaff1515g016285	Scaff1515g016285				Scaff1515g016285
	Scaff1515g016286	Scaff1515g016286	Scaff1515g016286	Scaff1515g016286	Scaff1515g016286	Scaff1515g016286
	Scaff15289g071985	Scaff15289g071985	Scaff15289g071985	Scaff15289g071985	Scaff15289g071985	Scaff15289g071985
	Scaff15289g071986	Scaff15289g071986	Scaff15289g071986	Scaff15289g071986	Scaff15289g071986	Scaff15289g071986
	Scaff18769g079434	Scaff18769g079434	Scaff18769g079434	Scaff18769g079434	Scaff18769g079434	Scaff18769g079434
	Scaff21123g083914	Scaff21123g083914	Scaff21123g083914	Scaff21123g083914	Scaff21123g083914	Scaff21123g083914
	Scaff21123g083915		Scaff21123g083915	Scaff21123g083915	Scaff21123g083915	Scaff21123g083915
	Scaff21949g085322	Scaff21949g085322	Scaff21949g085322	Scaff21949g085322	Scaff21949g085322	Scaff21949g085322
	Scaff2226g021515	Scaff2226g021515	Scaff2226g021515	Scaff2226g021515	Scaff2226g021515	Scaff2226g021515
	Scaff23110g087285	Scaff23110g087285	Scaff23110g087285	Scaff23110g087285	Scaff23110g087285	
	Scaff23110g087286	Scaff23110g087286	Scaff23110g087286	Scaff23110g087286	Scaff23110g087286	
	Scaff242g003894	Scaff242g003894	Scaff242g003894	Scaff242g003894	Scaff242g003894	Scaff242g003894
	Scaff242g003895	Scaff242g003895				Scaff242g003895
	Scaff25080g090352	Scaff25080g090352	Scaff25080g090352	Scaff25080g090352	Scaff25080g090352	Scaff25080g090352

Target genome/ Query sequence	<i>C. elegans vit-1</i>	<i>C. elegans vit-2</i>	<i>C. elegans vit-3</i>	<i>C. elegans vit-4</i>	<i>C. elegans vit-5</i>	<i>C. elegans vit-6</i>
	Scaff30898g098056					Scaff30898g098056
	Scaff349g005192	Scaff349g005192	Scaff349g005192	Scaff349g005192	Scaff349g005192	Scaff349g005192
	Scaff349g005193	Scaff349g005193	Scaff349g005193	Scaff349g005193	Scaff349g005193	Scaff349g005193
	Scaff3500g029487	Scaff3500g029487	Scaff3500g029487	Scaff3500g029487	Scaff3500g029487	Scaff3500g029487
	Scaff3804g031270	Scaff3804g031270	Scaff3804g031270	Scaff3804g031270	Scaff3804g031270	Scaff3804g031270
						Scaff4688g035870
	Scaff3804g031271		Scaff3804g031271	Scaff3804g031271	Scaff3804g031271	
	Scaff6405g043787	Scaff6405g043787	Scaff6405g043787	Scaff6405g043787	Scaff6405g043787	Scaff6405g043787
	Scaff6533g044375	Scaff6533g044375				Scaff6533g044375
						Scaff6533g044376
	Scaff6533g044377	Scaff6533g044377				
	Scaff6533g044379	Scaff6533g044379	Scaff6533g044379	Scaff6533g044379	Scaff6533g044379	Scaff6533g044379
	Scaff6533g044380	Scaff6533g044380	Scaff6533g044380	Scaff6533g044380	Scaff6533g044380	Scaff6533g044380
	Scaff6982g046249	Scaff6982g046249	Scaff6982g046249	Scaff6982g046249	Scaff6982g046249	Scaff6982g046249
	Scaff6982g046250	Scaff6982g046250	Scaff6982g046250	Scaff6982g046250	Scaff6982g046250	Scaff6982g046250
	Scaff7843g049685	Scaff7843g049685	Scaff7843g049685	Scaff7843g049685	Scaff7843g049685	Scaff7843g049685
	Scaff7843g049686	Scaff7843g049686	Scaff7843g049686	Scaff7843g049686	Scaff7843g049686	Scaff7843g049686
	Scaff8571g052471	Scaff8571g052471		Scaff8571g052471	Scaff8571g052471	Scaff8571g052471

Target genome/ Query sequence	<i>C. elegans vit-1</i>	<i>C. elegans vit-2</i>	<i>C. elegans vit-3</i>	<i>C. elegans vit-4</i>	<i>C. elegans vit-5</i>	<i>C. elegans vit-6</i>
	Scaff8571g0524 72	Scaff8571g0524 72	Scaff8571g0524 72	Scaff8571g0524 72	Scaff8571g0524 72	Scaff8571g0524 72
	Scaff9762g0566 27	Scaff9762g0566 27				Scaff9762g0566 27
						Scaff9762g0566 28
	Scaff9762g0566 29					
	Scaff9762g0566 31	Scaff9762g0566 31	Scaff9762g0566 31	Scaff9762g0566 31	Scaff9762g0566 31	Scaff9762g0566 31

3.3.2 Genomic Location of *vit* Genes

BLAST searches also enabled the genomic location of the *vit* genes to be considered (Supplementary File 3.1). The *C. elegans vit* 1-5 genes were located on the nematode's X chromosome, whereas the *vit*-6 gene was located on chromosome IV. The *C. briggsae* genes most closely related to the *C. elegans vit* 1-5 genes were also located on the X chromosome. Also, *C. briggsae* 006, which is highly like *C. elegans vit*-6, was also located on chromosome IV. The genomic location of the *G. pallida vit* genes varied along the scaffold sequence. Similarly, the position of *G. rostochiensis vit* genes were found on different contigs. The *H. glycines vit* genes 6510, 13367 and 23424, were found on chromosome 6, 7 and 8, respectively. The position of *H. schachtii vit* genes had varied locations on the genome scaffold. Finally, the *vit* genes of both *M. incognita* and *M. javanica* were located on different contigs.

3.3.3 Sequence and Pairwise Alignment

Once gene sequences had been obtained from BLAST searches or predicted, they were aligned using MUSCLE, as described in section 3.2. Sequence alignment of all genes included in the analysis are given in Supplementary File 3.3. The multiple sequence alignment showed that there was greater conservation of the sequences at amino acid positions: 1220-1580, 1850-2550 and 3050-3330.

Sequence pairwise alignments were calculated between genes to determine the similarity of sequences by percentage identity, from within a species and between different species. The pairwise alignment matrix of percentage identity is given in Table 3.3. The *C. briggsae* genes are most closely related to the *C. elegans vit* genes, ranging from 29-88.6% similarity. The *C. elegans vit*-6 gene has the most diverse sequence compared to other *vit* genes in the same species, of only 30% percentage identity, whereas it shares 88.6% of its sequence with *C. briggsae* 006. Therefore, the *C. elegans vit*-6 and *C. briggsae* 006 are likely to be structurally similar but are dissimilar to other *vit* genes in both *C. elegans* and *C. briggsae*.

In comparison of the *Caenorhabditis* species to all PPN *vit* genes, there was an average sequence similarity of 20-25%, therefore may be structurally dissimilar. In the cyst nematode group, the similarity between genes was quite varied, from 21.7 to 97.2%. Predicted vitellogenin gene *G. ellingtonae* 016 was shown to be most closely related to *G. rostochiensis* 2537, with a pairwise alignment percentage identity of 97.2%. This predicted gene was also shown to be closely related to *G. pallida* 0128, with a similarity of 92%. The second predicted gene, *G. ellingtonae* 060, was most closely related to *G. rostochiensis* 6238 and *G. pallida* 7024, with percentage identification of 95.3% and 82.3%, respectively. The cyst nematode group contains three outliers, *G. pallida* 0564, *H. glycines* 13367 and *H. schachtii* 19700, with

only approximately 25% identity to other *vit* genes within the group, therefore may have a different structure to other *vit* genes. There are apparent potential *vit* gene homologs between *G. pallida*, *G. rostochiensis* and *G. ellingtonae* with high percentage identities and separately between *H. schachtii* and *H. glycines* genes.

When comparing the root-knot nematode and the cyst nematode *vit* genes, the percentage identity ranges from 21.2 to 38.6%, therefore have a low likelihood of having a similar protein structure to cyst nematodes. The root-knot nematode *vit* genes are relatively highly conserved with an average percentage identity of 75%, with a range of 45.5-98.4%. This group contains three outliers: *M. hapla* 3631, *M. incognita* 11129 and *M. javanica* 3894, which have sequence similarity to each other, but contain a more diverse sequence than the rest of the group.

The predicted *P. scribneri* genes have an average percentage identity of 40%, when compared to the rest of the PPN group, likely having a similar protein structure. Excluding *vit* genes *P. scribneri* 001 and 032, the remaining genes are highly conserved within the species, with percentage identity ranging from 91 to 97%.

Table 3.3: Pairwise alignment of vitellogenin genes from *Caenorhabditis briggsae*, *C. elegans*, *Globodera pallida*, *G. rostochiensis*, *G. ellingtonae*, *Heterodera glycines*, *H. schachtii*, *Meloidogyne hapla*, *M. incognita*, *M. javanica* and *Pratylenchus scribneri*. Percentage identity (%) for each pairwise alignment is given; values are highlighted in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%). (1/5)

Gene	<i>C. briggsae</i> 002	<i>C. briggsae</i> 006	<i>C. briggsae</i> 14203	<i>C. briggsae</i> 14234	<i>C. briggsae</i> 16767	<i>C. elegans</i> vit-1	<i>C. elegans</i> vit-2	<i>C. elegans</i> vit-3	<i>C. elegans</i> vit-4	<i>C. elegans</i> vit-5	<i>C. elegans</i> vit-6
<i>C. briggsae</i> 002	X										
<i>C. briggsae</i> 006	30.7	X									
<i>C. briggsae</i> 14203	57.8	29.7	X								
<i>C. briggsae</i> 14234	57.3	29.5	93.1	X							
<i>C. briggsae</i> 16767	82.4	30.7	55.3	54.9	X						
<i>C. elegans</i> vit-1	78.9	30.8	56.5	56.3	73.5	X					
<i>C. elegans</i> vit-2	82.2	31.3	57.8	57.3	75.5	93.9	X				
<i>C. elegans</i> vit-3	57.9	30.9	80.7	80.2	56.1	56.3	57.3	X			
<i>C. elegans</i> vit-4	58.1	31.0	80.7	80.4	56.4	56.5	57.5	98.7	X		
<i>C. elegans</i> vit-5	57.9	31.4	81.1	80.5	56.2	56.5	57.5	97.9	98.1	X	
<i>C. elegans</i> vit-6	30.1	88.6	29.4	29.0	30.5	30.4	30.6	30.6	30.9	30.8	X
<i>G. ellingtonae</i> 016	22.0	23.7	23.2	23.5	22.1	22.4	22.2	23.3	23.4	23.7	23.4
<i>G. ellingtonae</i> 060	22.7	23.3	22.6	22.5	23.1	22.8	23.7	22.6	23.0	23.1	22.7
<i>G. pallida</i> 0128	21.2	22.5	22.5	23.0	22.0	21.5	21.6	22.2	22.3	22.9	22.7
<i>G. pallida</i> 0564	20.9	21.7	20.9	20.6	21.2	21.2	21.2	21.2	21.7	21.8	21.6
<i>G. pallida</i> 7024	22.0	23.7	23.0	23.7	22.5	22.1	22.2	22.4	22.6	22.8	23.4
<i>G. pallida</i> 9459	22.4	22.6	22.6	22.7	22.3	21.9	22.5	22.5	22.6	22.7	22.5
<i>G. rostochiensis</i> 2537	21.9	23.3	23.5	23.8	22.4	22.6	22.4	23.7	23.6	24.1	23.5
<i>G. rostochiensis</i> 2714	22.3	22.6	22.2	21.6	22.1	22.0	22.0	22.6	22.7	23.1	22.9
<i>G. rostochiensis</i> 6238	22.4	22.7	22.7	22.8	22.6	22.7	22.5	22.4	22.7	23.0	22.9
<i>H. glycines</i> 13367	22.2	21.4	21.1	21.0	21.9	21.7	21.4	20.7	20.9	20.9	20.8
<i>H. glycines</i> 23424	21.9	23.2	23.3	23.4	22.6	22.4	22.7	23.6	23.1	23.6	22.8
<i>H. glycines</i> 6510	21.7	22.4	21.6	21.6	21.6	22.4	22.6	21.1	21.2	21.5	22.3
<i>H. schachtii</i> 007	22.0	23.2	23.2	23.2	22.7	21.8	22.4	23.4	23.1	23.3	23.4
<i>H. schachtii</i> 19700	22.2	21.4	21.4	21.3	21.8	21.5	21.4	21.0	21.4	21.4	21.0
<i>M. hapla</i> 3631	21.8	21.9	22.1	21.9	21.8	21.6	21.3	22.5	22.6	22.6	22.1
<i>M. incognita</i> 11129	23.4	24.2	23.2	23.3	23.5	22.7	23.4	23.1	23.4	23.4	23.7
<i>M. incognita</i> 17578	24.0	22.2	23.7	23.9	23.3	23.8	23.2	23.1	23.4	22.9	22.9
<i>M. incognita</i> 24409	24.7	23.6	23.8	24.0	24.3	23.9	24.2	23.8	24.0	24.1	23.7
<i>M. incognita</i> 28653	22.9	22.6	22.7	23.0	23.5	23.3	23.3	22.3	22.5	22.8	22.7
<i>M. incognita</i> 30865	24.6	22.9	23.7	23.8	23.9	23.9	23.9	23.1	23.5	23.6	23.1
<i>M. incognita</i> 6902	24.9	23.4	24.6	24.9	24.7	24.8	24.8	23.6	24.2	24.7	23.9
<i>M. incognita</i> 9435	24.1	22.6	23.5	23.6	23.7	23.8	23.9	22.4	22.6	23.4	23.2
<i>M. javanica</i> 15377	24.2	23.1	23.5	23.8	23.2	24.2	24.1	22.7	23.3	23.4	23.5
<i>M. javanica</i> 21515	24.8	23.0	23.5	23.8	23.9	24.0	24.0	22.7	23.4	23.7	23.0
<i>M. javanica</i> 29487	22.38	22.12	22.1	22.06	22	22.38	22.38	20.58	20.8	21.13	22.82
<i>M. javanica</i> 3127	23.11	22.44	22.68	22.77	23.03	23.49	23.49	22.58	22.74	23.01	23.18
<i>M. javanica</i> 3894	22.86	24.36	22.99	22.8	23.2	22.7	23.08	23.39	23.77	23.77	23.44
<i>M. javanica</i> 43787	24.75	23.12	23.84	23.92	23.89	24.27	24.18	22.91	23.57	23.88	23.07
<i>M. javanica</i> 5192	24.32	22.95	23.61	23.31	23.69	24.09	24.1	22.34	22.95	23.29	22.43
<i>P. scribneri</i> 001	24.47	24.08	24.35	24.12	23.92	24.31	24.73	23.49	23.47	23.55	24.19
<i>P. scribneri</i> 017	23.74	22.65	22.88	22.49	23.43	24.12	23.99	22.22	22.61	22.71	22.23
<i>P. scribneri</i> 019	23.2	22.63	22.85	22.36	23.76	23.81	24.14	22.92	23.03	22.57	22.49
<i>P. scribneri</i> 064	24.39	22.19	23.44	22.98	23.47	24.16	24.1	23.26	22.98	23.07	22.05
<i>P. scribneri</i> 111	23.43	22.78	22.82	21.68	23.28	24.46	24.15	23.21	23.48	23.39	22.33
<i>P. scribneri</i> 147	24.55	23.25	23.78	23.49	23.49	24.64	24.25	23.98	24.28	24.28	23.04
<i>P. scribneri</i> 032	23.64	24.03	23.26	23.44	23.31	23.32	23.36	23.51	23.62	23.6	24.84

Table 3.3: Pairwise alignment of vitellogenin genes from *Caenorhabditis briggsae*, *C. elegans*, *Globodera pallida*, *G. rostochiensis*, *G. ellingtonae*, *Heterodera glycines*, *H. schachtii*, *Meloidogyne hapla*, *M. incognita*, *M. javanica* and *Pratylenchus scribneri*. Percentage identity (%) for each pairwise alignment is given; values are highlighted in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%). (2/5)

Gene	<i>G. ellingtonae</i> 016	<i>G. ellingtonae</i> 060	<i>G. pallida</i> 0128	<i>G. pallida</i> 0564	<i>G. pallida</i> 7024	<i>G. pallida</i> 9459	<i>G. rostochiensis</i> 2537	<i>G. rostochiensis</i> 2714	<i>G. rostochiensis</i> 6238	<i>H. glycines</i> 13367	<i>H. glycines</i> 23424
<i>C. briggsae</i> 002											
<i>C. briggsae</i> 006											
<i>C. briggsae</i> 14203											
<i>C. briggsae</i> 14234											
<i>C. briggsae</i> 16767											
<i>C. elegans</i> vit-1											
<i>C. elegans</i> vit-2											
<i>C. elegans</i> vit-3											
<i>C. elegans</i> vit-4											
<i>C. elegans</i> vit-5											
<i>C. elegans</i> vit-6											
<i>G. ellingtonae</i> 016	X										
<i>G. ellingtonae</i> 060	72.6	X									
<i>G. pallida</i> 0128	92.0	67.2	X								
<i>G. pallida</i> 0564	25.0	23.3	23.3	X							
<i>G. pallida</i> 7024	68.6	82.3	65.7	24.5	X						
<i>G. pallida</i> 9459	69.1	66.6	66.0	24.6	63.4	X					
<i>G. rostochiensis</i> 2537	97.2	72.2	79.5	24.9	68.3	68.4	X				
<i>G. rostochiensis</i> 2714	70.2	67.6	66.6	24.2	64.1	93.8	69.6	X			
<i>G. rostochiensis</i> 6238	74.1	95.3	70.5	24.8	87.0	68.3	74.0	69.1	X		
<i>H. glycines</i> 13367	25.1	25.4	24.5	58.1	24.9	24.1	25.1	23.9	25.2	X	
<i>H. glycines</i> 23424	73.9	67.8	70.8	23.7	64.7	65.2	73.1	66.7	70.0	25.0	X
<i>H. glycines</i> 6510	57.9	55.3	55.4	23.8	55.3	59.6	57.4	60.5	57.9	24.0	57.2
<i>H. schachtii</i> 007	74.5	67.1	71.3	23.6	64.1	65.3	73.5	66.7	69.5	25.1	96.1
<i>H. schachtii</i> 19700	25.4	25.4	24.5	58.1	24.9	24.7	25.6	25.0	25.4	95.9	24.6
<i>M. hapla</i> 3631	33.3	31.3	32.7	21.7	31.9	32.0	33.8	32.7	33.4	21.7	32.7
<i>M. incognita</i> 11129	35.8	34.2	34.5	22.0	34.1	35.0	35.8	35.4	35.4	23.3	35.4
<i>M. incognita</i> 17578	35.4	34.1	33.6	23.2	33.4	34.1	34.8	34.8	35.1	23.1	34.9
<i>M. incognita</i> 24409	36.0	34.9	33.8	24.3	33.6	35.0	35.8	34.8	35.6	24.6	35.0
<i>M. incognita</i> 28653	37.4	35.2	36.2	23.3	35.3	36.8	37.4	36.9	36.4	24.9	36.8
<i>M. incognita</i> 30865	36.6	34.1	35.1	23.1	34.7	35.7	36.1	36.2	36.0	23.7	35.9
<i>M. incognita</i> 6902	34.9	34.0	32.6	24.2	32.9	33.8	35.0	34.5	35.1	24.2	34.6
<i>M. incognita</i> 9435	37.0	34.4	35.5	23.6	35.4	36.4	36.9	36.8	36.4	24.9	36.5
<i>M. javanica</i> 15377	38.6	35.7	37.1	23.6	36.0	37.9	38.6	38.6	36.9	24.6	37.2
<i>M. javanica</i> 21515	36.4	33.8	34.8	22.7	34.5	35.6	36.0	35.9	35.9	23.3	35.8
<i>M. javanica</i> 29487	36.13	32.21	34.45	21.2	33.16	35.25	35.89	35.35	34.42	23.39	34.89
<i>M. javanica</i> 3127	37.53	35.14	36.09	23.52	35.54	37.31	37.57	37.5	36.6	24.82	37
<i>M. javanica</i> 3894	35.82	33.59	34.25	22.65	33.56	34.52	35.97	35.27	35.11	23.72	34.4
<i>M. javanica</i> 43787	36.25	33.71	34.73	23.19	34.33	35.59	36.17	36.1	35.89	23.68	35.8
<i>M. javanica</i> 5192	36.44	34.09	34.74	23.47	35.15	35.78	36.4	36.43	36.31	23.98	36.38
<i>P. scribneri</i> 001	42.62	38.66	41.62	24.14	41.01	41.64	43.13	42.18	41.54	24.47	40.97
<i>P. scribneri</i> 017	42.81	40.65	40.95	25.16	40.77	42.7	42.51	43.2	42.58	24.95	41.91
<i>P. scribneri</i> 019	43.47	42.26	41.67	25.12	41.77	43.84	43.21	43.92	43.58	25.68	42.45
<i>P. scribneri</i> 064	43.43	41.91	41.54	25.4	41.67	43.22	42.97	43.39	43.44	25.5	42.19
<i>P. scribneri</i> 111	42.79	41.55	40.49	25.51	40.49	42.3	42.21	43.13	42.72	25.48	41.44
<i>P. scribneri</i> 147	42.98	42.03	40.2	26.99	40.91	42.67	42.32	43.16	43.34	26.49	41.66
<i>P. scribneri</i> 032	40.84	39.1	39.53	23.16	39.36	39.4	41.52	39.99	39.99	23.08	39.39

Table 3.3: Pairwise alignment of vitellogenin genes from *Caenorhabditis briggsae*, *C. elegans*, *Globodera pallida*, *G. rostochiensis*, *G. ellingtonae*, *Heterodera glycines*, *H. schachtii*, *Meloidogyne hapla*, *M. incognita*, *M. javanica* and *Pratylenchus scribneri*. Percentage identity (%) for each pairwise alignment is given; values are highlighted in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%). (3/5)

Gene	<i>H. glycines</i> 6510	<i>H. schachtii</i> 007	<i>H. schachtii</i> 19700	<i>M. hapla</i> 3631	<i>M. incognita</i> 11129	<i>M. incognita</i> 17578	<i>M. incognita</i> 24409	<i>M. incognita</i> 28653	<i>M. incognita</i> 30865	<i>M. incognita</i> 6902	<i>M. incognita</i> 9435
<i>C. briggsae</i> 002											
<i>C. briggsae</i> 006											
<i>C. briggsae</i> 14203											
<i>C. briggsae</i> 14234											
<i>C. briggsae</i> 16767											
<i>C. elegans</i> vit-1											
<i>C. elegans</i> vit-2											
<i>C. elegans</i> vit-3											
<i>C. elegans</i> vit-4											
<i>C. elegans</i> vit-5											
<i>C. elegans</i> vit-6											
<i>G. ellingtonae</i> 016											
<i>G. ellingtonae</i> 060											
<i>G. pallida</i> 0128											
<i>G. pallida</i> 0564											
<i>G. pallida</i> 7024											
<i>G. pallida</i> 9459											
<i>G. rostochiensis</i> 2537											
<i>G. rostochiensis</i> 2714											
<i>G. rostochiensis</i> 6238											
<i>H. glycines</i> 13367											
<i>H. glycines</i> 23424											
<i>H. glycines</i> 6510	X										
<i>H. schachtii</i> 007	57.9	X									
<i>H. schachtii</i> 19700	24.2	24.8	X								
<i>M. hapla</i> 3631	31.1	32.4	21.6	X							
<i>M. incognita</i> 11129	33.8	35.9	23.5	73.2	X						
<i>M. incognita</i> 17578	33.6	34.6	24.2	46.4	46.9	X					
<i>M. incognita</i> 24409	34.3	34.5	25.2	47.4	48.3		X				
<i>M. incognita</i> 28653	35.6	36.5	25.0	48.7	50.2	84.8	97.1	X			
<i>M. incognita</i> 30865	34.8	35.7	24.0	47.9	49.0	96.3	87.4	88.2	X		
<i>M. incognita</i> 6902	33.7	34.7	24.5	46.9	47.6	95.5	86.9	86.8	95.7	X	
<i>M. incognita</i> 9435	35.5	36.3	24.9	48.7	50.1	84.3	96.5	97.0	87.5	86.4	X
<i>M. javanica</i> 15377	36.2	37.4	25.0	48.9	49.7	85.8	96.7	96.8	89.1	88.4	98.1
<i>M. javanica</i> 21515	35.1	35.9	23.6	47.9	48.9	92.8	86.3	87.7	96.0	94.5	86.7
<i>M. javanica</i> 29487	33.79	34.56	23.27	45.51	46.73	77.17	85.14	88.97	81.63	77.88	89.11
<i>M. javanica</i> 3127	35.93	37.4	24.95	47.87	49.94	84.59	96.41	98.38	87.88	87.02	96.88
<i>M. javanica</i> 3894	34.29	34.69	23.56	73.28	94.98	46.84	48.17	49.88	48.81	47.68	50.12
<i>M. javanica</i> 43787	34.95	35.83	24.09	47.98	49.34	93.51	87.41	88.21	96.47	96.62	87.41
<i>M. javanica</i> 5192	35.49	36.41	24.23	47.89	48.95	92.82	86.66	87.66	95.94	98.83	86.88
<i>P. scribneri</i> 001	37.36	40.62	24.82	35.78	35.34	37.62	38.13	39.46	38.78	38.1	39.37
<i>P. scribneri</i> 017	40.38	41.61	25.07	37.94	38.23	41.87	41.33	43.72	43.6	41.56	43.35
<i>P. scribneri</i> 019	40.83	42.34	25.77	38.78	39.4	42.73	43.55	44.83	44.23	43.37	44.44
<i>P. scribneri</i> 064	41.02	41.85	25.43	38.88	39.63	42.78	43.04	44.53	44.29	43.45	44.32
<i>P. scribneri</i> 111	40.01	41.29	25.8	38.46	39.19	42.45	43	44.72	44.24	42.84	43.74
<i>P. scribneri</i> 147	40.04	41.32	26.85	38.15	38.61	43.51	42.62	43.3	43.57	43.29	42.75
<i>P. scribneri</i> 032	35.98	39.82	23.33	34.26	33.68	37.95	38.14	39.02	38.32	38.11	38.96

Table 3.3: Pairwise alignment of vitellogenin genes from *Caenorhabditis briggsae*, *C. elegans*, *Globodera pallida*, *G. rostochiensis*, *G. ellingtonae*, *Heterodera glycines*, *H. schachtii*, *Meloidogyne hapla*, *M. incognita*, *M. javanica* and *Pratylenchus scribneri*. Percentage identity (%) for each pairwise alignment is given; values are highlighted in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%). (4/5)

Gene	<i>M. javanica</i> 15377	<i>M. javanica</i> 21515	<i>M. javanica</i> 29487	<i>M. javanica</i> 3127	<i>M. javanica</i> 3894	<i>M. javanica</i> 43787	<i>M. javanica</i> 5192	<i>P. scribneri</i> 001	<i>P. scribneri</i> 017	<i>P. scribneri</i> 019	<i>P. scribneri</i> 064
<i>C. briggsae</i> 002											
<i>C. briggsae</i> 006											
<i>C. briggsae</i> 14203											
<i>C. briggsae</i> 14234											
<i>C. briggsae</i> 16767											
<i>C. elegans vit-1</i>											
<i>C. elegans vit-2</i>											
<i>C. elegans vit-3</i>											
<i>C. elegans vit-4</i>											
<i>C. elegans vit-5</i>											
<i>C. elegans vit-6</i>											
<i>G. ellingtonae</i> 016											
<i>G. ellingtonae</i> 060											
<i>G. pallida</i> 0128											
<i>G. pallida</i> 0564											
<i>G. pallida</i> 7024											
<i>G. pallida</i> 9459											
<i>G. rostochiensis</i> 2537											
<i>G. rostochiensis</i> 2714											
<i>G. rostochiensis</i> 6238											
<i>H. glycines</i> 13367											
<i>H. glycines</i> 23424											
<i>H. glycines</i> 6510											
<i>H. schachtii</i> 007											
<i>H. schachtii</i> 19700											
<i>M. hapla</i> 3631											
<i>M. incognita</i> 11129											
<i>M. incognita</i> 17578											
<i>M. incognita</i> 24409											
<i>M. incognita</i> 28653											
<i>M. incognita</i> 30865											
<i>M. incognita</i> 6902											
<i>M. incognita</i> 9435											
<i>M. javanica</i> 15377	X										
<i>M. javanica</i> 21515	88.2	X									
<i>M. javanica</i> 29487	89.02	80.94	X								
<i>M. javanica</i> 3127	97.03	87.17	88.7	X							
<i>M. javanica</i> 3894	49.37	48.63	46.23	49.7	X						
<i>M. javanica</i> 43787	88.97	96.67	81.6	87.98	49.22	X					
<i>M. javanica</i> 5192	88.79	95.15	80.2	87.58	49.02	96.89	X				
<i>P. scribneri</i> 001	40.01	38.54	37.51	39.47	35.31	38.81	39.16	X			
<i>P. scribneri</i> 017	44.03	43.34	41.07	43.59	38.35	43.41	43.65	44.93	X		
<i>P. scribneri</i> 019	45.05	44.07	42.28	44.75	38.91	44.14	44.31	45.84	94.05	X	
<i>P. scribneri</i> 064	44.95	44.08	42.19	44.45	39.14	44.35	44.5	45.97	92.64	95.45	X
<i>P. scribneri</i> 111	45.06	43.91	42.38	44.63	38.92	44.14	43.99	44.68	91.08	95.44	92.9
<i>P. scribneri</i> 147	44.3	42.98	40.58	43.12	38.8	43.45	43.29	44.56	92.87	96.5	96.85
<i>P. scribneri</i> 032	39.61	38.01	36.84	38.93	34.07	38.12	39.09	88.47	42.99	44.01	44.02

Table 3.3: Pairwise alignment of vitellogenin genes from *Caenorhabditis briggsae*, *C. elegans*, *Globodera pallida*, *G. rostochiensis*, *G. ellingtonae*, *Heterodera glycines*, *H. schachtii*, *Meloidogyne hapla*, *M. incognita*, *M. javanica* and *Pratylenchus scribneri*. Percentage identity (%) for each pairwise alignment is given; values are highlighted in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%). (5/5)

Gene	<i>P. scribneri</i> 111	<i>P. scribneri</i> 147	<i>P. scribneri</i> 032
<i>C. briggsae</i> 002			
<i>C. briggsae</i> 006			
<i>C. briggsae</i> 14203			
<i>C. briggsae</i> 14234			
<i>C. briggsae</i> 16767			
<i>C. elegans</i> vit-1			
<i>C. elegans</i> vit-2			
<i>C. elegans</i> vit-3			
<i>C. elegans</i> vit-4			
<i>C. elegans</i> vit-5			
<i>C. elegans</i> vit-6			
<i>G. ellingtonae</i> 016			
<i>G. ellingtonae</i> 060			
<i>G. pallida</i> 0128			
<i>G. pallida</i> 0564			
<i>G. pallida</i> 7024			
<i>G. pallida</i> 9459			
<i>G. rostochiensis</i> 2537			
<i>G. rostochiensis</i> 2714			
<i>G. rostochiensis</i> 6238			
<i>H. glycines</i> 13367			
<i>H. glycines</i> 23424			
<i>H. glycines</i> 6510			
<i>H. schachtii</i> 007			
<i>H. schachtii</i> 19700			
<i>M. hapla</i> 3631			
<i>M. incognita</i> 11129			
<i>M. incognita</i> 17578			
<i>M. incognita</i> 24409			
<i>M. incognita</i> 28653			
<i>M. incognita</i> 30865			
<i>M. incognita</i> 6902			
<i>M. incognita</i> 9435			
<i>M. javanica</i> 15377			
<i>M. javanica</i> 21515			
<i>M. javanica</i> 29487			
<i>M. javanica</i> 3127			
<i>M. javanica</i> 3894			
<i>M. javanica</i> 43787			
<i>M. javanica</i> 5192			
<i>P. scribneri</i> 001			
<i>P. scribneri</i> 017			
<i>P. scribneri</i> 019			
<i>P. scribneri</i> 064			
<i>P. scribneri</i> 111	X		
<i>P. scribneri</i> 147	95.07	X	
<i>P. scribneri</i> 032	43.91	43.74	X

3.3.4 Phylogenetic Relationships of Vitellogenin

A phylogenetic tree was constructed to infer the relationships between vitellogenin genes in cyst nematodes, root-knot nematodes, a root-lesion nematode *P. scribneri* and Rhabditida species, *C. elegans* and *C. briggsae*. Figure 3.1 was adapted from the phylogenetic tree constructed in Wileman *et al.* (2023). The phylogenetic tree presented here (Figure 3.1) now includes *P. scribneri*. This tree is split into five groups: root-knot nematodes, *P. scribneri*, two clusters of cyst nematodes and a final group of *Caenorhabditis* species. The majority of PPN genes cluster as a large group separate to the *Caenorhabditis* species; however, the small second cluster of cyst nematode genes are closer to the *Caenorhabditis* species group. Within each cluster, there were mixed locations for each gene, and they were not clustered based on their species.

The larger cluster of cyst nematode genes contained small groupings of more closely related genes, that had relatively high percentage identity (Figure 3.1). For example, *H. schachtii* 007 and *H. glycines* 23424 had a percentage identity of 96.1% and were next to each other on the phylogenetic tree. A second cluster of cyst nematodes appeared to be an outgroup and was closer to *Caenorhabditis* species. This outgroup contained the genes: *H. glycines* 13367, *H. schachtii* 19700 and *G. pallida* 0564, each had relatively low sequence conservation when compared to *vit* genes within their own species.

The root-knot nematode cluster within Figure 3.1, comprising *M. hapla*, *M. incognita* and *M. javanica*, was split into two main groups. The small outgroup of this cluster included 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 majority of the recently included *P. scribneri* genes were in a cluster close to, but separate from, the root-knot nematodes, with two *P. scribneri* genes as an outgroup. Within the main cluster, this was separated into two groups of two and then three genes.

The divergence of genes between *C. elegans* and a closely related species, *C. briggsae*, was considered. Figure 3.1 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, *vit-1* and *vit-2*, from *C. elegans*, were more closely related to *C. briggsae* 002 and 16767, whereas *C. elegans* *vit-3*, *vit-4* and *vit-5* were more closely related to *C. briggsae* 14203 and 14234.

3.3.5 Protein Domains in Vitellogenin

Predicted vitellogenin protein domains illustrated in Figure 3.2, determined by Pfam (Bateman *et al.*, 2004), show the approximate location of each protein domain within the genes, and is adapted from Wileman, *et al.* (2023). The predicted amino acid position of each protein domain for each gene is given in Table 3.4. Pfam was used to determine the protein domains in all genes found in BLAST searches as given in Table 3.1. 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 3.2 clearly illustrates that *C. elegans* and *C. briggsae* vitellogenin domains are present at similar positions. Similarly, 11 genes from the cyst nematode species have protein domains in equivalent positions but with varied sequence length. 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 pairs of *M. incognita* vitellogenin genes (28653 and 30865) (24409 and 6902) have almost identical domain positions (Table 3.4). Many of the *M. javanica* predicted vitellogenin genes domains are in comparable positions, excluding *M. javanica* 5192.

The length of the vitellogenin genes in *C. elegans* and *C. briggsae* are very similar; however, the *C. elegans* vit-6 and *C. briggsae* 006 are approximately 50 amino acids longer than the other genes in this group. The *Caenorhabditis* species group of genes are generally smaller than those found in PPN. The average length of PPN vit genes was 1874 amino acids, whereas in the *Caenorhabditis* species the average length was 1625 amino acids. The smallest vit gene was *G. ellingtonae* 060 containing 1337 amino acids, whereas the largest vit gene was *G. pallida* 0564 with 2545 amino acids.

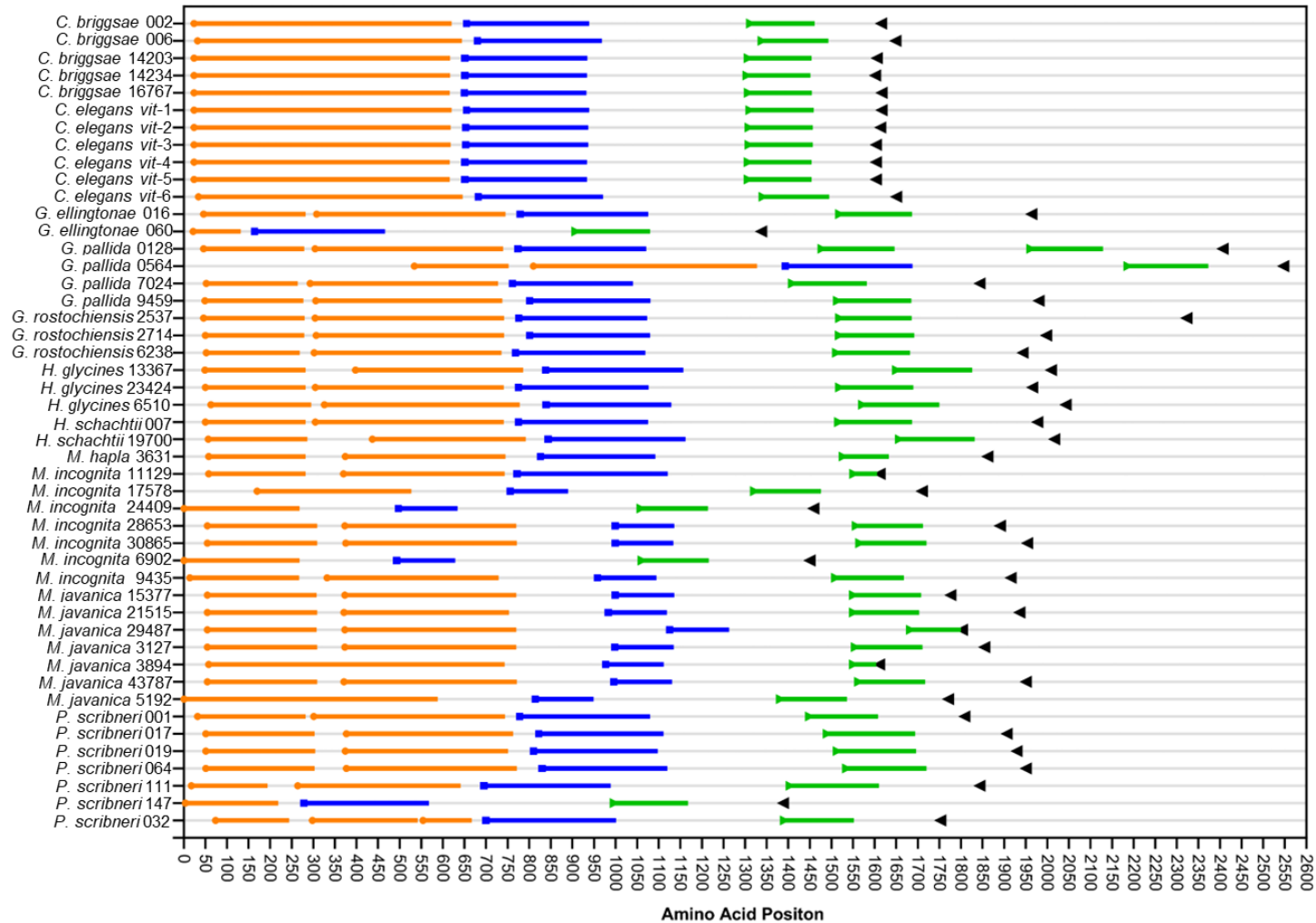


Figure 3.2: Schematic summary of domain structures of vitellogenin proteins of species of *Heterodera*, *Globodera*, *Meloidogyne*, *Pratylenchus* and *Caenorhabditis*. Orange boxes represent the domains of Vitellogenin_N; blue boxes represent vitellogenin open beta-sheet domains; green boxes represent von Willebrand factor type D domains and, black arrows represent the end of the sequence. The diagram is not drawn to scale.

Table 3.4: Amino acid position of protein domains in putative vitellogenin proteins in species of *Caenorhabditis*, *Globodera*, *Heterodera*, *Meloidogyne* and *Pratylenchus*. Domain position indicated by start and end point derived from Pfam (Bateman *et al.*, 2004), some genes were predicted to have the same domain in multiple positions, indicated by values. Full length of amino acid sequence given by total residues.

Protein Domain Amino Acid Position/ Gene Description	Vitellogenin _N		Vitellogenin _N		Vitellogenin _N		Open β - Sheet		von Willebrand factor, Type D		von Willebrand factor, Type D		Total amino acid residues
	Start	End	Start	End	Start	End	Start	End	Start	End	Start	End	
<i>C. briggsae</i> 002	24	620					655	939	1314	1461			1615
<i>C. briggsae</i> 006	32	644					680	968	1340	1493			1648
<i>C. briggsae</i> 14203	24	617					651	935	1308	1454			1605
<i>C. briggsae</i> 14234	24	617					651	934	1305	1451			1601
<i>C. briggsae</i> 16767	24	616					650	933	1309	1455			1616
<i>C. elegans vit-1</i>	24	620					655	939	1312	1459			1616
<i>C. elegans vit-2</i>	24	618					653	937	1310	1457			1613
<i>C. elegans vit-3</i>	24	618					653	937	1310	1457			1603
<i>C. elegans vit-4</i>	24	616					651	934	1308	1454			1603
<i>C. elegans vit-5</i>	24	616					651	934	1308	1454			1603
<i>C. elegans vit-6</i>	34	646					682	971	1342	1495			1650
<i>G. ellingtonae</i> 016	46	282	308	745			779	1076	1520	1687			1963
<i>G. ellingtonae</i> 060	22	132					164	466	909	1080			1337
<i>G. pallida</i> 0128	46	279	304	740			774	1071	1479	1646	1963	2129	2406
<i>G. pallida</i> 0564	534	752	810	1328			1393	1688	2188	2373			2546
<i>G. pallida</i> 7024	52	264	293	728			761	1040	1411	1582			1843
<i>G. pallida</i> 9459	49	277	306	738			801	1081	1515	1685			1980
<i>G. rostochiensis</i> 2537	46	280	304	742			776	1073	1520	1686			2322

Protein Domain	Vitellogenin _N		Vitellogenin _N		Vitellogenin _N		Open β - Sheet		von Willebrand factor, Type D		von Willebrand factor, Type D		Total amino acid residues
Amino Acid Position/ Gene Description	Start	End	Start	End	Start	End	Start	End	Start	End	Start	End	
<i>G. rostochiensis</i> 2714	50	279	307	742			801	1080	1519	1692			1997
<i>G. rostochiensis</i> 6238	52	269	302	736			768	1069	1513	1682			1943
<i>H. glycines</i> 13367	49	282	398	786			838	1157	1652	1826			2008
<i>H. glycines</i> 23424	50	282	305	741			775	1077	1520	1690			1965
<i>H. glycines</i> 6510	63	295	326	778			839	1129	1573	1750			2042
<i>H. schachtii</i> 007	50	282	305	741			775	1075	1518	1687			1977
<i>H. schachtii</i> 19700	57	286	437	792			844	1162	1658	1832			2016
<i>M. hapla</i> 3631	58	282	374	745			826	1092	1528	1633			1861
<i>M. incognita</i> 11129	58	282	370	743			772	1121	1553	1606			1611
<i>M. incognita</i> 17578	170	527					756	890	1323	1476			1709
<i>M. incognita</i> 24409	1	268					497	634	1060	1214			1458
<i>M. incognita</i> 28653	55	309	373	770			999	1136	1558	1712			1890
<i>M. incognita</i> 30865	55	309	375	771			999	1134	1567	1720			1953
<i>M. incognita</i> 6902	1	268					493	629	1063	1216			1449
<i>M. incognita</i> 9435	14	267	332	729			958	1095	1511	1668			1915
<i>M. javanica</i> 15377	55	308	373	770			999	1136	1552	1708			1776
<i>M. javanica</i> 21515	55	309	371	753			983	1119	1552	1703			1935
<i>M. javanica</i> 29487	55	308	373	770			1125	1263	1684	1739	1738	1800	1802
<i>M. javanica</i> 3127	55	309	373	770			998	1135	1556	1711			1854
<i>M. javanica</i> 3894	58	743					977	1112	1552	1604			1610
<i>M. javanica</i> 43787	55	309	371	771			996	1131	1564	1717			1950

Protein Domain	Vitellogenin _N		Vitellogenin _N		Vitellogenin _N		Open β - Sheet		von Willebrand factor, Type D		von Willebrand factor, Type D		Total amino acid residues
Amino Acid Position/ Gene Description	Start	End	Start	End	Start	End	Start	End	Start	End	Start	End	
<i>M. javanica</i> 5192	1	588					814	949	1383	1536			1770
<i>P. scribneri</i> 001	32	282	301	744			778	1080	1450	1608			1808
<i>P. scribneri</i> 017	51	303	377	763			822	1111	1491	1694			1906
<i>P. scribneri</i> 019	51	304	374	751			810	1098	1514	1696			1929
<i>P. scribneri</i> 032	74	244	298	542	554	667	700	1001	1392	1552			1752
<i>P. scribneri</i> 064	51	303	377	771			830	1120	1536	1720			1950
<i>P. scribneri</i> 111	18	194	264	641			695	989	1405	1610			1843
<i>P. scribneri</i> 147	4	219					278	568	998	1168			1388

3.3.6 Gene Ontology

Gene ontology (GO) is used to describe gene function and cellular location (Ashburner *et al.*, 2000; Aleksander *et al.*, 2023), based on the inputted gene sequence. The result of the gene ontology search is given in Table 3.5. Most of the putative vitellogenin genes have a GO description of lipid transporter activity and lipid transport (44/ 47 genes). The only genes not possessing these descriptions are the cyst nematode outgroups exhibited in Figure 3.1 (*G. pallida* 0564, *H. glycines* 13367 and *H. schachtii* 19700). Many sequences (41/ 47 genes) are described to have nutrient reservoir activity, excluding genes *G. pallida* 0564, *H. glycines* 13367, *H. schachtii* 19700, *M. javanica* 3894 and *M. javanica* 5192. Two closely related genes, *C. elegans vit-6* and *C. briggsae* 006, both have additional GO terms associated with their sequences, which predicted function includes the positive regulation of cholesterol import, negative regulation of protein oxidation and negative regulation of reactive oxygen species (ROS) biosynthetic process. The *C. briggsae* 006 gene was also described to have roles in intracellular protein transport and small GTPase binding. Despite that *G. pallida* 0564 contains vitellogenin related protein domains, associated GO descriptions show involvement in endocytic recycling and protein transport. By contrast, *H. glycines* 13367 and *H. schachtii* 19700, GO descriptions alluded to potential functions of cysteine-type deubiquitinase activity, protein deubiquitination and being involved in the ubiquitin-dependent protein catabolic process.

Some sequences have had GO terms associated with location in cellular components. All *C. elegans* and *C. briggsae vit* genes sequences have a GO term associated with the extracellular region; however, *C. elegans vit-2* is also described to be present in the vesicle lumen and yolk granule. Furthermore, *H. glycines* 13367 and *H. schachtii* 19700 have GO cellular components described as the cytosol and nucleus, whereas *G. pallida* 0564 GO term relating to the cellular component is an endosome.

Table 3.5: Putative *vit* genes with associated GO terms and descriptions, retrieved from InterPro using gene sequences.

Gene	GO Term	GO Description
<i>H. glycines</i> 13367	GO:0004197	Cysteine-type deubiquitinase activity
<i>H. schachtii</i> 19700	GO:0004197	Cysteine-type deubiquitinase activity
<i>H. glycines</i> 13367	GO:0004843	Cysteine-type deubiquitinase activity
<i>H. schachtii</i> 19700	GO:0004843	Cysteine-type deubiquitinase activity
<i>C. briggsae</i> 002	GO:0005319	Lipid transporter activity
<i>C. briggsae</i> 006	GO:0005319	Lipid transporter activity
<i>C. briggsae</i> 14203	GO:0005319	Lipid transporter activity
<i>C. briggsae</i> 14234	GO:0005319	Lipid transporter activity
<i>C. briggsae</i> 16767	GO:0005319	Lipid transporter activity
<i>C. elegans vit-1</i>	GO:0005319	Lipid transporter activity
<i>C. elegans vit-2</i>	GO:0005319	Lipid transporter activity
<i>C. elegans vit-3</i>	GO:0005319	Lipid transporter activity
<i>C. elegans vit-4</i>	GO:0005319	Lipid transporter activity
<i>C. elegans vit-5</i>	GO:0005319	Lipid transporter activity
<i>C. elegans vit-6</i>	GO:0005319	Lipid transporter activity
<i>G. ellingtonae</i> 016	GO:0005319	Lipid transporter activity
<i>G. ellingtonae</i> 060	GO:0005319	Lipid transporter activity
<i>G. pallida</i> 0128	GO:0005319	Lipid transporter activity
<i>G. pallida</i> 7024	GO:0005319	Lipid transporter activity
<i>G. pallida</i> 9459	GO:0005319	Lipid transporter activity
<i>G. rostochiensis</i> 2537	GO:0005319	Lipid transporter activity
<i>G. rostochiensis</i> 2714	GO:0005319	Lipid transporter activity
<i>G. rostochiensis</i> 6238	GO:0005319	Lipid transporter activity
<i>H. glycines</i> 23424	GO:0005319	Lipid transporter activity
<i>H. glycine</i> 6510	GO:0005319	Lipid transporter activity
<i>H. schachtii</i> 007	GO:0005319	Lipid transporter activity
<i>M. hapla</i> 3631	GO:0005319	Lipid transporter activity
<i>M. incognita</i> 11129	GO:0005319	Lipid transporter activity
<i>M. incognita</i> 17578	GO:0005319	Lipid transporter activity
<i>M. incognita</i> 24409	GO:0005319	Lipid transporter activity
<i>M. incognita</i> 28653	GO:0005319	Lipid transporter activity
<i>M. incognita</i> 30865	GO:0005319	Lipid transporter activity
<i>M. incognita</i> 6902	GO:0005319	Lipid transporter activity
<i>M. incognita</i> 9435	GO:0005319	Lipid transporter activity
<i>M. javanica</i> 15377	GO:0005319	Lipid transporter activity
<i>M. javanica</i> 21515	GO:0005319	Lipid transporter activity
<i>M. javanica</i> 29847	GO:0005319	Lipid transporter activity
<i>M. javanica</i> 3127	GO:0005319	Lipid transporter activity
<i>M. javanica</i> 3894	GO:0005319	Lipid transporter activity
<i>M. javanica</i> 43787	GO:0005319	Lipid transporter activity
<i>M. javanica</i> 5192	GO:0005319	Lipid transporter activity
<i>P. scribneri</i> 001	GO:0005319	Lipid transporter activity
<i>P. scribneri</i> 017	GO:0005319	Lipid transporter activity
<i>P. scribneri</i> 019	GO:0005319	Lipid transporter activity
<i>P. scribneri</i> 032	GO:0005319	Lipid transporter activity
<i>P. scribneri</i> 064	GO:0005319	Lipid transporter activity
<i>P. scribneri</i> 111	GO:0005319	Lipid transporter activity
<i>P. scribneri</i> 147	GO:0005319	Lipid transporter activity
<i>C. briggsae</i> 002	GO:0005576	Extracellular region

Gene	GO Term	GO Description
<i>C. briggsae</i> 006	GO:0005576	Extracellular region
<i>C. briggsae</i> 14203	GO:0005576	Extracellular region
<i>C. briggsae</i> 14234	GO:0005576	Extracellular region
<i>C. briggsae</i> 16767	GO:0005576	Extracellular region
<i>C. elegans vit-1</i>	GO:0005576	Extracellular region
<i>C. elegans vit-2</i>	GO:0005576	Extracellular region
<i>C. elegans vit-3</i>	GO:0005576	Extracellular region
<i>C. elegans vit-4</i>	GO:0005576	Extracellular region
<i>C. elegans vit-5</i>	GO:0005576	Extracellular region
<i>C. elegans vit-6</i>	GO:0005576	Extracellular region
<i>H. glycines</i> 13367	GO:0005634	Nucleus
<i>H. schachtii</i> 19700	GO:0005634	Nucleus
<i>G. pallida</i> 0564	GO:0005768	Endosome
<i>H. glycines</i> 13367	GO:0005829	Cytosol
<i>H. schachtii</i> 19700	GO:0005829	Cytosol
<i>H. glycines</i> 13367	GO:0006511	Ubiquitin-dependent protein catabolic process
<i>H. schachtii</i> 19700	GO:0006511	Ubiquitin-dependent protein catabolic process
<i>C. briggsae</i> 002	GO:0006869	Lipid transport
<i>C. briggsae</i> 006	GO:0006869	Lipid transport
<i>C. briggsae</i> 14203	GO:0006869	Lipid transport
<i>C. briggsae</i> 14234	GO:0006869	Lipid transport
<i>C. briggsae</i> 16767	GO:0006869	Lipid transport
<i>C. elegans vit-1</i>	GO:0006869	Lipid transport
<i>C. elegans vit-2</i>	GO:0006869	Lipid transport
<i>C. elegans vit-3</i>	GO:0006869	Lipid transport
<i>C. elegans vit-4</i>	GO:0006869	Lipid transport
<i>C. elegans vit-5</i>	GO:0006869	Lipid transport
<i>C. elegans vit-6</i>	GO:0006869	Lipid transport
<i>G. ellingtonae</i> 016	GO:0006869	Lipid transport
<i>G. ellingtonae</i> 060	GO:0006869	Lipid transport
<i>G. pallida</i> 0128	GO:0006869	Lipid transport
<i>G. pallida</i> 7024	GO:0006869	Lipid transport
<i>G. pallida</i> 9459	GO:0006869	Lipid transport
<i>G. rostochiensis</i> 2537	GO:0006869	Lipid transport
<i>G. rostochiensis</i> 2714	GO:0006869	Lipid transport
<i>G. rostochiensis</i> 6238	GO:0006869	Lipid transport
<i>H. glycines</i> 23424	GO:0006869	Lipid transport
<i>H. glycine</i> 6510	GO:0006869	Lipid transport
<i>H. schachtii</i> 007	GO:0006869	Lipid transport
<i>M. hapla</i> 3631	GO:0006869	Lipid transport
<i>M. incognita</i> 11129	GO:0006869	Lipid transport
<i>M. incognita</i> 17578	GO:0006869	Lipid transport
<i>M. incognita</i> 24409	GO:0006869	Lipid transport
<i>M. incognita</i> 28653	GO:0006869	Lipid transport
<i>M. incognita</i> 30865	GO:0006869	Lipid transport
<i>M. incognita</i> 6902	GO:0006869	Lipid transport
<i>M. incognita</i> 9435	GO:0006869	Lipid transport
<i>M. javanica</i> 15377	GO:0006869	Lipid transport
<i>M. javanica</i> 21515	GO:0006869	Lipid transport
<i>M. javanica</i> 29847	GO:0006869	Lipid transport
<i>M. javanica</i> 3127	GO:0006869	Lipid transport

Gene	GO Term	GO Description
<i>M. javanica</i> 3894	GO:0006869	Lipid transport
<i>M. javanica</i> 43787	GO:0006869	Lipid transport
<i>M. javanica</i> 5192	GO:0006869	Lipid transport
<i>P. scribneri</i> 001	GO:0006869	Lipid transport
<i>P. scribneri</i> 017	GO:0006869	Lipid transport
<i>P. scribneri</i> 019	GO:0006869	Lipid transport
<i>P. scribneri</i> 032	GO:0006869	Lipid transport
<i>P. scribneri</i> 064	GO:0006869	Lipid transport
<i>P. scribneri</i> 111	GO:0006869	Lipid transport
<i>P. scribneri</i> 147	GO:0006869	Lipid transport
<i>C. briggsae</i> 006	GO:0006886	Intracellular protein transport
<i>G. pallida</i> 0564	GO:0015031	Protein transport
<i>H. glycines</i> 13367	GO:0016579	Protein deubiquitination
<i>H. schachtii</i> 19700	GO:0016579	Protein deubiquitination
<i>C. briggsae</i> 006	GO:0031267	Small GTPase binding
<i>C. elegans vit-2</i>	GO:0031410	Cytoplasmic vesicle
<i>C. elegans vit-2</i>	GO:0031983	Vesicle lumen
<i>G. pallida</i> 0564	GO:0032456	Endocytic recycling
<i>C. elegans vit-2</i>	GO:0042718	Yolk granule
<i>C. briggsae</i> 002	GO:0045735	Nutrient reservoir activity
<i>C. briggsae</i> 006	GO:0045735	Nutrient reservoir activity
<i>C. briggsae</i> 14203	GO:0045735	Nutrient reservoir activity
<i>C. briggsae</i> 14234	GO:0045735	Nutrient reservoir activity
<i>C. briggsae</i> 16767	GO:0045735	Nutrient reservoir activity
<i>C. elegans vit-1</i>	GO:0045735	Nutrient reservoir activity
<i>C. elegans vit-2</i>	GO:0045735	Nutrient reservoir activity
<i>C. elegans vit-3</i>	GO:0045735	Nutrient reservoir activity
<i>C. elegans vit-4</i>	GO:0045735	Nutrient reservoir activity
<i>C. elegans vit-5</i>	GO:0045735	Nutrient reservoir activity
<i>C. elegans vit-6</i>	GO:0045735	Nutrient reservoir activity
<i>G. ellingtonae</i> 016	GO:0045735	Nutrient reservoir activity
<i>G. ellingtonae</i> 060	GO:0045735	Nutrient reservoir activity
<i>G. pallida</i> 0128	GO:0045735	Nutrient reservoir activity
<i>G. pallida</i> 7024	GO:0045735	Nutrient reservoir activity
<i>G. pallida</i> 9459	GO:0045735	Nutrient reservoir activity
<i>G. rostochiensis</i> 2537	GO:0045735	Nutrient reservoir activity
<i>G. rostochiensis</i> 2714	GO:0045735	Nutrient reservoir activity
<i>G. rostochiensis</i> 6238	GO:0045735	Nutrient reservoir activity
<i>H. glycines</i> 23424	GO:0045735	Nutrient reservoir activity
<i>H. schachtii</i> 007	GO:0045735	Nutrient reservoir activity
<i>M. hapla</i> 3631	GO:0045735	Nutrient reservoir activity
<i>M. incognita</i> 11129	GO:0045735	Nutrient reservoir activity
<i>M. incognita</i> 17578	GO:0045735	Nutrient reservoir activity
<i>M. incognita</i> 24409	GO:0045735	Nutrient reservoir activity
<i>M. incognita</i> 28653	GO:0045735	Nutrient reservoir activity
<i>M. incognita</i> 30865	GO:0045735	Nutrient reservoir activity
<i>M. incognita</i> 6902	GO:0045735	Nutrient reservoir activity
<i>M. incognita</i> 9435	GO:0045735	Nutrient reservoir activity
<i>M. javanica</i> 15377	GO:0045735	Nutrient reservoir activity
<i>M. javanica</i> 21515	GO:0045735	Nutrient reservoir activity
<i>M. javanica</i> 29847	GO:0045735	Nutrient reservoir activity

Gene	GO Term	GO Description
<i>M. javanica</i> 3127	GO:0045735	Nutrient reservoir activity
<i>M. javanica</i> 43787	GO:0045735	Nutrient reservoir activity
<i>P. scribneri</i> 001	GO:0045735	Nutrient reservoir activity
<i>P. scribneri</i> 017	GO:0045735	Nutrient reservoir activity
<i>P. scribneri</i> 019	GO:0045735	Nutrient reservoir activity
<i>P. scribneri</i> 032	GO:0045735	Nutrient reservoir activity
<i>P. scribneri</i> 064	GO:0045735	Nutrient reservoir activity
<i>P. scribneri</i> 111	GO:0045735	Nutrient reservoir activity
<i>P. scribneri</i> 147	GO:0045735	Nutrient reservoir activity
<i>C. briggsae</i> 006	GO:1903427	Negative regulation of reactive oxygen species biosynthetic process
<i>C. elegans vit-6</i>	GO:1903427	Negative regulation of reactive oxygen species biosynthetic process
<i>C. briggsae</i> 006	GO:1904109	Positive regulation of cholesterol import
<i>C. elegans vit-6</i>	GO:1904109	Positive regulation of cholesterol import
<i>C. briggsae</i> 006	GO:1904807	Negative regulation of protein oxidation
<i>C. elegans vit-6</i>	GO:1904807	Negative regulation of protein oxidation

3.3.7 Genome Availability

This study has been reliant on the data availability from whole genome sequencing of PPN genomes. Further genomes have subsequently become available for *G. pallida* and *G. rostochiensis* hosted on Wormbase Parasite. Three genomes are now available for *G. pallida* from Cotton *et al.* (2014), van Steenbrugge *et al.* (2023) and an unpublished genome (Bioproject: PRJNA702104). Additionally, there are now three genomes available for *G. rostochiensis* from Eves-van den Akker *et al.* (2016) and two from (van Steenbrugge *et al.*, 2021). These genomes were compared to determine whether putative *vit* genes from each genome were comparable. Table 3.6 illustrates the pairwise identity values of putative *vit* genes from three *G. pallida* genomes. All genes from the Cotton *et al.* (2014) genome had sequences resembling those found in other genomes. The gene *G. pallida* 9459 and *G. pallida* 0564 had 87.38% and 95.24% percentage identity to van Steenbrugge *et al.* (2023) genes, *G. pallida* 5978 and *G. pallida* 4789, respectively. Similarly, *G. pallida* 0128 shared 94.38% percentage identity with *G. pallida* 7045 (van Steenbrugge *et al.*, 2023) and 94.44% with *G. pallida* 5538, from the unpublished genome. Finally, *G. pallida* 7024 was shown to share 83.62% percentage identity to the predicted *G. pallida*_90 gene from the unpublished genome.

The three available genomes of *G. rostochiensis* were also compared (Table 3.7). Three genes from different genomes, *G. rostochiensis* 6238, 2260 and 16066, were all found to be 100% identical. Similarly, three further genes, *G. rostochiensis* 2537, 2396 and 8204, from different genomes had an average of 98.5% percentage identity. Also, *G. rostochiensis* 2714, 10898 and 8560, each from different genomes, shared an average of 99.16% percentage identity.

Table 3.6: Pairwise alignment of vitellogenin genes from three *G. pallida* genomes. Percentage identity (%) for each pairwise alignment is given; values given in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%).

	<i>G. pallida</i> 5538	<i>G. pallida</i> 9459	<i>G. pallida</i> 0128	<i>G. pallida</i> 7024	<i>G. pallida</i> 0564	<i>G. pallida</i> 4789	<i>G. pallida</i> 5968	<i>G. pallida</i> 7045	Predicted GP_90
<i>G. pallida</i> 5538	X								
<i>G. pallida</i> 9459	68.84	X							
<i>G. pallida</i> 0128	94.44	66.00	X						
<i>G. pallida</i> 7024	68.63	63.32	65.71	X					
<i>G. pallida</i> 0564	24.38	24.48	23.31	24.33	X				
<i>G. pallida</i> 4789	24.7	24.84	23.64	24.78	95.24	X			
<i>G. pallida</i> 5968	63.63	87.38	60.63	58.4	23.71	23.62	X		
<i>G. pallida</i> 7045	94.69	70.85	94.38	71.09	25.16	26.1	66.87	X	
Predicted GP_90	72.46	65.83	66.81	83.62	24.79	24.69	58.43	53.19	X

Table 3.7: Pairwise alignment of vitellogenin genes from three *G. rostochiensis* genomes. Percentage identity (%) for each pairwise alignment is given; values given in a relative colour scale of red for low similarity (20-29%), orange/ yellow for medium similarity (29-49%) and green for high similarity (>50%).

	<i>G. ros</i> 6238	<i>G. ros</i> 2537	<i>G. ros</i> 2714	<i>G. ros</i> 10898	<i>G. ros</i> 18657	<i>G. ros</i> 2260	<i>G. ros</i> 2396	<i>G. ros</i> 16066	<i>G. ros</i> 6962	<i>G. ros</i> 8204	<i>G. ros</i> 8560
<i>G. rostochiensis</i> 6238	X										
<i>G. rostochiensis</i> 2537	74.02	X									
<i>G. rostochiensis</i> 2714	68.99	69.64	X								
<i>G. rostochiensis</i> 10898	69.03	69.72	98.7	X							
<i>G. rostochiensis</i> 18657	25.31	25.33	25.25	25.19	X						
<i>G. rostochiensis</i> 2260	100	74.02	69.14	69.18	25.29	X					
<i>G. rostochiensis</i> 2396	74.09	98.78	69.55	69.57	25.13	74.09	X				
<i>G. rostochiensis</i> 16066	100	74.02	69.14	69.18	25.29	100	74.04	X			
<i>G. rostochiensis</i> 6962	25.31	25.33	25.25	25.19	99.55	25.31	25.21	25.31	X		
<i>G. rostochiensis</i> 8204	73.92	98.27	69.34	69.36	25.12	73.92	98.57	73.92	25.12	X	
<i>G. rostochiensis</i> 8560	72.27	72.74	99.62	100	27.58	72.27	72.11	72.27	27.58	71.7	X

Discussion

This work has highlighted the presence of *vit* genes within the genomes of cyst, root-knot and root-lesion nematodes that have not previously been extensively studied. The presence of *vit* 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 was conserved between the cyst, RKN and RLN, and *Caenorhabditis vit* 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 *vit* 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 and Lehner, 2019); it is possible that multiple genes in PPN are required to form similar complexes. Data presented here illustrated the

relationship between these paralogous genes, with each species having differing numbers of genes. Only one *vit* gene from the genome of *M. hapla* was identified that contained all three protein domains, whereas the genome of *M. incognita* had 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).

In phylogenetic analysis, the species were grouped based on their defining characteristics; the cyst nematodes were clustered separately to the RKN, RLN and *Caenorhabditis* sp. *vit* genes, but they were not clustered by species. It could be anticipated that the groups would form separate clusters as *Caenorhabditis* spp. is part of the order Rhabditida, whereas the PPN reported here are in the order Tylenchida. The phylogenetic analysis presented here supports the findings of Scholl and Bird (2005). This study used expressed sequence tag data to determine phylogenetic relationships between *C. elegans* and several PPN species; with phylogenetic trees also giving *C. elegans* as an outgroup and grouped RKN species separately to cyst nematodes. In work presented here, the *vit* genes were not clustered by species, this may be due to the variation required to encode different protein subunits.

The genome of *G. ellingtonae* contained two *vit* genes not previously predicted, 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 evaluate further the nematode species against other members of the *Globodera* family.

Similarly, the genome of *P. scribneri* contained seven *vit* genes that were not previously predicted. These genes formed their own cluster in the phylogenetic tree but had the most sequence homology to the root-knot nematodes. It has been suggested that the divergence between the root-knot and root-lesion nematodes occurred much later than the last common ancestor with the cyst nematodes (Qing *et al.*, 2023). Additionally, the *P. scribneri* *vit* genes demonstrated a high proportion of sequence homology between paralogous *vit* genes. Four genes had a percentage identity of greater than 91%. This is comparable to the relationships between *C. elegans* *vit*-3, *vit*-4 and *vit*-5, which are all expressed and encode the YP170A protein subunit. It is possible that these closely related *P. scribneri* genes are utilised to encode a subunit with a similar function.

The genomes of species within the cyst nematode genus *Heterodera* contain *vit* genes; this work highlights their presence within the genomes of *H. schachtii* and *H. glycines*. Two *H. schachtii* genes were shown to be closely related to those in *H. glycines*; however, a third *H. glycines* gene appeared 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 sequenced 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). In contrast to this finding, as described in Wileman *et al.*, (2023) and presented here, the *H. schachtii vit* genes *H. schachtii* 007 (Hsc_gene_6726) and *H. schachtii* 19700 (Hsc_gene_19700) contain an open beta-sheet domain, which was not previously annotated.

The phylogenetic analysis resulted in two clusters of cyst nematode *vit* genes, a larger cluster containing 11 genes and a smaller cluster of three genes. This second smaller cluster had less than 30% percentage identity to other cyst nematode *vit* genes. Interestingly, these genes with divergent sequences also contained protein domains found in *vit* genes; however, these proteins may not be functional as vitellogenin and may have alternative functions. In gene ontology searches, the predicted function of these genes was varied, *H. glycines* 13367 and *H. schachtii* 19700 were described as having protein deubiquitination activity and *G. pallida* 0564 is involved in endocytic recycling. It is possible that the divergence in sequence results in alternate secondary and tertiary structures that provide varied function.

The relationship between 15 *vit* genes from root-knot nematode species has been illustrated, showing a high level of sequence conservation between most of these genes, predominantly between the seven *M. incognita* and seven *M. javanica vit* genes. The only *vit* gene identified in *M. hapla* was an outlier of this group. This result is consistent with previous phylogenetic study utilising 18S rRNA gene sequencing (Álvarez-Ortega *et al.*, 2019), which found both *M. incognita* and *M. javanica* as part of the same cluster, whereas *M. hapla* was separated into a different cluster. Currently, no studies have highlighted the presence of vitellogenin within the genomes of *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 Bellafore *et al.* (2008), five vitellogenin genes were found in the secreted effectors of female pharyngeal glands using mass spectrometry. A similar study of the effectors produced by female pharyngeal glands in *G. rostochiensis* resulted in no vitellogenin genes being detected (Maier *et al.*, 2012).

As vitellogenin plays a key role in the development of the nematode, it could be anticipated that it would have a relatively conserved gene 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 and Blumenthal, 1989); therefore, *vit* genes in *C. elegans* and *C. briggsae* are highly conserved and are likely to have a similar protein structure. Furthermore, *C. elegans* and *C. briggsae* *vit* genes are located on the X chromosome, excluding *C. elegans* *vit-6* and *C. briggsae* 006, which are located on chromosome IV. By contrast, the location of PPN *vit* genes in their respective genomes are more varied.

Protein domains present in *C. elegans* vitellogenin appear within the identified *vit* genes of cyst, root-knot and root-lesion nematodes. Using Pfam to predict the location of protein domains, *C. elegans*, *C. briggsae*, cyst, root-knot and root-lesion nematode *vit* genes were shown to contain three protein domains: Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor type D domain. These domains were positioned in the same order for each gene but had varied positions along the length of the gene. Both the Vitellogenin_N domain and the open beta sheet domains are involved in transporting lipids (Smolenaars *et al.*, 2007), whereas the von Willebrand factor type D domain has adhesive properties (Baker, 1988a). The domain Vitellogenin_N has been described to form a cavity to allow the binding of lipids (Zhai *et al.*, 2024). The presence of the three domains within PPN *vit* genes may provide these functions to the encoded genes.

Gene ontology searches of *vit* gene sequences resulted in many genes being annotated with lipid transporter activity, lipid transport and nutrient reservoir activity. The annotation of GO:0006869 and GO:0005319 relates to lipid transport and is defined as the use of a transporter or pore to move lipids between cellular locations (Binns *et al.*, 2009). In *C. elegans*, vitellogenin proteins have been shown to be involved in the transport of cholesterol (Matyash *et al.*, 2001). Furthermore, *C. elegans* *vit* genes have also been shown to be homologous to the human apolipoprotein B-100 protein, a lipid transporter (Baker, 1988b). Intriguingly, the GO term of nutrient reserve was also used to describe PPN *vit* genes and has been described as a potential function in many species, including *C. elegans* (Perez and Lehner, 2019). The nutrient provision of amino acids to the developing progeny has been suggested by Winter *et al.* (1996). The role of a nutrient reserve may also explain the nematodes' requirement for several *vit* genes of related but non-homologous gene sequence to provide a greater diversity of amino acids.

The construction of phylogenetic trees and bioinformatic analysis is often reliant on the quality of the data available. With the increasing number of PPN genomes published, there is an

opportunity to utilise these data for phylogenetic study. In 2017, fewer than twenty PPN genomes had been published (Kikuchi *et al.*, 2017); however, a greater number of genomes are now being sequenced utilising techniques such as long read sequencing. The development of sequencing technologies has also made it easier for laboratories to sequence and reanalyse the genomes of organisms previously sequenced, such as *G. pallida* and *G. rostochiensis*. Therefore, it was possible to make a comparative study of *vit* genes of the same species from different genomes. The putative *vit* genes from earlier sequencing studies of *G. pallida* (Cotton *et al.*, 2014) and *G. rostochiensis* (Eves-van den Akker *et al.*, 2016) had high sequence similarity to gene sequences in more recently available genomes (van Steenbrugge *et al.*, 2021; van Steenbrugge *et al.*, 2023). This gave more confidence in the presence of these genes within the PCN genomes and in their respective genomic sequences.

This work set out to highlight the presence of and predict *vit* 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 a species of root-lesion nematode. The alignment and phylogenetic study of these genes could be used to infer relationships between the PPN species. Additionally, the protein domains of *vit* genes that appear in *C. elegans* were also present in the PPN. By determining the presence of *vit* 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.

Whilst the presence of *vit* 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 *vit* 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-specific expression of vitellogenin would help to understand their presence and role(s) in PPN.

4 Gene Expression of Vitellogenin in PPN

4.1 Introduction

Gene expression is the process of a gene being transcribed to become mRNA and translated into a protein product; however, not all mRNA is translated (Ramírez-Clavijo and Montoya-Ortíz, 2013). The quantity of mRNA within a cell can be analysed to determine the levels of transcription, its tissue location and if genes are expressed under certain conditions (National Centre for Biotechnology Information, 2017). By studying gene expression, genes of interest can be compared between different species and at different life stages. It is also a useful tool to infer further the role and biological functions of the synthesised proteins. Gene expression can be measured by using Northern blots, qPCR, microarrays and RNA-Seq (Singh *et al.*, 2018). Once data are collected from RNA-seq analysis by aligning the number of reads to each gene, it is commonly normalised to ensure that the values obtained reflect the true biological condition. Various methods of normalisation can be used; firstly, normalisation by library size, including reads per kilobase per million mapped reads (RPKM) and fragments per kilobase per million mapped reads (FPKM). These are calculated by dividing the read count by the total number of reads in the sample, this represents the amount of each gene in relation to the total mRNA (Equation 4.1) (Evans *et al.*, 2018). Another commonly reported normalisation method is gene expression defined as transcripts per million (TPM) (Equation 4.2). TPM is like RPKM and FPKM but for all samples the sum of all TPM values is the same (Zhao *et al.*, 2020). Alternatively, data can be normalised by distribution, by using DESeq or the Trimmed Mean of the M-values (TMM) methods. DESeq is performed by determining the ratio of each read count to the geometric mean of all read counts of that gene from all samples (Evans *et al.*, 2018). In contrast, TMM uses a single sample as a reference and calculates fold changes and expression levels of other samples as a relative value to this reference (Evans *et al.*, 2018).

Equation 4.1: Normalisation calculation for reads per kilobase per million mapped reads (RPKM) and fragments per kilobase per million mapped reads (FPKM) (Zhao *et al.*, 2020).

$$\text{RPKM or FPKM} = 10^9 \times \frac{\text{Reads mapped to transcript}}{\text{Total reads} \times \text{Transcript length}}$$

Equation 4.2: Normalisation calculation for transcripts per million (TPM) (Zhao *et al.*, 2020).

$$\text{TPM} = 10^6 \times \frac{\text{Reads mapped to transcript/Transcript length}}{\text{Sum (Reads mapped to transcript/Transcript length)}}$$

Vitellogenin gene expression has been widely studied in both vertebrates and invertebrates. The expression of vitellogenin in vertebrates has included research of African claw frogs (*Xenopus laevis*) (Baker and Shapiro, 1978), chickens (*Gallus gallus*) (Evans *et al.*, 1988) and the rainbow trout (*Salmo gairdneri*) (Le Guellec *et al.*, 1988). Furthermore, many studies have focused on invertebrates, mainly insects, including the ants *Temnothorax longispinosus* (Kohlmeier *et al.*, 2018) and *Formica fusca* (Morandin *et al.*, 2019); honey bees (*Apis mellifera*) (Colonello-Frattini *et al.*, 2010; Münch *et al.*, 2015); the Asian lady beetle (*Harmonia axyridis*) (Zhang *et al.*, 2017) and the beet armyworm (*Spodoptera exigua*) (Zhao *et al.*, 2018). Other studies centred their research on crustaceans, including the banana shrimp *Penaeus (Litopenaeus) merguensis* (Phiriyangkul and Utarabhand, 2006) and the sea urchin, *Strongylocentrotus purpuratus* (Shyu *et al.*, 1986).

The expression of vitellogenin (*vit*) genes in nematodes has primarily been studied in *Caenorhabditis elegans*. Work has shown that *vit* genes are primarily expressed in the hermaphrodite's intestine (Kimble and Sharrock, 1983). Other early studies into *vit* expression in *C. elegans* found that these genes are abundant in adult hermaphrodites but not in males or larval stages (Blumenthal *et al.*, 1984). Studies have also shown that analysing gene expression of *vit* genes in *C. elegans* is a useful biomarker for monitoring development and the impact of endocrine disrupting chemicals in water systems (Custodia *et al.*, 2001) and in pesticide usage (Moya *et al.*, 2022). The expression of *vit* genes has also been studied in animal-parasitic nematodes, including studies of *Haemonchus contortus* (Hartman *et al.*, 2001), *Trichostrongylus vitrinus* (Nisbet and Gasser, 2004) and *Oesophagostomum dentatum* (Cottee *et al.*, 2006).

No current studies have specifically focused on *vit* gene expression in plant-parasitic nematodes (PPN); however, as a result of investigations into other aspects of the nematode's life cycle or parasitism, some studies have made observations about *vit* gene expression. A study of the genome and life-stage transcriptome of *Globodera pallida* observed that *vit* genes were the most abundantly expressed genes at 28- and 35-days post infection (DPI), which corresponded to the female life stages (Cotton *et al.*, 2014). In a study of the *Pratylenchus penetrans* transcriptome, a high number of FPKM transcripts were associated with vitellogenin (Vieira *et al.*, 2015). Also, a study of expression in pharyngeal gland cells (Bakhetia *et al.*, 2007) utilised *vit* gene expression as a marker for female nematode maturity in *Heterodera glycines*, as the authors believed it was not expressed in second-stage juveniles (J2).

The aim of the research in this chapter is to obtain and utilise gene expression data to enhance the understanding of *vit* gene expression at different life stages in species of PPN. Earlier research which explored gene expression and transcriptomes of *C. elegans*, *C. briggsae* and

PPN, were studied to retrieve gene expression data where available of *vit* genes identified in Chapter 3. This information was also used to determine how they were expressed during different stages of the nematode life cycle.

4.2 Materials and Methods

4.2.1 Data Acquisition

Gene expression data were acquired by literature searches and exploring data deposited on Wormbase and WormBase Parasite. Inclusion criteria of studies were gene expression data collected at different stages in the nematode life cycle and from nematode species previously studied in Chapter 3. Gene expression data for each species were collected from the sources given in Table 4.1, the original study prior to the data deposition on these sites is also provided. Table 4.1 also lists the life stages studied and the type of data used in the analysis of vitellogenin genes. Gene expression levels were given in transcripts per million (TPM) or fragments per kilobase of transcript per million fragments mapped (FPKM) for most species; however, normalised expression or relative expression values were calculated in the original study.

Table 4.1: Summary table of gene expression studies of *Caenorhabditis* species and plant-parasitic nematodes that have included several stages of the nematode life cycle. Data source for analysis and data type from normalisation methods is given.

Species	Study	Data source used in analysis	Nematode life stage	Normalised expression method from data source
<i>C. briggsae</i>	Thomas <i>et al.</i> , 2012	Wormbase Parasite	Mixed L4 and adult pseudo-females and males	TPM
<i>C. briggsae</i>	Grün <i>et al.</i> , 2014	Wormbase Parasite	Mixed embryos, L1 larvae, and young adults	TPM
<i>C. elegans</i>	Hashimshony <i>et al.</i> , 2014	Wormbase, RNAseq FPKM gene search tool	30 - 850 minutes post first-cleavage in embryo	FPKM
<i>C. elegans</i>	Dillman <i>et al.</i> , 2015	Wormbase, RNAseq FPKM gene search tool	Embryo, L1 larvae, dauer and adult	FPKM
<i>C. elegans</i>	Grün <i>et al.</i> , 2014	Wormbase, RNAseq FPKM gene search tool	Embryo, L1 larvae, L2 larvae, L3 larvae, L4 larvae and young hermaphrodite female	FPKM

Species	Study	Data source used in analysis	Nematode life stage	Normalised expression method from data source
<i>G. pallida</i>	Cotton <i>et al.</i> , 2014	Wormbase Parasite	7, 14, 21, 27 and 35 DPI, adult males, J2 and eggs	TPM
<i>G. rostochiensis</i>	Duceppe <i>et al.</i> , 2017	Wormbase Parasite	Dry cysts, hydrated cysts in water for 7 days, hydrated cysts in potato root diffusate for 15 min, 1 h, 8 h, 24 h, 48 h, and 7 days	TPM
<i>G. rostochiensis</i>	Eves-van den Akker <i>et al.</i> , 2016	Paper additional file 7: File S1.	Dormant cysts, hydrated eggs, J2 and 14 DPI	Normalised expression
<i>H. glycines</i>	Gardner <i>et al.</i> , 2018	Wormbase Parasite	Infective J2 from susceptible and resistant hosts, and pre-parasitic J2	TPM
<i>H. schachtii</i>	Siddique <i>et al.</i> , 2022	Paper supplementary data 4	Dormant cysts, pre-infective J2, 10 and 48 HPI, males and females 12 DPI and females 24 DPI	Relative expression
<i>M. incognita</i>	Blanc-Mathieu <i>et al.</i> , 2017	Wormbase Parasite	Eggs, pre-parasitic J2, parasitic J3-J4, adult females and adult males	TPM
<i>M. incognita</i>	Choi <i>et al.</i> , 2017	Wormbase Parasite	Eggs, J2, J3, J4 and adult female	TPM

Each study used various methods to extract and quantify gene expression, from different nematode species. Table 4.2 provides a summary of the key methodology used in each study and how the data were analysed to provide gene expression values; full methodology can be found in the original published studies. Two studies researched *G. rostochiensis* gene expression, which was derived from two different populations; similarly, two studies also considered gene expression in *M. incognita*, which also came from two different populations. In some cases, the normalisation method used in the original study was different to the reported normalisation method provided by Wormbase or Wormbase Parasite; Table 4.1 lists the expression level derived from these sources, whereas Table 4.2 provides the method used in the original study.

Table 4.2: Summary of key methodology used in gene expression studies of *Caenorhabditis elegans*, *C. briggsae* and plant-parasitic nematode species; including the strain or population used, number of replicates, RNA extraction method, reference gene, sequencing method and data normalisation method used.

Species	Study	Strain/ Population	Replicates	RNA extraction method	Reference Genes	Sequencing Method	Normalisation method used in study
<i>C. briggsae</i>	Thomas <i>et al.</i> , 2012	Pseudo- female: AF16- derived <i>she-1</i> (v47). Male: AF16	Three bio reps.	TRIzol-reagent, phenol/ chloroform	Not provided	Illumina Genome Analyzer II or HiSeq 2000	FPKM
<i>C. briggsae</i>	Grün <i>et al.</i> , 2014	Wild type AF16	Two bio reps.	TRIzol-reagent, dynabead mRNA purification kit	Not provided	Illumina Genome Analyzer II	Reads per kilobase of gene locus per million mapped reads (RPKM)
<i>C. elegans</i>	Hashimshony <i>et al.</i> , 2014	Not provided	Three bio reps.	TRIzol-reagent, paramagnetic purification	Not provided	Illumina HiSeq 2000	TPM
<i>C. elegans</i>	Dillman <i>et al.</i> , 2015	Wild type N2	Not provided ¹	Phenol- chloroform extraction using TRIzol	Not provided	Illumina Genome Analyzer IIx	FPKM
<i>C. elegans</i>	Grün <i>et al.</i> , 2014	Wild type N2	Two bio reps.	TRIzol-reagent, dynabead mRNA purification kit	Not provided	Illumina Genome Analyzer II	RPKM
<i>G. pallida</i>	Cotton <i>et al.</i> , 2014	'Lindley' population, a Pa2/3 pathotype, James	Two bio reps.	RNeasy Mini Kit (Qiagen)	Not provided	Illumina Genome Analyser IIx or HiSeq	Mean RPKM

Species	Study	Strain/ Population	Replicates	RNA extraction method	Reference Genes	Sequencing Method	Normalisation method used in study
		Institute, Dundee					
<i>G. rostochien- sis</i>	Duceppe <i>et al.</i> , 2017	Soil samples, collected in 2011, Saint- Amable, Quebec, Canada	Two bio reps.	RNeasy Plus mini kit (Qiagen)	GR, PMP-3 and aaRS. Stable across life stages	Illumina HiSeq 2000	TPM
<i>G. rostochiensis</i>	Eves-van den Akker <i>et al.</i> , 2016	Pathotype Ro1, James Hutton Institute, Dundee	Two bio reps.	RNeasy Mini Kit (Qiagen)	Not provided	Illumina Hiseq	Trimmed Mean of M-values normalisation (TMM)
<i>H. glycines</i>	Gardner <i>et al.</i> , 2018	SCN inbred population PA3 (HG Type 0)	Three bio reps.	PerfectPure Fibrous Tissue Kit (5Prime)	Not provided	Illumina HiSeq. 2500	TPM
<i>H. schachtii</i>	Siddique <i>et al.</i> , 2022	Nematode stock culture, Bonn, Germany	Three bio reps.	RNeasy Plant Mini Kit (Qiagen)	Actin	Illumina NovaSeq 6000	Normalised with DESeq2 v1.22.2
<i>M. incognita</i>	Blanc-Mathieu <i>et al.</i> , 2017	Strain: Morelos, from INRA Sophia Antipolis collection	Three bio reps. ²	TRIzol Reagent (Invitrogen)	Not provided	Illumina NextSeq 500	TPM
<i>M. incognita</i>	Choi <i>et al.</i> , 2017	Population maintained on tomato plants	Three tech. reps.	LiCl ₂ method	Not provided	Illumina HiSeq 2500	Not recorded

1: Number of replicates was not provided; however, each life stage contained a various number of replicates when obtained from Wormbase Parasite.

2: Data available from Wormbase Parasite, only one value recorded.

4.2.2 Data Analysis

Data acquired from the sources listed in Table 4.1 were entered into GraphPad Prism software (version 10.2.3) to produce graphs of mean gene expression levels, with error bars plotted using standard error of the mean. Mean expression level was described as low, moderate, or high; low expression was described as ≤ 10 TPM/FPKM, moderate expression as 11 to 999 TPM/FPKM and high expression as > 1000 TPM/FPKM (EMBL-EBI, 2018).

Statistical analysis was performed using GraphPad Prism software (version 10.2.3), comparing single genes across different nematode life stages. The standard deviation (SD) of the mean was calculated and compared between each condition. The normality of the data was tested by assessing a QQ plot or performing a Shapiro-Wilk test. If the data demonstrated a normal distribution with similar SD, an ordinary one-way ANOVA was performed with a Tukey's multiple comparisons test. If the SD was dissimilar, a Brown-Forsythe and Welch ANOVA test with Dunnett's T3 multiple comparisons test was used. If the data was not normally distributed a nonparametric test, Kruskal-Wallis, was completed, with a Dunn's multiple comparisons test. Multiple comparison tests were performed by comparing the mean rank of each condition to the mean rank of every other condition. Gene expression data from Thomas *et al.* (2012) only contained two conditions, male and pseudo-female. These data were normally distributed but had unequal SD, therefore an unpaired *t*-test with Welch's correction for each gene was used. If the result of the statistical analyses were found to be significant, they were reported to have a *P* value of less than 0.05.

4.3 Results

4.3.1 Vitellogenin Gene Expression in *Caenorhabditis briggsae*

The expression of genes in *Caenorhabditis briggsae* males and pseudofemales was studied by Thomas *et al.* (2012), providing a gene dataset for both sexes, which included the putative vitellogenin genes. Figure 4.1 illustrates the expression level in transcripts per million (TPM), all plotted values are given in a logarithmic scale. The mean expression level in pseudofemales was significantly higher than in males for all *vit* genes ($P < 0.05$). Pseudofemales' mean expression level for all genes was greater than 1000 TPM, ranging from 1875 TPM for *C. briggsae* 14203 to 8537 TPM for *C. briggsae* 006. By contrast, the expression level in males was less than one for most genes, excluding *C. briggsae* 006, which was at 200 TPM. Therefore, pseudofemales highly expressed *vit* genes and males had low expression.

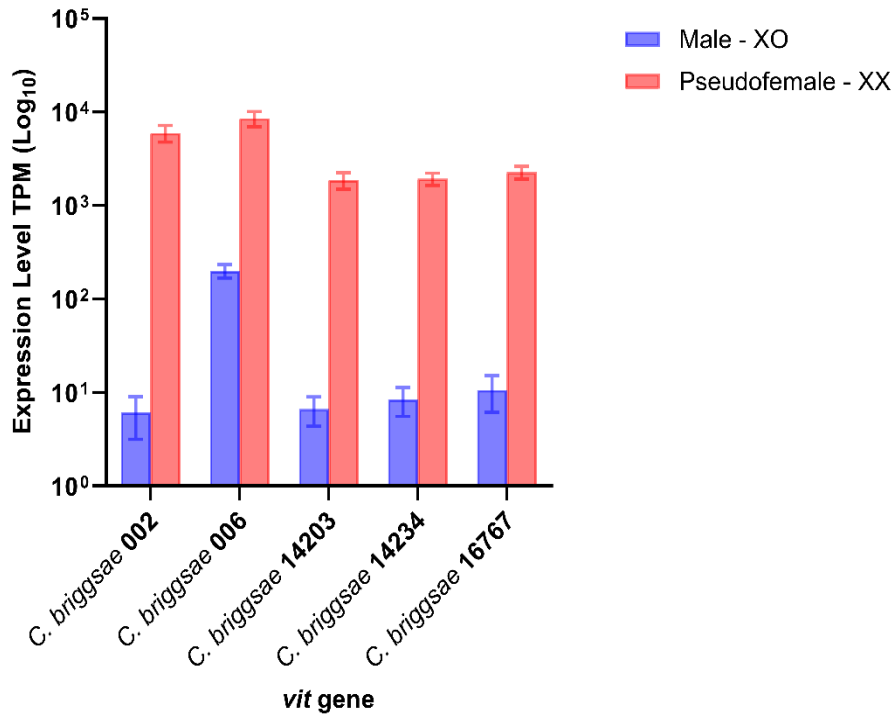


Figure 4.1: Expression level in transcripts per million (logarithmic scale) of *Caenorhabditis briggsae* vit genes in both males and pseudofemales. Produced using data available from Thomas *et al.* (2012). Error bars indicate standard error of the mean (n=3).

Figure 4.2 also presents vitellogenin gene expression in *C. briggsae* using data provided from Grün *et al.*, (2014). This study compared gene expression changes at different stages of the lifecycle including mixed embryos, first-stage larvae (L1), and young adults. In Figure 4.2, the mean expression of vit genes was greatest in the young adult life stage, with all genes having an expression level of greater than 1000 TPM. The gene *C. briggsae* 006 had the greatest expression level of 11,832 TPM. The expression level for mixed embryos and L1 was zero for genes *C. briggsae* 14234 and *C. briggsae* 16767, and therefore could not be plotted. The remainder of the vit genes in mixed embryos and L1 had low expression, at approximately 1.5 TPM or lower. The expression of *C. briggsae* 002, *C. briggsae* 006 and *C. briggsae* 14203 was significantly greater in the young adult stage than L1 and mixed embryos ($P < 0.05$). No significant difference in expression of *C. briggsae* 14234 and *C. briggsae* 16767 was observed between life stages studied.

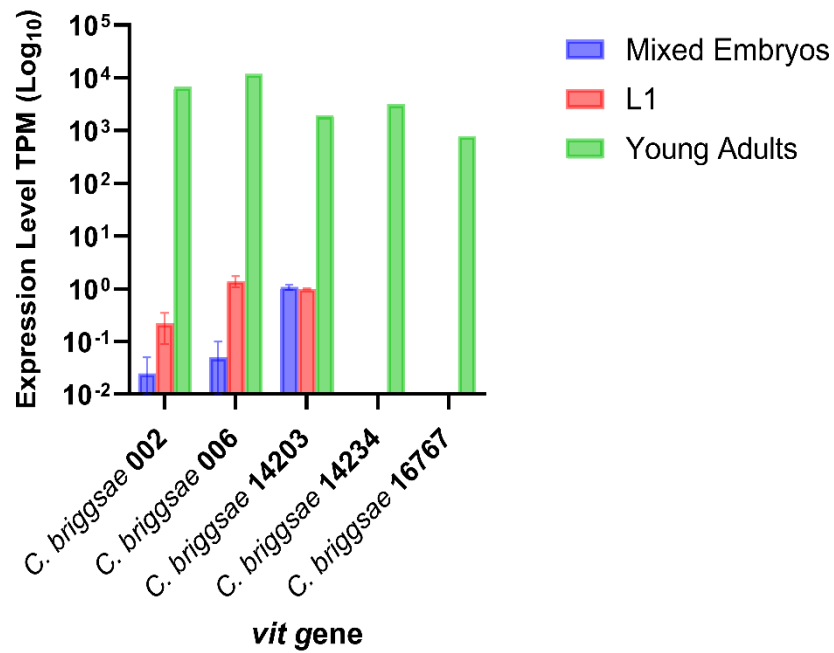


Figure 4.2: Expression level in transcripts per million (logarithmic scale) of *Caenorhabditis briggsae* *vit* genes in mixed embryos, L1 larvae and young adults. Produced using data available from Grün *et al.*, (2014). Error bars indicate standard error of the mean (n=2: L1 and Mixed Embryos; n=1: Young Adults).

4.3.2 Vitellogenin Gene Expression in *Caenorhabditis elegans*

The changes in gene expression during embryo development of *Caenorhabditis elegans* was studied by Hashimshony *et al.* (2014), included in these data was the *C. elegans* *vit* genes. Expression level of *vit* genes is shown in Figure 4.3, given in fragments per kilobase of transcript per million fragments mapped (FPKM), over a period of 850 min post first cell cleavage. All genes remain below 100 FPKM until approximately 320 min, at which point there is a small increase in expression. After 400 min, there are greater peaks in expression level of all genes, but the overall trend remains varied. The trend for all genes indicates an increase of expression over the time studied, with a moderate level of expression. The expression of *vit-1* remains relatively low when compared to the other *vit* genes, with the greatest increase in expression observed in *vit-3*, which is shown to have a high expression level (1020 TPM) at 850 min.

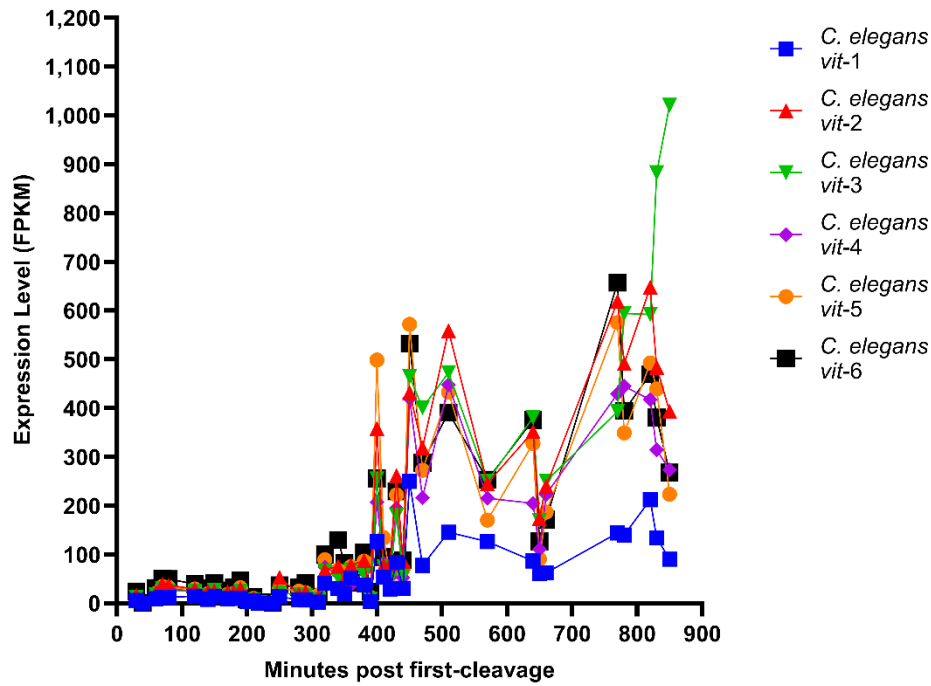


Figure 4.3: Expression level in fragments per kilobase of transcript per million fragments mapped of *Caenorhabditis elegans vit* genes in embryonic development. Expression level was recorded at timed intervals post first-cleavage. Produced using data available from Hashimshony *et al.* (2014) (n=1).

The gene expression levels at different *C. elegans* life stages were included in a study by Dillman *et al.* (2015), to compare to their nematode genomes of interest. Life stages included embryonic, L1, dauer and adult. In Figure 4.4, all six *vit* genes had the greatest expression during the adult life stage, with *vit-2* and *vit-6* having the highest expression level of 2867 FPKM and 3836 FPKM, respectively. The embryonic and L1 life stages had low levels of expression for most genes; however, the mean expression in the embryo of *vit-2* and *vit-6* was 10 and 11 FPKM, respectively. The dauer stage had a varied but moderate expression level of the *vit* genes. The statistical analysis suggested that there were no significant differences in gene expression at different life stages.

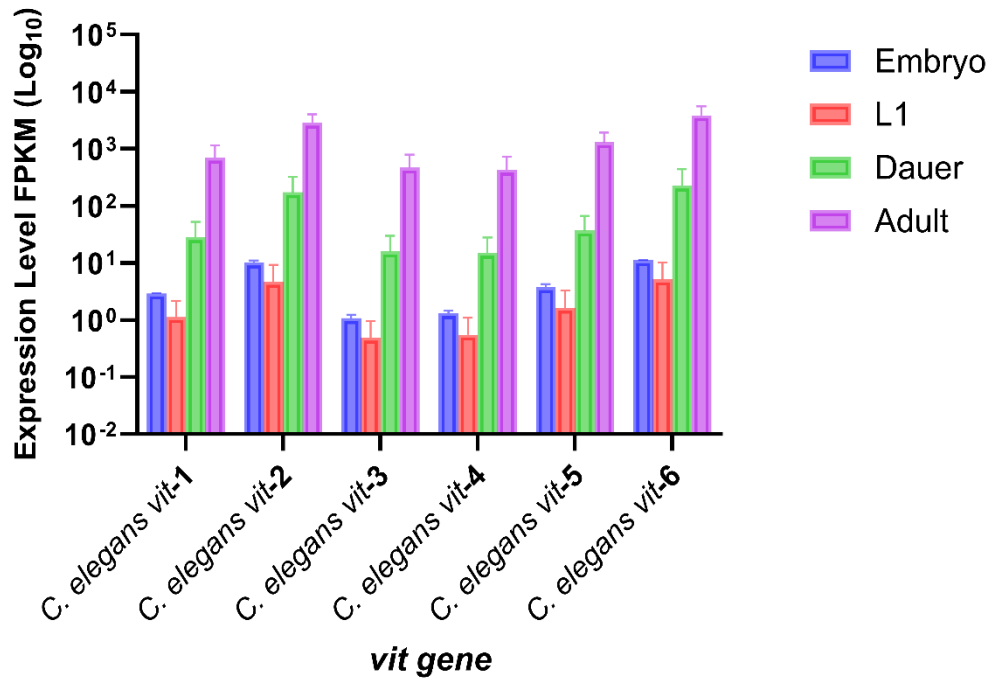


Figure 4.4: Expression level in fragments per kilobase of transcript per million fragments mapped (logarithmic scale), of *Caenorhabditis elegans* *vit* genes in the embryo, L1, dauer and adults. Produced using data available from Dillman *et al.* (2015). Error bars indicate standard error of the mean (n=2).

In addition to studying gene expression changes in *C. briggsae*, Grün *et al.* (2014), also studied *C. elegans* at various life stages including: embryonic, L1-L4 larval stages and young hermaphrodite females. In Figure 4.5, all six *vit* genes showed the highest levels of expression at the young adult hermaphrodite stage, with *vit-2* and *vit-6* having the highest expression level of 3271 and 4171 FPKM, respectively. The expression of all *vit* genes at the embryonic, L1 and L4 stages was low, and was less than 10 FPKM for each gene. Furthermore, gene expression in L2 of *vit-1*, *vit-3* and *vit-4* was low (9, 7 and 7 FPKM), and at a moderate level for *vit-2*, *vit-5*, and *vit-6* (205, 30 and 183 FPKM). Similarly, during the L3 life stage, gene expression of *vit-1*, *vit-3*, *vit-4* and *vit-5* was low, but moderate for *vit-2* and *vit-6*, with the same expression level of 31 FPKM for each. Statistical analysis showed that there were no significant differences in expression of *vit-1*, *vit-2*, *vit-3*, *vit-4* or *vit-6* between the different life stages studied. By contrast, statistically significant differences in gene expression levels of *vit-5* were observed between life stages. Expression level of *vit-5* in the adult hermaphrodite was significantly greater than all other stages studied ($P < 0.05$). Furthermore, gene expression of *vit-5* was significantly greater in L3 than in the embryo, L1 and L4 ($P < 0.05$). Finally, the gene expression of *vit-5* in L2 was significantly increased when compared to the embryo, L1, L3 and L4 ($P < 0.05$).

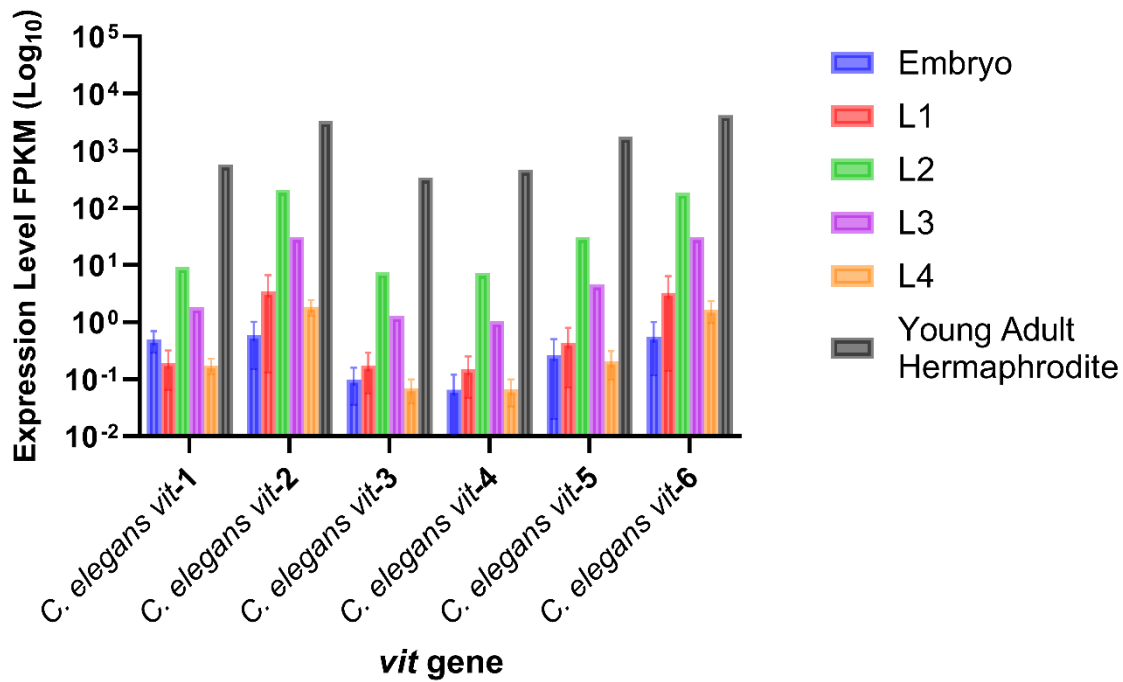


Figure 4.5: Expression level in fragments per kilobase of transcript per million fragments mapped (logarithmic scale), of *Caenorhabditis elegans* *vit* genes in the embryo, L1-L4 larvae and young adult hermaphrodites. Produced using data available from Grün *et al.*, 2014. Error bars indicate standard error of the mean (n=3: Embryo and L1; n=2: L4; n=1: L2, L3 and Young Adult Hermaphrodite).

4.3.3 Vitellogenin Gene Expression in *Globodera pallida*

The changes in gene expression during eight different stages of the *G. pallida* life cycle was studied by Cotton *et al.* (2014). Gene expression was measured at 7-, 14-, 21-, 28- and 35-days post infection (DPI), in adult males, second-stage juveniles (J2) and eggs. These data included the gene expression level in TPM of the four *G. pallida* *vit* genes and is presented in Figure 4.6. This graph indicates that there was a great increase in expression level of all genes from 7 DPI to the developed female at 35 DPI. The expression levels of all genes were low at 7 DPI, excluding *G. pallida* 0564 which had a moderate level of expression (15 TPM). Similarly, expression of all genes was low (below 10 TPM) at 14 DPI. Genes *G. pallida* 0128 and *G. pallida* 9459 had a high mean expression level at 21 DPI of 1451 and 2526 TPM, respectively, whereas the remaining two genes were only at a moderate level. The expression of all genes was high at 28 DPI and 35 DPI; the gene expression level at 35 DPI was greatest for *G. pallida* 0128 and *G. pallida* 9459, at 12,190 TPM and 28,344 TPM, respectively. Finally, expression of all genes in adult males, J2 and eggs was less than one TPM for all genes. These apparent substantial changes in expression level of the *vit* genes were not statistically significant.

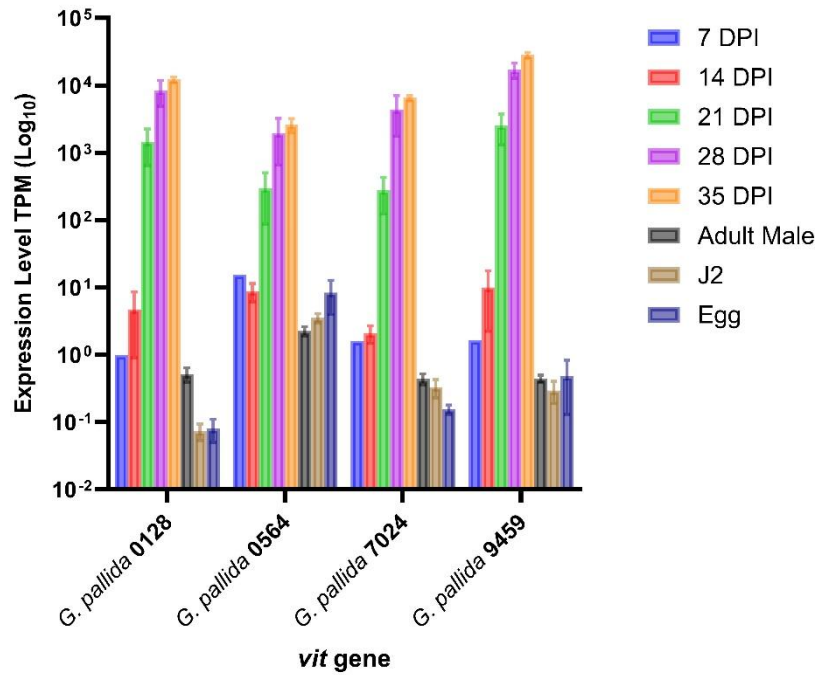


Figure 4.6: Expression level in transcripts per million (logarithmic scale) of *Globodera pallida* vit genes at 7, 14-, 21-, 28- and 35-days post infection, adult males, second-stage juveniles, and eggs. Produced using data available from Cotton *et al.*, 2014. Error bars indicate standard error of the mean (n=2, excluding 7 DPI where n=1).

4.3.4 Vitellogenin Gene Expression in *Globodera rostochiensis*

To determine the changes in gene expression from cyst dormancy to nematode hatch, Duceppe *et al.* (2017) studied the transcriptomes of *G. rostochiensis* dry cysts, hydrated cysts soaked in water for 7 days, hydrated cysts soaked in potato root diffusate for 15 min, 1 h, 8 h, 24 h, 48 h and 7 days, and hatched J2. The data available from this study for vit genes is presented in Figure 4.7. The mean expression levels for all conditions were less than one for genes *G. rostochiensis* 16066 and 8560. Furthermore, the expression level of *G. rostochiensis* 8204 was less than one for all conditions, excluding hatched J2, which was slightly higher at 2.3 TPM. The expression levels of *G. rostochiensis* 6962 was higher than other genes, with a moderate level of expression, ranging from 34 to 144 TPM. Notably, hatched J2 had a higher expression level of all genes than all other conditions studied; however, none of these differences were statistically significant.

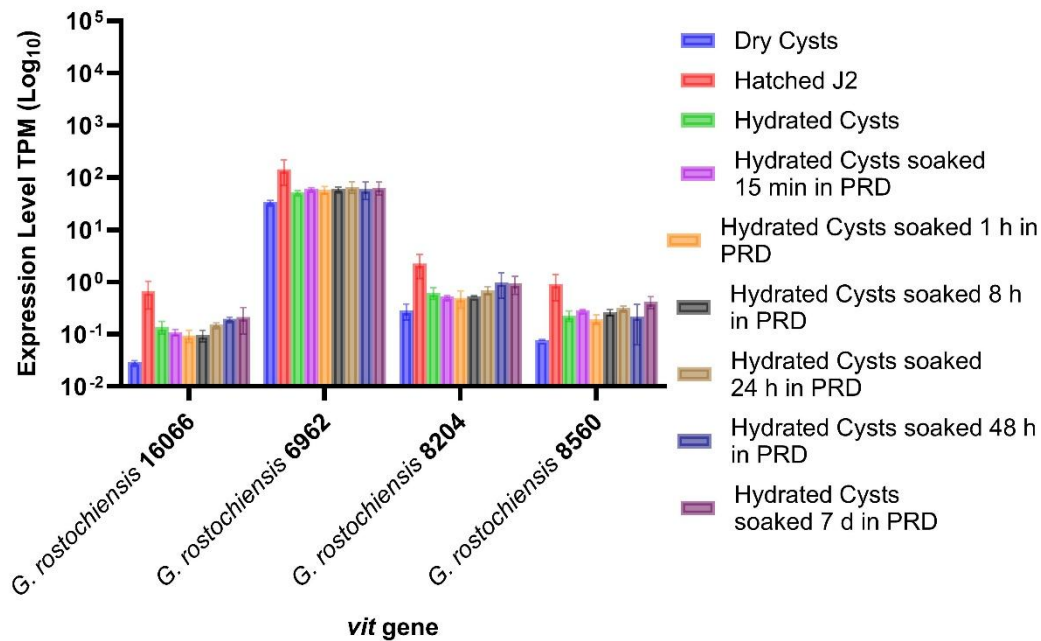


Figure 4.7: Expression level in transcripts per million (logarithmic scale) of *Globodera rostochiensis* *vit* genes under various conditions: dry cysts, hatched second-stage juveniles (J2), hydrated cysts and hydrated cysts soaked in potato root diffusate (PRD) for 15 min, 1 h, 8 h, 24 h, 48 h or 7 days. Produced using data available from Duceppe *et al.* (2017). Error bars indicate standard error of the mean (n=2).

Normalised expression of genes in *G. rostochiensis* was calculated for different conditions including dormant cysts, hydrated eggs, J2 and at 14 DPI (Eves-van den Akker *et al.*, 2016). Data acquired from this study allowed *vit* genes for each condition to be presented in Figure 4.8. Normalised expression levels of all genes were shown to be increased between the hydrated eggs and J2 and then 14 DPI conditions. The highest levels of expression were found in the 14 DPI condition; genes *G. rostochiensis* 2537 and *G. rostochiensis* 2714 had normalised expression levels of 103 and 102, whereas the expression level of *G. rostochiensis* 6238 was slightly lower at 14. The normalised expression counts of all genes in the dormant cysts, hydrated eggs and J2 were less than one; however, there was a significant difference in expression of *G. rostochiensis* 2537 between dormant cysts and hydrated eggs ($P < 0.05$). There was no significant difference in gene expression between the other life stages of *G. rostochiensis* 2537. Similarly, there were no significant difference in the expression of *G. rostochiensis* 2714 or *G. rostochiensis* 6238 between any life stages studied.

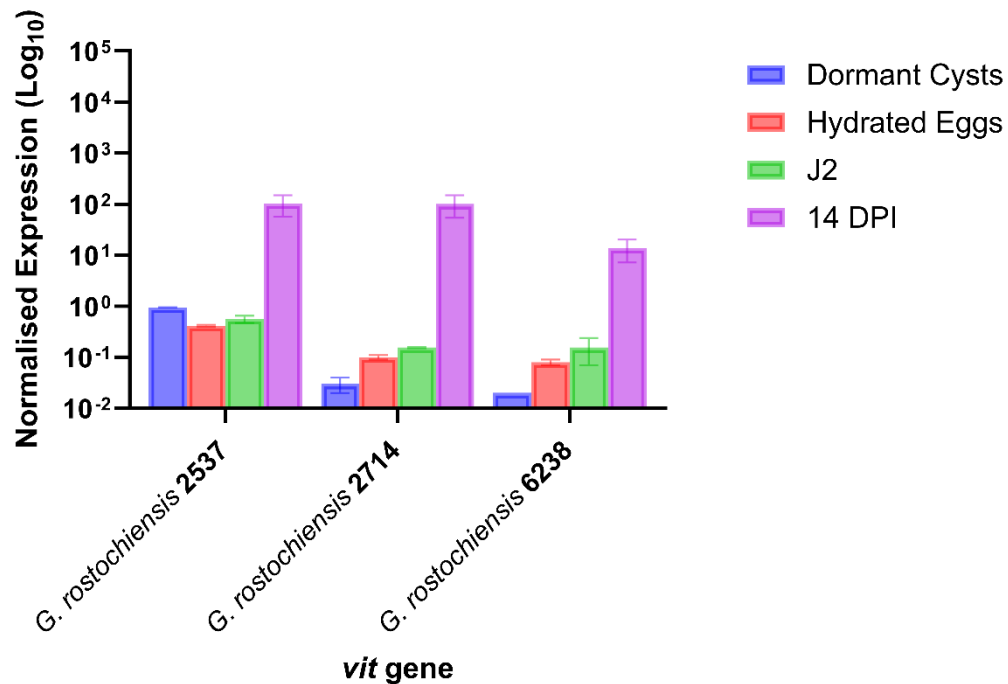


Figure 4.8: Normalised expression (logarithmic scale) of *Globodera rostochiensis vit* genes in dormant cysts, hydrated eggs, second-stage juveniles (J2) and 14 days post infection (DPI). Produced using data available from Eves-van den Akker *et al.*, (2016). Error bars indicate standard error of the mean (n=2).

4.3.5 Vitellogenin Gene Expression in *Heterodera glycines*

Heterodera glycines transcriptomes in the presence and absence of a susceptible host was studied by Gardner *et al.* (2018). This study provided a database of gene expression values for pre-parasitic J2 and infective J2 from compatible and resistant hosts. The *vit* genes expression levels from these data are presented in Figure 4.9. The mean expression level of most genes in all conditions included in this study were low; however, the expression of the gene *H. glycines* 13367 in the pre-parasitic J2 and the parasitic J2 (resistant host) was at a moderate level of 11.5 and 10.6 TPM, respectively. The expression level of *H. glycines* 13367 in the parasitic J2 (compatible host) was significantly less than in the parasitic J2 (resistant host) and pre-parasitic J2 ($P < 0.05$). Conversely, the expression of *H. glycines* 6510 was significantly lower in the pre-parasitic J2, when compared to the other conditions ($P < 0.05$). No statistically significant difference in the expression of *H. glycines* 23424 between conditions was observed.

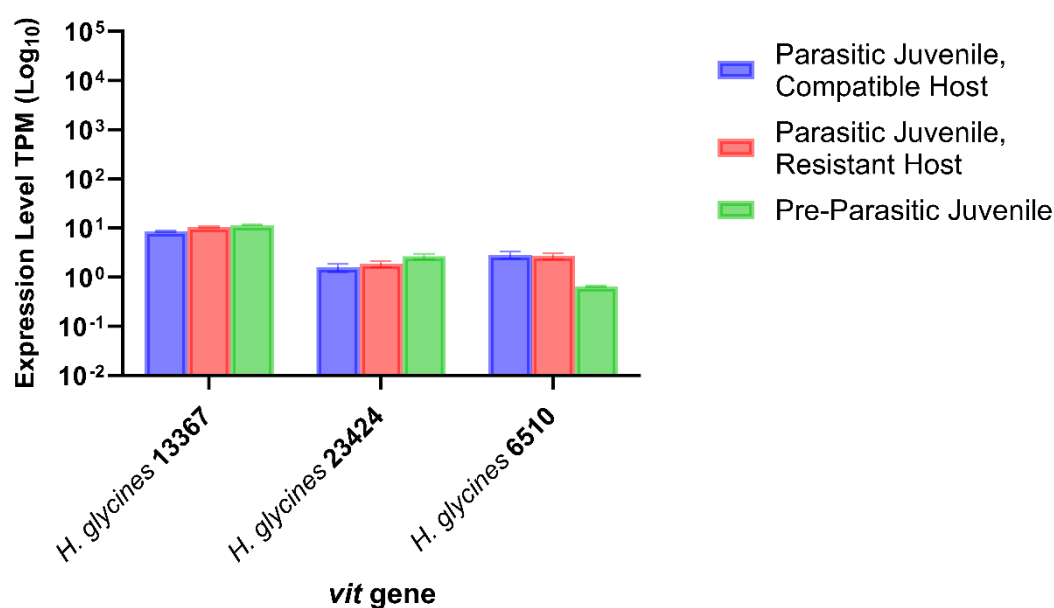


Figure 4.9: Expression level in transcripts per million (logarithmic scale) of *Heterodera glycines* *vit* genes under different conditions: parasitic juveniles from a compatible or resistant host and pre-parasitic juveniles. Produced using data available from Gardner *et al.* (2018). Error bars indicate standard error of the mean (n=3).

4.3.6 Vitellogenin Gene Expression in *Heterodera schachtii*

The life stage transcriptomes of *Heterodera schachtii* were studied by Siddique *et al.* (2022), including dormant cysts, pre-infective J2, nematodes 10 hours post infection (HPI), 48 HPI, females 12 DPI, males 12 DPI and females 24 DPI. Figure 4.10 presents the results obtained for the *H. schachtii* *vit* genes that were available in the data set provided by Siddique *et al.* (2022). The relative expression of *H. schachtii* 007 increased from 10 HPI with a mean expression of 186, this then increased to 335 for 48 HPI and further increased to 110,231 in female 12 DPI and 856519 in the female 24 DPI condition. The male 12 DPI expression level remained at a moderate 801. The relative expression in the pre-infective J2 was similar to 10 HPI, with a value of 208. The lowest relative expression of this gene was seen in the dormant cyst, at 79. The expression of *H. schachtii* 007 was significantly greater in the male 12 DPI than dormant cysts, pre-infective J2 and at 10 HPI ($P < 0.05$). The statistical analysis resulted in no further significant differences in expression in the remaining life stages studied.

In addition, Figure 4.10 illustrates that the expression of *H. schachtii* 19700 followed a similar trend to *H. schachtii* 007; where the expression level greatly increased after 48 HPI from 81 to 16,333 in female 12 DPI and to 83,253 at 24 DPI. Like *H. schachtii* 007, the *H. schachtii* 19700 gene expression was lowest in the dormant cyst. There was a gradual decrease in the expression of *H. schachtii* 19700 between conditions pre-infective J2, 10 HPI and 48 HPI. There were significant differences in expression of *H. schachtii* 19700 between the male 12 DPI life stage and dormant cysts, at 10 HPI and at 48 HPI ($P < 0.05$). No significant differences in expression were identified between other life stages.

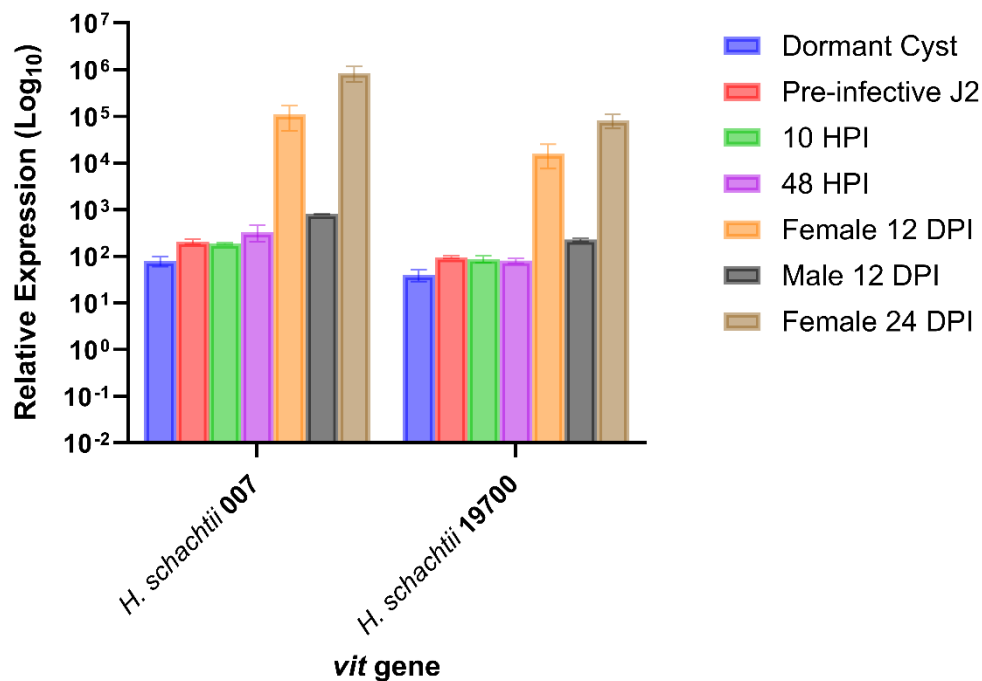


Figure 4.10: Relative expression (logarithmic scale) of *Heterodera schachtii* vit genes at different stages of the nematode life cycle, including dormant cyst, pre-infective J2, 10 and 48 hours post infection (HPI), females and males 12 days post infection (DPI) and females 24 DPI. Produced using data available from Siddique *et al.* (2022). Error bars indicate standard error of the mean (n=3).

Utilising the dataset provided by Siddique *et al.* (2022), the expression of *H. schachtii* genes that were found to be homologous to vit genes in *C. elegans* that did not contain the full protein domains could be investigated. In Figure 4.11, all genes presented contained the Vitellogenin_N domain; however, Hsc_gene_6737 also contained the open beta-sheet domain. Several life stages were shown not to express Hsc_gene_25699, including the dormant cyst, 10 HPI, 48 HPI and males 12 DPI. There were no significant differences between the life stages in expression of Hsc_gene_25699. Similarly, Hsc_gene_15837 was not expressed in the dormant cyst and Hsc_gene_5564 was not expressed at 10 HPI. Interestingly, Hsc_gene_6737 was shown to have high levels of expression for most life

stages: dormant cyst at 1202, 10 HPI at 1568, 48 HPI at 1420, female 12 DPI at 1865 and female 24 DPI at 9039, whereas in the pre-infective J2 and male 12 DPI conditions moderate levels of the gene were expressed. There was a significant difference between the expression of Hsc_gene_6737 in male 12 DPI and female 24 DPI ($P < 0.05$); however, there was no significant difference between the remaining life stages. The gene Hsc_gene_5564 also showed high levels of expression in the female 12 DPI and 24 DPI conditions, at 17,347 and 129,018, respectively; however, there were no significant differences in the expression of the gene between these life stages and other stages studied. Hsc_gene_15837 was also expressed at all stages, with the relative expression level increasing in the female 12 and 24 DPI conditions, at a moderate level of 49 and 284, respectively. There was a significant difference in the expression of Hsc_gene_15837, between the dormant cyst and female 24 DPI ($P < 0.05$); however, no further significant differences in expression were observed.

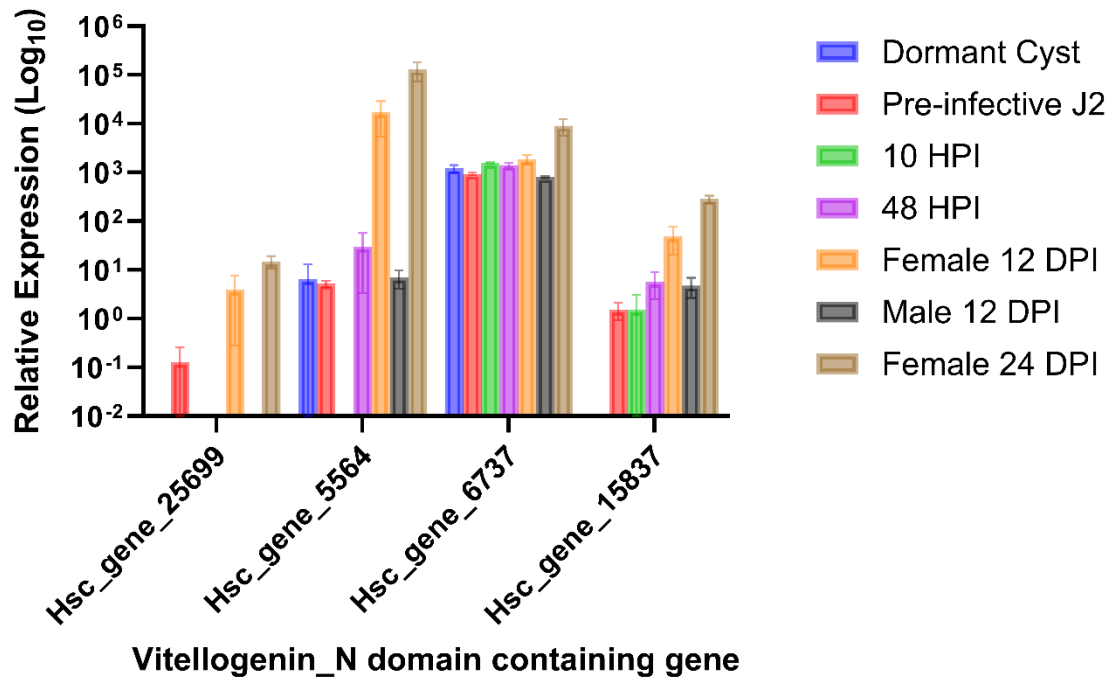


Figure 4.11: Relative expression (logarithmic scale) of *Heterodera schachtii* genes containing Vitellogenin_N domain, that showed homology to *Caenorhabditis elegans* vit genes. Relative expression calculated at different stages of the nematode life cycle, including dormant cyst, pre-infective J2, 10 and 48 hours post infection (HPI), females and males 12 days post infection (DPI) and females 24 DPI. Data not available for Male 24 DPI. Produced using data available from Siddique *et al.* (2022). Error bars indicate standard error of the mean (n=3).

4.3.7 Vitellogenin Gene Expression in *Meloidogyne incognita*

Two studies utilised RNA-seq approaches to determine gene expression at different points in the *M. incognita* life cycle. Blanc-Mathieu *et al.* (2017) studied gene expression from the egg, pre-parasitic J2, parasitic J3-J4, adult females and adult males. Data available for the *vit* genes is included in Figure 4.12. Gene expression of *M. incognita* 17578 in the egg and adult male could not be plotted, as both values equalled zero. The expression levels of genes *M. incognita* 11129, *M. incognita* 17578, *M. incognita* 24409, *M. incognita* 28653 and *M. incognita* 9435, increased as the nematode developed through its life cycle, from pre-infective J2 to an adult female. The expression level of these genes decreased from the parasitic J3-J4 to the adult male. Genes *M. incognita* 30865 and *M. incognita* 6902 follow a similar trend; however, the expression level decreased between pre-parasitic J2 and parasitic J3-J4. The expression level of all genes was highest in adult females. The gene *M. incognita* 9435 was shown to have a high expression level in adult females at 1610 TPM, whereas expression of *M. incognita* 17578 was low during this stage (3.7 TPM). All other genes had a moderate level of expression in the adult female condition. Statistical analysis could not be successfully performed on these data as the sample size of each condition equalled one.

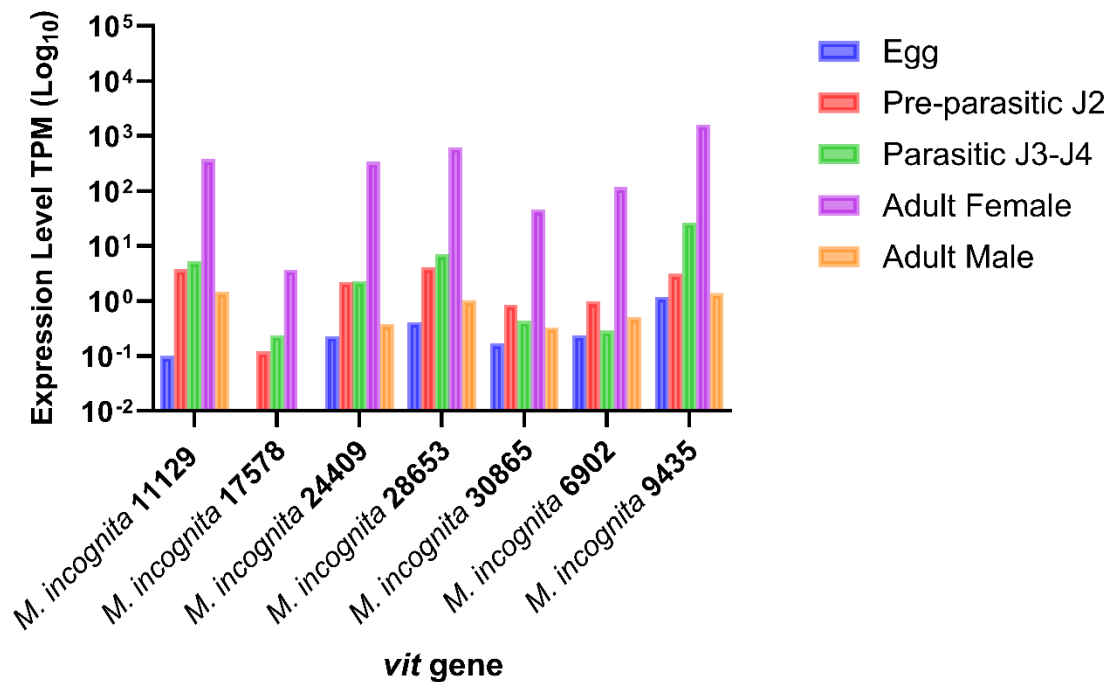


Figure 4.12: Expression level in transcripts per million (logarithmic scale) of *Meloidogyne incognita vit* genes at different stages of the nematode life cycle, including the egg, pre-parasitic second-stage juveniles (J2), parasitic third and fourth-stage juveniles (J3-J4), adult females and adult males. Produced using data available from Blanc-Mathieu *et al.* (2017) (n=1).

Figure 4.13 presents the *vit* gene expression in *M. incognita* using data available from Choi *et al.* (2017). Life stages included in their study were the egg, J2, J3, J4 and adult female. There was a low level of expression of the *vit* genes in the egg and J2, of less than 1.5 TPM for all genes. As the nematode developed from J2 to an adult female, the expression level of the *vit* genes increased; however, in some cases there was a slight decrease in expression from J3 to J4. This was demonstrated in the expression of genes *M. incognita* 11129 and *M. incognita* 6902. There was also a low level of expression of all genes in J3 and J4 life stages, excluding gene *M. incognita* 9435 where there was a moderate level of expression, at 12 TPM (J3) and 20 TPM (J4). The mean expression level was the greatest for the adult females in all *vit* genes, when compared to other life stages. Most *vit* genes had a moderate expression in the adult female stage; however, *M. incognita* 17578 had a low level of expression (8 TPM).

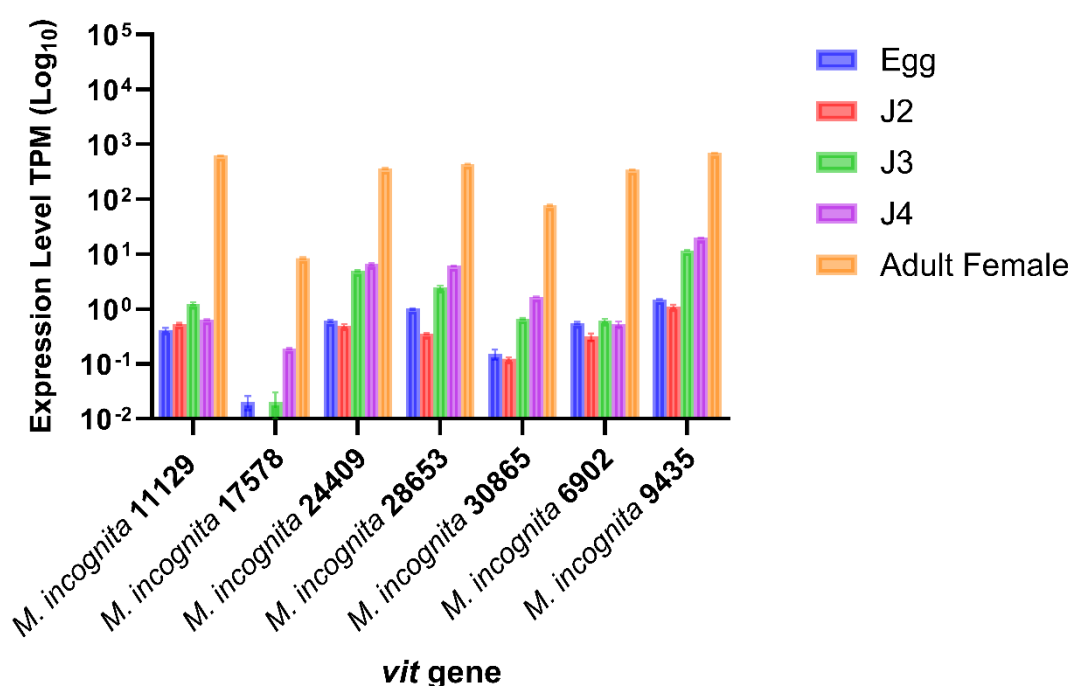


Figure 4.13: Expression level in transcripts per million (logarithmic scale) of *Meloidogyne incognita vit* genes at different stages of the nematode life cycle, including the egg, second-stage juvenile (J2), third-stage juvenile (J3), fourth-stage juvenile (J4) and adult female. Produced using data available from Choi *et al.* (2017). Error bars indicate standard error of the mean (n=3).

Results of a statistical analysis of *vit* gene expression using the data from Choi *et al.* (2017) and presented in Figure 4.13, demonstrated that the expression of *M. incognita* 11129 was significantly increased in the adult female when compared to the egg ($P < 0.05$), whereas the expression of *M. incognita* 17578 in the adult female was significantly greater than in J2 ($P < 0.05$). There were no further statistically significant differences in the expression of both genes between other life stages. Statistical analysis of the expression of *M. incognita* 24409 showed

that the expression in the adult female was significantly greater than in the egg, J2, J3 and J4 ($P < 0.05$). Also, expression of this gene was significantly increased in J4 than in the J2 and egg ($P < 0.05$). Additionally, the expression of *M. incognita* 24409 was significantly higher in J3 when compared to J2 and the egg ($p < 0.05$). Equally, the same life stages were significantly different in the expression of *M. incognita* 28653 ($P < 0.05$); however, there was also a significant difference between the expression in the egg and J2 ($P < 0.05$). Again, there was a significant difference in the expression of *M. incognita* 30865 between these life stages ($P < 0.05$); however, there was no significant difference between the egg and J2, instead there was a significant difference in expression between the egg and J3 ($P < 0.05$). In the adult female, the expression of *M. incognita* 6902 was significantly greater than in all other conditions ($P < 0.05$). Also, there was a significant difference in the expression of this gene between J2 and J3. Finally, in statistical analysis of the expression of *M. incognita* 9435, there was a significant difference between the adult female and J2 ($P < 0.05$); however, no further statistically significant observations were made.

4.4 Discussion

Vitellogenin gene expression has been widely studied in various species; however, little is known about the expression of PPN *vit* genes. Using the *vit* genes obtained in Chapter 3, the expression of these genes was investigated by extracting gene expression data from published data sets, available in the public domain. This work has enabled a better understanding of when in the life cycle of several species of PPN these genes are expressed. Vitellogenin gene expression for most species was greatest in females or at the time point which corresponded to this life stage. The general trend for most PPN species indicated that the expression levels of the *vit* genes increased from the egg or J2 to developed females. For example, in *M. incognita* females, the expression of some *vit* genes were statistically significant in their increased level of expression when compared to earlier life stages (egg, J2, J3 and J4). Similarly, the expression of three *vit* genes (*C. briggsae* 002, 006 and 14203) in *C. briggsae* were significantly greater in young adult hermaphrodites when compared to early life stages of mixed embryos and L1. A trend was observed in the expression of *C. elegans vit* genes, where the expression decreased from L2 to L4, with a significant reduction in the expression of *vit-5*. A possible cause in the reduction of *vit* gene expression at L4 may be due to the switch to spermatogenesis, prior to oocyte production on the onset of adulthood in hermaphrodites (Davis *et al.*, 2023).

An increase in the expression of *vit* genes in adult female nematodes may be due to the great importance placed on yolk and egg production at this point in the nematode life cycle. In the *C. elegans* embryo, it has been approximated that yolk constitutes 37% of total protein present

(Kimble and Sharrock, 1983). Furthermore, it has been shown that the yolk protein complex encoded by the *vit* 1-5 genes in *C. elegans* makes up 25% of all *de novo* protein synthesis in the young adult hermaphrodite intestine (Kimble and Sharrock, 1983). The relative high expression of *vit* genes in females across different species summarised here is supported by findings from van Nostrand *et al.* (2013), who determined that *C. elegans vit* genes were among the most highly expressed genes in young adults and corresponded to approximately 3% of all mapped reads. In addition, a study of *G. pallida* life stage transcriptomes found that *vit* genes were the most highly expressed genes at 28 and 35 DPI, corresponding to the adult female life stage (Cotton *et al.*, 2014). Additionally, an increased expression of PPN *vit* genes in females may be due to the requirement of abundant egg production at this life stage. Female cyst nematodes can produce several hundred eggs; for example, a single potato cyst nematode can produce up to 600 eggs (AHDB, 2024), whereas one root-knot nematode female can produce an egg mass containing several hundred eggs in favourable conditions (Price *et al.*, 2024), with reports of up to 1000 eggs (Subedi *et al.*, 2020).

It may be expected due to the predicted function of nutrient provision to the developing nematode within the egg, that expression of *vit* genes would be restricted to the female; however, *vit* genes were also expressed at other stages and in males of some species. Data illustrating the gene expression in *H. schachtii* and *C. briggsae* showed that once the nematodes had been sexually differentiated, *vit* genes were also moderately expressed in males. Low expression of *vit* genes in *M. incognita* males was shown. The expression at a low to moderate level of *vit* genes may allude to an alternative function in males. Alternatively, it may be that these genes are expressed, but not synthesised. According to Shen and Hodgkin (1988), vitellogenin proteins are not found in *C. elegans* larval stages and adult males. They believe this is due to the activity of *mab-3* preventing the production of yolk proteins; when mutated *mab-3* causes male nematodes to accumulate the vitellogenin protein in the body cavity. Orthologous genes to *mab-3* are present within the genomes of PPN. Bakhietia *et al.*, (2007) suggest that *vit* genes are not expressed in J2 of *H. glycines*; however, results shown here indicate that *vit* genes are expressed but at low to moderate levels. This is supported in a study of *M. hapla*, where it was indicated that vitellogenin proteins are present in J2 (Mbeunkui *et al.*, 2010). In BLAST searches, *H. schachtii* genes were identified that only contained the Vitellogenin_N or Vitellogenin_N and open beta-sheet domains, the expression of these genes was investigated. Some of these genes appear to be expressed; however, it may be possible that they are expressed but not synthesised; although, this would need to be confirmed experimentally.

Inside the developing embryo of *C. elegans*, *vit* gene expression was shown to gradually increase after 400 min post-first cell cleavage. After fertilisation, from 0 to 350 min, cells begin to proliferate and go through gastrulation. At approximately 150 min, the hermaphrodite nematode lays the eggs. Meanwhile, the developing embryo begins the morphogenesis stage from 330 min to hatch (approximately 800-850 min). Within morphogenesis, at 350-390 min epidermal enclosure takes place and at approximately 400 min elongation of the animal begins. Also, at 430 min, the animal's muscles start to have twitch-like movements and show sexual dimorphism at 510 min. Finally, at 760 min the animal begins pharyngeal pumping and hatches from the egg at approximately 800-850 min (Hall *et al.*, 2017). The increase in *vit* gene expression corresponds approximately to when the animal is elongating and begins to demonstrate some movement.

Various methodologies were used to determine the gene expression in different species of PPN, focusing on different life stages, enabling *vit* gene expression data to be available for study. These data sets cannot be directly compared due to the different methodology used, as shown in Table 4.2. Each study used slightly different methods of RNA extraction, sequencing, and data normalisation. The method used for RNA extraction will have had variation in the concentration and purity of the samples; however, most studies mentioned how the quality of their RNA samples was controlled. Also, few studies mentioned the use of a reference gene; if studies had utilised the same reference genes, it may have been possible to compare the levels of expression across different studies. Furthermore, there are limitations in the use of normalisation techniques, as they are dependent on assumptions made about the data collected. Methods such as RPKM and FPKM rely on total expression being the same across different conditions, whereas data normalised by DESeq controls for this by normalising data by distribution across all samples (Evans *et al.*, 2018). Therefore, data originating from sources using DESeq normalisation can be seen as more reliable. Comparisons between *vit* gene expression in different species and life stages may have been possible if multiple studies had utilised DESeq normalisation. To compare these studies, a statistical model titled p-value combination could be used for differential analysis (Rau *et al.*, 2014). In addition, technical or biological variation in these studies may have caused a large spread in data between some replicates, resulting in large calculated standard deviations in the expression level of some genes. This may have resulted in the differences in gene expression between some groups to not be significant.

This analysis into the expression of *vit* genes in PPN has shown that these genes are expressed in these species and vary in expression level at different life stages, with female nematodes demonstrating the highest level of expression. By utilising data sets available the

expression of *vit* genes could begin to be understood, including in the cyst nematodes, *H. schachtii*, *G. pallida* and *G. rostochiensis*. Unfortunately, gene expression data were not available for all species included in Chapter 3, therefore life stage expression of *vit* genes is yet to be determined in *G. ellingtonae*, *M. hapla*, *M. javanica* and *P. scribneri*. Further work could be completed to understand the life stage expression in other species of PPN that have not yet been studied. Furthermore, in *C. elegans*, these expressed genes are synthesised into protein complexes; however, there is no current understanding if these proteins are synthesised in cyst nematodes.

5 Vitellogenin Protein in Cyst Nematodes

5.1 Introduction

The work presented in Chapter 3 and Chapter 4 illustrates that vitellogenin genes are present within plant-parasitic nematode genomes and are shown to be expressed. The molecular weight of the vitellogenin proteins (VIT) if synthesised are yet to be determined in PPN, including the cyst nematodes: *Heterodera schachtii*, *Globodera pallida* and *G. rostochiensis*.

The molecular weight of the vitellogenin protein has been determined in several species across the animal kingdom. Early work studying the African claw frog, *Xenopus laevis*, demonstrated that vitellogenin was a precursor protein with an estimated molecular weight of 200 kDa (Bergink and Wallace, 1974). This study suggested that the smaller lipovitellin (31 kDa and 120 kDa) and phosvitin (35 kDa and 25 kDa) proteins were products of vitellogenin. The molecular weights of vitellogenin proteins have also been estimated in several other species including: honeybee (*Apis mellifera*) at 180 kDa (Wheeler and Kawooya, 1990); mosquito (*Aedes aegypti*) at 337 kDa (Borovsky and Whitney, 1987); freshwater crab (*Potamon potamios*) at 551 kDa, which became three polypeptides of 115 kDa, 105 kDa and 85 kDa under denaturing conditions (Pateraki and Stratakis, 1997); and the chicken (*Gallus gallus domesticus*) at 450-500 kDa (Deeley *et al.*, 1975).

Within the phylum Nematoda, there have been several studies that explored the molecular weight of VIT proteins. *Caenorhabditis elegans* has been the primary model for investigating vitellogenin in nematodes. Vitellogenin in *C. elegans* consist of four polypeptides: two large, yp170A and yp170b, both 170 kDa; and two smaller polypeptides, yp115 and yp88, at 115 kDa and 88 kDa, respectively (Perez and Lehner, 2019). Another rhabditid nematode, *Oscheius tipulae*, a prospective alternative model organism to *C. elegans*, has been shown to have three vitellogenin polypeptides. These polypeptides named VT1, VT2 and VT3 had estimated molecular weights of 175 kDa, 107 kDa and 82 kDa, respectively (Almenara *et al.*, 2013).

Masler (1999) investigated proteins that had been deposited into the eggs of the PPN species, *Heterodera glycines*. Two proteins were identified to be the most abundant, constituting over 50% of all proteins present within the egg. These proteins were labelled MEP-I and MEP-II, with molecular weights of 190 kDa and 180 kDa, respectively. These proteins may not have been from eggshell origin, as most proteins previously attributed to the *H. glycines* eggshell were of 70 kDa or less (Kennedy *et al.*, 1997). Masler (1999) suggested that these proteins could potentially be vitellin, due to similar molecular weights identified to vitellogenin proteins found in *C. elegans*. Antibodies were later generated against MEP-I and MEP-II, which only

recognised proteins in samples containing females and eggs, but not J2 (Masler, 2002). Due to their larger size and comparative molecular weight to *C. elegans* and *O. tipulae* VIT proteins, these MEP proteins may in fact be the precursor protein vitellogenin, not vitellin as proposed.

In antibody production, membrane bound immunoglobulins on naïve B-cells serve as receptors for antigens. Once an antigen is recognised by a receptor, these antigens are internalised by the cell and are presented as smaller peptides by MHC class II molecules. Next, helper T-cells recognise the complex and activate the B-cell, by the interaction between CD40 on the B-cell and CD40L on the helper T-cell (Janeway *et al.*, 2001). Once activated, B-cells will only produce antibodies that contain a unique antigen binding site (Alberts *et al.*, 2002). Epitopes recognised by an antibody are typically five to seventeen amino acids in length (Flaherty, 2012), therefore peptides of longer length can produce multiple epitopes for antibody production. Once activated, the B-cell proliferates and may become an effector cell or a memory cell. The effector cell secretes antibodies, then matures to become a plasma cell, producing large quantities of antibodies (Alberts *et al.*, 2002). Alternatively, the B-cell will become a memory B-cell, to elicit an immune response if the host is again exposed to the same antigen (Alberts *et al.*, 2002).

In studies assessing the molecular weight of a specific protein, antibodies may be produced to investigate the protein. Antibodies can be raised against protein targets by using animal hosts, such as rabbits, mice, donkeys and goats. In polyclonal antibody production, target proteins contain multiple epitopes which would give rise to a range of antibodies. To produce polyclonal antibodies, several immunisations would be made in the animal and then an adjuvant may be added to enhance an immune response. Animal serum can then be collected to obtain the polyclonal antibodies (Quintero-Ronderos *et al.*, 2013). Conversely, monoclonal antibodies target a single epitope. Monoclonal antibodies are typically produced by inoculating mice with the chosen antigen, once the antigen titre in the blood is sufficient, B-cells are isolated from the spleen and fused with myeloma cells, to form hybridomas, from which antibodies can be purified (Quintero-Ronderos *et al.*, 2013).

At present no further studies have considered VIT proteins in PPN; therefore, the aim of this work was to determine the molecular weight of these proteins in the cyst nematodes *H. schachtii*, *G. pallida* and *G. rostochiensis*. Additionally, the molecular weight of VIT protein in second-stage juveniles of *H. schachtii* was studied.

5.2 Materials and Methods

5.2.1 Antibody Design and Synthesis

With the aim to produce antibodies against the vitellogenin protein in *Heterodera schachtii*, *Globodera pallida* and *G. rostochiensis*, a peptide sequence for synthetic peptide synthesis was required. The synthetic peptide was then used as inoculum for antibody production, as part of the custom antibody production service provided by ThermoFisher Scientific Inc.

Putative vitellogenin genes GPLIN_000012800 (*G. pallida*), GROS_g02537 (*G. rostochiensis*) and Hsc_gene_6726 (*H. schachtii*) were aligned using the MUSCLE function in Jalview software (Waterhouse *et al.*, 2009) using the default settings. The selection of the peptide sequence for synthesis was chosen following ThermoFisher Scientific Inc. standard protocol, whereby they advised on suitable sequences for optimal peptide synthesis for custom antibody production. These peptides were chosen to be easy to conjugate to carrier proteins, safe and could produce a high titre of antibodies. These peptides were also verified using the antigen profiler peptide tool (ThermoFisher Scientific, 2025).

Two regions of the cyst nematode *vit* gene alignment with high percentage of homology between the three sequences were selected, resulting in two synthetic peptides being produced (A and B). Peptide A was derived from a 22 aa sequence that was the same in both *G. pallida* and *G. rostochiensis* (Table 5.1). Peptide B was instead derived from a 16 aa sequence, using the *G. pallida* sequence as a template (Table 5.1). Synthetic peptides were synthesised by ThermoFisher Scientific Inc. producing 5 mg of peptides A and B, which were reconstituted to a working concentration of 1 mg/mL.

Table 5.1: Position of peptide sequences of *Heterodera schachtii*, *Globodera rostochiensis* and *G. pallida* from sequence alignment for synthetic peptide design. Peptide synthesised indicated (*).

Species	Peptide Sequence	Position	Length
Peptide A			
<i>H. schachtii</i>	DEQQLRSSGKWEFSLFNAEERE	514:535	22
<i>G. rostochiensis</i>	DEQQLRSSGKWQYSVFNAEERE*	769:790	22
<i>G. pallida</i>	DEQQLRSSGKWQYSVFNAEERE*	767:788	22
Peptide B			
<i>H. schachtii</i>	DDSDDQGEELRMANGQ	1314:1329	16
<i>G. rostochiensis</i>	DDSDDQEEELRMANGQ	1672:1687	16
<i>G. pallida</i>	DDSDDQGEELRMANGR*	1632:1647	16

A proportion of the synthetic peptides were retained by Thermofisher Scientific Inc. to be used in custom polyclonal antibody production. Two rabbits were used in antibody production per peptide, resulting in two batches of crude serum raised against each peptide. Antibodies raised using peptide A were named anti-VIT A1 and anti-VIT A2; similarly, antibodies raised against peptide B were given as anti-VIT B1 and anti-VIT B2.

Animals were subject to a pre-bleed in both peptide conditions to provide pre-immune serum. The animals were then inoculated on day 0, with boost injections on day 14 and 28; initial bleeds were taken on day 35 and provided as crude serum. The animals again received boost injections on day 42, final samples of crude serum were obtained on day 58.

5.2.2 Protein Extraction

Female *H. schachtii* were extracted as described in Chapter 2. Dependent on the concentration of the sample required, different quantities of cysts were placed in Tris Buffered Saline with 0.1% Tween 20 (TBST). For assays requiring a concentrated sample of cyst protein, one cyst per microlitre was placed in a 1.5 mL Eppendorf tube, to a total of 120 μ L. The cysts were then homogenised in buffer using an Eppendorf tissue grinder. Samples were then frozen at -20°C for later use. Homogenates of *G. pallida* and *G. rostochiensis* were used to compare profiles to *H. schachtii*. *Heterodera schachtii* eggs were collected as described in Chapter 2. The eggs from 20 or 30 cysts were collected in an Eppendorf tube containing 60 μ L TBST and homogenised using an Eppendorf tissue grinder, rotating approximately 50 times. Second-stage juveniles (J2) of *H. schachtii* were collected as described in Chapter 2, until approximately 1000 J2 were contained in 500 μ L of water. The sample was then homogenised using an Eppendorf tissue grinder, rotating approximately 100 times. Sonication was then used to further disrupt the J2 sample, using an Ultrasonic Processor XL 2015 (Qsonica). Samples were kept on ice and pulsed for 30 s on/off for a total of 5 min. Samples were then visually checked under a stereomicroscope at 30 $\times$ magnification and reduced to 100 μ L.

5.2.3 Protein Quantification

Proteins from homogenised samples were quantified using the Pierce Bicinchoninic acid (BCA) protein assay kit, following the manufacturer's recommended protocol for the microplate procedure. This assay had a working range of 125-2000 μ g/mL. A standard curve was prepared using bovine serum albumin as a standard (0-2000 μ g/mL). In the microplate procedure, 10 μ L of sample or standard was inoculated in duplicate, to which 200 μ L of working reagent was added. The plate was covered and mixed on a plate shaker for 30s then incubated at 37°C for 1h. Once cooled to room temperature, the absorbance was measured at 562nm. The standard curve was used to determine the final concentration of the samples.

5.2.4 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins based on their molecular weight. Two types of polyacrylamide gels were used either Mini-PROTEAN® TGX™ Precast Protein Gels (10% or 12%) or made within the laboratory, consisting of a 12.5% resolving gel and 4% stacking gel (Table 5.2). Molecular weight makers used were PageRuler™ Plus Prestained Protein Ladder and Invitrogen™ iBright™ Prestained Protein Ladder, 2 µL of either ladder was loaded onto the gel. Samples were diluted to a known concentration in Tris Buffered Saline, then further diluted 1:1 with 2X sample buffer. The sample buffer used was either prepared in the laboratory (Table 5.2) or 2X Laemmli sample buffer (Biorad Inc.) containing 2 mM dithiothreitol. Samples were then loaded onto the gel. The tank was filled with 2X electrophoresis buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). The gels were run at a constant voltage of 200V, the initial current was set at 150mA and were run until the gel front was approximately 2 cm from the bottom of the cassette, with an estimated run time of 45 min.

Initially, gels were stained with Coomassie blue 0.1% stain; however, to achieve a more sensitive stain, silver staining was used. The Pierce Silver Stain Kit was used following the manufacturer's recommendation; the background of some gels was reduced using Pierce™ Silver Stain Rescue Reagent, following the manufacturer's recommended protocol. For samples to be sent for protein identification, Coomassie-G250 staining was performed using Invitrogen™ SimplyBlue™ SafeStain, following the manufacturer's recommended maximum sensitivity staining protocol.

Table 5.2: Reagents required for casting two gels and 1X sample buffer.

Reagent	Resolving Gel (12%)	Stacking Gel (4%)	1X Sample buffer
Acrylamide stock 30%	4.00 mL	1.33 mL	-
Tris-HCl, 1.5 M, pH 8.8	2.50 mL	-	-
Tris-HCl, 0.5 M, pH 6.8	-	1.25 mL	1.25 mL
Water	3.35 mL	2.34 mL	5.5 mL
SDS, 10%	100 µL	50 µL	2 mL
APS, 10%	50 µL	25 µL	-
TEMED	5 µL	5 µL	-
Glycerol	-	-	1 mL
Bromophenol blue, 1%	-	-	200 µL
Dithiothreitol (DTT), 2M	-	-	50 µL

5.2.5 Dot Blot Assay

Dot blot assays were used to determine the appropriate concentration of primary and secondary antibodies to be used in further dot blot and western blot analysis. This method was also used to determine whether a selected primary antibody would successfully bind to any protein in the homogenised cyst nematode samples. Either polyvinylidene fluoride (PVDF)

or nitrocellulose membranes were used for the immunoblotting assays. PVDF membranes were activated by soaking in methanol for 15 s and were allowed to dry completely, nitrocellulose membranes required no activation and performed better for dot blot assays. Dependant on the assay performed, 5 μ L of the primary antibody, synthetic peptide or the supernatant of homogenised cyst samples was inoculated onto the membrane. This was left to soak into the membrane for 30 min, until the membrane was dry. The membrane was then blocked with 5% dried milk in TBST for 60 min. The membrane was then incubated with the primary antibody on a shaker for 60 min. This was then washed three times in TBST for 10 min each. The membrane was then incubated with the secondary antibody on a shaker for 60 min and then washed three times in TBST, for 5 min each. All secondary antibodies used were raised in goat against either rabbit or rat and were horseradish peroxidase (HRP) conjugated. The membrane was then visualised using Pierce™ DAB Substrate Kit, following the manufacturer's recommended protocol, or later using 1-Step™ Ultra TMB-Blotting Solution (Thermo Scientific™), as a safer and more sensitive alternative. This blotting solution stained the proteins in the membrane for a maximum of 30 min and was then washed twice with ultrapure water.

Three preliminary dot blot assays were completed. An initial dot blot assay was a dilution series of day 35 primary antibodies: anti-VIT A1 and A2, anti-VIT B1 and B2 and the pre-immune serum obtained from these animals. Each anti-*vif* primary antibody or pre-immune serum was inoculated onto a membrane in a serial dilution from 1:250 to 1:2000. The membranes were incubated in goat anti-rabbit secondary antibody at 1:250, 1:500 or 1:1000.

The second dot blot assay experiment was a serial dilution of the synthetic peptides A and B (1 mg/mL), from 1:125 to 1:2000, inoculated onto a PVDF membrane as described above. The membranes were incubated in pre-immune serum, anti-VIT A or anti-VIT B day 35 serum, at a concentration of 1:1000 or 1:2000. The concentration of the goat anti-rabbit secondary antibody was 1:250.

The third dot blot assay consisted of a serial dilution of the synthetic peptide A (1 mg/mL), homogenised cysts of BCN (166 μ g/mL) or PCN (164 μ g/mL). The samples were inoculated onto the nitrocellulose membrane from neat to 1:100. The membrane was then blocked and incubated in the anti-*vif* antibodies (A1 or A2), using the day 58 serum, at either 1:250, 1:500 or 1:1000 antibody concentration. The membranes were then incubated in a goat anti-rabbit secondary antibody at 1:250.

5.2.6 Western Blot Assay

Western blots were carried out using fully wet apparatus after the SDS-PAGE protocol had been completed. The protein was transferred using either nitrocellulose or PVDF membranes that were activated by soaking in methanol for 15 s; the performance of these membranes was compared. The gels were placed into the cassette with an ice block and transfer buffer (25 mM Tris, 192 mM glycine, 10% ethanol, pH 8.3). The tank was immersed in ice and set on top of a magnetic stirrer. The transfer was run with a constant voltage of 100 V for 60 min.

In initial runs, to determine whether the protein had transferred, the PVDF membrane was stained with 2% Ponceau S in 10% acetic acid and destained using ultrapure water. For a more sensitive stain, black Indian ink was used. This protocol required the membrane to be washed three times in phosphate buffered saline with 0.1% Tween 20 (PBST) for 30 min each, twice in an orbital shaker at 37°C and once at room temperature. The membrane was then placed in Indian ink solution (0.1%) overnight and destained using small quantities of PBST.

When the membranes were incubated with antibodies, the protocol using black Indian ink was not followed. After the transfer was complete all steps were completed on a shaker. The membrane was placed in the blocking buffer, 5% dried milk in TBST, for 60 min. After blocking, the membrane was then incubated in the primary antibody, overnight at 4°C. The next day, the membrane was washed three times in TBST, for 10 min each. It was then incubated in the secondary antibody for 60 min and again rinsed three times in TBST, for 5 min each. The bands were then visualised using 3 mL of 1-Step™ Ultra TMB-Blotting Solution (Thermo Scientific™) to cover the membrane for up to 30 min and then rinsed twice in ultrapure water.

Western blot analysis was used to determine the concentration of sample that could be visualised. This was performed by preparing a serial dilution of the synthetic peptide A, from 1 mg/mL to 0.0078 mg/mL. The western blot was completed as described above and incubated in either day 35 anti-VIT A1 or A2 antibodies or day 58 anti-VIT A1 or A2 antibodies. In subsequent western blots the synthetic peptide A (0.0625 mg/mL) was used as a positive control and homogenised *C. elegans* eggs (150 µg/mL) as a likely negative control.

Western blot analysis and SDS-PAGE was also used to determine the molecular weight of proteins in *H. schachtii*, *G. pallida* and *G. rostochiensis* recognised by the anti-VIT A. Homogenised cyst and egg samples were adjusted to 200 µg/mL, in most gel images presented, unless otherwise stated. Second-stage juvenile samples were used as neat due to the low concentration of the sample (25 µg/mL). In some cases, the same protein concentration across samples could not be achieved due to the low concentration in the original sample.

5.2.7 Analysis of Polyacrylamide Gels and Membranes

Polyacrylamide gels, dot and western blot membranes were imaged using Syngene InGenius3 gel documentation with Genesys software and then a newly available iBright CL1500 Imaging System with incorporated iBright Analysis Software, for better visualisation of western blots and stained gels. Stained gels and membranes were additionally digitally captured using a 50 Megapixel camera. To determine the approximate molecular weights of bands from stained gels and western blot membranes, the images were loaded into imageJ (Schneider *et al.*, 2012), an image processing and analysis program. The relative migration distance (R_f) of each molecular weight marker, protein bands of interest and the gel front was measured. The R_f was calculated using the equation given below:

$$R_f = \frac{\text{migration distance of protein of interest}}{\text{migration distance of gel front}}$$

Once the R_f for each gel was calculated, a standard curve of log molecular weight and respective R_f values of the molecular weight markers were plotted. The equation of this line was used to determine the approximate molecular weight of the proteins of interest.

Dot blot membranes and selected western blot membranes were also analysed for the intensity of colour of the dot or band produced. Dot and band intensity was measured using the gel analysis function of ImageJ. This analysis gave a peak on a graph for each dot or band, that could be used to calculate the peak area. The peak area was proportional to the colour intensity of the dots or bands measured and described as area under the curve of either dot or band intensity and measured as arbitrary units.

5.2.8 Protein Identification

In western blot analysis, the antibodies used recognised proteins at various molecular weights, and further analysis was required to confirm the identity of these proteins. Protein bands produced by performing a western blot were compared to bands produced by Coomassie-G250 staining. The theoretical molecular weights of vitellogenin proteins were predicted using ExPASy Compute pI/MW tool (Gasteiger *et al.*, 2005). Corresponding bands were cut out of the stained gel, containing one biological replicate of the species investigated and sent to Metabolomics and Proteomics Technology Facility, University of York, where they were destained and digested with trypsin. Tandem and MALDI-TOF mass spectrometry were used to identify peptide amino acid sequences and to give a preliminary protein identification. Further samples from a single biological replicate from *H. schachtii*, *G. rostochiensis* and *G. pallida* cysts were then also analysed by liquid chromatography mass spectrometry.

5.2.9 Data Presentation and Statistical Analysis

In Section 5.3, dot and western blot analysis was presented as the original photograph of the membrane used. The mean dot or band intensities were also presented on a graph with an associated data table. For ease of interpretation data has been presented in several formats.

Statistical analysis was performed using GraphPad Prism software (version 10.2.3), comparing the dot intensity from different samples at different concentrations of primary and secondary antibody. The standard deviation (SD) of the mean was calculated and compared between each condition. A two-way ANOVA was used with Tukey's multiple comparison test, to compare each cell mean with every other cell mean on that row or column. The confidence interval used was 0.05.

5.3 Results

5.3.1 Comparison of Cyst Nematode Proteins using SDS-PAGE

Silver and Coomassie staining techniques were used as described in Section 5.2, to determine the protein abundance required to visualise proteins from cyst nematodes. Once samples were able to be visualised, staining was used to compare the molecular weights of proteins within the cysts and eggs of *G. pallida* and *H. schachtii*.

The SDS-PAGE gel in Figure 5.1 has demonstrated that silver staining protein homogenates derived from ten cysts gave the clearest image, whereas when protein was extracted from twenty cysts it was too saturated and difficult to visualise. As expected, the protein bands visualised from 1 or 10 homogenised cysts demonstrate the same molecular weight (Table 5.3). The resolution of this gel shows that there were 24 bands recognised that could be measured; however, there may be more protein bands that are less abundant, therefore not visualised.

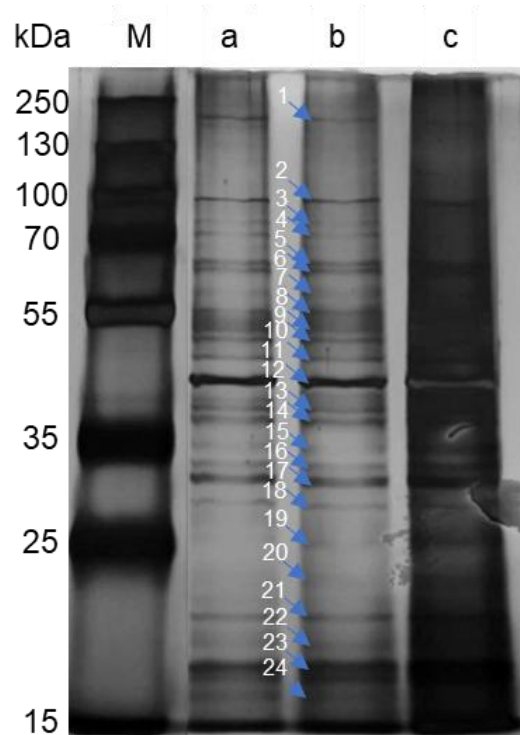


Figure 5.1: SDS-PAGE gel of *Globodera pallida* cyst homogenates, stained by silver staining. The first lane contained the molecular weight standard ladder (M). The lanes marked a, b, and c contained protein extracts from 1, 10 or 20 homogenised *G. pallida* cysts, respectively. Blue arrows and associated number indicate measured band, the molecular weights are given in Table 5.3.

Table 5.3: Approximate molecular weight (kDa) of protein bands measured in Figure 5.1.

Band number	1	2	3	4	5	6	7	8	9	10
Molecular weight (kDa)	194	107	97	69	65	63	58	54	50	47
Band number	11	12	13	14	15	16	17	18	19	20
Molecular weight (kDa)	44	40	39	36	34	32	30	27	23	22
Band number	21	22	23	24						
Molecular weight (kDa)	21	19	18	16						

SDS-PAGE was then used to compare the protein profiles of *G. pallida* (PCN) and *H. schachtii* (BCN cysts, stained by silver staining (Figure 5.2). This figure illustrates that there were many protein bands in common; there were 14 protein bands with approximately the same molecular weight, as indicated by the blue arrows. Additionally, each species had proteins of a unique molecular weight not observed in the other, as indicated by the yellow arrows. The PCN cyst sample had six unique bands, whereas the BCN cyst sample only contained four.

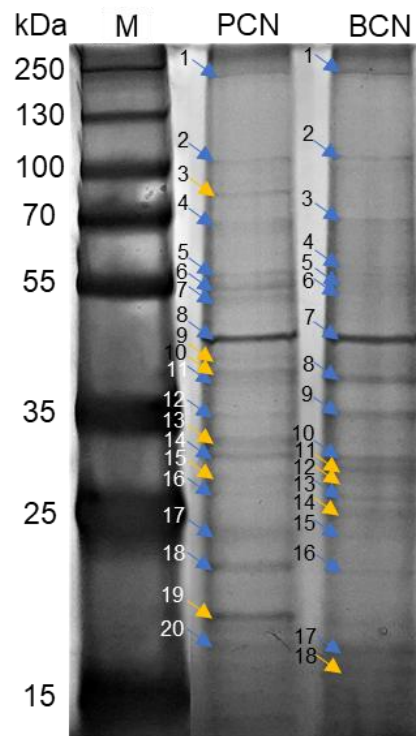


Figure 5.2: SDS-PAGE gel of cyst homogenates of *Globodera pallida* (PCN) and *Heterodera schachtii* (BCN), stained by silver staining. The first lane contains the molecular weight standard ladder (M). The lane marked PCN contains a protein extract from 10 homogenised *Globodera pallida* cysts and the lane marked BCN contains protein extracted from five homogenised *Heterodera schachtii* cysts. Blue arrows indicate bands in common between samples, whereas yellow arrows indicate bands only appearing in one sample.

The use of SDS-PAGE and Coomassie staining was a requirement of visualisation and extraction of bands for further protein analysis by mass spectrometry. The SDS-PAGE gel presented in Figure 5.3 was performed to optimise protein concentration for effective visualisation of the protein bands of interest. This figure shows the decreasing concentration of protein from cyst homogenates of *H. schachtii* and *G. pallida*; from 800 µg/ mL to 200 µg/ mL. It is clearly illustrated that a higher concentration of protein is required for better visualisation when using this methodology, including the use of SimplyBlue™ safe stain. Therefore, for bands to be visualised effectively by Coomassie staining the protein concentration should be at least 800 µg/ mL.

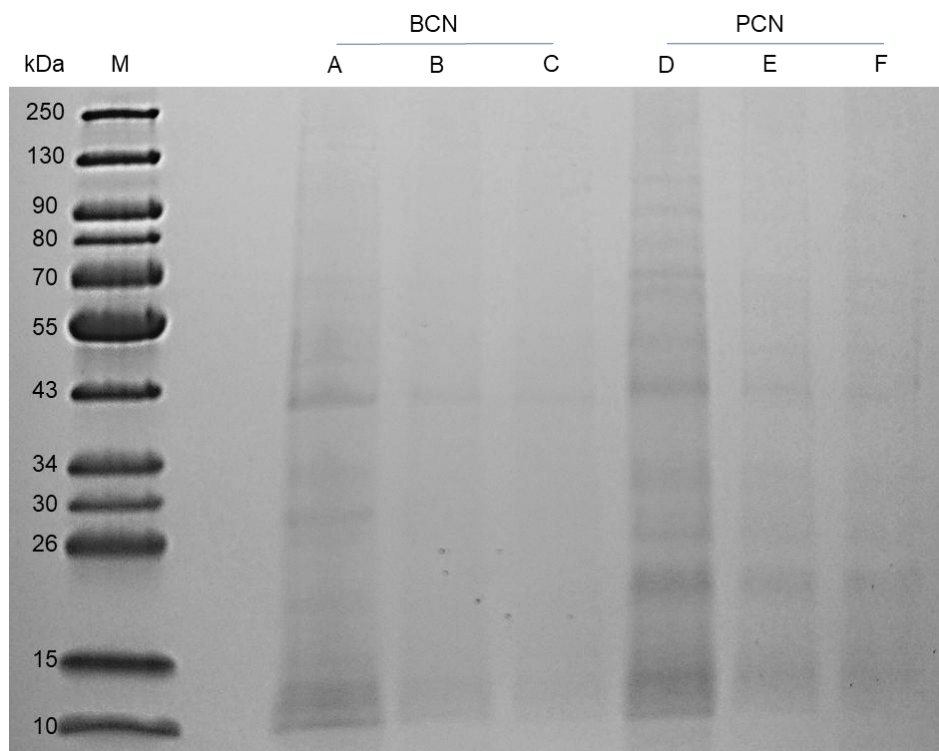


Figure 5.3: SDS-PAGE gel to compare concentration of cyst nematode protein extracts required for detectable protein by Coomassie staining. SDS-PAGE gel contains protein extracts from cysts of *Heterodera schachtii* (BCN) and *Globodera pallida* (PCN), stained by SimplyBlue™ safe stain, a Coomassie stain. The first lane contains the molecular weight standard ladder (M). Samples loaded into lanes A and D had a concentration of 800 µg/ mL, B and E at 400 µg/ mL and C and F at 200 µg/ mL.

5.3.2 Antibody Design and Synthesis

In order to produce antibodies against the vitellogenin protein in *H. schachtii*, *G. pallida* and *G. rostochiensis*, *vit* gene sequences from each genome were selected and aligned (Figure 5.4). Two regions of the alignment were chosen that showed homology between the sequences; these regions were also chosen because they were suitable for synthetic peptide production and to raise antibodies. The first synthetic peptide (A) was 22 amino acids in length and started from amino acid position 775 in the sequence alignment in Figure 5.4. The second synthetic peptide (B) was 16 amino acids in length and instead started from the amino acid at 1688 in the sequence alignment (Figure 5.4). Synthetic peptides were synthesised by ThermoFisher Scientific Inc. and used to immunise two rabbits for each peptide, to generate anti-vitellogenin (anti-VIT) antibodies.



Figure 5.4: Sequence alignment of *vit* genes *G. pallida* 0128, *G. rostochiensis* 2537 and *H. schachtii* 007 from *Globodera pallida*, *G. rostochiensis* and *H. schachtii*, with amino acid sequence of synthetic peptides highlighted. Synthetic peptide A was made from the sequence at amino acid positions 775-796; whereas synthetic peptide B was made from the sequence at amino acid positions 1688-1702. Black dotted boxes indicate the position of these sequences in the alignment.

5.3.3 Dot Blot Assays

5.3.3.1 Test of Primary Antibodies and Pre-Immune Serum

Dot blot assays were a useful tool to determine whether a primary antibody would interact with any part of the sample prior to performing a western blot. Initially, the interaction between the anti-VIT antibodies and the secondary antibodies were tested using dot blots. Figure 5.5 (a) gives a schematic drawing of the dot blot assays, indicating the position of each sample and the concentration of primary antibody from the serial dilution, of either anti-VIT A, anti-VIT B, or the associated pre-immune serums. Also, in Figure 5.5 (b-d) the successful recognition of the primary antibodies and pre-immune serum by the anti-rabbit secondary antibody is shown.

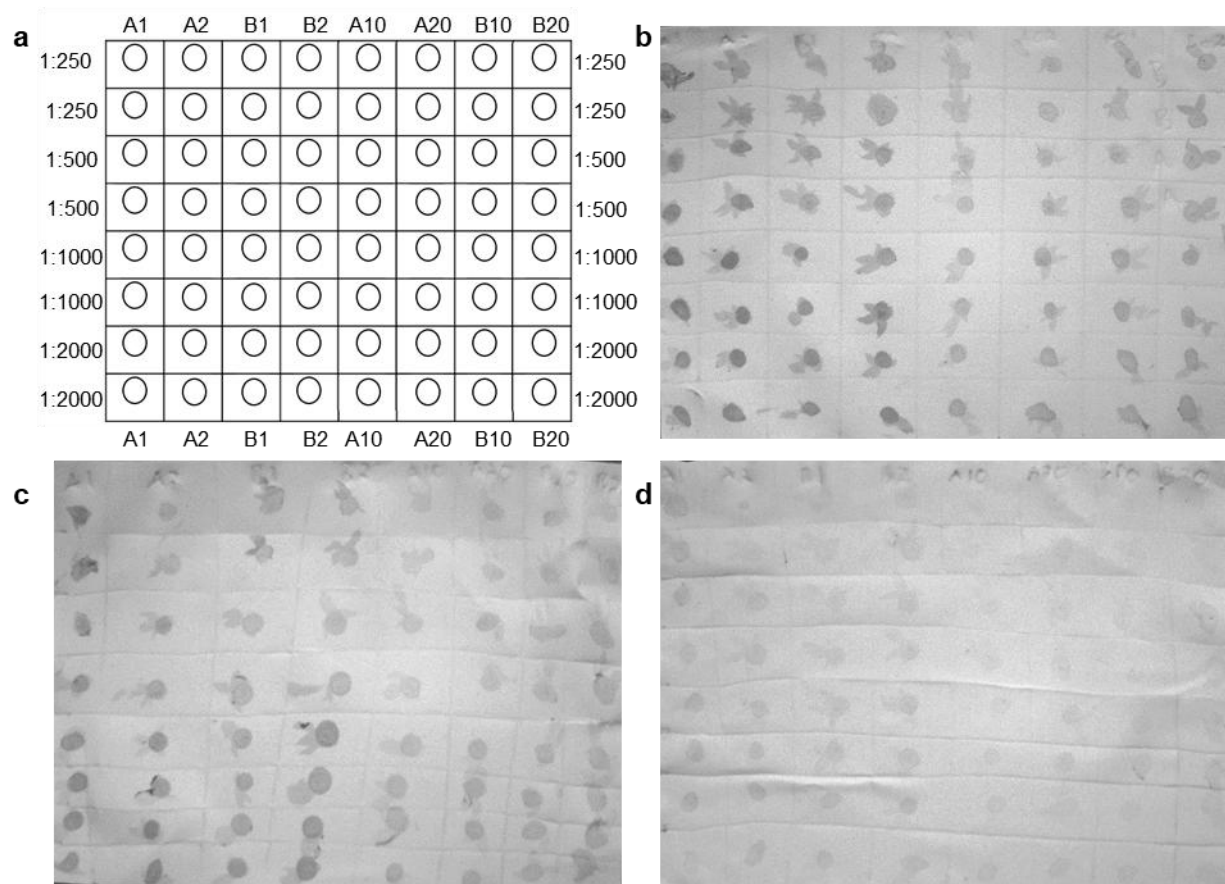


Figure 5.5: Dot blot assays of dilution series using primary antibodies produced against synthetic peptides from *vit* gene sequences. a: schematic drawing of dot blot assay presented in b, c and d, indicating sample position and dilution from 1:250 to 1:2000, from a 1 mg/ mL stock. Images b-d: images of dot blot assay results at different concentrations of goat anti-rabbit secondary antibody, at either 1:250 (b), 1:500 (c) or 1:1000 (d). A1 and A2: antibody A produced from two different rabbits. B1 and B2: antibody B produced from two different rabbits. A1-0, A2-0, B1-0 and B2-0: pre-immune serum from day zero rabbit bleeds. Each inoculation onto PVDF membrane was performed in duplicate and was visualised using DAB substrate.

The dot intensity of each dot in Figure 5.5 was measured, and the mean value was recorded (Table 5.4), these data are represented in Figure 5.6. The observed dot intensity was greatest when the secondary antibody was used at 1:250 concentration, decreasing in the use of 1:500 and then 1:1000. When the secondary antibody concentration was 1:250, many dilutions of primary antibody had significantly greater dot intensity when compared to the 1:1000 concentration of secondary antibody; this was observed in primary antibodies A1, A2 and B1 at the 1:250 dilution ($p < 0.05$). The dot intensity of A2 and B2 at 1:1000 dilution showed significant differences between the 1:250 and 1:1000 concentration of secondary antibodies ($p < 0.05$). Again, there were significant differences between the 1:250 and 1:1000 secondary antibody concentrations, in the dot intensity of the 1:2000 dilution of A1 and B1 ($p < 0.05$). There were also some significant differences in dot intensity of the primary antibody dilutions when comparing the use of the secondary antibodies at either 1:250 and 1:500. These significant results were seen in the 1:1000 dilution of B2 and the 1:2000 dilution of B1 ($p < 0.05$). There were limited significant differences in dot intensity between the concentrations of secondary antibody, when comparing the dilutions of pre-immune serum; only one significant difference was recorded, between the 1:500 and 1:1000 concentration of secondary antibody for the B2 and B2-0 dilutions at 1: 250 ($p < 0.05$).

The secondary antibody recognised the primary antibodies and pre-immune serum from all bleeds giving a visual dot; however, the reaction appeared to be stronger for the primary antibodies than the pre-immune serum. When using the secondary antibody at 1:250, significant differences in the dot intensity between the dilution series of primary antibodies and pre-immune serum were observed. When the dilution of primary antibodies and pre-immune serum was 1:250, there were significant differences between A1 and A1-0, and A2 and A2-0 ($p < 0.05$). Furthermore, the dilution of 1:1000, showed significant differences between A1 and A1-0, A2 and A2-0, and B2 and B2-0 dot intensities ($p < 0.05$). Finally, when the dilution was 1:2000, there were significant differences between the A1 and A1-0, and B1 and B1-0 ($p < 0.05$). There were no further significant differences observed between the primary antibodies and pre-immune serum dot intensities.

Table 5.4: Mean dot blot intensity of dilution series using primary antibodies produced against synthetic peptides, designed from *vit* gene sequences. Each primary antibody (A1, A2, B1, B2) or pre-immune serum (A1-0, A2-0, B1-0 and B2-0) had a concentration of 1:250, 1:500, 1:1000 or 1:2000, diluted from 1 mg/ mL. Each primary antibody dilution was then incubated in goat anti-rabbit HRP conjugated secondary antibody at 1:250, 1:500 or 1:1000. Area under the curve of dot intensity of the dot blot assay was then measured using imageJ, area determined in arbitrary units.

Secondary Antibody concentration	Primary Antibody concentration	Primary Antibody				Pre-immune serum			
		A1	A2	B1	B2	A1-0	A2-0	B1-0	B2-0
1:250	1:250	8618	7763	6533	6473	4314	3576	4338	5203
	1:500	4728	6102	5677	7106	3805	3812	4900	5838
	1:1000	8011	8855	6335	9740	3516	3757	4287	3135
	1:2000	8840	7539	9217	5996	3278	3977	3247	4377
1:500	1:250	5202	4698	4780	5608	2671	2714	5429	6278
	1:500	5312	3566	3818	4337	2939	3380	4646	6647
	1:1000	6576	5097	4383	5272	2837	2049	3461	2669
	1:2000	5856	5308	3692	4834	3441	3083	5022	4798
1:1000	1:250	2879	2329	1993	3019	1744	4386	1398	1835
	1:500	2720	4124	4307	5371	2969	2630	2239	4889
	1:1000	3867	2986	3918	3645	1915	2542	4384	4086
	1:2000	3813	3688	3191	4350	2584	2285	1362	1886

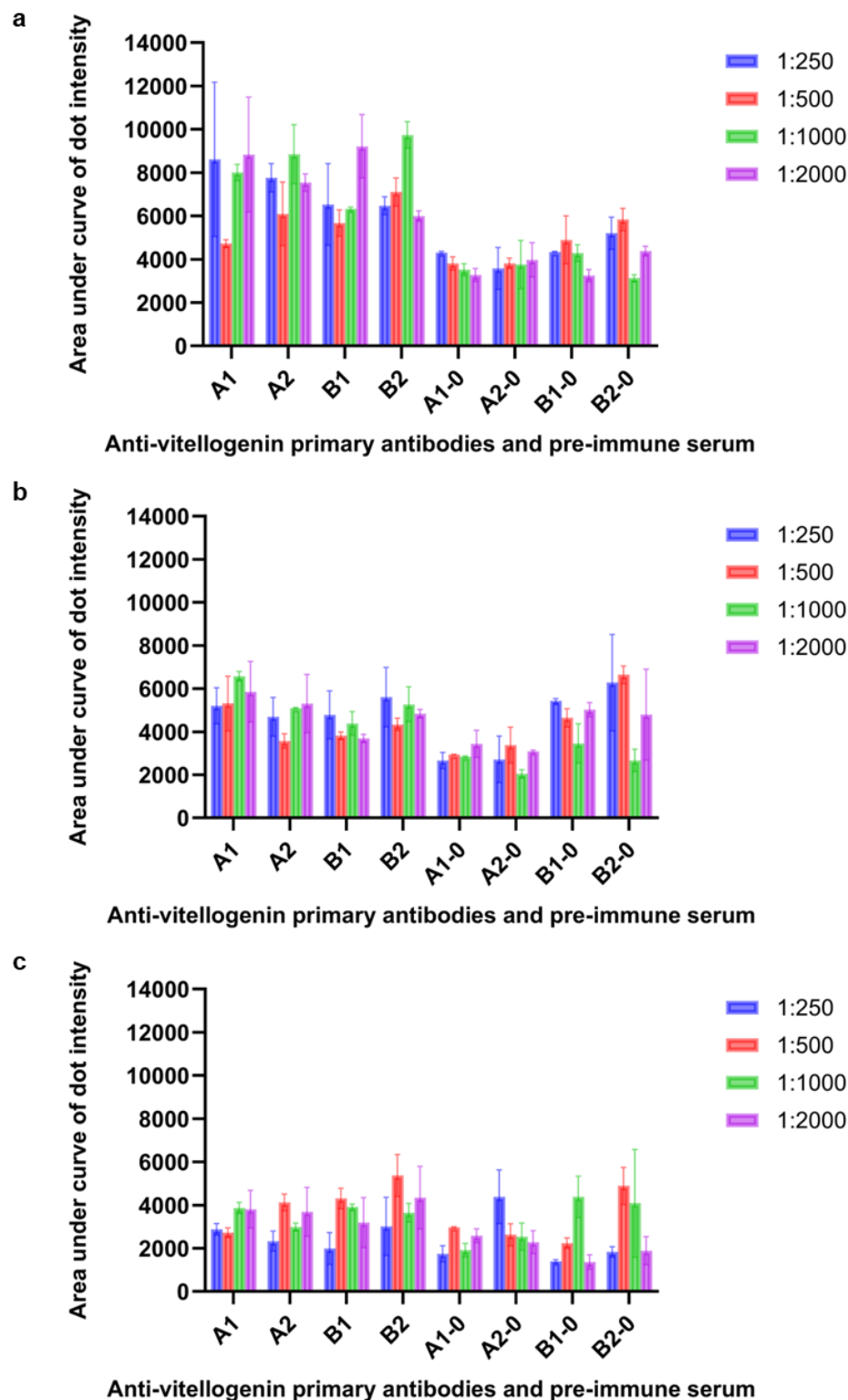


Figure 5.6: Area under the curve of mean dot intensity in dot blot assay, of dilution series using primary antibodies produced against synthetic peptides from *vit* gene sequences. A1 and A2: antibody one (A) produced from two different rabbits. B1 and B2: antibody two (B) produced from two different rabbits. A1-0, A2-0, B1-0 and B2-0: pre-immune serum from day zero rabbit bleeds. Incubated in goat anti-rabbit secondary antibody at 1:250 (a), 1:500 (b) or 1:1000 (c). Error bars represent standard error of the mean (n=2).

5.3.3.2 Test of Anti-VIT Recognition of Synthetic Peptides

Once it was established that there was an interaction between the primary and secondary antibodies, dot blot assays were used to determine whether the anti-VIT antibodies could successfully recognise the synthetic peptides (A and B). A serial dilution of each peptide was inoculated onto a membrane and was incubated in either anti-VIT A (Figure 5.8), anti-VIT B (Figure 5.9) or pre-immune serum. Each dot blot was performed in duplicated by utilising the antibodies produced by two different rabbits, which was used as two separate batches of antibodies, 1 or 2. In Figure 5.7, a schematic drawing of the layout of each dot blot is given, indicating sample used and concentration.

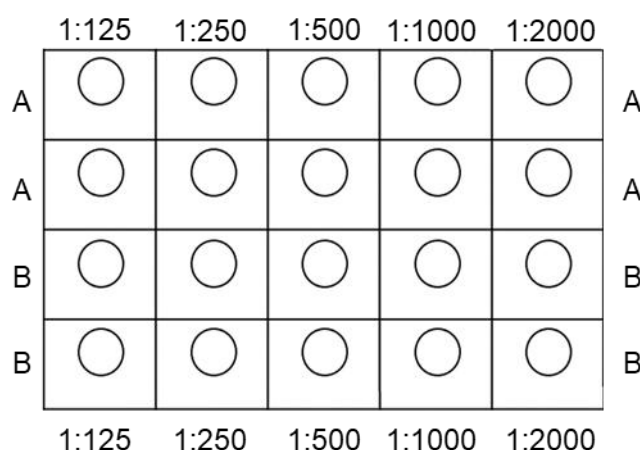


Figure 5.7: Schematic drawing of dot blot assays presented in Figure 5.5, indicating sample position of serial dilution using synthetic peptides A and B, from 1:125 to 1:2000.

In Figure 5.8, the synthetic peptide A (pA) has successfully been recognised by the anti-VIT A antibodies, at both concentrations of primary antibody used, 1:1000 and 1:2000. The anti-VIT A antibodies do not recognise synthetic peptide B. Dot blot assays using anti-VIT B antibody only recognised the synthetic peptide B at 1:125, for one of the antibody replicates (B1) (Figure 5.9) and was not successful in recognising protein in any other dilution tested. The anti-VIT B antibody did not recognise synthetic peptide A.

Membranes were also incubated in pre-immune serum from day zero bleeds from either of two rabbits for each synthetic peptide inoculum. These dot blot assays were performed to determine whether there was non-specific recognition of the synthetic peptides by the rabbit serum collected prior to antibody production. There were no interactions between the pre-immune serum from any animal with either synthetic peptide.

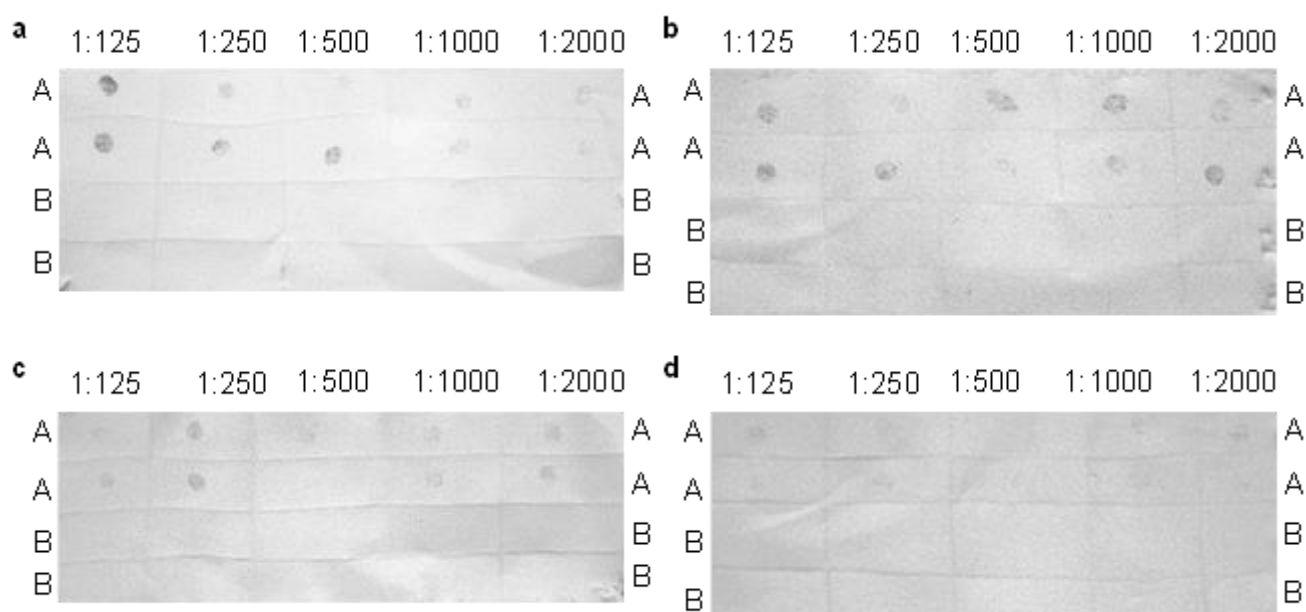


Figure 5.8: Dot blot assays of dilution series of synthetic peptides (A & B) from the *vit* gene sequences, on PVDF membrane. Incubated in anti-vitellogenin primary antibody A. a: A1 at 1:1000, b: A1 at 1:2000, c: A2 at 1:1000, and d: A2 at 1:2000, from rabbit day 35 serum. Incubated in goat anti-rabbit secondary antibody at 1:250, visualised using DAB substrate.

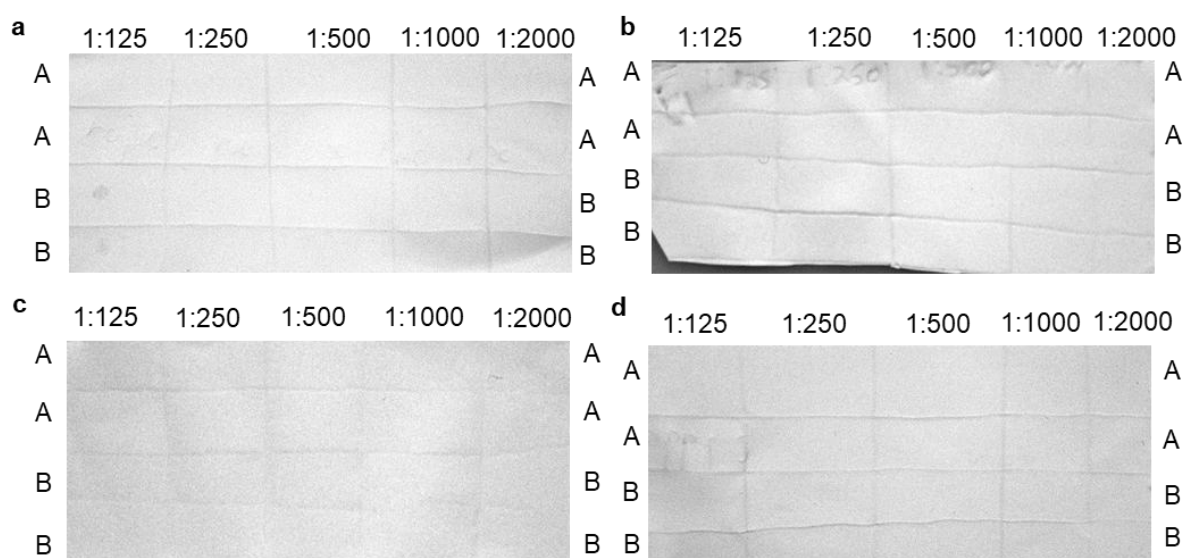


Figure 5.9: Dot blot assays of dilution series of synthetic peptides (A and B) from *vit* gene sequences, on PVDF membrane. Incubated in anti-vitellogenin primary antibody B, a: B1 at 1:1000, b: B1 at 1:2000, c: B2 at 1:1000 and d: B2 at 1:2000, from rabbit day 35 serum. Incubated in goat anti-rabbit secondary antibody at 1:250, visualised using DAB substrate.

The dot intensity of each dot was measured and is reported in Table 5.5 and illustrated in Figure 5.10. As shown in Table 5.5, the mean dot intensity was shown to be greater in the 1:1000 antibody concentrations than 1:2000 of all antibodies. Each dilution of the synthetic peptides was compared between the two concentrations of primary antibody tested (1:1000 and 1:2000). There was a significant difference between the A2 primary antibody concentrations in dot intensity of the 1:125 and 1:250 dilutions of pA. Also, the dot intensity of the 1:250 dilution of pA was significantly greater in the 1:1000 A1 primary antibody concentration than the 1:2000 concentration ($p < 0.05$). As no dot was detected in the 1:2000 concentration of the B1 antibody, the dot observed in the 1:125 dilution at the 1:1000 antibody concentration was shown to be significantly different ($p < 0.05$).

The difference in dot intensity of the synthetic peptide dilution at each primary antibody concentration could also be considered. When using the A1 primary antibody at 1:1000, there was a significant difference in the dot intensity between the 1:125 and 1:1000 dilutions of pA ($p < 0.05$), whereas, when using the A2 primary antibody at 1:1000, the dot intensity of pA at dilution 1:125, was significantly greater than the 1:250 and 1:2000 dilutions ($p < 0.05$). Also, the 1:250 dilution of pA had a dot intensity that was significantly greater than the 1:500, 1:1000, and 1:2000 dilutions ($p < 0.05$). Finally, the dot intensity was also significantly different between the 1:1000 and 1:2000 dilutions ($p < 0.05$). Next, the change in dot intensity of the dilution series when incubated in the 1:2000 primary antibodies were determined. When the A1 and A2 antibodies were used at 1:2000, the dot intensity of the pA 1:125 dilution was significantly greater than 1:250 ($p < 0.05$). Using the A2 antibody also showed a significant difference between the 1:125 and the 1:500, 1:1000 and 1:2000 dilutions ($p < 0.05$). Furthermore, the dot intensity of the 1:250 dilution of pA was significantly greater than the 1:500, 1:1000 and 1:2000 dilutions, when using the A1 antibody at 1:2000 ($p < 0.05$). As no dots were visualised with the A antibodies against synthetic peptide B, all differences in dot intensity between synthetic peptides A and B were significant.

The anti-VIT B1 antibody only recognised the synthetic peptide B at 1:125; however, no other interactions between the peptide and antibodies were observed and subsequently could not be measured. Similarly, the anti-VIT B antibody did not recognise pA. No reaction was observed when samples were incubated in pre-immune serum, therefore the mean dot intensity was zero and could not be plotted in Figure 5.10.

Table 5.5: Area under the curve of mean dot intensity of dot blot assays using synthetic peptides (A and B) from *vit* gene sequences. Samples were incubated in anti-vitellogenin A or anti-vitellogenin B, from Day 35 bleeds of two rabbits per antibody or incubated in pre-immune serum, from day zero bleeds, at 1:1000 or 1:2000. Samples were then incubated in anti-rabbit secondary antibody at 1:250. Area under the curve of dot intensity of the dot blot assays were then measured using imageJ, measured in arbitrary units.

Antibody; Secondary antibody dilution	Peptide A					Peptide B				
	1:125	1:250	1:500	1:1000	1:2000	1:125	1:250	1:500	1:1000	1:2000
Anti-VIT A1; 1:1000	9071	7685	7897	7256	8700	0	0	0	0	0
Anti-VIT A1; 1:2000	7824	4643	7595	7648	7511	0	0	0	0	0
Anti-VIT A2; 1:1000	3110	7072	2692	4272	1192	0	0	0	0	0
Anti-VIT A2; 1:2000	6172	2677	4061	2963	3035	0	0	0	0	0
Anti-VIT B1; 1:1000	0	0	0	0	0	4724	0	0	0	0
Anti-VIT B1; 1:2000	0	0	0	0	0	0	0	0	0	0
Anti-VIT B2; 1:1000	0	0	0	0	0	0	0	0	0	0
Anti-VIT B2; 1:2000	0	0	0	0	0	0	0	0	0	0
Pre-immune serum A1; 1:1000	0	0	0	0	0	0	0	0	0	0
Pre-immune serum A1; 1:2000	0	0	0	0	0	0	0	0	0	0
Pre-immune serum A2; 1:1000	0	0	0	0	0	0	0	0	0	0
Pre-immune serum A2; 1:2000	0	0	0	0	0	0	0	0	0	0
Pre-immune serum B1; 1:1000	0	0	0	0	0	0	0	0	0	0
Pre-immune serum B1; 1:2000	0	0	0	0	0	0	0	0	0	0
Pre-immune serum B2; 1:1000	0	0	0	0	0	0	0	0	0	0
Pre-immune serum B2; 1:2000	0	0	0	0	0	0	0	0	0	0

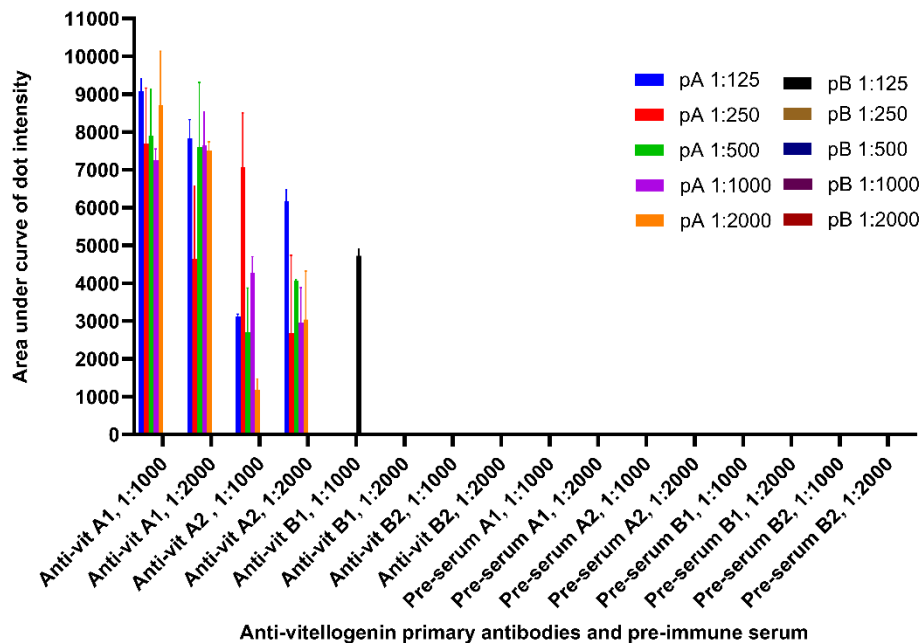


Figure 5.10: Area under the curve of mean dot intensity from dot blot assay of dilution series using synthetic peptides (A + B) from *vit* gene sequences, from 1:125 to 1:2000. Membranes were incubated in either anti-vitellogenin A or anti-vitellogenin B, each produced in duplicate from day 35 bleeds from two rabbits per antibody, at 1:1000 or 1:2000; or incubated in pre-immune serum from bleeds at Day 0. Samples were then incubated in anti-rabbit secondary antibody at 1:250. Error bars represent standard error of the mean (n=2).

5.3.3.3 Detection of Cyst Nematodes using Anti-VIT A

The anti-VIT A antibody successfully recognised synthetic peptide A in dot blot assays. Once this had been established as a positive control, the antibody could be used to determine whether it could successfully recognise cyst nematode protein extracts from cysts of *H. schachtii* and *G. pallida*. The schematic diagram in Figure 5.11 represents the layout for each dot blot assay. A serial dilution of the cyst nematode sample was inoculated onto the membrane and incubated at either 1:250, 1:500 or 1:1000 concentration of anti-VIT A antibody.

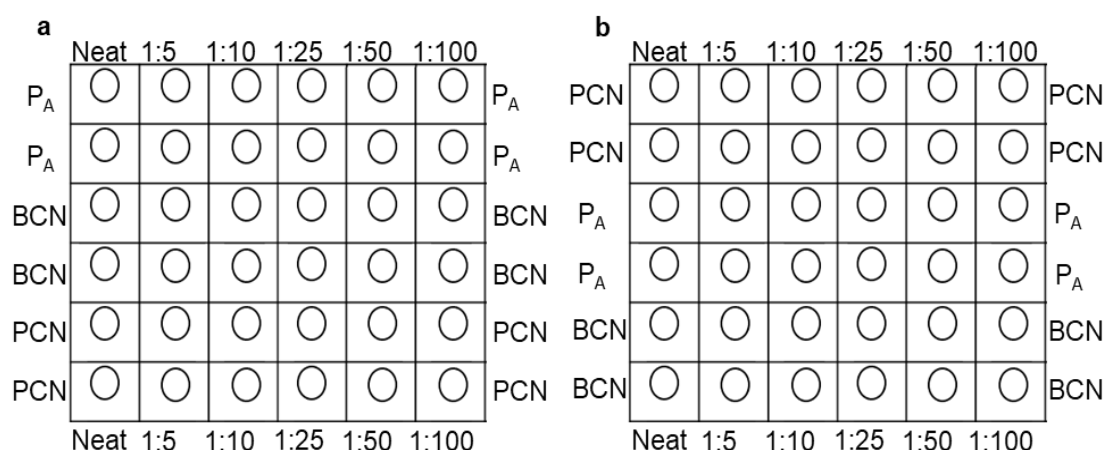


Figure 5.11: Schematic drawing of dot blot assays presented in Figure 5.12. Layout (a) was used in dot blots a, c and e; whereas layout (b) was used in dot blots b, d and f. Schematic drawings indicate sample position of serial dilution using synthetic peptide *A* (P_A), protein extracts of *Heterodera schachtii* cysts (BCN) and *Globodera pallida* cysts (PCN), from 1:125 to 1:2000.

Dot blot assays shown in Figure 5.12 present the successful recognition of *H. schachtii* and *G. pallida* cyst homogenates and pA by anti-VIT A antibodies at 1:250, 1:500 and 1:1000. The dots are easiest to visualise when neat samples are used (150 µg/ mL). It is clearly illustrated that when the concentration of the sample decreases, the intensity of the dot also decreases. When using the A1 anti-VIT antibody, all samples are visible to the 1:10 dilution (15 µg/ mL); however, for the A2 replicate, cyst nematode samples could only be clearly visualised when neat. The anti-VIT A antibody recognises the synthetic peptide at lower concentrations than the cyst homogenates of *H. schachtii* or *G. pallida*, and can be visualised at 1:100 dilution (1.5 µg/ mL) in most blots in Figure 5.13, excluding images a and f.

The measured values for dot intensity are given in Table 5.6 and graphically represented in Figure 5.13. The graphs in Figure 5.13 show the effect of decreasing the concentration of the sample on mean dot intensity of all samples. As shown in the values given in Table 5.6, the dot intensity steadily decreases for the synthetic peptide but has a more dramatic decline in dot intensity in the BCN and PCN samples. The antibody recognition of PCN is greater than BCN, as indicated by the mean dot intensity values.

The use of the A2 antibody gave mixed results in statistical analysis using two-way ANOVA. Similar to the A1 antibody, at all concentrations of the primary antibody, the neat BCN samples had a greater dot intensity than all other dilutions tested ($p < 0.05$). This trend was also observed in the PCN samples ($p < 0.05$); however, the comparison of neat samples to 1:5 was not significant. Additionally, when the primary antibody at either 1:250 or 1:500 was used, the 1:5 dilution of PCN was statistically different to all other dilutions tested ($p < 0.05$).

Across all three antibody concentrations tested, results demonstrated a significant difference between the neat, 1:5 and 1:10 dilutions, with the 1:100 dilution of pA ($p < 0.05$). Additionally, in the use of the 1:250 concentration of A2, neat samples had a significantly greater dot intensity than 1:50 and 1:100 ($p < 0.05$). Next, the use of 1:500 concentration of A2 showed significant differences of the 1:25 and 1:50 dilutions to the 1:100 dilution. Finally, in the use of the 1:1000 concentration of A2, significant differences between the neat samples and 1:10, 1:25 and 1:50 were observed ($p < 0.05$); this was also seen between the 1:5 dilution with 1:25 and 1:50 dilutions ($p < 0.05$); and for the 1:10 dilution with 1:25 and 1:50 dilutions ($p < 0.05$).

The difference in dot intensity could also be assessed between the different samples at each dilution and differing primary antibody concentrations. The results of a two-way ANOVA showed that when using the A1 antibody at 1:250, there was a significant difference in dot intensity of all dilutions of pA when compared to both BCN and PCN, excluding the 1:100 dilution ($p < 0.05$). There was also a significant difference shown between the intensity in BCN and PCN when samples were either neat or 1:10 ($p < 0.05$). In addition, in the use of antibody A2 at 1:250, there were significant differences in dot intensity of pA at all dilutions when compared to BCN and PCN ($p < 0.05$); however, this excluded the neat samples. These data also show that PCN had a significantly increased dot intensity in the neat and 1:5 samples when compared to BCN ($p < 0.05$). Furthermore, when the A1 antibody had a concentration of 1:500 or 1:1000, there were significant differences in dot intensity between pA and both BCN and PCN, from the neat samples to the 1:50 dilution; however, there was no significant difference between the 1:100 dilution at 1:500 concentration. The difference between the dot intensity of pA and both BCN and PCN at the 1:100 dilution, was instead shown to be significant at 1:1000. Alternatively, using the A2 antibody at 1:500, there were no significant differences between each of the neat samples; however, the dot intensity was significantly greater in pA than BCN at all other concentrations tested ($p < 0.05$). Also, there were significant differences between pA and PCN in the 1:10 dilution ($p < 0.05$). Moreover, the dot intensity from the 1:5 dilution of PCN was significantly greater than that of BCN, although all other dilutions were not significant. As fewer dots were identified in the use of the A2 antibody at 1:1000, only significant differences were shown in the sample dilutions of 1:5 and 1:10, where pA had a significantly greater intensity than both BCN and PCN.

Furthermore, the two-way ANOVA analysis with a Tukey post-hoc test resulted in statistical differences between the primary antibody concentrations (A1 and A2) at different dilutions of the samples. The dot intensity of the neat sample of pA was significantly greater in the 1:1000 A1 dilution than 1:500 ($p < 0.05$). Additionally, the dot intensity of 1:10 and 1:25 sample dilutions of pA was significantly greater in the 1:1000 A1 antibody dilution than the 1:250 and 1:500 ($p < 0.05$). There was also a significant difference in the dot intensity of the pA sample

at concentrations of 1:25 and 1:50, when the A2 antibody concentration was 1:1000 instead of 1:250 and 1:500, which gave significantly greater dot intensities ($p < 0.05$). The only significant difference in the dot intensity of a nematode sample was the PCN 1:5 dilution; the A2 antibody at a 1:1000 dilution gave significantly greater dot intensities than both the 1:250 and 1:500 dilutions ($p < 0.05$).

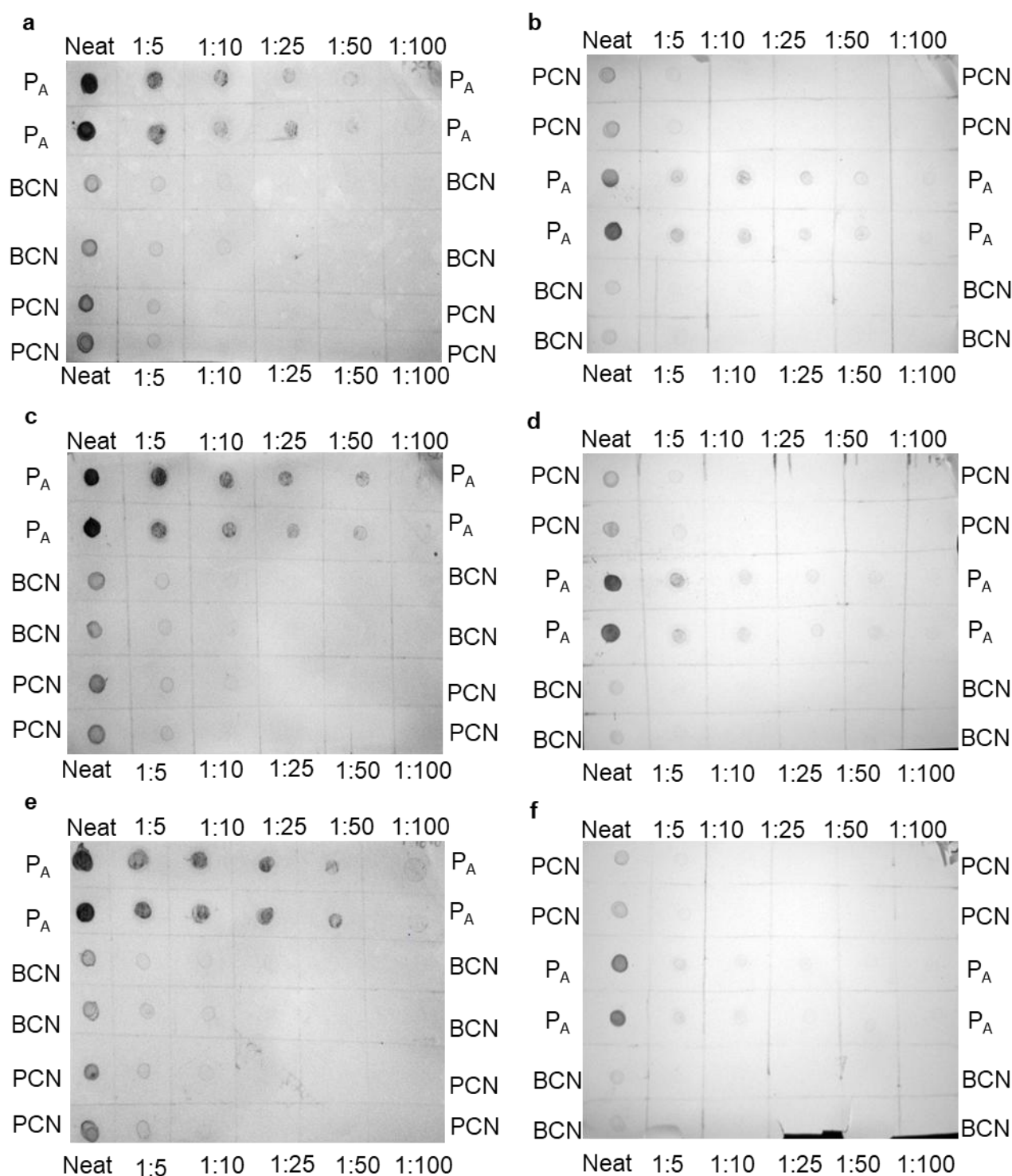


Figure 5.12: Dot blot assays of dilution series of the synthetic peptide A (PA), protein homogenates of *Heterodera schachtii* cysts (BCN) or *Globodera pallida* cysts (PCN), protein concentration adjusted to 150 µg/ mL, as neat and diluted to 1:100. The nitrocellulose membrane was incubated in anti-vitellogenin primary antibodies, from the day 70 serum from two rabbits, 'A1' and 'A2'. a) A1 at 1:250; b) A2 at 1:250; c) A1 at 1:500; d) A2 at 1:500; e) A1 at 1:1000; f) A2 at 1:1000. The membrane was also incubated in anti-rabbit secondary antibody at 1:250, visualised using 1-Step™ Ultra TMB-Blotting Solution.

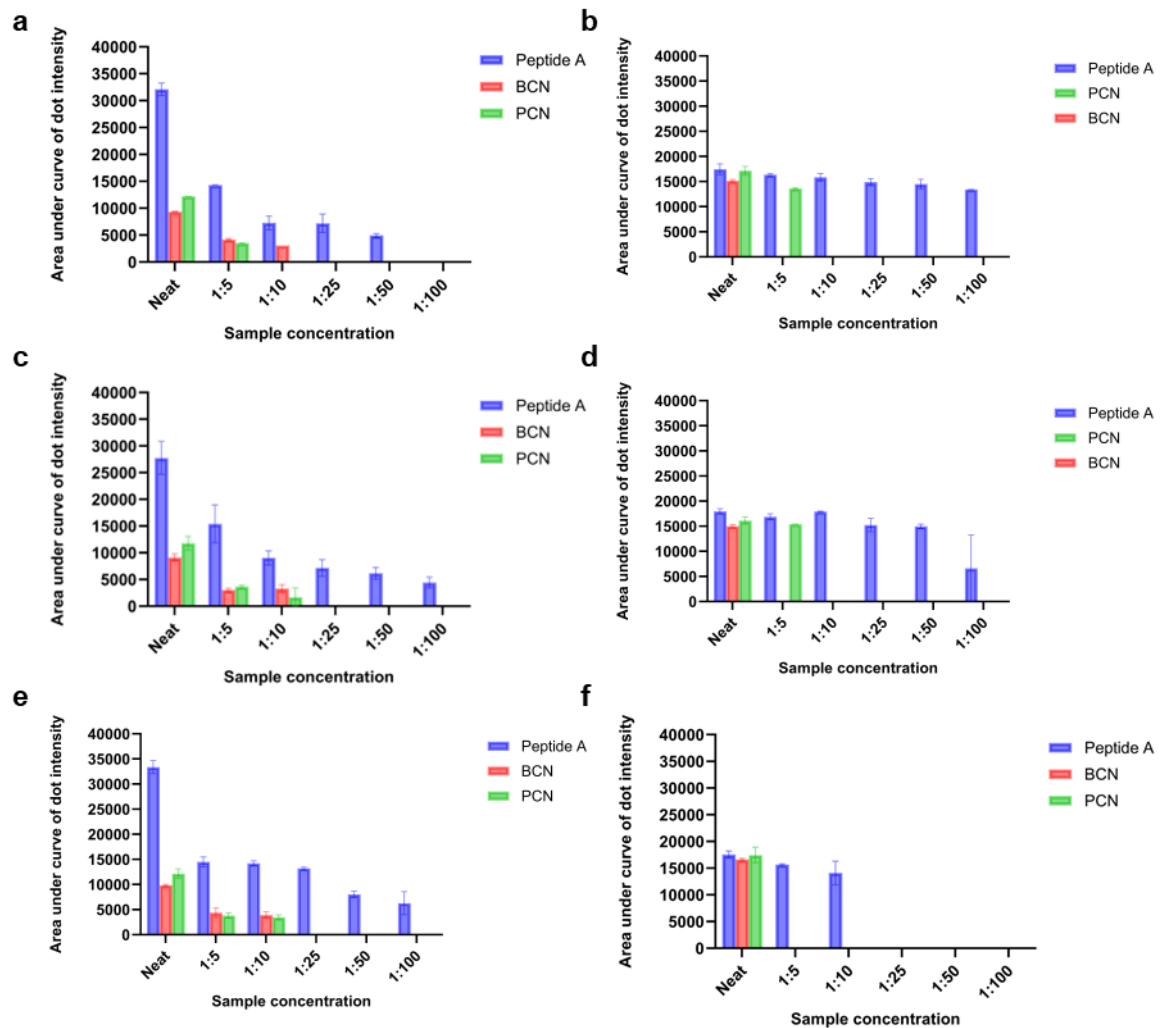


Figure 5.13: Area under the curve of mean dot intensity from dot blot assay of dilution series (neat to 1:100) using synthetic peptide A, protein homogenates adjusted to 150 µg/ mL of *Heterodera schachtii* cysts (BCN) or *Globodera pallida* cysts (PCN) (Figure 5.12), each dot was performed in duplicate. These samples were incubated in anti-vitellogenin primary antibodies from the day 70 serum from two rabbits, 'A1' and 'A2'. a) A1 at 1:250; b) A2 at 1:250; c) A1 at 1:500; d) A2 at 1:500; e) A1 at 1:1000; f) A2 at 1:1000. Samples were incubated in anti-rabbit secondary antibody at 1:250. Error bars represent standard error of the mean (n=2).

Table 5.6: Area under the curve of mean dot intensity of dot blot assays of dilution series using synthetic peptide A (pA), protein homogenates of *Heterodera schachtii* cysts and *Globodera pallida* cysts, adjusted to 150 µg/ mL (Figures 5.12). Samples were incubated in anti-vitellogenin primary antibodies (A1 or A2), from rabbit day 70 serum, at 1:250, 1:500 or 1:1000. Samples were then incubated in anti-rabbit secondary antibody at 1:250. Area under the curve of dot intensity of the dot blot assay was then measured using imageJ.

Primary Antibody Conc.		1:250			1:500			1:1000		
Sample Conc.	Ab.	pA	BCN	PCN	pA	BCN	PCN	pA	BCN	PCN
Neat	A1	32117	9365	12205	27765	9103	11792	33377	9849	12155
	A2	17462	15151	17135	17962	15017	16119	17544	16608	17454
1:5	A1	14322	4183	3510	15413	3019	3660	14510	4374	3827
	A2	16352	0	13610	16844	0	15428	15643	0	0
1:10	A1	7280	3082	0	9049	3282	1685	14209	3931	3457
	A2	15836	0	0	17900	0	0	14105	0	0
1:25	A1	7209	0	0	7182	0	0	13202	0	0
	A2	14871	0	0	15251	0	0	0	0	0
1:50	A1	4936	0	0	6153	0	0	8020	0	0
	A2	14516	0	0	14908	0	0	0	0	0
1:100	A1	0	0	0	4414	0	0	6281	0	0
	A2	13401	0	0	6624	0	0	0	0	0

Statistical differences in the dot intensity of the sample dilution series, at each concentration of the primary antibody could be explored. At each A1 primary antibody concentration, the intensity of neat samples was significantly greater than all other dilutions tested ($p < 0.05$). At the 1:250 and 1:500 concentration of the A1 antibody, the dot intensity from the 1:5 dilution of pA was significantly greater than all other dilutions ($p < 0.05$); however, when the antibody concentration was 1:1000, only differences between the 1:5 dilution and 1:50, and 1:100 were shown to be significant ($p < 0.05$). Additionally, when the A1 antibody concentration was 1:250; there were further significant differences in dot intensity of pA dilutions, between 1:10 and 1:100; 1:25 and 1:100; 1:50 and 1:100 ($p < 0.05$). Whereas, when the A1 concentration was 1:1000, the dot intensity of the pA 1:10 dilution was significantly greater than 1:50 and 1:100 ($p < 0.05$). There were also significant differences between the 1:25 dilution and 1:50, and 1:100 ($p < 0.05$). Next, when using the A1 antibody concentration of 1:250 and 1:1000, there were significant differences between the dilutions of the BCN samples, when diluted to either 1:5 and 1:10, against all other dilutions tested ($p < 0.05$). Finally, the dot intensity of the PCN 1:5 dilution was shown to be significantly greater than the 1:25, 1:50, and 1:100 dilutions ($p < 0.05$).

5.3.4 Western Blot Assays

5.3.4.1 Serial Dilution of Synthetic Peptides

Following the successful recognition of proteins from PCN and BCN, as well as synthetic peptide A by anti-VIT A antibody using dot blot assays, western blots were used to determine the size of the proteins recognised by this antibody. In Figure 5.14, four western blot membranes are presented that have been incubated in A1 or A2 antibodies from either day 35 or day 70 of anti-VIT antibody production. All images show that the western blot had been successfully transferred and the anti-VIT A antibody also recognises the synthetic peptide A when performing this assay, at approximately 12 kDa. Figure 5.14 reveals in each case that there has been a steady decrease in protein visualised on the membrane as the protein concentration loaded onto the gel decreases. In western blot analysis, a serial dilution of the synthetic peptide B was also performed and incubated in anti-VIT B; however, no reaction was observed.

The bands visualised in Figure 5.14 were measured using image J to give an intensity value (Table 5.7). The graph in Figure 5.15 represents these values and shows that there was a decrease in band intensity with the reduction of peptide concentration, which can also be observed in Figure 5.14. The day 70, A1 replicate of anti-VIT A antibodies had the highest band intensity whereas the day 35 bands had the lowest band intensity. It can also be seen in this graph that the day 70 antibodies recognised protein at lower concentrations (0.008 mg/mL), compared to no visual band when using the day 35 antibodies. The different timed animal bleeds had similar band intensities at the higher concentration of 1 mg/mL and 0.5 mg/mL; however, below 0.25 mg/mL the day 70 antibodies gave higher band intensities and were therefore more sensitive to the protein present. The ability of the antibodies to recognise protein samples at lower concentrations was useful due to the relatively low concentration of protein in some cyst nematode samples used in this work.

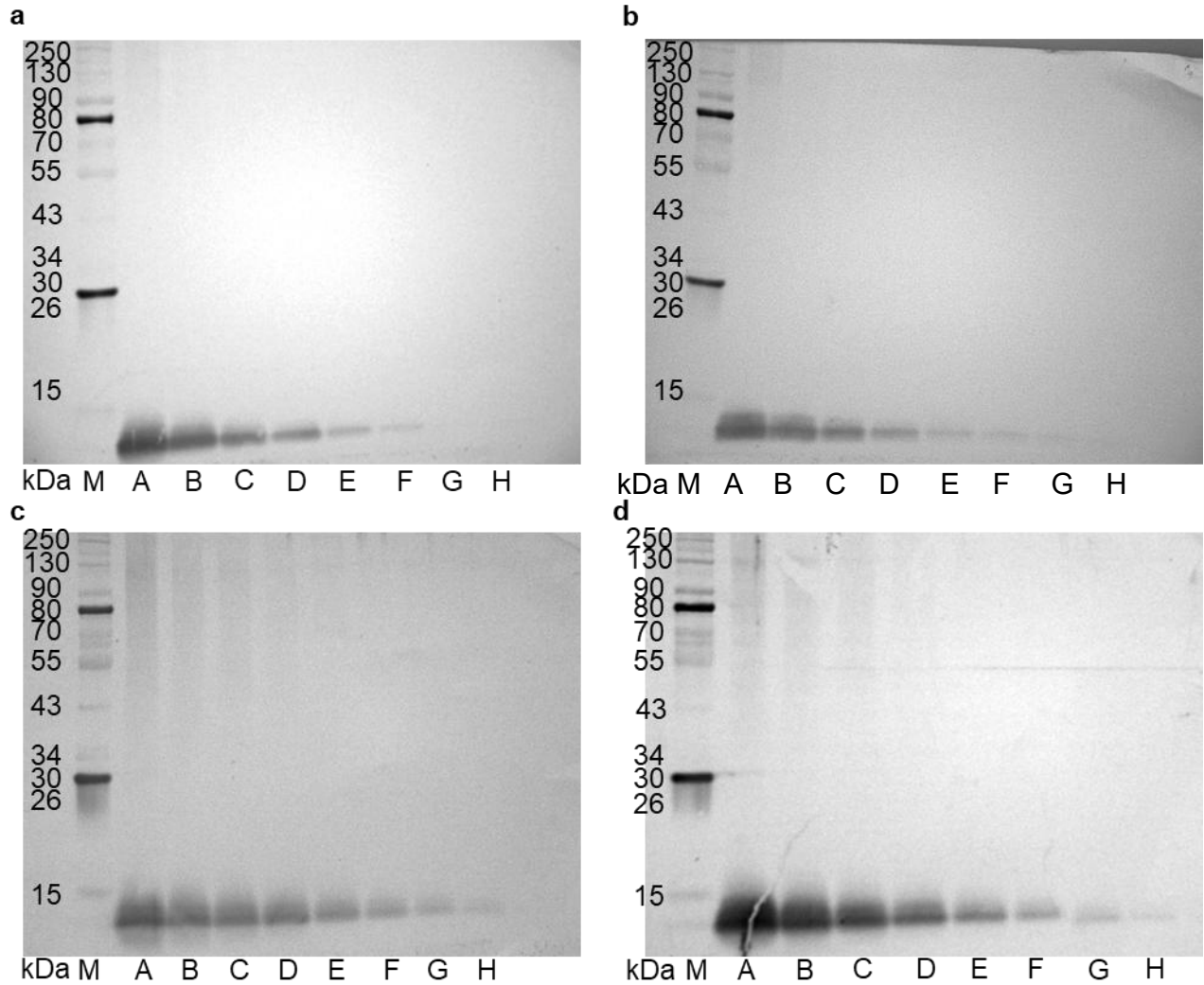


Figure 5.14: Western blot analysis of dilution series of synthetic peptide A. Proteins were transferred onto a PVDF membrane and incubated in anti-VIT antibodies, from bleeds at day 35 (a and b) or from day 70 (c and d). Replicate A1: a and c; replicate A2: b and d. Membranes were then incubated in goat anti-rabbit secondary antibody at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; A: 1 mg/ mL; B: 0.5 mg/ mL; C: 0.25 mg/ mL; D: 0.15 mg/ mL; E: 0.0625 mg/ mL; F: 0.0313 mg/ mL; G: 0.0156 mg/ mL; H: 0.0078 mg/ mL.

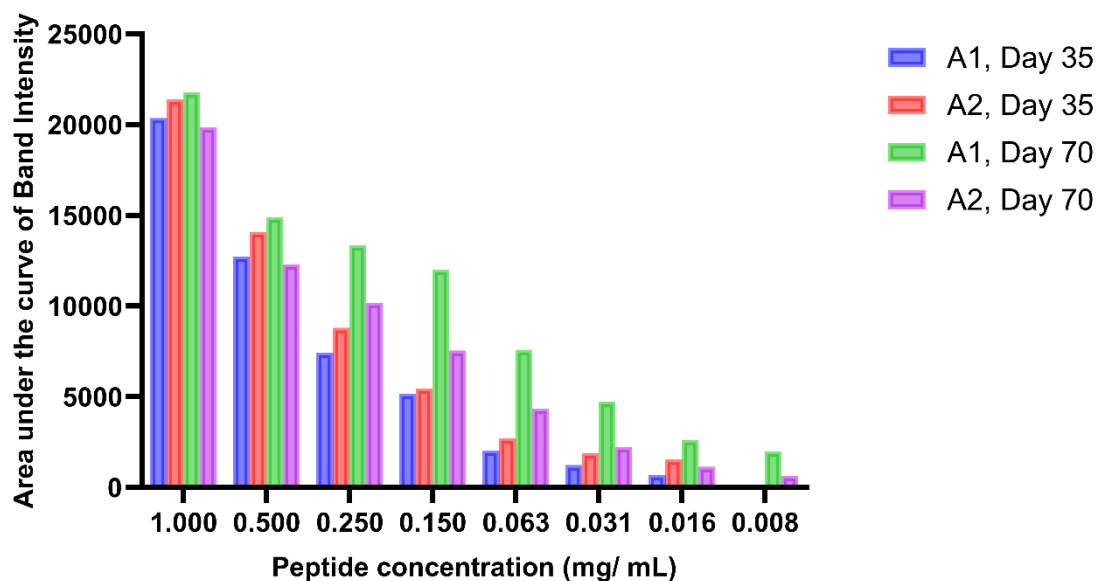


Figure 5.15: Area under the curve of band intensity in western blot analysis of dilution series of synthetic peptide A. Incubated in anti-vitellogenin antibodies (A1 or A2) at either day 35 or 70 of production, with a primary antibody concentration of 1:250 and incubated in goat anti-rabbit secondary antibody at 1:250 (n=1).

Table 5.7: Area under the curve of band intensity in western blot analysis of dilution series of synthetic peptide A. Incubated in anti-vitellogenin antibodies (A1 or A2) at either day 35 or 70 of production, at 1:250 primary antibody concentration and 1:250 secondary antibody concentration.

Peptide concentration (mg/ mL)	Primary Antibody			
	A1, Day 35	A2, Day 35	A1, Day 70	A2, Day 70
1	20369	21378	21776	19860
0.5	12726	14096	14875	12307
0.25	7429	8798	13366	10169
0.15	5178	5472	12016	7540
0.063	2024	2710	7576	4341
0.031	1232	1888	4708	2208
0.016	674	1530	2629	1142
0.008	0	0	1969	621

5.3.5 Western Blot Analysis of Cyst Nematode Proteins Recognised by Anti-VIT A

After a series of experiments to optimise the transfer of cyst nematode proteins onto membranes and incubation in anti-VIT A antibodies, several experiments were conducted to determine the molecular weight of proteins recognised by the anti-VIT antibodies. Figure 5.16 and Figure 5.17 present two replicates utilising the same cyst nematode samples, with synthetic peptide A utilised as a positive control. These figures show a comparison between cyst homogenate samples of *G. pallida*, *G. rostochiensis* and *H. schachtii*. Furthermore, Figure 5.18 and Figure 5.19 extend this work to consider the recognition of proteins in J2 of *H. schachtii*. Finally, in Figure 5.20 and 5.21, this work was broadened to include samples containing eggs from *G. pallida*, *G. rostochiensis* and *H. schachtii*, and again the J2 of *H. schachtii*.

In Figure 5.16, six protein bands have been recognised by the anti-VIT A antibody, in cyst samples from *G. pallida*, these have an approximate molecular weight of 180 kDa, 138 kDa, 108 kDa, 89 kDa, 77 kDa and 45 kDa. Alternatively, from cyst samples of *G. rostochiensis*, there are four protein bands recognised by anti-VIT A at 174 kDa, 152 kDa, 108 kDa and 12 kDa. Finally, in the *H. schachtii* cyst samples, six protein bands have been more successfully recognised. The anti-VIT A antibody recognises protein bands at 174 kDa, 108 kDa, 89 kDa, 45 kDa, 33 kDa and 17 kDa. The synthetic peptide A was also recognised by the anti-VIT A antibody in this western blot, with a molecular weight of 12 kDa. Homogenised *C. elegans* eggs were used as a prospective negative control and resulted in no proteins from this sample being recognised by the antibody. It is apparent that the antibody recognises proteins with similar molecular weights between *G. rostochiensis* and *H. schachtii* at 174 kDa. Similarly, a protein band was recognised at approximately 89 kDa in both *G. pallida* and *H. schachtii*. Finally, protein bands were in common across all species at approximately 108 kDa.

A western blot was also conducted using the same samples as shown in Figure 5.16, this replicate blot is given in Figure 5.17. Due to the variation in running SDS-Page gels, there was some slight variation in the molecular weight designated to the protein bands present between these two gels. In the *G. pallida* cysts samples, proteins were measured at 169 kDa, 141 kDa, 108 kDa, 89 kDa, 79 kDa, 55 kDa, 51 kDa and 46 kDa. The recognition of protein bands in *G. rostochiensis* was less successful with only three bands recognised at 178 kDa, 46 kDa and 13 kDa. Furthermore, in *H. Schachtii* six protein bands were recognised at 199 kDa, 93 kDa, 51 kDa, 46 kDa, 34 kDa and 18 kDa. The synthetic peptide used as a positive control had a molecular weight of 13 kDa. When comparing the recognised protein bands of *G. pallida* in Figure 5.16 and Figure 5.17, five of the protein bands are in approximately the same position, with two additional bands at 51 kDa and 55 kDa. Alternatively, only two bands are in similar

positions from *G. rostochiensis*. Finally, in *H. schachtii* four protein bands are in similar positions. In this western blot, all species share a protein band of the same molecular weight at 46 kDa.

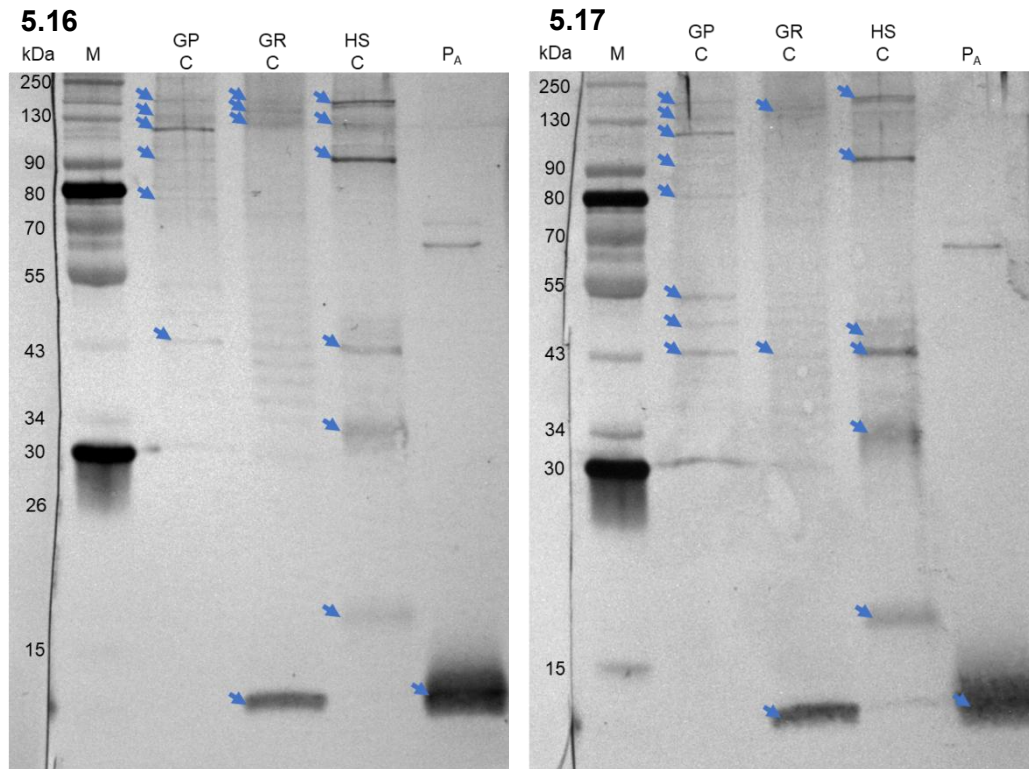


Figure 5.16 & 5.17: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Protein concentrations were adjusted to 200 µg/ mL and transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts; GR C: *G. rostochiensis* cysts; HS C: *Heterodera schachtii* cysts; P_A: synthetic peptide A (62.6 µg/ mL); *Caenorhabditis elegans* eggs used as negative control (200 µg/ mL). Blue arrows indicate measured bands.

In Figure 5.18, a western blot comparing *G. pallida* cysts and *H. schachtii* cysts and J2, proteins were partially transferred onto a membrane, possibly due to poor contact between the lower half of the gel and the membrane. Protein bands recognised in *G. pallida* had molecular weights of 156 kDa, 81 kDa, 63 kDa, 61 kDa, 55 kDa, 54 kDa, 53 kDa and 45 kDa. The *H. schachtii* cyst sample had three protein bands at 194 kDa, 148 kDa and 91 kDa; two of these bands approximately correspond to two recognised in Figure 5.17, at 199 kDa and 93 kDa. Two faint bands were recognised in the J2 of *H. schachtii* at approximately 209 kDa and 148 kDa. The smaller of the two bands observed in J2 of *H. schachtii* is also present in the cysts of this species. Again, the synthetic peptide was a successful positive control at approximately 11 kDa.

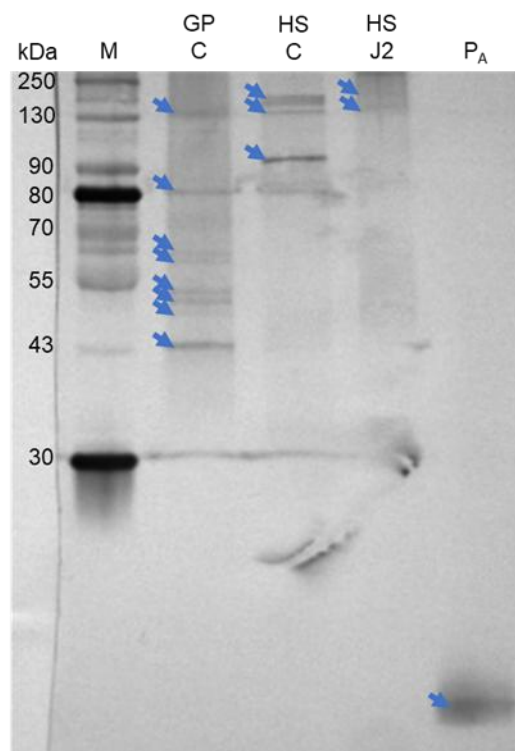


Figure 5.18: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Proteins were transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts (362 µg/ mL); HS C: *Heterodera schachtii* cysts (194 µg/ mL); HS J2: *H. schachtii* second-stage juveniles (30.43 µg/ mL); P_A: synthetic peptide A (62.6 µg/ mL); *Caenorhabditis elegans* eggs used as negative control (150 µg/ mL). Blue arrows indicate measured bands.

Another western blot which compared the cyst samples of *G. pallida*, *G. rostochiensis* and *H. schachtii* and the J2 of *H. schachtii* is given in Figure 5.19. In this western blot, only the lower molecular weight proteins were recognised by the anti-VIT A antibody. Figure 5.19 indicates that in all three cysts samples, a protein was recognised at 35 kDa. Furthermore, the *G. rostochiensis* and *H. schachtii* cyst samples also possessed protein bands at 19 kDa and 14 kDa. The synthetic peptide A was also measured to be 14 kDa.

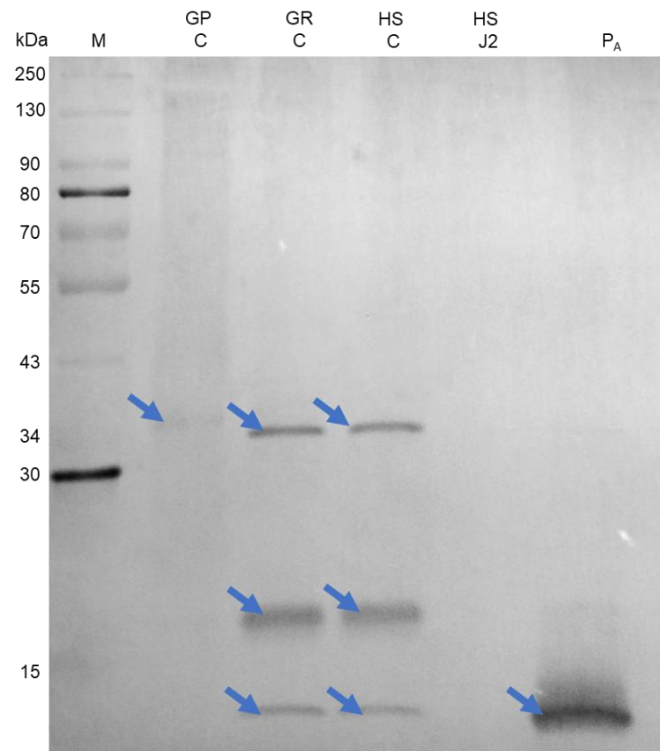


Figure 5.19: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Proteins were adjusted to 250 $\mu\text{g}/\text{mL}$, transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts; GR C: *G. rostochiensis* cysts; HS C: *Heterodera schachtii* cysts; HS J2: *H. schachtii* second-stage juveniles; P_A: synthetic peptide A (62.6 $\mu\text{g}/\text{mL}$). Blue arrows indicate measured bands.

The proteins recognised by the anti-VIT A antibody were then compared between cysts and eggs of *G. pallida*, *G. rostochiensis* and *H. schachtii*. Protein extracts of *H. schachtii* J2 were also used and are presented in Figure 5.20. It was observed that the proteins recognised by the anti-VIT A antibody had shared molecular weights between the cysts and eggs of *G. pallida*, at approximately 158 kDa and 101 kDa. Seven protein bands were recognised from cysts of *G. rostochiensis*; however, only five were visualised in their eggs. Approximate molecular weights of proteins shared between the cysts and eggs of *G. rostochiensis* were 145 kDa, 133 kDa, 95 kDa, 80 kDa and 75 kDa. Additional protein bands in cysts had approximate molecular weights of 112 kDa and 48 kDa. The anti-VIT A antibody recognised proteins in both cysts and eggs of *H. schachtii* with shared molecular weights of 144 kDa, 96 kDa, 85 kDa, 48 kDa and 34 kDa. The eggs of *H. schachtii* had an additional band at 246 kDa, which was also shared with the J2 sample. The J2 presented a further band with a molecular weight of 113 kDa. The synthetic peptide was recognised and had a molecular weight of 12 kDa.

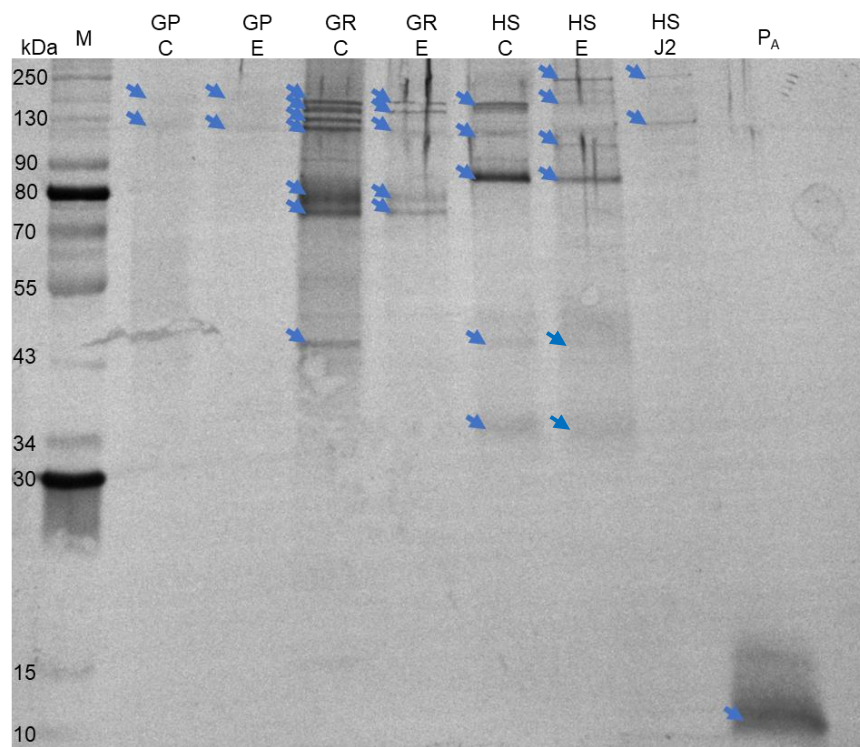


Figure 5.20: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Proteins were transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts (79.63 µg/ mL); GP E: *G. pallida* eggs (155.25 µg/ mL); GR C: *G. rostochiensis* cysts (561.5 µg/ mL); GR E: *G. rostochiensis* eggs (214.63 µg/ mL); HS C- *Heterodera schachtii* cysts (219.63 µg/ mL); HS E: *H. schachtii* eggs (59.63 µg/ mL); HS J2- *H. schachtii* second-stage juveniles (concentration below detection limit); P_A: synthetic peptide A (62.6 µg/ mL). Blue arrows indicate measured bands.

The final western blot (Figure 5.21) had the same design as Figure 5.20 but used alternative samples for the cysts of *G. pallida*, *G. rostochiensis* and *H. schachtii*; the same samples for eggs were used as previously studied. In this western blot no protein was visible from the *G. pallida* eggs. Although, the anti-VIT A antibody did recognise proteins in *G. pallida* cysts. Two proteins of the same molecular weight were shared between the *G. pallida* and *G. rostochiensis* cysts at 109 kDa and 83 kDa; however, two further bands were visible in *G. rostochiensis* cysts at 78 kDa and 11 kDa. Similarly, the 109 kDa band was also visible in the *G. rostochiensis* egg sample, with a further shared band between the cysts and eggs at 78 kDa. The egg sample also contained a visible band at 74 kDa. When comparing the *H. schachtii* cysts and eggs, they shared five protein bands with the same molecular weight, at 186 kDa, 113 kDa, 94 kDa, 50 kDa and 32 kDa. In the J2 sample, two bands were also visible

at 113 kDa and 50 kDa. The synthetic peptide was recognised with a molecular weight of 11 kDa.

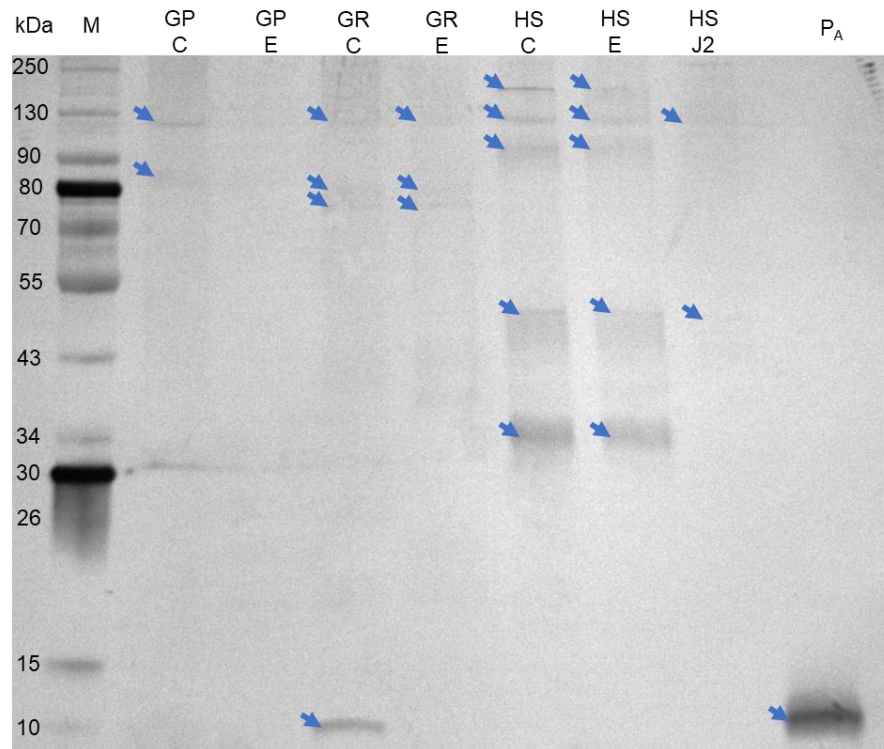


Figure 5.21: Western blot analysis of cyst nematode proteins recognised by anti-vitellogenin antibody A. Proteins were adjusted to 70.5 µg/ mL, before being transferred onto a PVDF membrane and incubated in anti-vitellogenin antibody A, at 1:250. An anti-rabbit secondary antibody was used at 1:250 and visualised using 1-Step™ Ultra TMB-Blotting Solution. M: Molecular weight marker; GP C: *Globodera pallida* cysts; GP E: *G. pallida* eggs; GR C: *G. rostochiensis* cysts; GR E: *G. rostochiensis* eggs; HS C- *Heterodera schachtii* cysts; HS E: *H. schachtii* eggs; HS J2- *H. schachtii* second-stage juveniles; P_A: synthetic peptide A at 280 µg/ mL. Blue arrows indicate measured bands.

As shown in Table 5.8, western blots containing *G. pallida* cysts, from five biological replicates, comprised of six proteins of different molecular weights identified by the anti-VIT A antibody. These proteins had a molecular weight of 157 kDa, 107 kDa, 81 kDa, 55 kDa, 52 kDa and 46 kDa; however, there were seven additional proteins recognised that were not identified from multiple biological replicates. The anti-VIT A antibody recognised five protein bands from four biological replicates of *G. rostochiensis*, with molecular weights of 110 kDa, 82 kDa, 77 kDa, 47 kDa and 13 kDa. Again, seven protein bands were recognised from different western blots that only appeared in one biological replicate. Next, the five biological replicates of *H. schachtii* cysts samples contained six protein bands recognised by the anti-VIT A antibody, with molecular weights of 146 kDa, 95 kDa, 89 kDa, 47 kDa, 34 kDa and 19 kDa. The *H. schachtii* J2 were taken from three separate samples; however, there was not consistency in the molecular weight these proteins recognised. Therefore, from the three biological replicates used, there were five protein bands recognised. The egg samples of all species were only taken from one biological replicate each; with two proteins recognised from *G. pallida* eggs, five from *G. rostochiensis* eggs and eight from *H. schachtii* eggs. The average molecular weight of the positive control, synthetic peptide A, was 12 kDa.

Table 5.8: Summary table of western blots presented in Figure 5.16 - 5.21. Molecular weight (MW) values (kDa) for each band measured from each figure are given. The MW of visualised bands observed in both technical replicates of Figures 5.16 and 5.17 are given. Figure 5.18 - Figure 5.21 represent different biological replicates for cyst samples. Molecular weights were calculated for *H. schachtii* second-stage juvenile samples, different biological replicates are presented in Figures 5.18 and 5.19, the same sample was used in Figures 5.20 and 5.21, the MW of bands visualised in both samples are also given. The MW of proteins from egg samples were derived from one biological replicate. The mean values from samples deriving from different biological replicates were then calculated.

Sample	MW (kDa) values found in both gel replicates. Fig. 5.16 & Fig. 5.17	Figure 5.18, MW (kDa)	Figure 5.19, MW (kDa)	Figure 5.20, MW (kDa)	Figure 5.21, MW (kDa)	MW (kDa) values found in both gel replicates. for <i>H. schachtii</i> J2, Fig. 5.20 & 5.21	Mean MW from biological replicates (kDa)
<i>G. pallida</i> Cysts		156		158			157
	140						
	108			101	109		107
	89						
	78	81			83		81
		63					
		61					
	55	55					55
	51	53/ 54					52
	46	45					46
			35				

Sample	MW (kDa) values found in both gel replicates. Fig. 5.16 & Fig. 5.17	Figure 5.18, MW (kDa)	Figure 5.19, MW (kDa)	Figure 5.20, MW (kDa)	Figure 5.21, MW (kDa)	MW (kDa) values found in both gel replicates. for <i>H. schachtii</i> J2, Fig. 5.20 & 5.21	Mean MW from biological replicates (kDa)
<i>G. rostochiensis</i> Cysts	176						
				145			
				133			
				112	109		110
				95			
				80	83		82
				75	78		77
				48			47
			35				
			19				
	13		14		11		13
<i>H. schachtii</i> Cysts		194					
					186		
		148		144			146
					113		
				96	94		95
	91	91		85			89
					50		
	46			48			47
	34		35	34	32		34
	18		19				19
			14				

Sample	MW (kDa) values found in both gel replicates. Fig. 5.16 & Fig. 5.17	Figure 5.18, MW (kDa)	Figure 5.19, MW (kDa)	Figure 5.20, MW (kDa)	Figure 5.21, MW (kDa)	MW (kDa) values found in both gel replicates. for <i>H. schachtii</i> J2, Fig. 5.20 & 5.21	Mean MW from biological replicates (kDa)
<i>H. schachtii</i> J2				246			
		209					
		148					
				113	114	114	
					50		
<i>G. pallida</i> Eggs				158			
				101			
<i>G. rostochiensis</i> Eggs				145			
				133			
					109		
				95			
				80	78		
				75	74		
<i>H. schachtii</i> Eggs				246			
					186		
				144			
					113		
				96	94		
				85			
				48	50		
				34	32		

Sample	MW (kDa) values found in both gel replicates. Fig. 5.16 & Fig. 5.17	Figure 5.18, MW (kDa)	Figure 5.19, MW (kDa)	Figure 5.20, MW (kDa)	Figure 5.21, MW (kDa)	MW (kDa) values found in both gel replicates. for <i>H. schachtii</i> J2, Fig. 5.20 & 5.21	Mean MW from biological replicates (kDa)
Synthetic peptide A	13	11	14	12	11		12

5.3.6 Sequencing of Vitellogenin Proteins

In order to confirm the identity of *H. schachtii* proteins detected by the anti-VIT antibodies, protein bands were cut out and processed as described in Section 5.2. Bands were removed from the Coomassie stained SDS-PAGE, of one biological replicate, at approximately 200-100 kDa, 46 kDa, 34 kDa and 18 kDa. The MALDI-MS/MS results were used to search the Uniprot database. The regions at approximately 200-100 kDa and 34 kDa, resulted in no significant peptide spectral matches. The band corresponding to the 46 kDa protein was given a provisional identification of *Arabidopsis thaliana* 12S seed storage protein, whereas the 18 kDa protein band was given a provisional identification of Actin 1. There was a low confidence in these results; therefore, the 34 kDa protein was also processed by liquid chromatography-mass spectrometry. Unfortunately, this analysis resulted in many known contaminants; however, four potential protein identifications were made, but also had a low confidence. The putative proteins identified were glutathione peroxidase (29 kDa), actin (25 kDa), beta-1,4-endoglucanase 2 (39 kDa) and ubiquitin extension protein 2 (9 kDa).

5.4 Discussion

The molecular weight of proteins from the cysts and eggs of *H. schachtii*, *G. pallida* and *G. rostochiensis* were determined using an antibody raised against vitellogenin. When cyst homogenates were incubated in the anti-VIT A antibody, mean molecular weights from biological replicates gave six proteins from *H. schachtii*, six proteins from *G. pallida* and five proteins from *G. rostochiensis* cyst samples. The calculated molecular weights of the *H. schachtii* proteins were 146 kDa, 95 kDa, 89 kDa, 47 kDa, 34 kDa and 19 kDa. Furthermore, the molecular weights of the *G. pallida* cyst proteins were 157 kDa, 107 kDa, 81 kDa, 55 kDa, 52 kDa and 46 kDa. Finally, the five proteins detected in *G. rostochiensis* had molecular weights of 110 kDa, 82 kDa, 77 kDa, 47 kDa and 13 kDa. Previous studies have also determined the molecular weight of VIT proteins in various other species, with a range of molecular weights of the studied polypeptides. The cyst nematode *H. glycines* was shown to have two major egg proteins, with molecular weights of 190 kDa and 180 kDa (Masler, 1999). These proteins have a comparable molecular weight to the larger protein recognised in *H. schachtii* (197 kDa). As shown in Chapter 3, vitellogenin genes were highly conserved between *H. glycines* and *H. schachtii* but may have further variation in the protein to account for the slight difference in molecular weight. Also, *C. elegans* contained four vitellogenin polypeptides, two of 170 kDa, 115 kDa and 88 kDa (Perez and Lehner, 2019). Some of the proteins from *G. pallida* (107 kDa) and *G. rostochiensis* (110 kDa and 82 kDa) correlate approximately to the smaller polypeptides from *C. elegans*.

Multiple proteins were recognised by anti-VIT A in each sample by western blot analysis. These proteins may be multiple polypeptides of vitellogenin, which has been demonstrated in other species. As described previously, *C. elegans* contains four polypeptides (Perez and Lehner, 2019). A study describing *Dolichorhabditis* sp. obtained three polypeptides from eggs, with estimated molecular weights of 175.2 kDa, 107 kDa and 83 kDa (Winter, 1992). Similarly, *Oscheius tipulae* was shown to contain three vitellogenin polypeptides (Almenara *et al.*, 2013).

The antibody raised to target vitellogenin recognised multiple proteins by western blotting, these proteins were from *H. schachtii* of cyst, egg and J2 origin. It was originally hypothesised that the antibody would recognise the protein from within the egg samples and therefore would also be present within the cysts. This is supported by the understanding that *C. elegans* VIT proteins are taken into the developing oocyte by receptor mediated endocytosis (Perez and Lehner, 2019) and are still present within the egg prior to nematode hatch (Kimble and Sharrock, 1983). Evidence from the work presented here demonstrated that the anti-VIT antibody detected proteins from cysts, eggs and J2 of *H. schachtii*, when assessing all protein bands visible from all blots. Within the cyst samples there were five proteins detected that were not present in either the eggs or J2 samples and, therefore, may have been associated with the female. There were a further four proteins that were shared between the cysts and eggs, which may have been localised to the egg only. Next, three proteins appeared in the cysts, eggs and J2, with a further protein shared between the J2 and egg; it is possible that these proteins are of J2 origin. This is contrary to the findings of Masler (2002), whose antibodies raised against major egg proteins in *H. glycines* did not detect proteins in J2 samples. It is possible that the method used to obtain the antibodies by Masler (2002) provided a higher specificity than the methodology used here, leading to fewer proteins detected. By contrast, in a study of the proteome of *Meloidogyne hapla* VIT proteins were discovered by mass spectrometry in J2 (Mbeunkui *et al.*, 2010). Interestingly, these authors postulate that due to the lack of oogenesis at this stage, the proteins are instead being stored in the J2 intestine. Two proteins of comparable molecular weight were also present in both cysts and eggs of *G. pallida*. Additionally, there were six proteins in the *G. rostochiensis* samples that were present in both cysts and eggs, with a further six proteins detected in the cysts, which again may be specific to the females. It is also possible that the same protein is present in all life stages tested and is not specific to the female, egg or J2.

In dot blot assays that were used to optimise the sample and antibody concentrations for future western blot analysis, the calculated dot intensity of *G. pallida* samples was greater than in *H. schachtii* samples; however, cyst homogenates from both samples were within a detectable range. This is likely due to the original sequence used for synthetic peptide design, which was taken from the *G. pallida* gene sequence. The aligned *H. schachtii* gene sequence contained

three different amino acids, which may have caused the variation in specificity. Higher dot intensities were measured when using the synthetic peptide, which is likely due to the size and conformation of the peptide sequence being less complex than the proteins from cyst nematode samples, even under denaturing conditions. It is also possible that the populations used in this study have a varied amino acid composition compared to the reference amino acid sequences used as the basis for synthetic peptide design. Furthermore, due to the nature of polyclonal antibody production, antibodies may have been raised to multiple epitopes from the synthesised peptide, reducing antibody specificity. The multiple epitopes may have given rise to antibody targets of different sizes, as detected in western blotting, due to antibodies only recognising approximately five to seventeen amino acids (Flaherty, 2012).

In many species, vitellogenin has been described as a precursor protein (Gerber-Huber *et al.*, 1987). Therefore, the smaller proteins recognised by the anti-VIT antibodies may be the final yolk protein products of vitellogenin. The variation in molecular weight of the proteins may also be due to cleavage of the proteins or post-translational modifications. Vitellogenin is a large lipo-glyco-phosphoprotein (Perez and Lehner, 2019), therefore variations in the molecular weight of the proteins detected may be a result of the degree of post-translational modifications present within the protein. Proteins recognised in western blotting may have been a result of protein degradation or non-specific binding. This may have occurred due to a limitation of polyclonal antibody production, leading to increased cross-reactivity of these antibodies which can cause a reduction in specificity. Alternatively, the production of a monoclonal antibody against vitellogenin may have given better specificity to the target antigen. To better understand the specificity of the polyclonal antibodies additional positive controls could have been implemented. An additional positive control could be derived from protein purified from the use of expression vectors. The full target *vit* gene sequence selected for synthetic peptide design could be cloned and then inserted into an expression vector to produce the target protein. Once purified, the protein sample could then be used as a control in further immunoblots. This would have enabled simpler optimisation of the western blot protocol, to determine antibody concentrations, and to assess antibody specificity; as this control would have been more representative of the target protein. Furthermore, the size and conformation of the synthetic peptide positive control used likely had increased affinity of the antibody, greater than the antibody interaction with nematode samples. The use of a positive control created by cloning and insertion into an expression vector may have better represented the complexity of a biological sample.

Antibodies raised against synthetic peptide A were successful in the recognition of dilutions of the peptide; however, antibodies derived from synthetic peptide B were less successful. It is possible that synthetic peptide B did not elicit a strong immune response in the generation of

antibodies, therefore resulting in the lack of protein recognition in both dot and western blot assays. This work could be expanded to develop synthetic peptides based on other *vit* gene sequences; the anti-VIT antibodies presented are raised to peptides designed from one cluster of closely related *vit* gene sequences from *H. schachtii*, *G. pallida* and *G. rostochiensis*.

The use of SDS-PAGE with silver staining demonstrated differences between proteins within the cysts of *G. pallida* and *H. schachtii*. There was some variation in the number and positioning of protein bands when comparing these species, indicating that there were differences in protein composition. These differences were again highlighted by the variation in the molecular weight of proteins recognised by the anti-VIT antibody; however, this may have been reduced if the protein concentration was sufficient in all samples to standardise the concentration across all samples after using a BCA assay. Unfortunately, this was limited by the protein concentration available within the original samples. The optimisation of SDS-PAGE and Coomassie staining allowed proteins bands to be isolated and processed for protein identification. Unfortunately, the resulting protein identifications did not detect vitellogenin. The initial protein identification was taken from MALDI-TOF and compared to a UniProt database (Bateman *et al.*, 2023). A limitation of this method was that there were limited proteins deposited on this database; only 66 protein sequences were available for *H. schachtii*. Furthermore, the amount of protein present within the extracted band may not have been at a detectable level for MALDI-TOF analysis. The 34 kDa protein from *H. schachtii* was also processed by a more sensitive technique, LCMS; however, again it did not retrieve vitellogenin as an identified protein. It is possible that the protein band recognised at this molecular weight is not vitellogenin or its products and is instead a result of non-specific binding.

This work has demonstrated that the antibody synthesised to recognise the gene sequence of vitellogenin has recognised proteins within the cysts and eggs of PPN species: *G. pallida*, *G. rostochiensis* and *H. schachtii*. This work has provided evidence that *vit* genes are present within the genomes of these species and are likely translated into proteins, at various life stages. The use of this antibody has detected proteins within the cyst samples with molecular weights similar to the theoretical value for vitellogenin and equate to the molecular weight of vitellogenin in other nematode species; however, the identity of the antibody target has not yet been confirmed.

Western blot analysis has now been able to show that proteins are present in different cyst nematode life stages, but no current study has used immunolocalisation techniques to visualise the tissue location of the antibody target in these species. Chapter 6 examines if the antibody developed to target vitellogenin could be used in immunofluorescent study of *H.*

schachtii, *G. pallida* and *G. rostochiensis* for the localisation of proteins in cysts, J2 and eggs. Chapter 6 also explores the use of other antibodies raised against proteins present within the eggs of cyst nematodes.

6 Test of Antibodies with Specificity to Cyst Nematode Egg Proteins

6.1 Introduction

The vitellogenin protein may be present within the eggs of cyst nematodes; however, it is not the only protein to be found within the egg or egg protein that has been the interest of proteomic study using immunoblots. In previous studies, antibodies have been raised against whole eggs or cyst homogenates, and it has been predicted that the protein target is within the perivitelline fluid (Evans *et al.*, 1995). The antibody target may be vitellogenin, as an abundant protein within the perivitelline fluid; however, the target may be other proteins also present.

The egg proteins of cyst nematodes have previously been investigated using proteomic techniques, including differentiating the proteins by size, isoelectric focusing or producing antibodies against the egg or whole cyst sample. Some of these antibodies were initially developed for the differentiation of cyst nematode species but were shown to target egg proteins. This work was mostly directed towards PCN species, *Globodera rostochiensis* and *G. pallida* (Schots *et al.*, 1987; Schots *et al.*, 1990; Schots *et al.*, 1992); with some focus on other cyst nematodes such as *Heterodera schachtii* (Webster and Hooper, 1968). Early work comparing populations of PCN established that it was composed of two species, *G. rostochiensis* and *G. pallida*, instead of one species *Globodera rostochiensis* (Schots *et al.*, 1987). To differentiate PCN species, isoelectric focusing from single cysts was performed by Fleming and Marks (1983). This resulted in differing pI values; 5.9 for *G. rostochiensis* and 5.7 for *G. pallida*. These proteins were subsequently discovered to be of the same molecular weight, 34 kDa (see Robinson *et al.*, 1993 below).

In the study by Robinson *et al.* (1993), monoclonal antibodies were raised that recognised the 34 kDa proteins with differential pI values as recognised by Fleming and Marks (1983), in both PCN species. Additionally, polyclonal antibodies were raised against *G. rostochiensis* and *G. pallida* homogenates of eggs and J2. The monoclonal antibodies were labelled MAC 356 and MAC 357, and polyclonal antibodies were labelled PC 266 (against *G. pallida*) and PC 267 (against *G. rostochiensis*). MAC 356 showed strong reactivity with *G. pallida*, whereas MAC 357 gave a stronger response with *G. rostochiensis*; however, both antibodies showed weak cross-reactivity with both PCN species and *G. tabacum*. The authors hypothesised that these proteins could have a similar role in both species and may be found in a similar location and concentrations. Furthermore, the polyclonal antibodies PC 266 and PC 267 were strongly

cross-reactive with homogenates of both PCN species and showed some cross-reactivity with other PPN (including *H. schachtii*) and free-living nematodes tested.

Further work by Evans *et al.* (1995), using MAC 356 and MAC 357 discovered that these antibodies only recognised viable eggs and could be used to identify PCN species in mixed populations from soil extracts. This study hypothesised that the antibody was interacting with a protein within the perivitelline fluid, due to a large proportion of the antigen being detected in the supernatant when whole cysts or eggs were homogenised. It was also shown that little antigenicity came from the free J2, eggshells or cyst wall. In a subsequent study using MAC 356, the antigen was readily detected in solution once the eggs had been ruptured (Curtis *et al.*, 1998).

The 34 kDa protein was further studied by Curtis *et al.*, (1996), by raising both monoclonal and polyclonal antibodies against the protein; the polyclonal antibody was named PC320. This antibody only recognised live eggs, which agrees with the work of Robinson *et al.* (1993), who raised antibodies (MAC 356) against the same protein and was also able to show that the antibody only recognised live eggs. The 34 kDa protein was found in all pathotypes of *G. pallida* and *G. rostochiensis* (Curtis *et al.*, 1996). Immunolocalisation techniques were used to determine the location of antibody interaction. The monoclonal antibody recognised the protein within the amphids of the nematode. The amphids are located on either side of the nematode's mouth and are a chemosensory organ (Perry, 1996). Curtis *et al.*, (1996) suggested that the 34 kDa protein was not found in *Heterodera* species; however, the polyclonal antibody, PC320, raised against this protein was cross-reactive with *Heterodera* species tested.

Further work was conducted to localise the target of the PC320 antibody (Fioretti *et al.*, 2001) The PC320 antibody was shown to recognise both secretory and excretory products of *G. pallida*. In immunolocalisation studies, the protein recognised was localised to the amphids, genital primordium and muscles surrounding the metacarpus pump chamber. In the amphids the protein was specifically localised to a region in the amphidal canal and pores. The protein was shown to be present at all stages of the *G. pallida* life cycle but had the highest concentration in eggs and was not present in potato root diffusate. Western blot analysis using PC320 resulted in bands at 35, 30 and 23 kDa. A cDNA expression library from mixed stages (J2 and gravid females) of *G. pallida* was screened with the PC320 to determine the target protein identity. The positive clone that was amplified by PCR gave a putative identity of annexin, when searched using the EMBL database. Annexins are a family of proteins that bind phospholipids when calcium levels are high (Gerke *et al.*, 2024). The genes encoding for an annexin protein have been identified in *Caenorhabditis elegans* as *nex-1*, *nex-2* and *nex-3*. The protein encoded by *nex-1* encodes a 35 kDa protein and has been localised to the pharynx

(secretory gland cells and cuticle formation sites of the grinder), yolk granules within the oocytes, uterine wall, vulva and membranes of the spermathecal valve (Creutz *et al.*, 1996).

Further to the identification of annexin in *G. pallida*, annexin has recently been localised to the eggshell of *G. rostochiensis* (Price *et al.*, 2023). Antibodies raised to the identified annexin bound to the eggshell and recognised a recombinantly produced GROS_g03104, which was believed to encode annexin in *G. rostochiensis*. The antibody had no antigenicity to J2 samples tested. In western blot analysis, the antibodies recognised the annexin protein between 35-40 kDa.

In this chapter, the aim of the work was to confirm the identification of the protein identified by MAC 356 and PC320. Also, to determine whether these antibodies were recognising the same target protein, by western blot analysis and initial immunolocalisation studies. Furthermore, the aim of this work was also to expand previous studies to determine if these proteins were also present within populations of *H. schachtii*. Finally, initial immunolocalisation studies to determine the target proteins of the anti-vitellogenin antibody were conducted.

6.2 Materials and Methods

6.2.1 Immunoblotting of MAC 356 and PC320

Antibodies have previously been raised to proteins in cyst nematodes of egg origin. A monoclonal antibody, MAC 356, was raised against *G. pallida* to target a 34 kDa unidentified protein within the nematode egg (Robinson *et al.*, 1993). Subsequently available for study was a monospecific polyclonal antibody, PC320, with specificity targeted towards the same 34 kDa protein (Curtis *et al.*, 1996).

Immunoblot assays were used to determine whether either MAC 356 or PC320 recognised protein homogenates of cyst nematode samples from *H. schachtii*, *G. pallida* or *G. rostochiensis*. Dot blot assays were initially used to confirm whether secondary antibodies would recognise the primary antibodies tested and to determine whether the antibodies would successfully recognise any protein in the cyst samples. Dot blot assays were conducted as described in Section 5.2.5. A serial dilution of the MAC356 antibody was inoculated onto a PVDF membrane from 1:5 to 1:800 and incubated in anti-rat antibody at 1:500 or 1:1000. Extracted protein from cyst samples of 5, 10 or 20 *G. pallida* cysts were then inoculated onto a PVDF membrane and incubated in MAC 356 at 1:101 and then in anti-rat secondary antibodies at 1:500. Subsequently, a serial dilution of PC320 from 1:100 to 1:2000 was inoculated onto a nitrocellulose membrane and then incubated in either 1:250, 1:500 or 1:1000 concentration of anti-rabbit antibodies.

Next, western blot analysis was performed as described in Section 5.2.6, to determine whether the MAC356 and PC320 antibodies also detected protein at approximately 34 kDa, within the cyst nematode samples. In western blot analysis using MAC 356, the concentration of the primary antibody used was 1:500, whereas the anti-rat secondary antibody concentration was 1:250. In western blot analysis using PC320, protein samples were adjusted to 250 µL, incubated in PC320 antibodies at 1:500 and anti-rabbit secondary antibodies at 1:250. Western blot membranes were analysed as described in Section 5.2.7.

6.2.2 Protein Identification

Protein sequencing was performed as described in Section 5.2.8. Peptide sequences obtained by Tandem and MALDI-TOF mass spectrometry were subject to BLAST searches using tBLASTn against the genomes of *H. schachtii*, *G. pallida* and *G. rostochiensis*. BLAST searches were completed using default settings, with a maximum of 100 target sequences and an expect threshold of 10. Furthermore, to establish the protein identification of the 34 kDa band recognised by the MAC 356 and PC320 antibodies, a database of *G. pallida* and *G. rostochiensis* proteins was assembled. This database contained protein sequences with molecular weights between 30 and 35 kDa, that were available from UniProt (Bateman *et al.*, 2023). BLAST+ (Camacho *et al.*, 2009), a command line application, was then used to perform BLASTp searches against this database, using the peptide sequences as the search query with an expected threshold of 100. The 34 kDa protein was again analysed for protein identification, but instead by liquid-chromatography mass spectrometry, as described in Section 5.2.8.

6.2.3 Preparation of Biological Material for Immunofluorescence

Hatched *H. schachtii* second-stage juveniles (J2) were collected as described in Chapter 2, reduced to 100 µL total volume and placed at 90 °C for 1 min to be killed. The J2 were then placed on a glass slide and cut in half with a blade and collected in a tube. Samples were washed twice in M9 buffer (22 mM KH₂PO₄, 42 mM Na₂HPO₄, 20.5 mM NaCl and 1 mM MgSO₄) and then reconstituted in 100 µL Accustain (Sigma-Aldrich), a formalin-free fixative.

Cyst samples were collected as described in Chapter 2. Singular cysts of *H. schachtii*, *G. pallida* or *G. rostochiensis* were placed in 1.5 mL Eppendorf tubes and washed twice in M9 buffer before being reconstituted in 100 µL of Accustain. This was performed in triplicate. Similarly, egg samples from 30 cysts were collected as described in Chapter 2, before being washed twice in M9 buffer and reconstituted in 100 µL of Accustain. Once placed in the fixative, all samples were stored at -80°C for a minimum of 16 h and until use.

6.2.4 Immunofluorescence of Second-Stage Juveniles

Immunofluorescence was used to immunolocalise proteins recognised by the Vit-A and PC320 antibodies in *H. schachtii* J2 samples. These samples were removed from -80 °C and defrosted at room temperature for 50 min. The tubes were inverted five times at 30 and 45 min. The tubes were then centrifuged at 500 g for 2 min and washed twice for 10 min in 100 µL phosphate buffered saline with 2% Tween20 (v/v) (2% PBST). Between washes samples were placed in the centrifuge at 500 g for 2 min. After the second wash, samples were mixed and inoculated onto 8-well poly-L-lysine coated slides, 10 µL per well, and allowed to dry.

Samples were then blocked in 1% bovine serum albumin (BSA) in 2% PBST for 1 h at room temperature with agitation. Once blocked, 10 µL of the primary antibody at 1:250 or 1:500 concentration, diluted in PBS was added to the appropriate well and incubated overnight at 4°C with agitation on a platform shaker. After incubation, the slides were kept at room temperature for 1 h. Slides were washed by dipping the slides in two separate beakers of 2% PBST for 10 s each, patted with towel and dried for 10 min. This step was repeated twice. Samples were then incubated for 1 h at room temperature in 10 µL of goat anti-rabbit Fluorescein isothiocyanate (FITC) conjugated secondary antibody at 1:500, diluted in PBS. Slides were again washed twice for 10 s in 2% PBST and dried for 10 min. Samples were then mounted using 10 µL of Vectashield antifade mounting medium with DAPI, whilst being protected from light.

6.2.5 Immunofluorescence of Cysts and Eggs

Immunofluorescence of cyst and egg samples of *H. schachtii*, *G. pallida* and *G. rostochiensis* was performed using Vit-A, PC320 and MAC 356 antibodies or Vit-A and PC320 pre-immune serum, using a modified method of Sperling and Eves-van den Akker (2023). The MAC 356 antibody was not tested on *H. schachtii* samples as this has previously been shown to give no result in dot and western blot experiments.

Cyst and egg samples were removed from -80 °C and defrosted at room temperature for 50 min, and tubes were inverted five times at 30 and 45 min. The samples were then centrifuged at 500 g for 2 min and washed twice for 10 min in 2% PBST, with centrifugation at 500 g for 2 min between wash steps. The samples were then partially dissolved in 1 mL of 1% sodium hypochlorite (v/v) in deionised water for 2 min, then centrifuged at 800 g for 30 s. Samples were washed twice for 2 min in 100 µL of 2% PBST, with centrifugation at 500 g for 2 min between wash steps. The samples were then digested using 50 µL of 10 mg/mL collagenase in 50 mM TES Buffer (10 mM Tris (pH 7.5), 10 mM ethylenediaminetetraacetic acid (EDTA; pH 8.0) and 5% sodium dodecyl sulphate (SDS)), supplemented with 36 mM CaCl₂, incubating for 60 min at 37 °C. Tubes were inverted five times after 15 min of

incubation. Samples were further digested by adding 50 μ L of 10 mg/mL proteinase K in M9 buffer, incubating at 37 °C for 60 min. Again, the tubes were inverted five times after 15 min. Next, samples were washed with M9 buffer and centrifuged at 500 *g* for 3 min. Buffer was removed and samples were frozen overnight at -80 °C.

The next day, samples were defrosted and 1 mL of -20 °C 100% methanol was added to each tube and incubated at -80 °C for 30 s. Samples were centrifuged at 13,000 *g* for 30 s and the methanol was removed. Next, 1 mL of -20 °C 100% acetone was added, and samples were incubated at -80 °C for 1 min before centrifugation at 13,000 *g* for 30 s. All but approximately 100 μ L of acetone was removed from the tubes and then 25 μ L of deionised water was added every 10 min until the volume reached approximately 200 μ L. To each tube, 100 μ L of 2% PBST was added and incubated for 5 min at room temperature. Samples were then centrifuged for 30 s at 4,000 *g* and washed twice with 2% PBST, with centrifugation at 500 *g* for 3 min between wash steps. Next, samples were incubated in 1 mL of 2 mg/mL glycine in deionised water for 15 min on ice. After incubation, samples were washed once with 2% PBST and once in 0.1% PBST, again centrifuging at 500 *g* for 3 min between washes. Samples were blocked using 1% BSA in 2% PBST for 1 h at room temperature with agitation. Egg samples were then aliquoted into separate tubes to be incubated in different primary antibodies. Blocking buffer was removed using a pipette and the samples were incubated in 200 μ L of the primary antibody, overnight at 4 °C, with agitation. All primary antibodies and pre-immune serums were applied at 1:500 concentration, diluted in PBS.

After overnight incubation, samples were moved to room temperature for 1 h, then washed twice for 10 min in 2% PBST, centrifuging at 500 *g* for 3 min between washes. Samples were then incubated in either goat anti-rabbit FITC conjugated or goat anti-rat FITC conjugated secondary antibodies at 1:500 concentration, for 1h, at room temperature. Following incubation, samples were washed with 2% PBST, centrifuged at 500 *g* for 3 min and washed with 0.1% PBST. Egg samples were then placed on poly-L-lysine coated eight well slides and the cyst samples were picked from the tubes and placed in the centre of standard poly-L-lysine coated slides. Samples were protected from light whilst 10 μ L of Vectashield antifade mounting medium with DAPI was added to cover the sample. Slides containing cyst samples were covered with a coverslip and sealed with clear nail varnish.

6.2.6 Visualisation of Samples

All samples were protected from light prior to visualisation. To visualise samples, Invitrogen™ EVOS™ M7000 Imaging System was used. Images were captured using the 'brightfield', 'DAPI' and 'GFP' filters. The 'GFP' filter was selected due to the LED light cube within the microscope being suitable for fluorophores with excitation and emission peaks of

approximately 482 nm and 524 nm, respectively. FITC is reported to have an excitation peak of 498 nm and emission peak of 525 nm, therefore could be visualised using the 'GFP' filter.

6.3 Results

6.3.1 Dot Blot Assays

Dot blot assays were used to confirm that the anti-rabbit secondary antibody recognised PC320 when inoculated directly onto a membrane, at several concentrations. The interaction between these two antibodies was successful and produced visual dots on the membrane. This process was also completed for MAC 356, where the anti-rat secondary antibody recognised the monoclonal antibody at all dilutions tested. Furthermore, a dot blot assay was used successfully to recognise different concentrations of protein extracts from *G. pallida* cyst homogenates to give dots on the membranes at all dilutions tested of PC320 and MAC 356.

6.3.2 Western Blot Analysis

Western blot analysis using the monoclonal antibody MAC 356 resulted in protein bands recognised across seven technical replicates, with molecular weights of 33-35 kDa in *G. pallida* cyst samples (Figure 6.1). The antibody MAC 356 also recognised a protein band at 34 kDa in *G. rostochiensis* cyst samples (Figure 6.2) but had no specificity towards *H. schachtii*. Figure 6.2 also demonstrated that the MAC 356 antibody recognised a 34 kDa protein in *G. pallida* cyst homogenates. Protein extracts from eggs of *G. pallida* and *H. schachtii* were also incubated in MAC 356. This resulted in a protein band in the same position as the cyst samples in *G. pallida* but no recognition of the *H. schachtii* eggs.

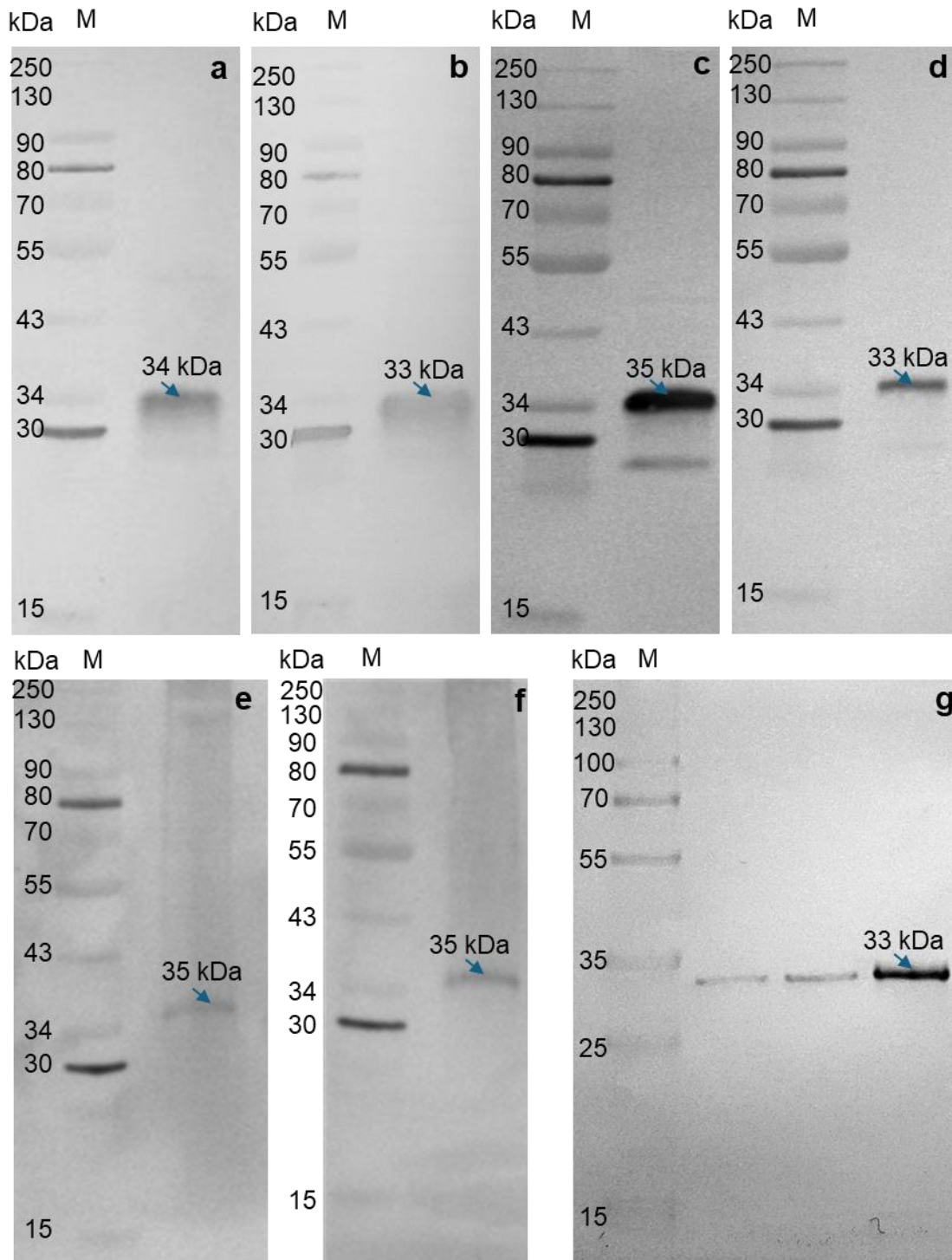


Figure 6.1: Photographs of seven western blot membranes using the MAC 356 monoclonal antibody (a-g). In each image, a protein band has been recognised in the *Globodera pallida* cyst homogenate samples. M: Molecular weight marker. Measured size of band is indicated by the blue arrow and the molecular weight value above it. In (g), three samples of increasing concentration of protein (protein extract from 1, 10 or 20 homogenised *G. pallida* cysts) were used and gave bands in the same position, of the same molecular weight (33 kDa).

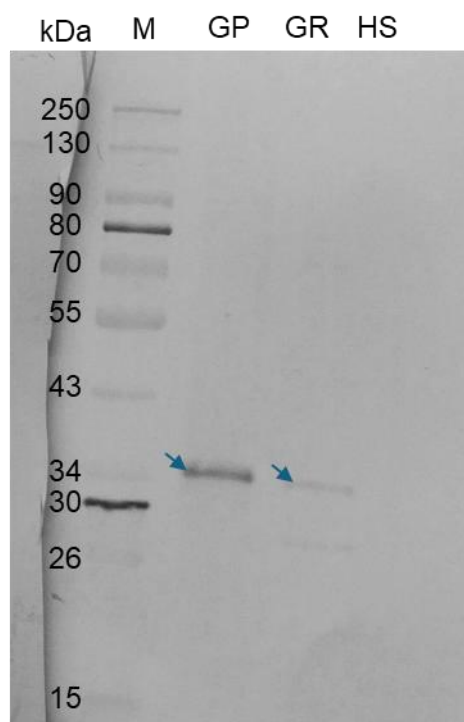


Figure 6.2: Photograph of western blot membrane incubated in MAC 356 monoclonal antibody against *Globodera pallida* (GP), *G. rostochiensis* (GR) and *Heterodera schachtii* (HS) cyst homogenate samples. M: Molecular weight ladder. Blue arrows indicate measured bands, each with a molecular weight of 34 kDa.

Once available, western blot analysis using PC320 showed the successful recognition of proteins in *G. pallida* and *H. schachtii* cyst samples (Figure 6.3). In a subsequent western blot, protein bands were also recognised in *G. rostochiensis* cyst homogenate samples (Figure 6.4).

In Figure 6.3, protein bands were recognised in *G. pallida* cyst homogenates at 293 kDa, 126 kDa, 108 kDa, 90 kDa, 78 kDa, 48 kDa, 35 kDa, 30 kDa, 27 kDa, 23 kDa and 19 kDa. The PC320 antibody recognised protein bands in *H. schachtii* cyst samples at 290 kDa, 123 kDa, 102 kDa, 92 kDa, 63 kDa, 50 kDa, 32 kDa, 22 kDa and 20 kDa. Six of the proteins in the *G. pallida* and *H. schachtii* samples had approximately the same molecular weight; therefore, the antibody may have detected the same protein in both samples.

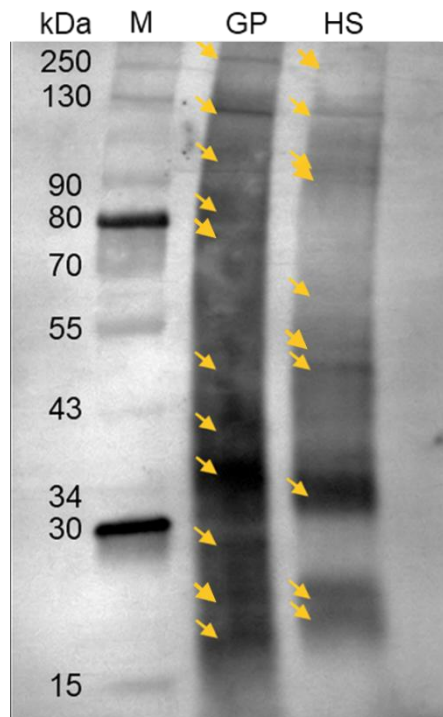


Figure 6.3: Photograph of western blot analysis using PC320 to recognise proteins in cyst homogenates of *Globodera pallida* (GP) and *Heterodera schachtii* (HS). M: Molecular weight marker. Blue arrow indicates measured band.

Figure 6.4 presents a comparison between *G. pallida*, *G. rostochiensis* and *H. schachtii* cyst homogenates incubated in PC320. Protein bands were recognised in the *G. pallida* sample at approximately 152 kDa, 97 kDa, 95 kDa, 89 kDa, 39 kDa, 35 kDa and 21 kDa, whereas in *G. rostochiensis* protein bands were measured at approximately 156 kDa, 81 kDa, 39 kDa, 36 kDa, 28 kDa and 20 kDa. Finally, the protein bands recognised in *H. schachtii* were approximately 141 kDa, 91 kDa, 53 kDa, 40 kDa, 36 kDa and 27 kDa. Four protein bands of similar molecular weight were shared between *G. pallida* and *G. rostochiensis*. Two protein bands of estimated molecular weights of 39/40 kDa and 35/36 kDa were shared between all three species.

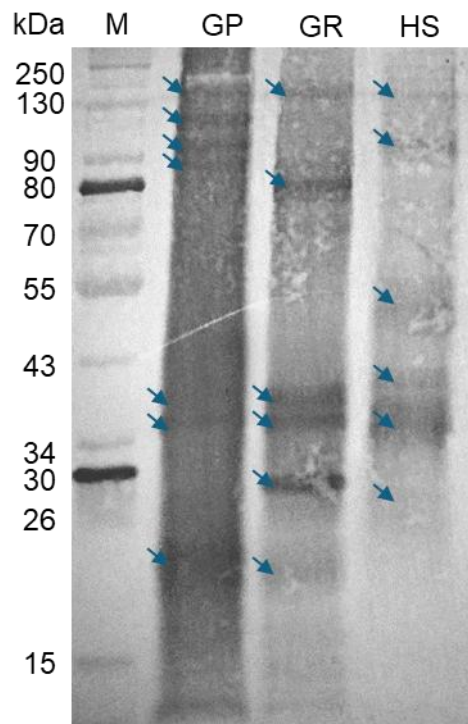


Figure 6.4: Western blot analysis of *Globodera pallida* (GP), *G. rostochiensis* (GR) and *Heterodera schachtii* (HS) incubated in monospecific polyclonal antibody, PC320. M: Molecular weight marker. Measured bands indicated by blue arrows.

6.3.3 Silver Stain of *G. pallida* Cysts

A silver stain of *G. pallida* cyst protein homogenates highlighted the presence of several protein bands with molecular weights between 25 kDa and 35 kDa (Figure 6.5). Proteins detected in this range may be the protein recognised by either MAC 356 or PC320 antibodies. Five proteins were visible in this range, with molecular weights of 27 kDa, 29 kDa, 31 kDa, 32 kDa and 35 kDa.

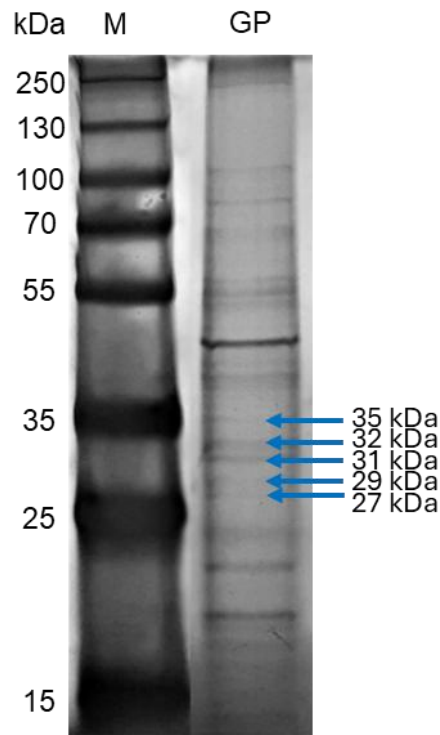


Figure 6.5: Silver stain image of cyst homogenates from *Globodera pallida* (GP). M: Molecular weight marker. Bands measured with molecular weights between 25 kDa and 35 kDa are indicated with blue arrows and molecular weight value.

6.3.4 Protein Identification

Identification of protein bands at approximately 34 kDa by MALDI-TOF, performed in triplicate, resulted in three putative protein identifications (Table 6.1). The first band resulted in a peptide identified that corresponded to the 14_3_3 domain-containing protein. The second band also gave a single peptide identified with a putative protein identification of Triosephosphate isomerase. Finally, the third band processed resulted in three peptides identified, all corresponding to the GST N-terminal domain-containing protein.

Table 6.1: Peptide sequences identified by MALDI-TOF spectrometry and corresponding protein identified through mascot searches with UniProt accession number.

Identified Protein	Peptides Identified
14_3_3 domain-containing protein (A0A183C635)	1. KSQQSYQEAFDIKD
Triosephosphate isomerase (A0A183C1W7)	1. KDVGATHVIIGHSEER
GST N-terminal domain-containing protein (A0A183CB11)	1. KLYYFDARG 2. RLLHYVGQQFEDVRF 3. KQLGQSTTILRF

Peptides from MALDI-TOF analysis (Table 6.1) were used as query sequences to search a curated database of *G. pallida* and *G. rostochiensis* proteins between 30 and 35 kDa, derived from UniProt databases (Table 6.2). The peptide with sequence homology to 14_3_3 domain

containing protein gave the same protein in the restricted database; however, in *G. rostochiensis* an F-box domain containing protein also had sequence similarity. The peptide query corresponding to triosephosphate isomerase obtained from the MASCOT search, also gave this protein identification when searching the restricted database for *G. pallida*. This protein was not identified in the database of *G. rostochiensis* proteins, when the E-value was less than ten. The MASCOT search of the third protein band processed gave three resulting peptide sequences with a protein identification of GST N-terminal domain-containing protein. These peptide sequences were then used to search the restricted molecular weight databases. In *G. pallida*, the first peptide sequence had sequence homology to triosephosphate isomerase and an unidentified protein, whereas when using the *G. rostochiensis* database, the protein glutathione S-transferase was identified. The database search of the second peptide, resulted in a protein identification of glutathione transferase, GST N-terminal domain-containing protein, an uncharacterised protein and SAC domain-containing protein. In *G. rostochiensis*, the second peptide was predicted to be glutathione S-transferase and an uncharacterised protein. Finally, the third peptide had a protein identification of glutathione transferase in both *G. pallida* and *G. rostochiensis*; however, an additional identification was made in *G. rostochiensis* of an uncharacterised protein.

Table 6.2: Results of UniProt database search of *Globodera pallida* and *G. rostochiensis* proteins, restricted to molecular weights of 30-35 kDa. Sequences with significant alignments and associated UniProt accession numbers are recorded for alignments with an E-value less than 10.

Peptide query	Species	UniProt accession number	Sequences producing significant alignments	E-value
14_3_3 domain-containing protein	<i>G. pallida</i>	A0A183C8Y5	14_3_3 domain-containing protein	0.00
		A0A183CQZ3	14_3_3 domain-containing protein	0.00
		A0A183C635	14_3_3 domain-containing protein	0.00
		A0A183BRP9	14_3_3 domain-containing protein	0.03
	<i>G. rostochiensis</i>	A0A914IEM9	14_3_3 domain-containing protein	0.00
		A0A914HY31	14_3_3 domain-containing protein	0.00
		A0A914HE10	14_3_3 domain-containing protein	0.02
		A0A914HF67	14_3_3 domain-containing protein	0.03
		A0A914HF77	14_3_3 domain-containing protein	0.24

Peptide query	Species	UniProt accession number	Sequences producing significant alignments	E-value
		A0A914HEW4	F-box domain-containing protein	6.20
Triosephosphate isomerase	<i>G. pallida</i>	A0A183C1W7	Triosephosphate isomerase	0.00
	<i>G. rostochiensis</i>	N/A		>10
GST N-terminal domain-containing protein, peptide 1	<i>G. pallida</i>	A0A183C1W7	Triosephosphate isomerase	0.00
		A0A183C189	Uncharacterised protein	4.30
	<i>G. rostochiensis</i>	A0A914HJM7	Glutathione S-transferase	0.73
		A0A914HKB9	Glutathione S-transferase	1.10
		A0A914I4S8	Glutathione S-transferase	8.80
GST N-terminal domain-containing protein, peptide 2	<i>G. pallida</i>	A0A183BP69	Glutathione transferase	0.00
		A0A183BI77	Glutathione transferase	0.00
		A0A183CGH9	Glutathione transferase	0.00
		A0A183BLP9	Glutathione transferase	0.04
		A0A183BZH6	Glutathione transferase	0.15
		A0A183CIB2	GST N-terminal domain-containing protein	0.31
		A0A183CPE3	Uncharacterised protein	7.40
		A0A183C0F4	SAC domain-containing protein	10.00
	<i>G. rostochiensis</i>	A0A914HKB9	Glutathione S-transferase	0.00
		A0A914HJM7	Glutathione S-transferase	0.00
		A0A914I4S8	Glutathione S-transferase	0.05
		A0A914GZU3	Glutathione S-transferase	0.07
		A0A914HV16	Glutathione S-transferase	0.19
		A0A914HAN0	Uncharacterised protein	7.70
		A0A914HNC8	Uncharacterised protein	9.10
GST N-terminal domain-	<i>G. pallida</i>	A0A183BI77	Glutathione transferase	0.01

Peptide query	Species	UniProt accession number	Sequences producing significant alignments	E-value
containing protein, peptide 3				
		A0A183CGH9	Glutathione transferase	0.01
		A0A183CGI0	Glutathione transferase	0.06
		A0A183BP69	Glutathione transferase	0.10
	<i>G. rostochiensis</i>	A0A914HKB9	Glutathione S-transferase	0.00
		A0A914HJM7	Glutathione S-transferase	0.00
		A0A914I4S8	Glutathione S-transferase	0.05
		A0A914GZU3	Glutathione S-transferase	0.07
		A0A914HV16	Glutathione S-transferase	0.19
		A0A914HAN0	Uncharacterised protein	7.70
		A0A914HNC8	Uncharacterised protein	9.10

A 34 kDa protein band from a Coomassie stained gel separated by SDS-PAGE was divided in two and identified using LCMS. This resulted in the protein identifications given in Table 6.3. Actin is a common contaminant of LCMS methodology (Hodge *et al.*, 2013) and appears in the identifications given in Table 6.3. Two proteins had molecular weights of approximately 35 kDa, annexin and Ig-like domain containing protein. The most significant match was given as a 166 kDa annexin protein.

Table 6.3: Protein identification of 34 kDa protein band from *Globodera rostochiensis* using liquid chromatography mass spectrometry. The UniProt accession number, description and mass (Da) of the identified protein are given. Significant peptide spectral matches and the number of sequences are shown, the higher value of significant matches indicated likelihood of protein identification. Identifications with the highest significant matches are highlighted in red.

UniProt Accession	Description	Mass (Da)	No. of significant matches	No. of significant sequences
Replicate 1				
A0A914HY21	A0A914HY21/ Annexin	166136	80	48
Q8IT74	Q8IT74/ Actin	42150	64	16
A0A914HJT9	A0A914HJT9/ Myosin heavy chain	227491	60	43
A0A914IC55	A0A914IC55/ Heat shock protein	71202	53	26

UniProt Accession	Description	Mass (Da)	No. of significant matches	No. of significant sequences
A0A914GWF3	A0A914GWF3/ Ig-like domain-containing protein	35821	52	20
A0A914H5T6	A0A914H5T6/ Annexin	35822	50	20
A0A914H6X0	A0A914H6X0/ Actin	42050	48	14
A0A914H8I8	A0A914H8I8/ Tr-type G domain-containing protein	183112	45	24
A0A914GW54	A0A914GW54/ Coenzyme Q-binding protein COQ10 START domain-containing protein	87463	44	24
A0A914HA34	A0A914HA34/ Galectin	47394	38	16
Replicate 2				
A0A914HY21	A0A914HY21/ Annexin	166136	87	44
A0A914H5T6	A0A914H5T6/ Annexin	35822	69	19
Q8IT74	Q8IT74/ Actin	42150	62	17
A0A914GW54	A0A914GW54/ Coenzyme Q-binding protein COQ10 START domain-containing protein	87463	61	26
A0A914HJT9	A0A914HJT9/ Myosin heavy chain	227491	56	44
A0A914H6X0	A0A914H6X0/ Actin	42050	48	15
A0A914GWF3	A0A914GWF3/ Ig-like domain-containing protein	35821	47	19
A0A914H8I8	A0A914H8I8/ Tr-type G domain-containing protein	183112	45	23
A0A023NCZ9	A0A023NCZ9/ Glutathione peroxidase (Fragment)	29397	37	11
A0A914IC55	A0A914IC55/ Heat shock protein 70	71202	35	20

6.3.5 Immunofluorescence

6.3.5.1 PC320

Immunofluorescence was used to determine the location of antibody interaction within cyst nematode samples including female cysts, eggs, J2 that were manually removed from eggs or hatched J2. Immunofluorescence techniques were performed using the antibodies previously investigated through western blot analysis in Chapter 5, the anti-vitellogenin (VIT) A antibodies, and Chapter 6, MAC 356 and PC320.

Female cysts of *H. schachtii*, *G. pallida* and *G. rostochiensis*, eggs of *H. schachtii* and J2 removed from the eggs of *G. rostochiensis* were incubated in PC320 antibodies (Figure 6.6-6.10). In Figures 6.8 and 6.10, eggshells appear to be highlighted by the PC320 antibody. Additionally, the anterior and posterior ends of the female cyst in Figure 6.6 are also highlighted. The posterior end of the J2 in Figure 6.9 (b) was shown to have a greater fluorescence signal than the nematode treated with the pre-immune serum (a). Figure 6.10 illustrates that the eggshell was recognised by the PC320 antibodies. The mean fluorescence intensity of each sample area was measured (Table 6.4). The measured fluorescence intensity of samples incubated in PC320 (b) was greater than the pre-immune serum (a) in Figure 6.8, 6.9 and 6.10; however, in Figure 6.6 and 6.7, the measured fluorescence was greater in pre-immune treated samples, possibly due to autofluorescence.

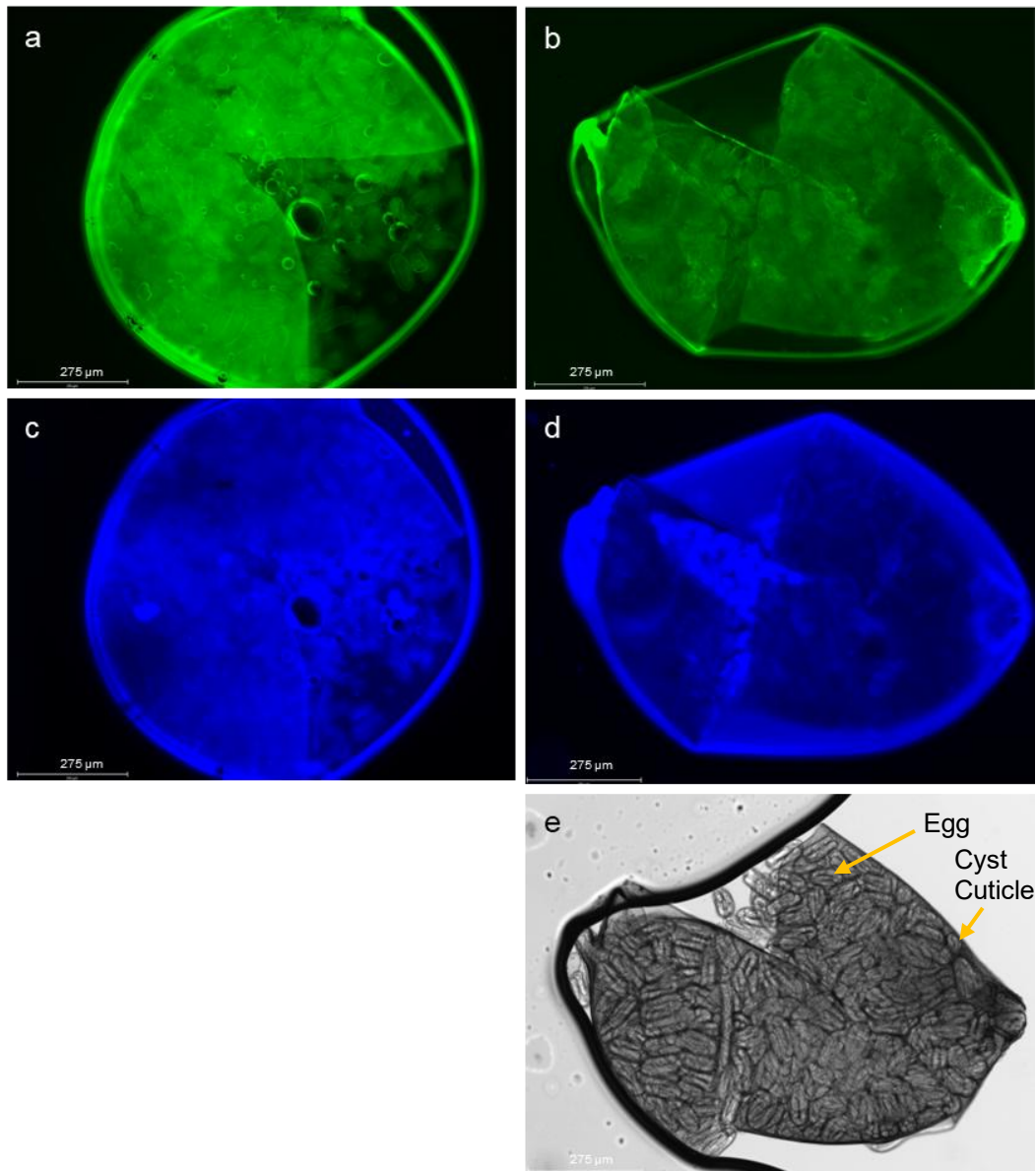


Figure 6.6: Immunofluorescent images of *Heterodera schachtii* female cysts, incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). A brightfield image of cyst shown in (b) and (d), is given in (e).

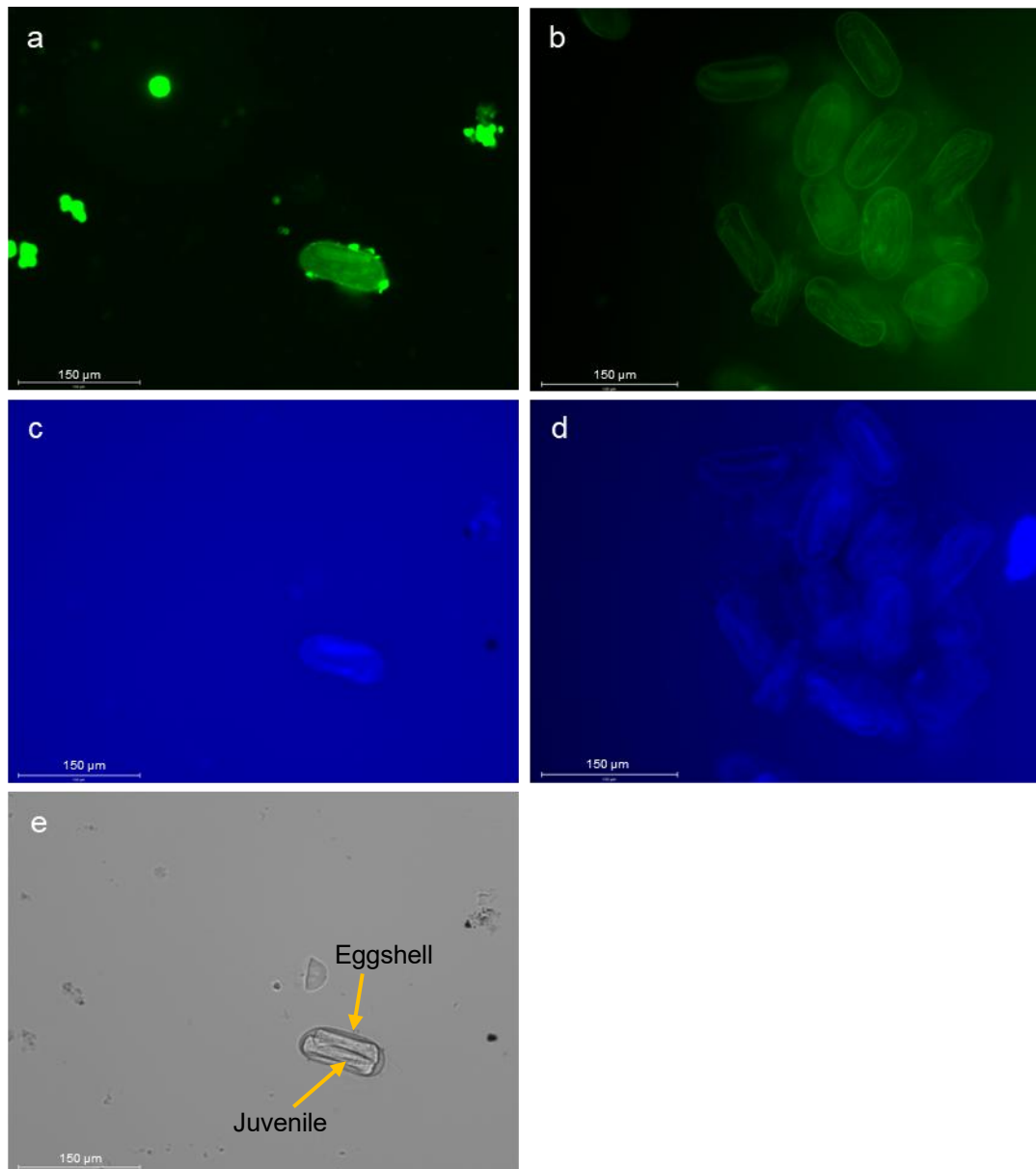


Figure 6.7: Immunofluorescent images of *Heterodera schachtii* eggs incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield image of egg shown in (a) and (c), given in (e).

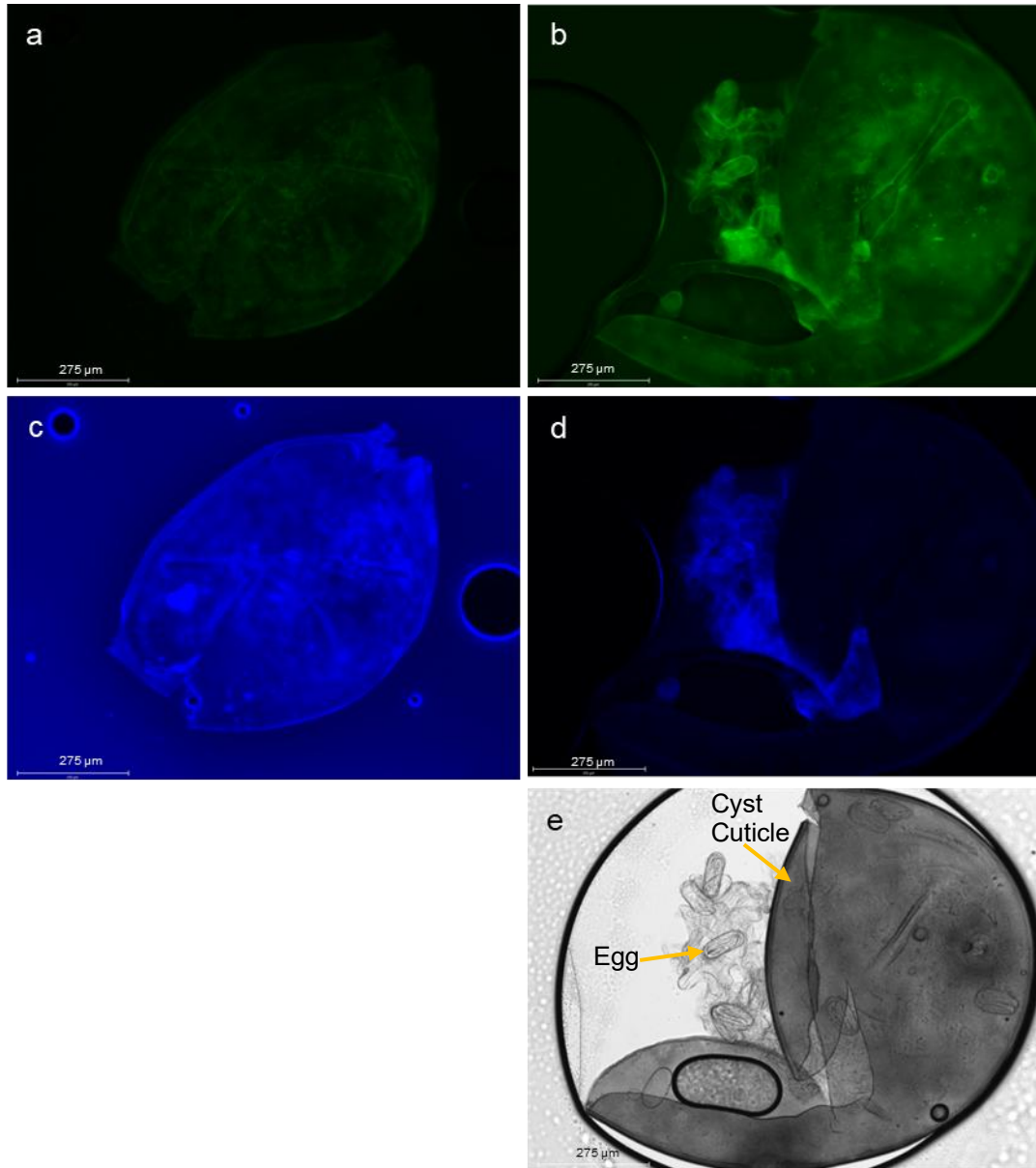


Figure 6.8: Immunofluorescent images of *Globodera pallida* cysts incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies, at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield image of cysts shown in (b) and (d) is given in (e).

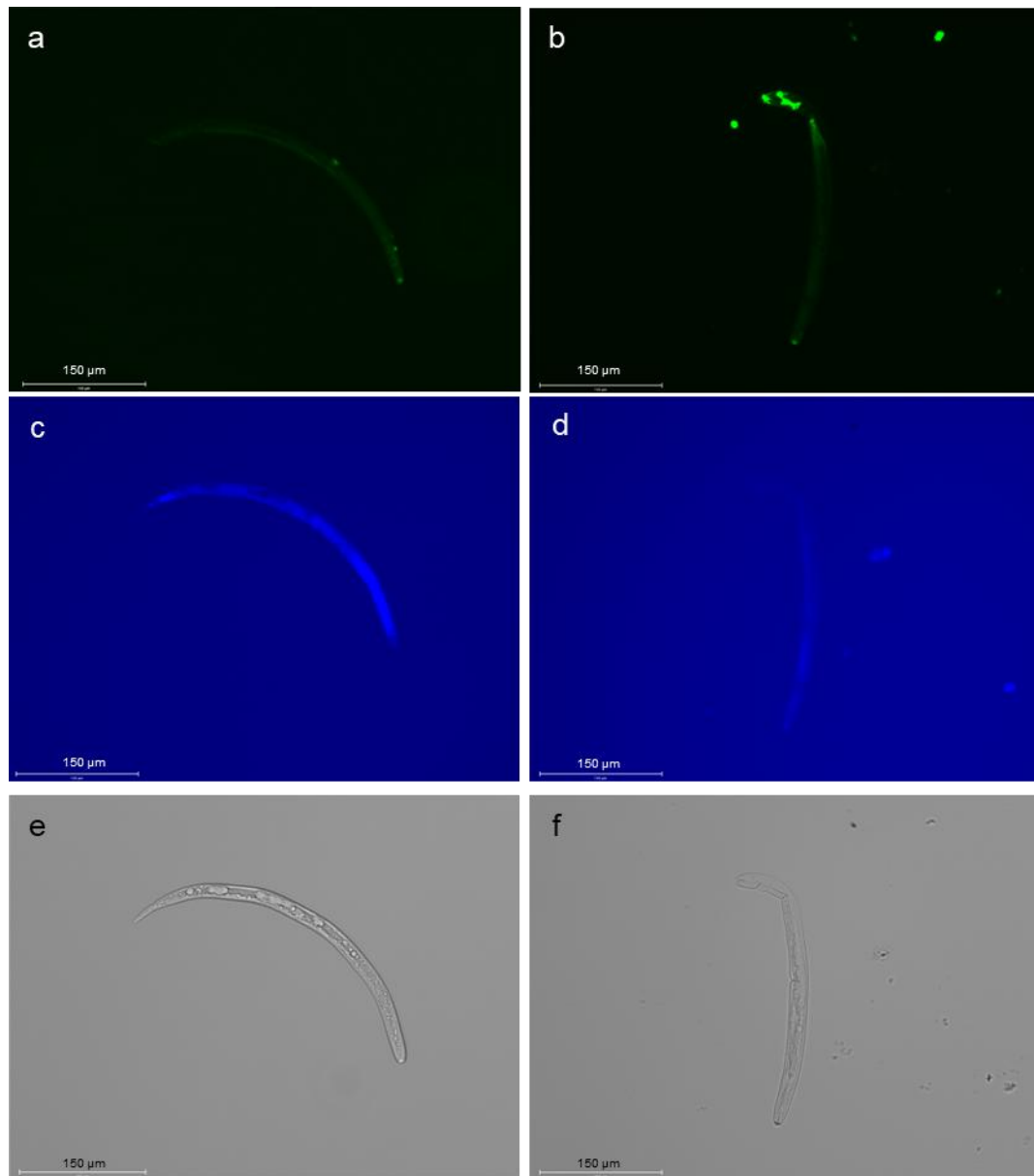


Figure 6.9: Immunofluorescent images of *Globodera rostochiensis* second-stage juveniles extracted from the egg, incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of the J2 are given in (e) and (f).

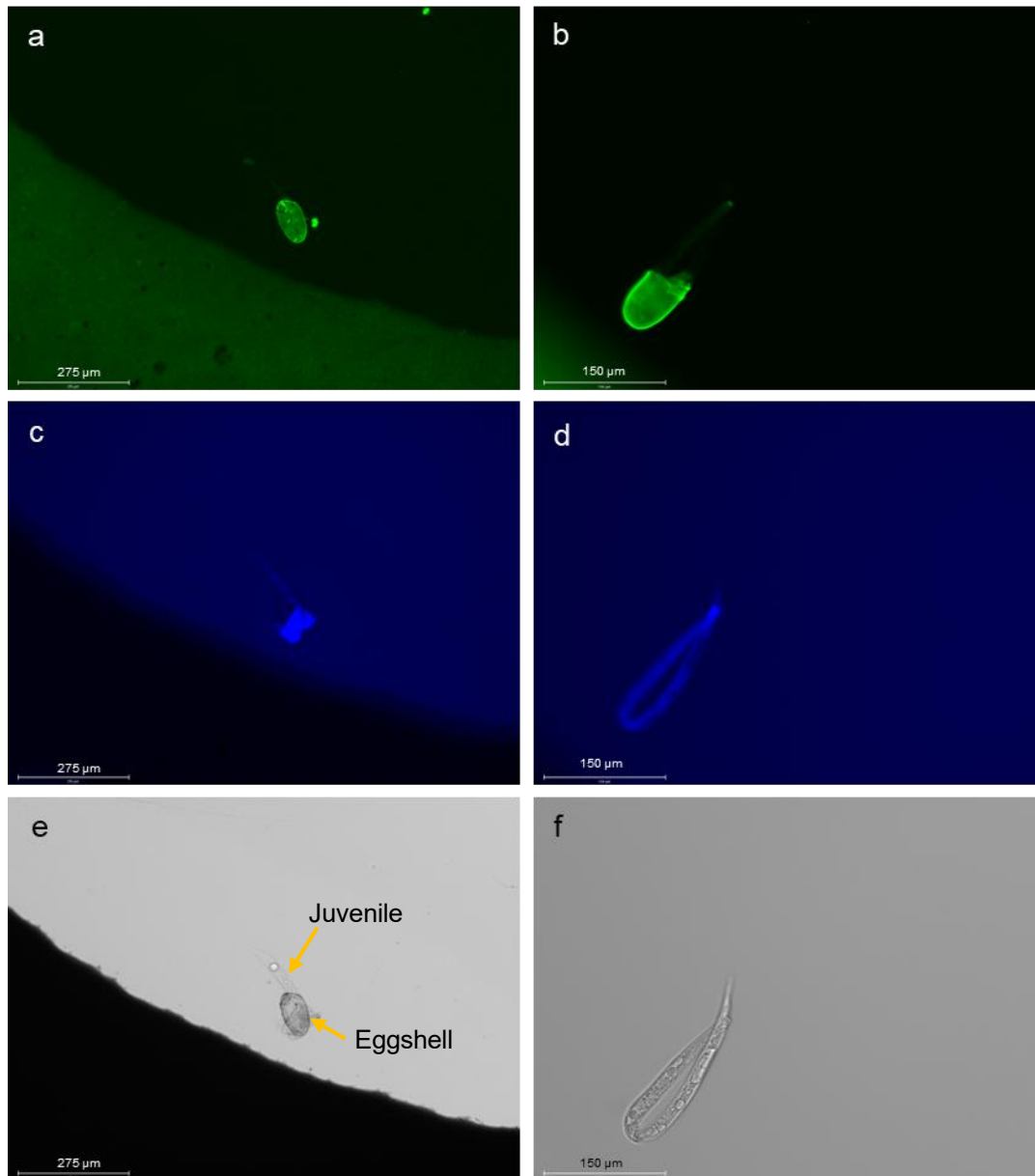


Figure 6.10: Immunofluorescent images of *Globodera rostochiensis* eggs and second-stage juveniles contained within the egg incubated in (a) pre-immune serum of PC320 antibodies or (b) PC320 antibodies at 1:500 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of egg are given in (e) and (f).

Table 6.4: Mean fluorescence intensity measured by imageJ of cyst nematode samples: *Globodera pallida*, *G. rostochiensis* and *Heterodera schachtii* as shown in Figure 6.6- 6.10. The mean fluorescent value of pre-immune serum or PC320 antibody at a concentration of 1:500, is given in arbitrary units (a.u).

Figure Number	Pre-Immune Serum Mean Fluorescence Intensity (a.u.)	PC320 Antibody (1:500) Mean Fluorescence Intensity (a.u.)
6.6	12.95	10.92
6.7	37.26	7.99
6.8	3.21	8.81
6.9	0.65	0.68
6.10	21.56	33.85

6.3.5.2 MAC 356 Antibodies

Female cysts and eggs were incubated in the monoclonal antibody, MAC 356. The localised response of the antibody was compared between *G. pallida* and *G. rostochiensis* samples.

The eggs of *G. pallida* appeared to have a greater response to the MAC 356 antibody than the eggs of *G. rostochiensis* (Figure 6.11 and Figure 6.12). The mean fluorescence of the *G. pallida* cyst sample in Figure 6.11 (a) was 7.36 a.u.; however, the mean fluorescence measurement for the *G. rostochiensis* cyst sample (b) was 14.98 a.u. When comparing the egg samples in Figure 6.12, the *G. pallida* egg had a greater mean fluorescence of 37.63 a.u. and was more defined, whereas the *G. rostochiensis* egg was slightly lower at 34.06 a.u.

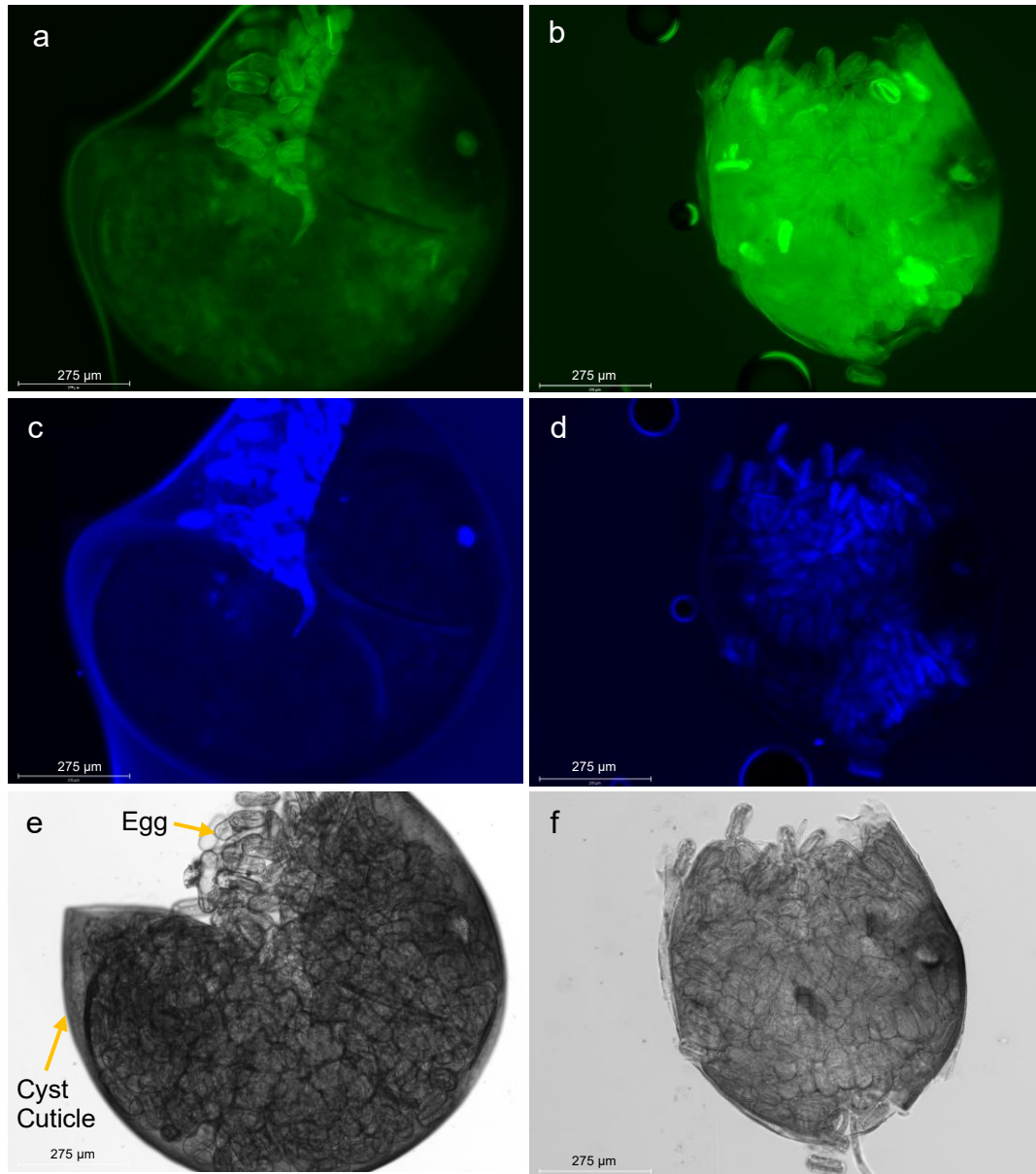


Figure 6.11: Immunofluorescent images of *Globodera pallida* (a) and *Globodera rostochiensis* (b) cysts incubated in MAC 356 antibodies at 1:250 concentration. Samples were then incubated in anti-rat secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of cysts shown are given in (e) and (f).

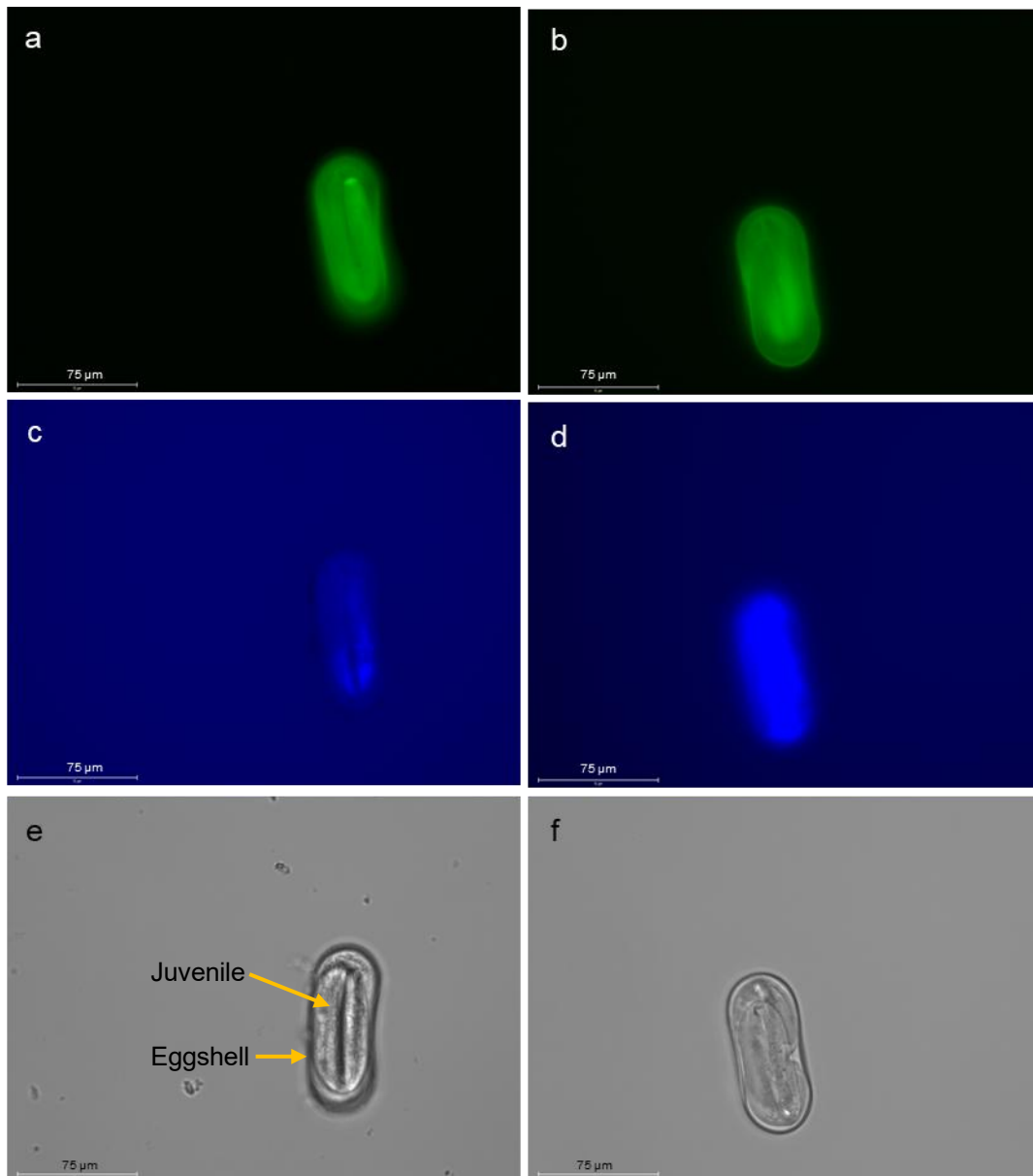


Figure 6.12: Immunofluorescent images of *Globodera pallida* (a) and *Globodera rostochiensis* (b) eggs incubated in MAC 356 antibodies at 1:500 concentration. Samples were then incubated in anti-rat secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of eggs are shown in (e) and (f).

6.3.5.3 Anti-Vitellogenin Antibodies

The localised reaction to anti-VIT antibodies in cyst nematode samples was determined. The fluorescence of each sample was measured and compared to the pre-immune serum.

The female cysts, eggs and J2 heads of *Heterodera schachtii* were incubated in either pre-immune serum or anti-VIT antibodies (Figure 6.13- 6.18). The measured mean fluorescence intensity values are given in Table 6.5. In Figure 6.13, the eggs contained within the cyst appear to fluoresce when using anti-VIT A antibodies when compared to the pre-immune serum. The measured fluorescence of the sample incubated in anti-VIT A was 9.09 a.u., whereas the pre-immune sample was 2.77 a.u. Similarly, in Figure 6.14, eggs incubated in anti-VIT A antibodies had a mean fluorescence level greater than the pre-immune serum treated samples, of 12.51 a.u. and 4.49 a.u., respectively. In Figure 6.15, the head of a J2 in (b), incubated in anti-VIT A antibodies, highlighted cells in the cuticle of the nematode and had a measured fluorescence of 12.55 a.u. The J2 sample incubated in pre-immune serum had a lower fluorescence of 9.30 a.u.

The female cysts of *G. pallida* were incubated in anti-VIT antibodies; unfortunately, the sample incubated in these antibodies contained few eggs. The sample incubated in pre-immune serum contained more eggs and showed a little autofluorescence. Although the antibody incubated sample (b) had few eggs, it had a greater mean fluorescence (7.2 a.u.) than the pre-immune serum samples (1.83 a.u.).

The female cysts and J2 of *G. rostochiensis* were also incubated in anti-VIT antibodies and pre-immune serum (Figure 6.17 and 6.18). Both female cyst samples contained many eggs; the sample incubated in the anti-VIT antibodies had clear illumination of the space surrounding the second-stage juveniles within the egg (Figure 6.17, b), whereas this was less defined in the pre-immune treated samples (a). The mean fluorescence of the anti-VIT treated samples was 14.49 a.u., whereas the mean fluorescence of the pre-immune serum was 5.21 a.u. In Figure 6.18, there was little difference between the two samples of J2; however, the head of the nematode in (b) had a marginal greater fluorescence. Furthermore, the mean fluorescence of the antibody treated sample was only slightly increased when compared to the pre-immune serum sample (Table 6.5).

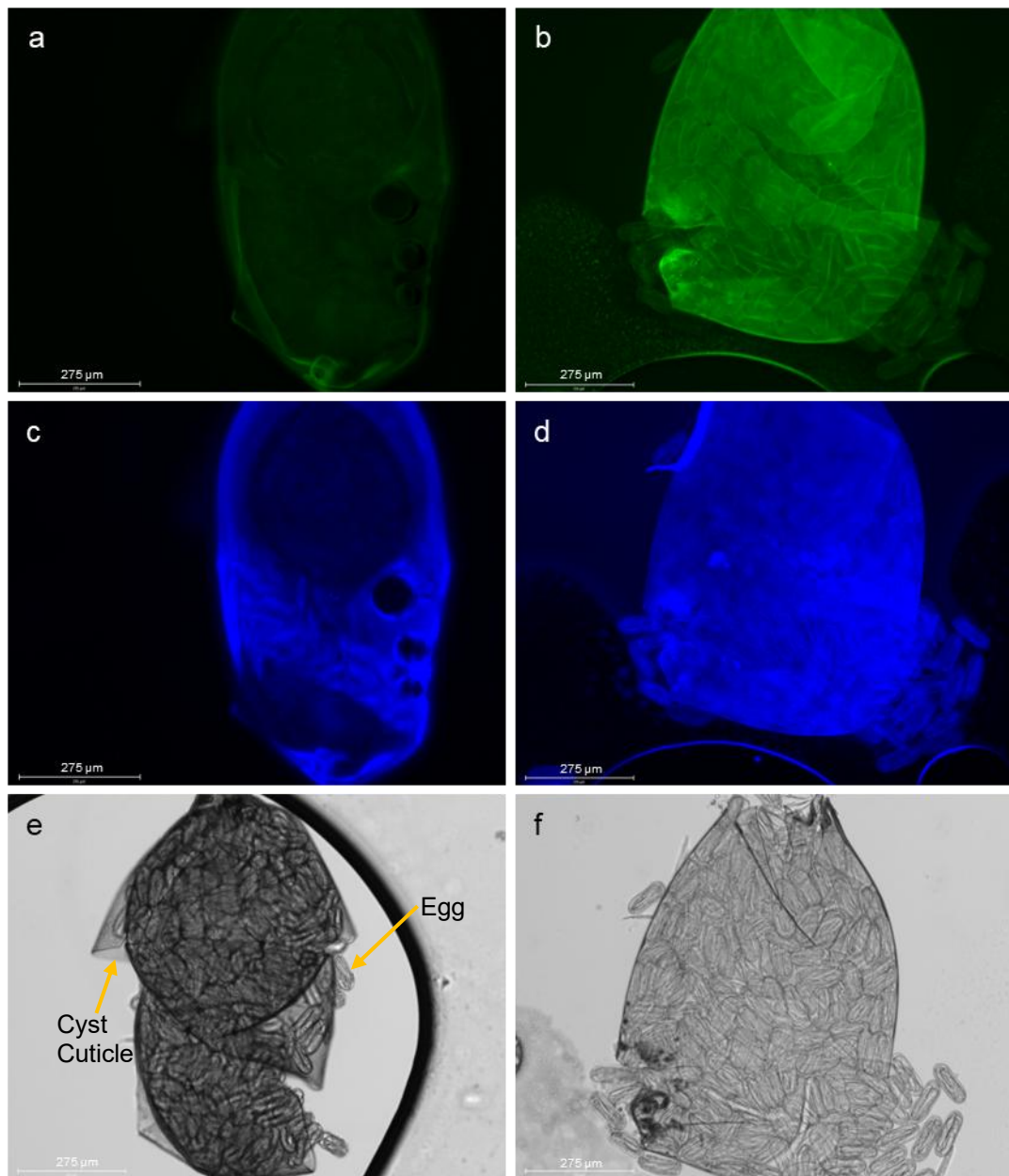


Figure 6.13: Immunofluorescent images of *Heterodera schachtii* cysts incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-VIT antibodies at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of cysts shown are given in (e) and (f).

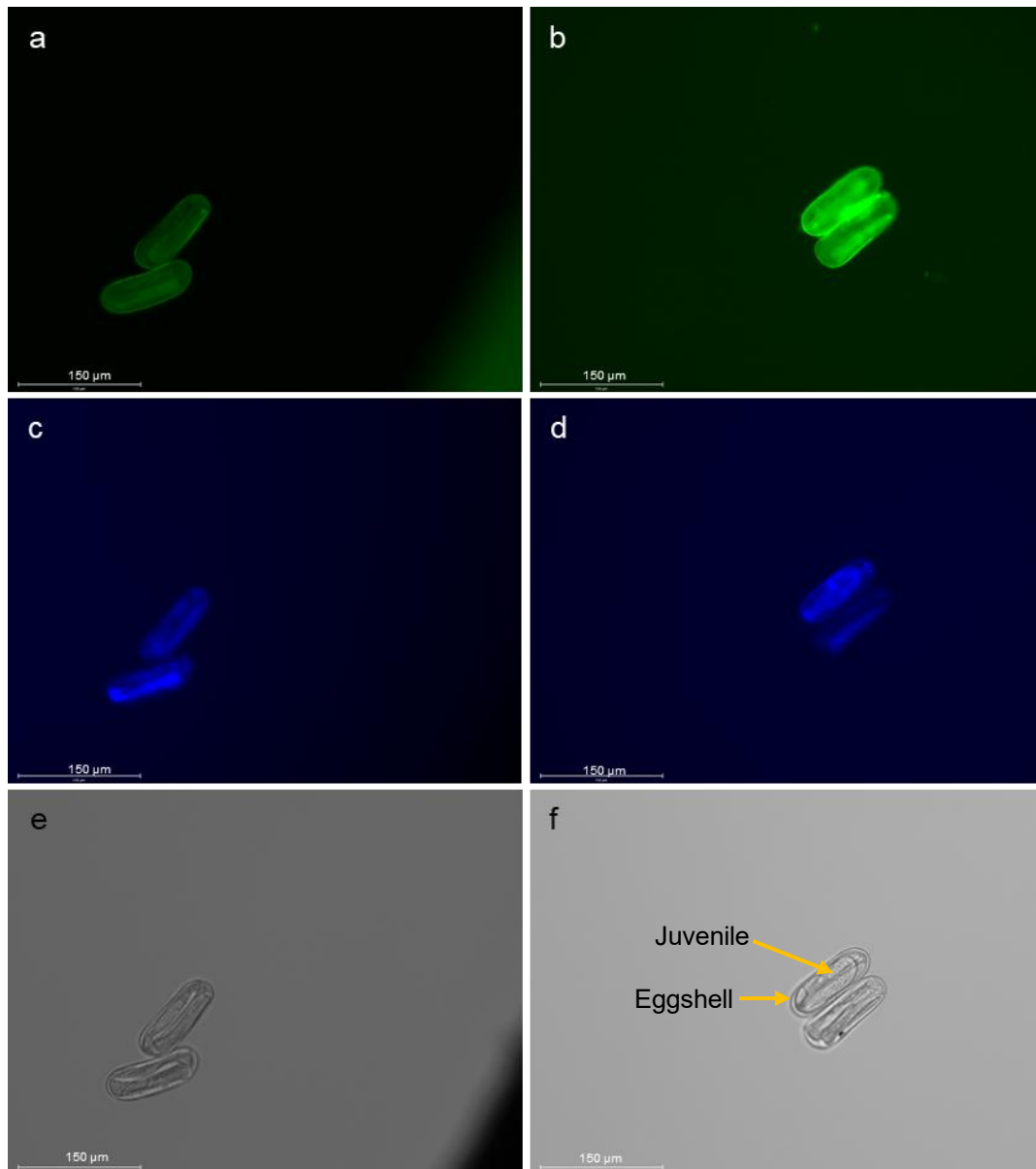


Figure 6.14: Immunofluorescent images of *Heterodera schachtii* eggs incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-VIT antibodies, at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of eggs shown are given in (e) and (f).

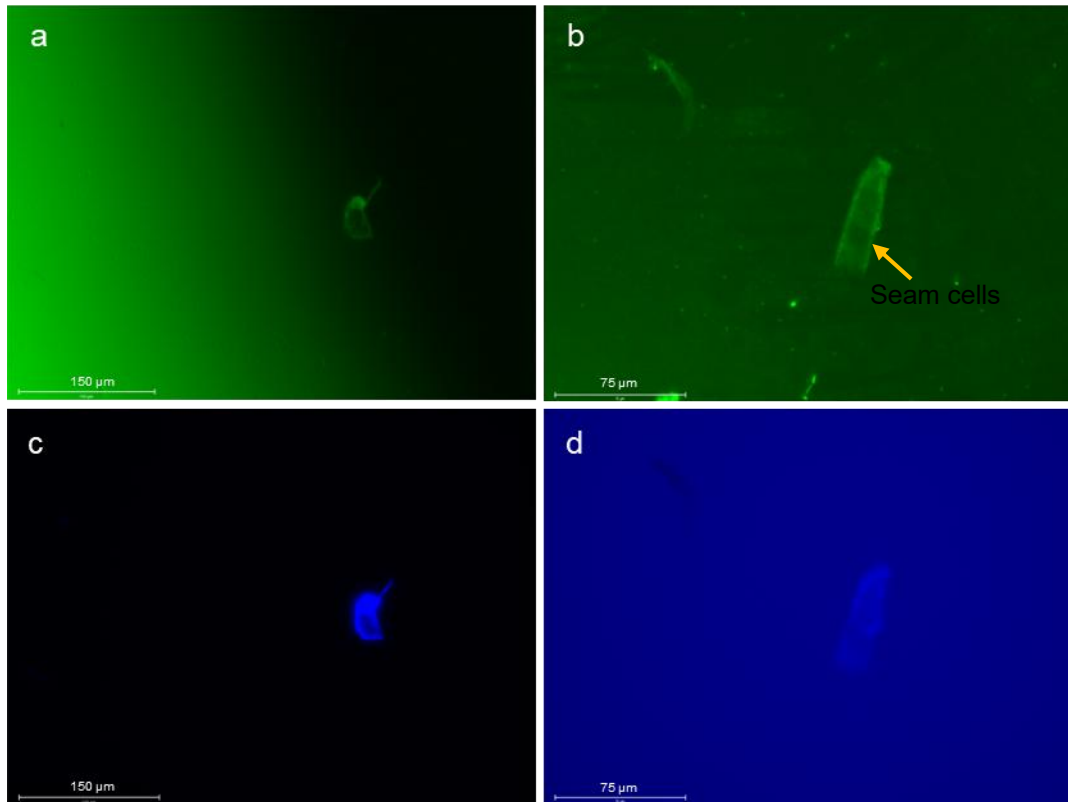


Figure 6.15: Immunofluorescent images of *Heterodera schachtii* second-stage juvenile heads incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-VIT antibodies, at 1:250 concentration, then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d).

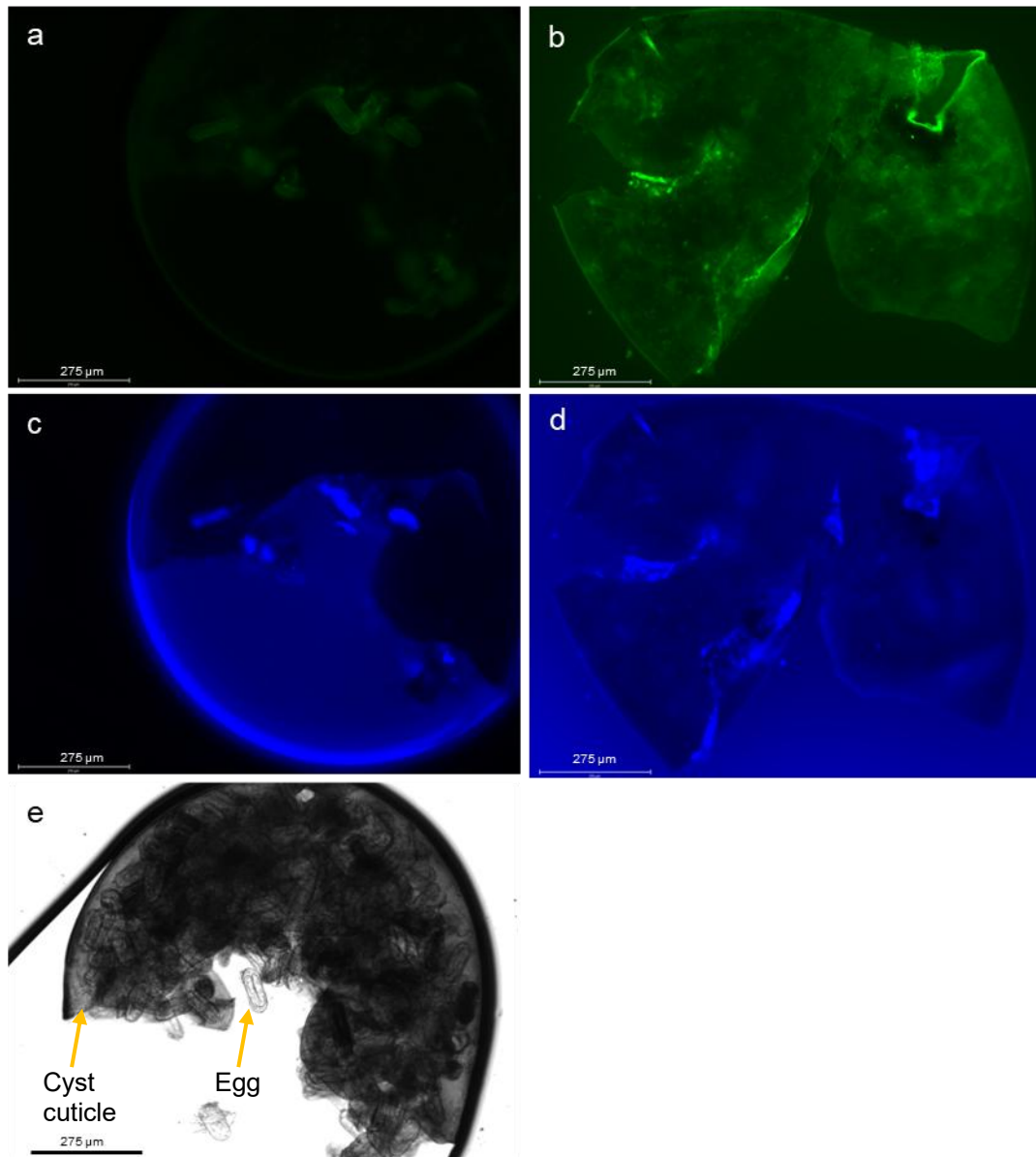


Figure 6.16: Immunofluorescent images of *Globodera pallida* cysts incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-vitellogenin antibodies at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield image of cyst shown in (a) and (c), given in (e).

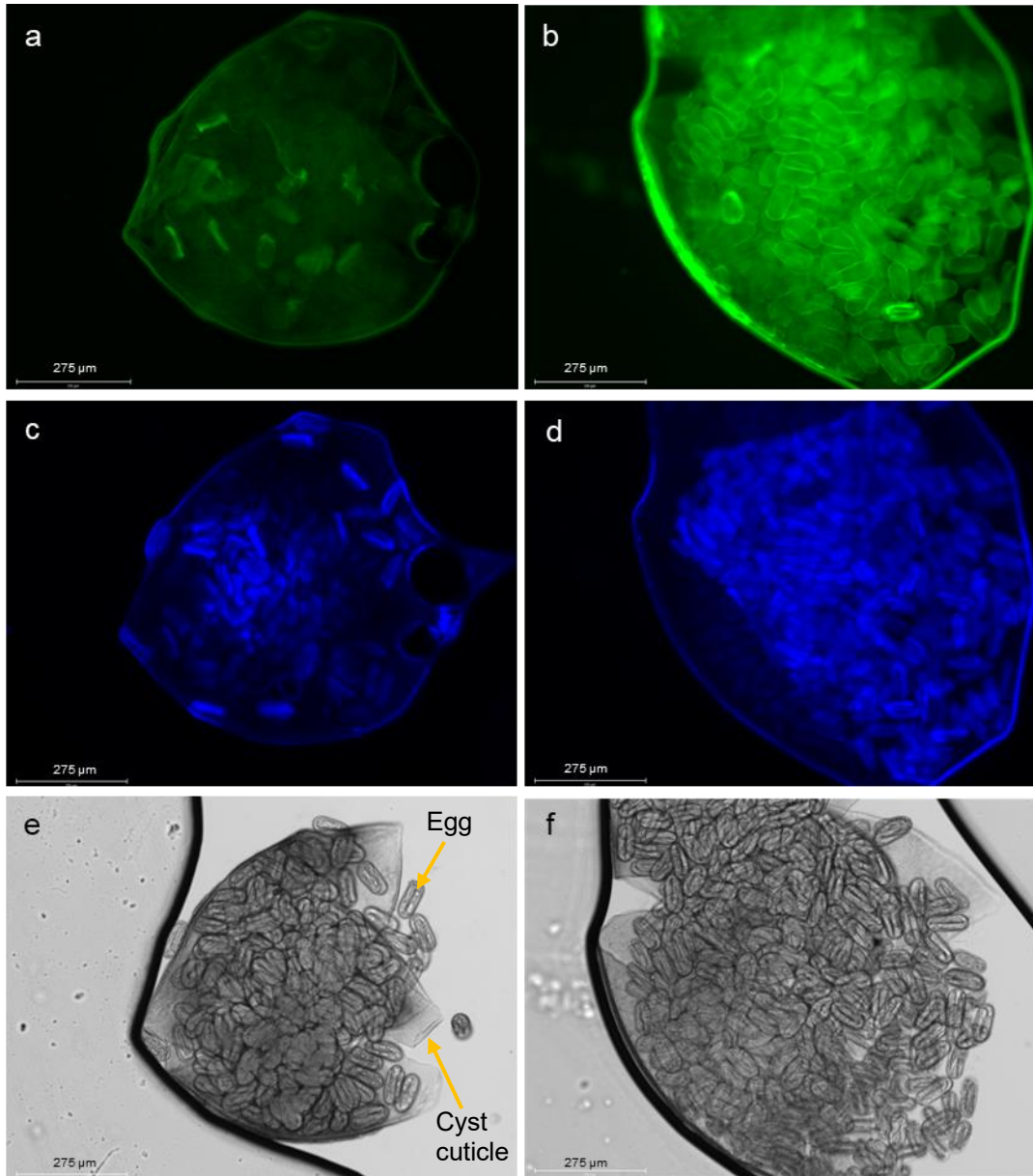


Figure 6.17: Immunofluorescent images of *Globodera rostochiensis* cysts incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-vitellogenin antibodies at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of cysts are given in (e) and (f).

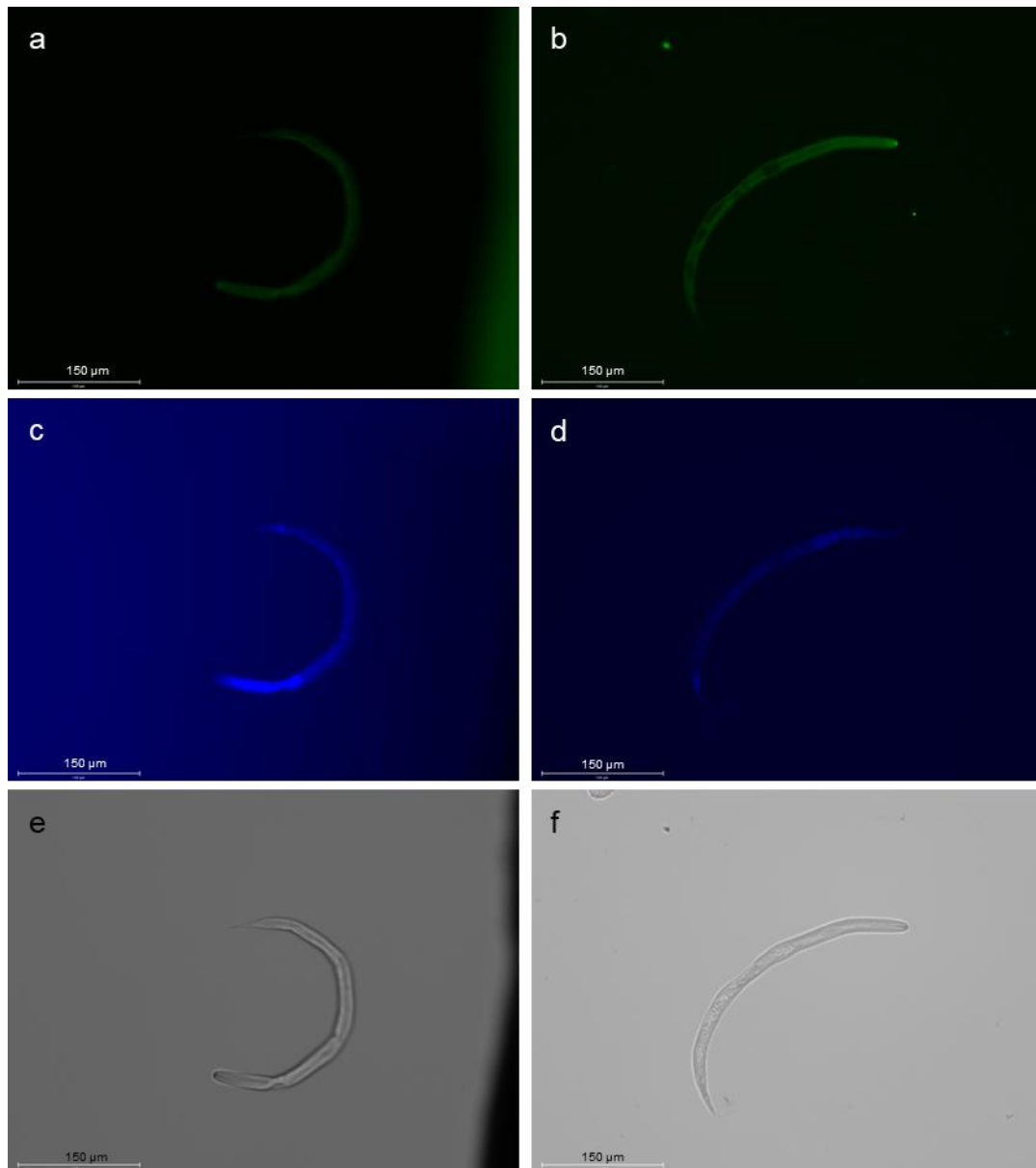


Figure 6.18: Immunofluorescent images of *Globodera rostochiensis* second-stage juvenile manually released from an egg; incubated in (a) pre-immune serum of anti-vitellogenin antibodies or (b) anti-vitellogenin antibodies at 1:250 concentration. Samples were then incubated in anti-rabbit secondary antibody, FITC conjugated, at 1:500. Samples were mounted in Vectashield antifade mounting medium with DAPI (c) and (d). Brightfield images of second-stage juveniles shown in (e) and (f).

Table 6.5: Mean fluorescence intensity measured by imageJ of cyst nematode samples, *Globodera pallida*, *G. rostochiensis* and *Heterodera schachtii* as given in Figure 6.13- 6.18. The mean fluorescent value of pre-immune serum or anti-VIT A antibody at a concentration of 1:500, is given in arbitrary units is given in arbitrary units (a.u).

Figure Number	Pre-Immune Serum Mean Fluorescence Intensity (a.u.)	Anti-VIT A Antibody (1:500) Mean Fluorescence Intensity (a.u.)
6.13	2.77	9.09
6.14	4.49	12.51
6.15	9.30	12.55
6.16	1.83	7.20
6.17	5.21	14.49
6.18	0.54	0.88

6.4 Discussion

The monoclonal antibody MAC 356 recognised a protein within *G. pallida* samples with an approximate molecular weight of 34 kDa. This finding is consistent with that of Robinson *et al.* (1993) who raised this antibody against a protein in *G. pallida*, which had an isoelectric focusing point of 5.7 and molecular weight of 34 kDa. Robinson *et al.* (1993) suggested that there was slight cross-reactivity of the antibody with *G. rostochiensis*. In western blot analysis and initial immunofluorescence study this cross-reactivity was confirmed and was clearly visible. The work conducted here also agrees with previous findings that the monoclonal antibody does not recognise *H. schachtii* samples.

The monospecific polyclonal antibody, PC320, recognised multiple proteins in *G. pallida*, *G. rostochiensis* and *H. schachtii*. A greater number of proteins were visualised in western blot analysis than discussed by either Curtis *et al.* (1996) or Fioretti *et al.* (2001). It has also been previously suggested that *H. schachtii* did not contain the same 34 kDa protein; however, in use of the PC320 antibody, a protein of a similar molecular weight was detected at 32 kDa and 36 kDa, giving an average molecular weight of 34 kDa in homogenised cyst samples. This protein may be a homologue of the 34 kDa protein in *G. pallida* and *G. rostochiensis*.

The use of the MAC 356 antibody resulted in fewer proteins detected in western blotting than PC320 and is likely due to antibody specificity. MAC 356 is a monoclonal antibody, whereas PC320 is a polyclonal antibody; monoclonal antibodies are likely to be more specific than polyclonal antibodies due to their production methods. The PC320 antibody would have been produced using methods for polyclonal antibody production like those described in Chapter 5. Alternatively, monoclonal antibodies are produced by immunising mice with the target antigen at 2–3-week intervals. Once antibodies specific to the target antigen have reached a sufficient

titre within the serum, the mice are euthanised and the antibody containing spleen cells are harvested. These cells are fused with myeloma cells to form a hybridoma, which is then screened for the antibody and cloned. Next, the hybridoma cells are proliferated by either injecting a second mouse to produce an ascites tumour or in vitro by tissue culture. These methods, once purified, provide antibodies against a single epitope (National Research Council (US) Committee on Methods of Producing Monoclonal Antibodies, 1999). Conversely, polyclonal antibodies may recognise multiple epitopes leading to cross-reactivity.

Initial protein identification using MALDI-TOF gave several outputs with low confidence, for the identification of the 34 kDa protein in *G. rostochiensis*. The 14_3_3 domain containing protein identified in PPN is thought to be involved in root invasion and feeding cell formation (Klink *et al.*, 2009) but has also been found in the secretome of *Meloidogyne incognita* (Bellafiore *et al.*, 2008). The preliminary identified triosephosphate isomerase protein is an enzyme found in the glycolytic pathway and has been sequenced in *Caenorhabditis elegans* (Logsdon *et al.*, 1995). Finally, the glutathione S-transferase protein is believed to be essential for parasitism of plant tissue and expressed in the pharyngeal secretory glands of *M. incognita* (Dubreuil *et al.*, 2007). A wider search of peptides retrieved from the Uniprot database search resulted in the same proteins identified across both PCN species.

The protein identification method liquid chromatography mass spectrometry, with its greater sensitivity, was able to give an increased confidence in the identification of the protein detected in *G. rostochiensis*. Unfortunately, these methods were not able to determine the identification of the 34 kDa protein in *H. schachtii*. It is possible that the protein concentration was too low for this detection method; however, results with greater confidence were obtained for *G. rostochiensis*. In both replicates, an annexin protein of a high molecular weight (166 kDa) was identified with the most significant matches and sequences; however, a protein with this molecular weight was not identified in western blot analysis. Also identified was an annexin with a molecular weight of 35.8 kDa, which may be the true protein identification due to the calculated molecular weight being close to the 34 kDa protein of interest. Another protein with the similar molecular weight in both replicates was the Ig-like domain-containing protein, with a molecular weight of 35.8 kDa.

It is likely that the identification of the 34 kDa protein with differential isoelectric focusing points in *G. pallida* and *G. rostochiensis* is an annexin protein. Annexins are proteins that bind membranes in high calcium conditions; it is believed that they regulate membrane transport, organise lipids on cellular membranes and regulate cellular stress (Gerke *et al.*, 2024). In LCMS protein sequencing, the 35.8 kDa *G. rostochiensis* protein detected, corresponded to a protein deposited on Uniprot. This identified protein, annexin, had a highly similar protein

sequence to the sequence determined by Fioretti *et al.* (2001). The protein sequence derived by Fioretti *et al.* (2001), was a result of screening a cDNA library with a monospecific polyclonal antibody, PC320, raised against the purified protein from *G. rostochiensis* with an isoelectric focusing point of 5.9 (Curtis *et al.*, 1996). This is supported by western blot analysis using PC320 which detected a protein at 36 kDa in *G. rostochiensis*. Furthermore, when the protein was isolated for protein sequencing by LCMS, this resulted in a protein identification of annexin. Interestingly, the annexin protein identified by LCMS, and previously sequenced by Fioretti *et al.* (2001), was also recently identified in egg samples of *G. rostochiensis* (Price *et al.*, 2023). In addition, an annexin homologue is present within the *H. schachtii* genome, encoded by Hsc_gene_5009, with a predicted molecular weight of 36.8 kDa (Siddique *et al.*, 2022), but has not experimentally been confirmed.

Preliminary investigations using immunofluorescence techniques were conducted to compare the localisation of proteins recognised by MAC 356 and PC320, in *H. schachtii*, *G. pallida* and *G. rostochiensis*. It could be expected that MAC 356 and PC320 would recognise proteins in similar locations due to being raised against the same target protein; however, PC320 may be less specific, as a polyclonal antibody. Furthermore, any differences in localisation may be due to species specificity. MAC 356 was shown to have greater specificity to *G. pallida* (Robinson *et al.*, 1993), supported by western blot findings, whereas PC320 was raised against a *G. rostochiensis* protein (Curtis *et al.*, 1996).

In this preliminary immunofluorescence work, incubating samples in PC320 resulted in greater fluorescence of *G. pallida* eggs when compared to the control. The use of MAC 356 also demonstrated fluorescence of the eggs in both *G. pallida* and *G. rostochiensis*. This supports the findings of Evans *et al.* (1995) and Curtis *et al.*, (1998) who suggested that the target protein of MAC 356 was available from the egg. Furthermore, Fioretti *et al.* (2001) also demonstrated that the greatest response of the PC320 antibody was to eggs. The PC320 antibody also indicated fluorescence of the eggshell in *G. rostochiensis*. If the target protein is annexin, this finding agrees with Price *et al.* (2023), whose work on the annexin protein localised it to the eggshell. The finding of annexin localisation to the eggshell by Price *et al.* (2023) is not consistent with Evans *et al.* (1995), whose findings stated that there was little antigenicity to eggshells by the monoclonal antibodies used at that time, which are now hypothesised to be targeting the same protein. Greater fluorescence was measured when *G. pallida* samples were incubated in MAC 356 than *G. rostochiensis*. This result is not surprising, as the original findings showed that MAC 356 was more specific to *G. pallida* (Robinson *et al.*, 1993) and was also demonstrated in western blot analysis.

Some fluorescence of J2 was visualised when incubating *G. rostochiensis* in PC320. Further work is required to fully understand the localisation of the target protein of MAC 356 so this can be successfully compared to PC320, which has previously been localised to the amphids, genital primordium and muscles surrounding the metacarpus pump chamber (Fioretti *et al.*, 2001).

Further results presented here represent a preliminary investigation into the immunolocalisation of the target protein of anti-vitellogenin antibodies. Greater fluorescence was observed in the eggs of *H. schachtii* and *G. rostochiensis*, but this may be a result of vitellogenin still being present within the egg prior to hatch, which has previously been demonstrated in the eggs of *C. elegans* (Sharrock, 1983). Interestingly, some fluorescence was shown in alternate segments along the cuticle of the *H. schachtii* J2 nematodes head, similar to the observations seen in the seam cells of surface mutants of *C. elegans* conferring resistance to bacterial pathogens (O'Rourke *et al.*, 2023). The location of antibody binding appears to show the annulation of the cuticle; however, little fluorescence was shown in *G. rostochiensis* J2, which were manually extracted from the egg. Mass spectrometry analysis of female *M. incognita* pharyngeal glands resulted in *vit* genes being detected in the secreted effectors (Bellafiore, *et al.*, 2020). It is possible that these secretions coated the outer cuticle of the *H. schachtii* nematode's head; however, in a similar study no *vit* genes were identified in *G. rostochiensis* female pharyngeal glands (Maier *et al.*, 2012). Furthermore, the difference in immunofluorescence response may be attributed to differences in protein deposition and localisation between the pre-infective J2 inside the egg and the hatched J2; however, this requires further investigation (Maier *et al.*, 2012). Limited research has associated vitellogenin with the cuticle of nematodes; however, intriguingly, in *C. elegans*, a regulator of *vit* gene expression, UNC-62, is expressed in the seam cells and hypodermis (van de Walle *et al.*, 2019), as well as in the intestine where it is believed to bind directly to *vit* gene promoters (Van Nostrand *et al.*, 2013).

Work presented here gives preliminary results of utilising immunofluorescence to localise the target proteins of MAC 356, PC320 and anti-vitellogenin antibodies. Future work could expand these results by including higher magnification of samples to develop further understanding of tissue-specific localisation. Also, additional life stages of *G. pallida*, *G. rostochiensis* and *H. schachtii* could be investigated to further establish the life stage specificity of the antibodies and the localisation of these proteins at each life stage.

7 Amplification of Genomic Region Containing *vit* Genes

7.1 Introduction

The development of molecular based technologies and accessibility to genome sequencing has aided the study of plant-parasitic nematodes (PPN). This technology has allowed advancements in phylogenetic studies, the detection of different species of PPN and understanding population dynamics, as well as the investigation of genes key to parasitism, host interaction and resistance.

One area of investigation utilising molecular techniques is the detection and identification of different species of PPN. With the use of polymerase chain reaction (PCR), primers that are species-specific can aid in distinguishing closely related species based on their DNA (de Oliveira *et al.*, 2011). This method can detect specific DNA sequences in eggs, second-stage juveniles (J2) and adult nematodes (Cameiro *et al.*, 2017). Species-specific primers have been developed for multiple species of cyst, root-knot and root-lesion nematodes, predominantly based on the ITS rRNA, but have also been developed to recognise other sequences including 18S rRNA, SCAR, COI or 28S rRNA (Subbotin, 2021). Cyst nematode species with species-specific primers include: *Globodera rostochiensis* and *G. pallida* (Mulholland *et al.*, 1996; Bulman and Marshall, 1997), *Heterodera schachtii* (Amiri *et al.*, 2002; Ko *et al.*, 2017) and *H. glycines* (Subbotin *et al.*, 2001; Ou *et al.*, 2008; Ko *et al.*, 2017). Also, the use of quantitative PCR allows specific genes to be amplified within a mixture; for example, a sample containing multiple organisms and soil (de Oliveira *et al.*, 2011). This method was utilised to detect *Pratylenchus scribneri* from field samples using primers specific to the ITS region (Arora *et al.*, 2020). Loop mediated isothermal amplification (LAMP) is a DNA based technology that can be used to amplify and detect PPN DNA from soil samples, plant and nematode tissues (Ahuja and Somvanshi, 2021). Studies have developed LAMP assays to detect *Meloidogyne* spp. (Niu *et al.*, 2011), *G. rostochiensis* (Ahuja *et al.*, 2021), *G. pallida* (Camacho *et al.*, 2021) and *Meloidogyne graminicola* (He *et al.*, 2021).

In order to amplify DNA from nematodes a variety of DNA extraction methods can be used. As described in Subbotin (2021), DNA extraction from J2 involves the breakdown of nematode tissues, and separation and isolation of DNA. Also described in Subbotin (2021) were several protocols for DNA extraction, including the use of proteinase K and worm lysis buffer (WLB) (Waeyenberge *et al.*, 2000), proteinase K with 10X PCR buffer (Subbotin *et al.*, 2018), NaOH

(Floyd *et al.*, 2002) and commercial kits. These methods can be compared to determine which method produces the largest quantity of high-quality DNA for further applications.

DNA sequencing has been a useful resource in nematology research. Sanger sequencing is a process of generating fragments of DNA of different lengths by adding nucleotides until a labelled chain-terminating dideoxy nucleotide (ddNTP) is added to the sequence (Sanger *et al.*, 1977). This method originally utilised gel electrophoresis for sequence determination; however, with the use of fluorescently labelled ddNTPs, capillary electrophoresis with laser detection can be used to automate the sequencing process (Eren *et al.*, 2022). A limitation of Sanger sequencing is only sequences of 800-1000 bp can be determined in one reaction (Crossley *et al.*, 2020); however, the genome of *M. incognita* was sequenced using Sanger sequencing (Abad *et al.*, 2008).

Next generation sequencing encompasses several sequencing technologies including Illumina, Pacific Bioscience (PacBio) and Oxford Nanopore methods (Eren *et al.*, 2022). The Illumina method is like Sanger sequencing; however, it has a higher throughput and is less time consuming. A DNA library is generated from fragmented genomic DNA sequences, that can be amplified and sequenced in parallel, to obtain results more quickly (Thermes, 2014). Illumina sequencing has been used in the RNA sequencing of *M. incognita* (Choi *et al.*, 2017). Both PacBio and Oxford Nanopore sequencing methods do not rely on amplification or fragmentation of the DNA sample and are less time consuming and more reliable than other earlier sequencing methods (Eren *et al.*, 2022). The sequencing of the *H. schachtii* genome used a combination of PacBio and Illumina sequencing (Siddique *et al.*, 2022). By contrast, the genome of *M. graminicola* was sequenced using Oxford Nanopore and Illumina sequencing (Phan *et al.*, 2020).

Molecular techniques have also been used to investigate genes of interest in PPN; however, no current studies have focused on the amplification of *vit* genes or vitellogenin nucleotide sequences in PPN. The only recognition of vitellogenin in PPN species has been a result of annotation of genomes and transcriptomes post-sequencing. The nucleotide sequences of *vit* genes have been obtained for several species including the chicken, with a nucleotide sequence of 20,342 base pairs (bp). The first nucleotide sequence to be determined for *Caenorhabditis elegans* vitellogenin was *vit-5*, that was shown to be 4869 bp (Spieth *et al.*, 1985a). The nucleotide sequences were then partially derived for five of the six *vit* genes (*vit-1*, *vit-2*, *vit-4*, *vit-5* and *vit-6*) (Spieth *et al.*, 1985b). The full nucleotide sequence was then determined and compared for *vit-2*, *vit-5* and *vit-6* (Spieth and Blumenthal, 1985).

Different populations of the same species of PPN can have varying DNA sequences. Therefore, determining the DNA sequence of the *H. schachtii* genomic DNA containing a *vit* gene may be useful in future applications of the populations studied or in future vitellogenin research. Furthermore, no current studies have developed specific primers for the amplification of vitellogenin in *H. schachtii*. Therefore, the aim of this work was to amplify and clone genomic DNA from *H. schachtii* that contained the *vit* gene Hsc_gene_6726.

7.2 Materials & Methods

7.2.1 DNA Extraction

In order to optimise DNA extraction methods, several protocols were compared to obtain samples with a high yield of DNA and good purity. In the first method, four *Heterodera schachtii* second-stage juveniles (J2) were placed in 10 µL of water within a 0.2 mL micro-Eppendorf tube and homogenised using sterile scissors. The samples were then washed through sequential steps of centrifugation at 2000 *g* for 1 min, removal of 5 µL of supernatant, addition of 5 µL of nuclease free water (NFW) and then vortexed. This was repeated several times until 5 µL of the solution remained within the tube. An amended protocol as described by Subbotin *et al.* (2018) was then followed; 2 µL of 10× PCR buffer (*Taq* PCR core kit, Qiagen) and 3 µL of proteinase K (600 µg/mL) was added to each tube. The samples were then incubated at 65°C for 1 h, then 95°C for 15 min.

DNA extraction methods were also compared between the use of 10× PCR buffer with proteinase K to worm lysis buffer (WLB) (50 mM KCl, 10mM Tris pH 8.2, 2.5mM MgCl₂, 0.45% NP-40, 0.45% Tween 20, 0.01% Gelatin) with proteinase K. Two J2 were placed in 5 µL of NFW and then were either crushed with a sterile tip or left whole. These samples were left overnight to allow the water to evaporate. Lysis conditions were followed as given in Table 7.1.

Table 7.1: Lysis conditions of samples in DNA extraction method optimisation.

Sample	Lysis Condition
1	10 µL 10× PCR buffer, NFW and proteinase K (120 µg/mL)
2	10 µL 10× PCR buffer, NFW and proteinase K (200 µg/mL)
3	10 µL WLB with proteinase K (23 µg/mL)
4	10 µL WLB with proteinase K (38 µg/mL)
5	10 µL WLB with proteinase K (383 µg/mL)

All samples were then placed at -80°C for 30 min. One sample from each lysis condition was incubated at 65°C for either 1 h, 2 h or 18 h, followed by incubation at 95°C for 15 min.

The quantity of DNA extracted and its purity was measured using a NanoDrop microvolume spectrophotometer. The target 260/280 ratio was 1.7-2.0, for DNA with higher purity. The machine was adjusted to zero using NFW, then each sample was measured.

7.2.2 Primer Design

The region of the *H. schachtii* genome (accession number: JAHGVF010000007.1) containing a putative *vit* gene, HSC_gene_6726, was used as a basis for primer design to amplify the genomic sequence of this gene (Figure 7.1). Three primer pairs were designed and tested (Table 7.2). Primer pairs A and C were designed using a combination of Primer3 (Untergasser *et al.*, 2012) and Primer-BLAST (Ye *et al.*, 2012), utilising the default parameters provided, with a product size range greater than 6000 bp. Primer pair B were chosen to represent the start and end of the target gene. Sequences were subject to BLAST searches to ensure that the primers were complementary to the target sequence, using default parameters. In subsequent analysis, the potential for the primers to self-anneal and form of secondary structures was evaluated using Oligo-Evaluator™ (Sigma-Aldrich, 2014).

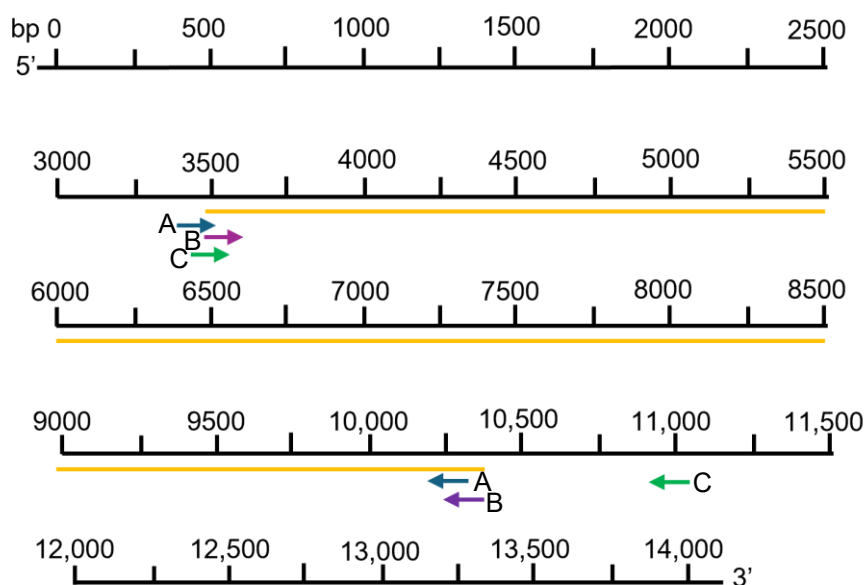


Figure 7.1: Line diagram of *Heterodera schachtii* genome sequence (accession number: JAHGVF010000007.1) utilised for primer design, containing a putative *vit* gene, Hsc_gene_6726. Orange line indicates gene location. Primer pairs: A (blue arrow), B (purple arrow) and C (green arrow).

Table 7.2: Primer design for use in PCR reactions to amplify the genomic region of the *H. schachtii* gene HSC_gene_6726.

Primer	Sequence (5'-3')	GC content (%)	Tm (°C)	Expected amplicon size (bp)
Primer A, forward	ATCGGACCGTATAAATAG	38.89	53.5	6953
Primer A, reverse	AAGGAGGCGCCGCG	78.57	65	
Primer B, forward	ATGCGCACTTTCGCGCTTTT	50	62.73	6856
Primer B, reverse	TCAGTAGCGTCTGCGCCA	61.37	61.11	
Primer C, forward	CCATCCCCTCCTGTGGAGTA	60	60.03	7652
Primer C, reverse	TTTCCTCTCGCGCTTTCTGT	50	59.97	

7.2.3 Polymerase Chain Reaction

To optimise PCR reactions, several approaches were taken to amplify the target DNA. PCR reactions were initially completed using the *Taq* PCR core kit (Qiagen). The master mix for PCR reactions included: 10× PCR reaction buffer, 50 mM MgCl₂, 10 mM dNTP, *Taq* polymerase (5 units/ µL), 25 µM forward primer A, 25 µM reverse primer A and NFW to a total of 23-24 µL per sample. The master mix was vortexed and aliquoted into 0.2 mL Eppendorf tubes, where 1-2 µL of DNA template was added. Nuclease free water was used as a negative control in each reaction. Samples were then centrifuged at 2000 g for 1 min. The PCR reaction was completed in a thermocycler, with the following profile: an initial denaturation at 94°C for 4 min, 35 cycles of 1 min at 94°C, 1.5 min at 55°C, and 2 min at 72°C, and a final step of 10 min at 72°C. Samples were then held at 4°C. Unfortunately, this reaction resulted in no DNA bands of the target amplicon size.

Next, a gradient PCR approach was used to determine the appropriate annealing temperatures for the reaction, using the master mix previously described and primer pair A. Two DNA samples extracted from *H. schachtii* were used, with nuclease free water as a negative control. The annealing temperatures across the thermocycler plate were set to eight set temperatures, ranging from 62.3°C to 47°C. PCR samples were then subject to agarose gel electrophoresis. Samples with an annealing temperature of 59.2°C gave an intense band on the gel. Subsequently, a PCR reaction comparing the master mix previously described to the same master mix with the inclusion of Q solution from the *Taq* PCR core kit (Qiagen) was investigated, using the new annealing set point. In this reaction some contamination of the negative control was observed and nematode DNA samples contained multiple bands not of the expected amplicon size.

New primer pairs (B and C) were designed as described in section 7.2.2. The new primer pairs were tested using the *Taq* PCR core kit (Qiagen), utilising the same samples for comparison between primers. The PCR protocol previously described was followed, with nuclease free water was used as a negative control, resulting in several bands amplified by both primer pairs B and C. Gradient PCR reactions were completed using primer pairs B and C, at eight annealing temperature set points. When using primer pair B, the annealing temperatures used were 65.6°C to 53.6°C, whereas when using primer pair C, the temperature ranged from 63°C to 52.8°C. This reaction was repeated using primer pair C, instead the annealing temperature set points were 64°C to 56.6°C. All reactions resulted in no bands visualised. The annealing temperature for both primer pairs was recalculated and a PCR reaction at a lower annealing temperature of 55°C was performed, resulting in several bands amplified.

Due to the size of the amplicons, the use of long-range PCR was investigated. Platinum SuperFi II PCR Master Mix (Thermo Fisher Scientific) was used, which contained Platinum SuperFi II DNA Polymerase, Platinum SuperFi II Buffer, and dNTPs. This product enabled universal annealing of the primers at 60°C. A master mix was made containing 2× Platinum SuperFi II PCR Master Mix, 10 µM forward primer, 10 µM reverse primer and NFW to a total of 18 µL, per sample. The master mix was vortexed and aliquoted into 0.2 mL Eppendorf tubes, 2 µL of DNA sample was added to each tube. The samples were placed in a thermocycler, using the following thermal profile: 98°C for 5 min; 35 cycles of 10 s at 98°C, 10 s at 60°C, and 3.5 min at 72°C, and a final step of 5 min at 72°C. The samples were then held at 4°C.

7.2.4 Agarose Gel Electrophoresis

The successful amplification of the sample DNA was confirmed with the use of agarose gel electrophoresis, which separates DNA fragments by their size. Due to the larger size of the amplicon, 0.5% agarose gels were cast. Gels were made by combining 30 mL of TBE buffer (130 mM tris, 45 mM boric acid and 2.5 mM EDTA) with 0.15 g of agarose for small gels or 70 mL of TBE buffer with 0.35 g of agarose for large gels. The agarose was dissolved in the microwave until the solution was clear. Once the solution had cooled slightly, depending on the size of the gel, 3 µL or 7 µL of GelRed® DNA Gel Stain was added to the solution. The solution was gently mixed and poured into a casting tray, with a comb inserted and allowed to set for 30 min. Once set, the comb was removed, and the tray was filled with sufficient TBE buffer to cover the gel. Meanwhile, 5 µL of the PCR products were added to a 0.2 mL Eppendorf tube with 1 µL of loading dye (20% sucrose, 10% Ficoll, 10mM EDTA, 0.25% bromophenol blue with TBE buffer). To determine the size of the DNA fragments, 5 µL of a ladder was loaded onto the gel, either Quick-Load® 1 Kb Extend DNA Ladder (Qiagen) or GeneRuler 1 Kb DNA Ladder (Thermo Fisher Scientific). Samples (5 µL) were also loaded

into the wells. The gel electrophoresis was then run for approximately 45 min at 100 V, until the gel front had migrated two-thirds through the gel.

DNA agarose gels were visualised similarly to methods described for western blots and SDS-PAGE gels (Section 5.2); however, UV was used to visualise the bands in the Syngene InGenius3 gel documentation system and a green LED transilluminator was used in the iBright CL1500 Imaging System. To determine the approximate length of the PCR product, the images were loaded into imageJ. The relative migration distance (R_f) of each marker, the DNA bands and the gel front was measured. The relative size of the DNA bands was calculated as previously described (Section 5.2).

Bands of interest were extracted from the agarose gels using the iBright CL1500 Imaging System to visualise the band, then using a sterile scalpel, the band was cut out of the gel and placed in a 1.5 mL Eppendorf tube. The Monarch DNA Gel Extraction Kit (New England Biolabs Inc.) was used to obtain purified DNA from the extracted band, following the manufacturer's recommended protocol. In brief, the gel slice was dissolved using gel dissolving buffer and heated to 50°C for up to 10 min. The solution was transferred into a spin column and centrifuged at 13,000 *g* for 1 min. DNA wash buffer was added to the tube and centrifuged at 13,000 *g* for 1 min and repeated. The spin column was then transferred into a new 1.5 mL Eppendorf tube; 20 μ L of elution buffer was added to the tube and incubated at room temperature for 1 min. The sample was then centrifuged at 13,000 *g* for 1 min. Purified DNA samples were measured using the NanoDrop spectrophotometer for DNA concentration (ng/ μ L) and DNA purity (260/280 ratio).

The Purified DNA was used as template DNA in a PCR reaction to obtain a higher yield of DNA. The PCR product from this reaction was used in agarose gel electrophoresis again to obtain a band of interest. Once a single band was obtained, this purified DNA was used in ligation reactions.

7.2.5 Ligation and Transformation

A blunt-end cloning protocol was used to insert the purified DNA into the pJET 1.2/blunt cloning vector, this was done using the CloneJET PCR Cloning Kit (Thermo Scientific™), following the manufacturer's recommended protocol for blunt-end DNA fragments generated using the SuperFi II DNA polymerase. A molar ratio of 3:1 of DNA fragment to pJET 1.2/blunt is recommended; however, in some cases due to a DNA concentration of less than 250 ng/ μ L, smaller ratios were used. The ligation reaction was set up on ice, as given in Table 7.3. The ligation mixture was then incubated at room temperature for 30 min and stored at -20°C until required for transformation.

Table 7.3: Reaction mixture for ligation reaction of DNA fragments from second-stage juveniles of *Heterodera schachtii* using the pJET 1.2/blunt cloning vector. Sample E2 is a replicate of sample E.

Reagent	DNA Fragment/ Reaction Volume							
Sample (Sample: Vector)	E (1:3)	F (1:6)	G (1:2.5)	J (14:1)	K (16:1)	H (2:1)	I (3:1)	E2 (2:1)
2× Reaction Buffer (μL)	10	10	10	10	10	10	10	10
Purified PCR Product (μL)	5	5	5	1	1	7.44	5.74	8
pJET1.2/blunt Cloning Vector (50 ng/μL) (μL)	1	1	1	1	1	1	1	1
Nuclease Free Water (μL)	3	3	3	7	7	0.56	2.26	0
T4 DNA Ligase (μL)	1	1	1	1	1	1	1	1
Total Volume per reaction (μL)	20							

Transformation reactions were conducted to introduce the plasmid construct (Figure 7.2) into bacterial cells for later plasmid purification. DH10B *Escherichia coli* competent cells (Thermo Scientific™) were thawed on ice and 50 μL aliquots were made in chilled 1.5 mL Eppendorf tubes. The ligation mixture (5 μL) was inoculated into the tube containing competent cells, gently mixed, and incubated for 30 min on ice. The cells were then heat shocked for 30 s at 42°C and again incubated on ice for 2 min. The cells were recovered using 250 μL S.O.C medium. The tubes were placed in a shaking incubator at 225 rpm for 1 h at 37°C. Lysogeny broth agar supplemented with ampicillin (100 μg/mL) (LB-A) was used to grow the transformed bacteria. Two agar plates were inoculated with 200 μL or 20 μL of neat bacterial suspension, and one plate was inoculated with 20 μL of the bacterial suspension diluted 1:10 in S.O.C medium. Bacterial cultures were incubated overnight at 37°C, then colonies from each plate were counted. The plasmid construct contained an ampicillin resistant gene, resulting in only resistant bacteria being able to grow. Additionally, if the construct lacked an insert, the lethal restriction enzyme present in the plasmid would kill the *E. coli* in the transformation reaction.

The initial transformation and ligation reaction for several samples was unsuccessful, therefore this process was repeated. Samples included in the initial reaction were E, F, G, J and K, whereas samples H and I were used in a subsequent reaction with a repeated reaction using sample E. Control DNA of a known sequence length was also used in this reaction to confirm the successful ligation and transformation. Furthermore, pUC 19 control was used to determine the success of the transformation reaction.

7.2.6 Colony PCR

Prior to plasmid purification, the presence and size of the inserted DNA were determined by performing colony PCR. A small amount of bacteria was picked from one colony using a sterile toothpick and resuspended in 12 μL NFW in a 0.2 mL Eppendorf tube. The sample was then heated at 98°C for 10 min. The PCR master mix was prepared using the Platinum SuperFi II PCR master mix, as described in Section 7.2.2, instead using the primers supplied with the CloneJET PCR Cloning Kit (Table 7.4). The master mix was vortexed and 15 μL aliquots were made into the required number of tubes. The colony suspension was added to the master mix, vortexed briefly and placed in the thermocycler. The thermal profile used was an initial denaturation at 98°C for 5 min, 35 cycles of 10 s at 98°C, 10 s at 60°C, and 4 min at 72°C, with a final step of 5 min at 72°C. Samples were then held at 4°C.

Table 7.4: pJET1.2 forward and reverse sequencing primers.

Primer	Sequence (5'-3')
Forward	CGACTCACTATAGGGAGAGCGGC
Reverse	AAGAACATCGATTTTCCATGGCAG

To determine the size of the PCR product amplified (plasmid and insert construct), agarose gel electrophoresis was used to separate the fragments by their size. The gels were then visualised and measured using methods described in Section 7.2.3.

7.2.7 Plasmid Purification

Plasmid DNA was purified from colonies selected from the colony PCR analysis using the GeneJET Genomic DNA Purification Kit (Thermo Scientific™), according to the manufacturer's recommended protocol. Initially, selected colonies were inoculated onto fresh LB-A agar plates and incubated for 24 h at 37°C. A single colony was then inoculated into 5 mL LB-A broth and incubated at 37°C for 16 h, in a shaking incubator at 200 rpm. The bacteria were then harvested by centrifugation at 6800 g for 2 min at room temperature and the supernatant was removed. All purification steps were then carried out at room temperature and all centrifugation steps were performed at 13,000 g . Pelleted cells were resuspended in 250 μL of the resuspension solution and vortexed. Subsequently, 250 μL of lysis solution was added to the tube and inverted six times. Next, 350 μL of the neutralisation solution was added and the tube was again inverted six times. The sample was centrifuged for 5 min to form a pellet of cell debris and the supernatant was transferred to the GeneJET spin column, then centrifuged for 60 s. Wash solution (500 μL) was added to the column and centrifuged for 60 s. This step was repeated with an additional centrifugation of 60 s. The spin column was transferred to a new 1.5 mL Eppendorf tube and 50 μL of elution buffer was added to the column, incubated for 2 min at room temperature and centrifuged for 2 min. An additional

elution step was then performed to collect residual DNA from the membrane. The purified plasmid DNA was quantified using the NanoDrop spectrophotometer and stored at -20°C.

7.2.8 DNA Sequencing

Purified plasmid DNA was diluted to 100 ng/μL and sent to Eurofins Genomics, Ebersberg, Germany using their TubeSeq Supreme service, a Sanger sequencing service. Primers used in the Sanger sequencing are given in Table 7.4. Due to a limitation of this technique, only the first 1100 bp of each sample could be sequenced. Chromatograms were viewed and trimmed to obtain a DNA sequence that contained sharp peaks and distinguishable nucleotides. The plasmid DNA was also removed from this sequence. The DNA sequences obtained were aligned against the reference DNA sequence to determine regions of homology. Additionally, BLASTn searches were done to determine homologous regions in the *H. schachtii* genome, to confirm whether the genomic sequence containing the Hsc_gene_6726 gene had been successfully cloned. The BLASTn search used default settings, with an expect value of 10.

The purified plasmid DNA samples that showed regions of similarity to the *H. schachtii* gene Hsc_gene_6726 were quantified using the Qubit dsDNA broad range assay kit, following the manufacturers recommended guidelines. These samples were then sent to a plasmid sequencing service provided by GeneWiz from Azenta, Takely, UK. A schematic of the plasmid vector and the *H. schachtii* genomic DNA insert is given in Figure 7.2.

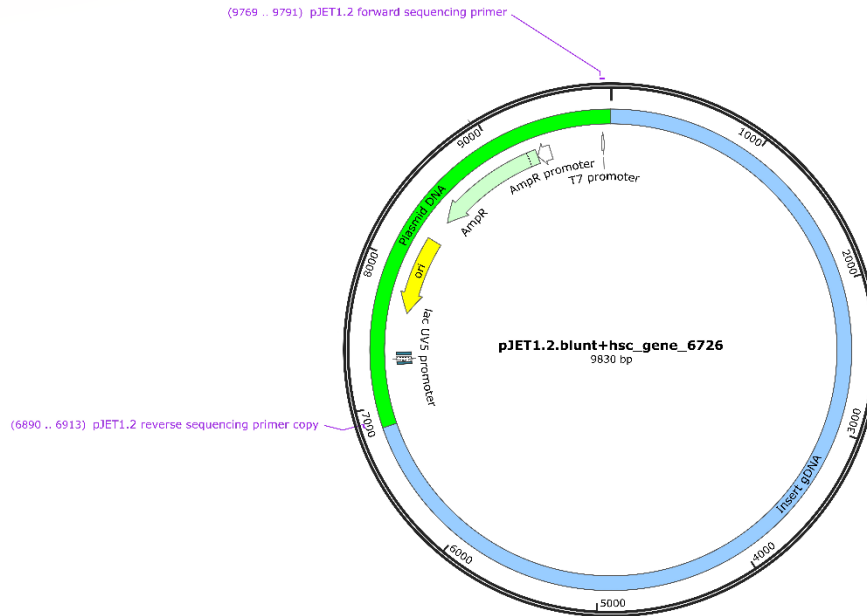


Figure 7.2: Diagram of plasmid construct: pJET 1.2 blunt plasmid DNA (green) with genomic DNA sequence of *Heterodera schachtii* gene Hsc_gene_6726 (blue). Origin of replication (yellow), ampicillin resistance site (pale green) and sequencing primer sites (purple) are labelled.

Once whole plasmid sequences had been obtained by next generation sequencing, these sequences were aligned to the reference sequence of the plasmid and insert construct. Next, the plasmid sequence was removed, only retaining the *H. schachtii* genomic sequence. A percentage identity of the pairwise alignment of these sequences was then calculated. Furthermore, a consensus sequence was generated and compared to the reference sequence, again a percentage identity score of this alignment was calculated. The consensus sequence was then used as a query for BLAST searches against the *H. schachtii* genome available from Wormbase Parasite, with expect values of 0.1-100.

7.3 Results

Polymerase chain reaction was used to amplify the region of the *Heterodera schachtii* genome that contained vitellogenin gene (Hsc_gene_6726) encoding DNA. Methodology was optimised by comparing several methods of DNA extraction from *H. schachtii* J2 and several primer pairs (Figure 7.3). To obtain bands of the approximate amplicon size expected and concentration required, PCR reactions were repeated (Figure 7.4, 7.5 and 7.6). After ligation and cloning of the target DNA into a plasmid vector, the presence of the plasmid was confirmed by colony PCR (Figure 7.7 and 7.8). Samples were then either sequenced by Sanger sequencing or next generation sequencing.

7.3.1 Comparison of DNA Extraction Methods

The use of primer A did not result in any amplified DNA of the expected amplicon size of 6953 base pairs (bp), instead multiple bands of various sizes were visualised, possibly a result of non-specific binding (Figure 7.3). By contrast, the use of primer B gave two amplicons with a higher sequence length; the band marked 'A' was approximately 7440 bp and band 'B' was approximately 8000 bp. The use of primer B also resulted in two bands with a greater signal at approximately 880 bp and 510 bp.

DNA extraction methods were compared (Figure 7.3); samples in the agarose gel lanes S1-S5 were prepared under different lysis conditions, either 10X PCR buffer or worm lysis buffer was used with several concentrations of proteinase K. Lane S6 contained a sample prepared under the same conditions as the sample in lane S1. Figure 7.3 illustrated that the 10X PCR buffer with a lower concentration of proteinase K gave bands of greater sequence length; however, this band was also present in Lane S2, prepared using worm lysis buffer with a lower concentration of proteinase K. The number of bands visualised in the gel reduced with an increased concentration of proteinase K.

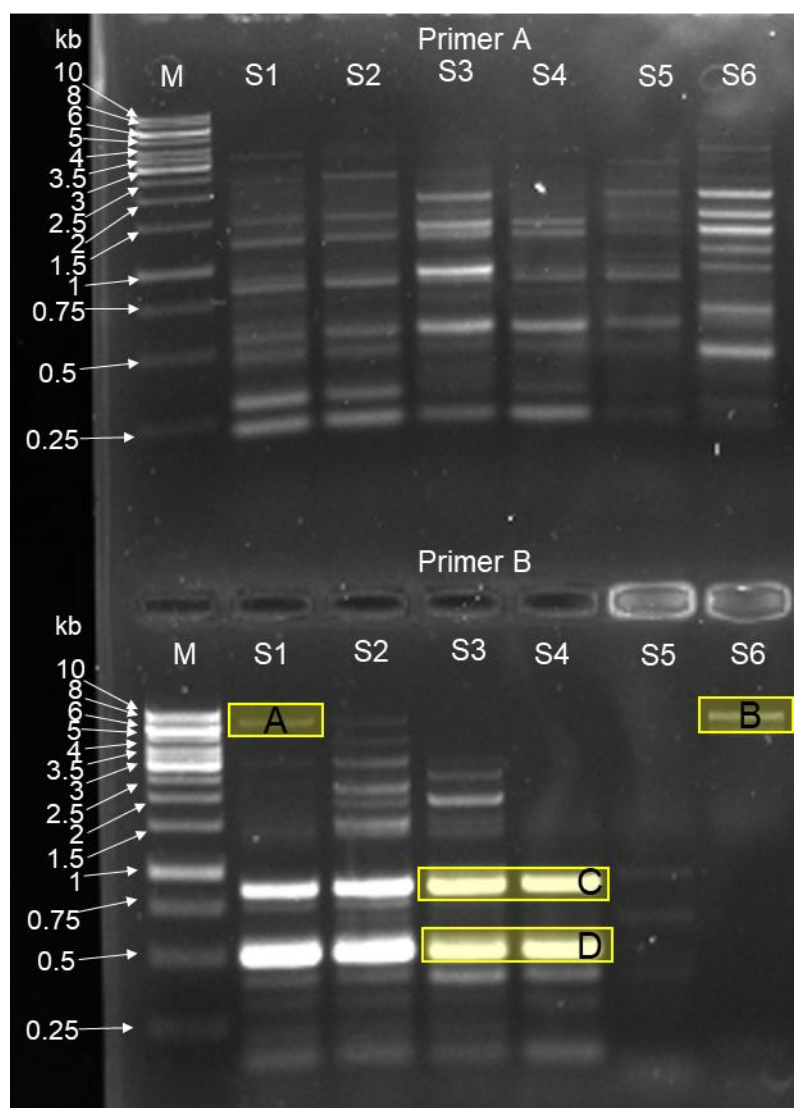


Figure 7.3: Comparison of two primers (A and B) for the amplification of *Heterodera schachtii* vitellogenin containing genomic DNA. M: DNA ladder (kilobase pairs). S1 – S6 represent different lysis conditions during sample preparation, described in section 7.2. Bands excised for further DNA amplification are highlighted (A-D).

7.3.2 Amplification of Genomic DNA

DNA bands were extracted from the agarose gel and purified to obtain DNA from bands labelled A, B, C and D in Figure 7.3. The DNA obtained from these bands had a low concentration (0.6- 3.7 ng/ μ L), therefore they were subject to a series of PCR reactions to increase the DNA concentration, using primer pair B (not shown). Figure 7.4 shows the resulting gel electrophoresis from a series of amplifications of DNA extracted from band A in Figure 7.3. The final reaction resulted in three DNA bands (E, F and G) with differing amplicon sizes of approximately 6900 bp, 6800 bp and 5600 bp, respectively. Purified DNA from these bands had a concentration of 7.4 ng/ μ L, 3.9 ng/ μ L and 7.6 ng/ μ L, with 260/280 ratios of 2.13- 2.43.

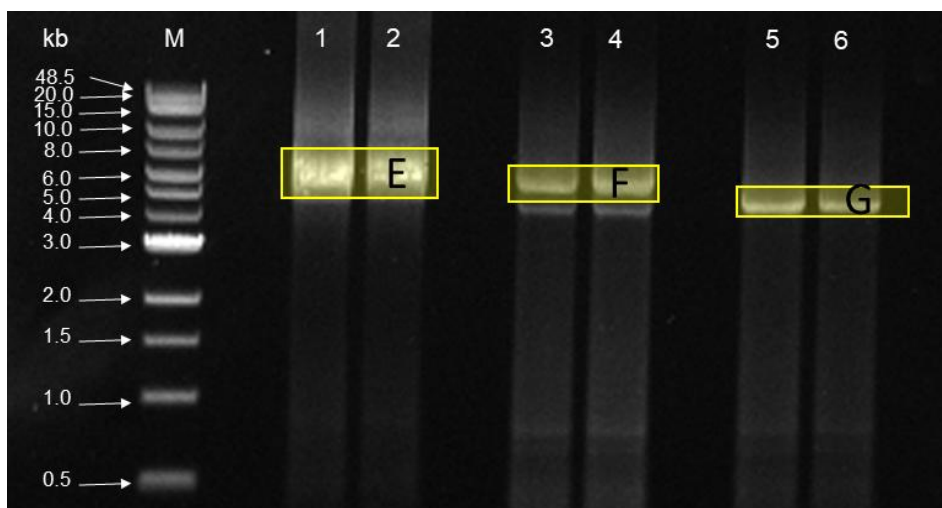


Figure 7.4: Amplification of *Heterodera schachtii* genomic DNA containing the vitellogenin gene encoding region, primers used were specific to this region. Samples 1-6 were obtained from DNA extracted from agarose gel slice of band A in Figure 7.3. M: DNA ladder (kilobase pairs). Bands excised for DNA sequencing are highlighted (E, F and G).

Next, purified DNA extracted from band B in Figure 7.3 was amplified in a PCR reaction. In Figure 7.5, different volumes of DNA were applied to the agarose gel, either 0.5 μL , 1 μL or 2 μL . Each volume of DNA used resulted in a DNA band recognised at approximately 4500 bp. DNA bands labelled H and I were extracted from the gel and purified, giving a DNA concentration of 3.5 ng/ μL and 6.1 ng/ μL , with 260/280 ratio of 2.14 and 1.93, respectively.

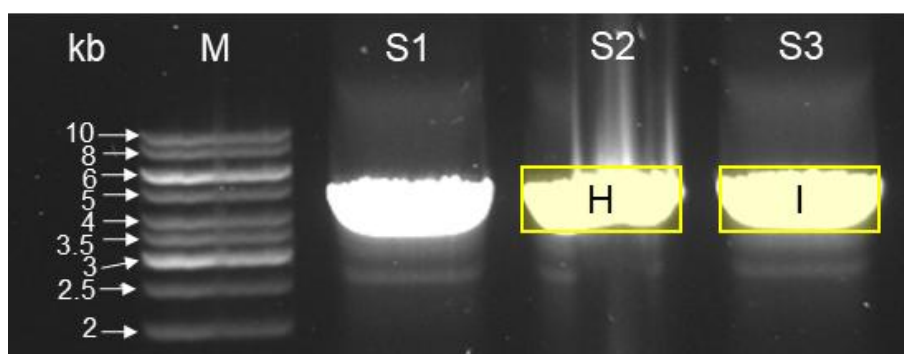


Figure 7.5: Amplification of *Heterodera schachtii* genomic DNA using primers specific to sequences surrounding the vitellogenin gene encoding region. The DNA sample used in the PCR reaction was obtained from agarose gel slice of band B in Figure 7.3. Samples S1-S3 represent different volumes of PCR product used in gel electrophoresis; 0.5 μL , 1 μL and 2 μL . M: DNA ladder (kilobase pairs). Bands excised for DNA sequencing are highlighted (H and I).

Finally, DNA was purified from bands C and D in Figure 7.3 and re-amplified in a PCR reaction. This resulted in clear bands with no non-specific binding, labelled J and K in Figure 7.6. The band labelled J had an amplicon size of approximately 700 bp and band K had an approximate band size of 400 bp. DNA was purified from the agarose gel with sample concentrations of bands J and K of 35.2 ng/ μ L and 21.1 ng/ μ L, respectively. The 260/280 ratio of band J was 1.85 and band K was 1.84.

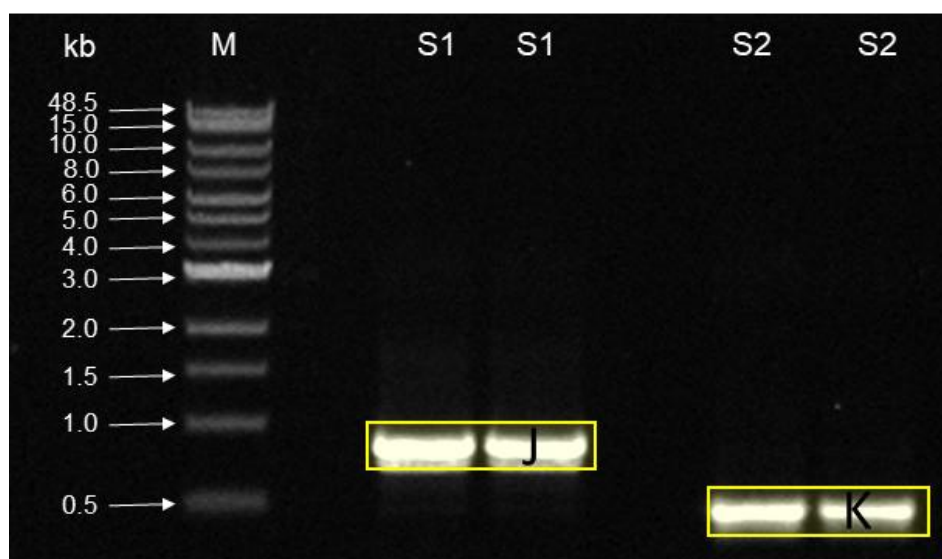


Figure 7.6: Amplification of *Heterodera schachtii* DNA isolated from agarose gel. Primers used were specific for the *H. schachtii* genomic DNA containing the vitellogenin gene sequence. M: DNA ladder (kilobase pairs). S1 and S2 were samples of DNA extracted from gel extracts of bands C and D highlighted in Figure 7.3. Bands excised for DNA sequencing are highlighted (J and K).

7.3.3 Cloning of Genomic DNA

Purified DNA samples extracted from highlighted bands E, F and G in Figure 7.4, and DNA from bands highlighted J and K in Figure 7.6, were ligated into the pJET 1.2 blunt plasmid vector. DH10B *E. coli* competent cells were transformed to uptake this plasmid. After incubation, each sample was inoculated onto lysogeny broth agar supplemented with ampicillin using three different volumes, 200 μ L and 20 μ L and at a 1:10 dilution of the broth, to give several dilutions of the cells. The total number of colonies were calculated and are given in Table 7.5. No colonies were visible from samples E- G; therefore, it is possible that these bacteria did not contain the target DNA. By contrast, samples J and K resulted in visible colonies; however, even with a higher inoculated volume the total number of colonies was low. The presence of the plasmid within these cells was confirmed by colony PCR (Figure 7.7).

Table 7.5: *Heterodera schachtii* genomic DNA was ligated into a plasmid vector (pJET 1.2). Transformed DH10B *Escherichia coli* competent cells were inoculated onto lysogeny broth agar supplemented with ampicillin, using different volumes (200 μ L, 20 μ L and 20 μ L of a 1:10 dilution). Total number of bacterial colony forming units were counted. DNA samples were obtained by purified DNA from gel bands in Figure 7.4 and Figure 7.6.

DNA sample	Volume inoculated (μ L)	Total colony count
E	200	0
E	20	0
E 1:10 Dilution	20	0
F	200	0
F	20	0
F 1:10 Dilution	20	0
G	200	0
G	20	0
G 1:10 Dilution	20	0
J	200	2
J	20	0
J 1:10 Dilution	20	0
K	200	19
K	20	2
K 1:10 Dilution	20	0
Control	200	7
Control	20	0
Control 1:10 Dilution	20	0

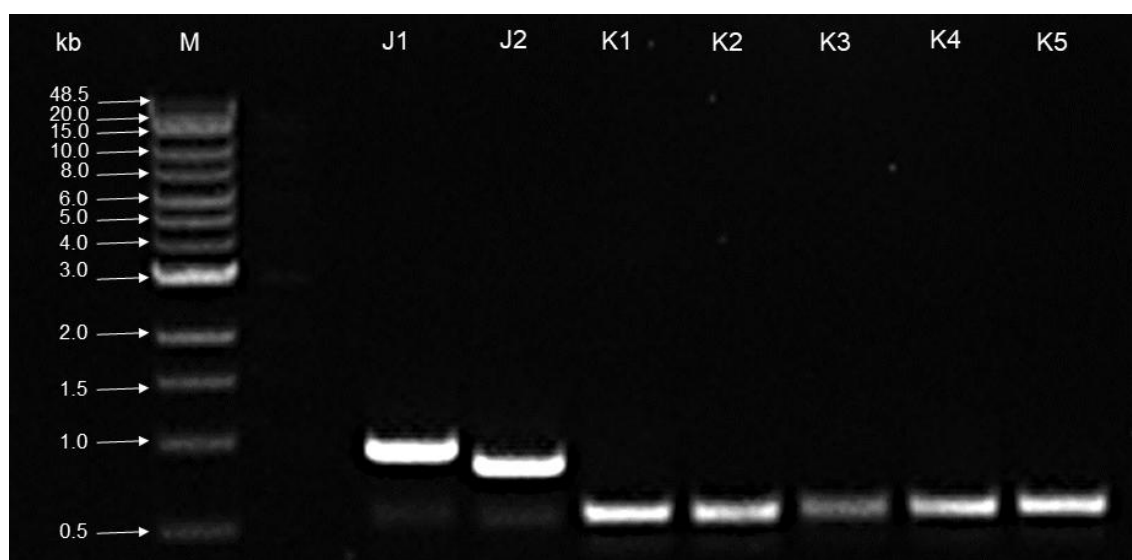


Figure 7.7: Colony PCR to determine the successful ligation and transformation of *Heterodera schachtii* genomic DNA into the pJET 1.2 plasmid construct. Primers used were provided with the CloneJET PCR cloning kit. M: DNA ladder (kilobase pairs). Samples J1- J2 represent amplification from two colonies, whereas K1- K5 were five colonies selected from the bacterial transformation.

The colony PCR presented in Figure 7.7 illustrates that the pJET 1.2 blunt plasmid was present within the bacterial colonies. The plasmid and insert from DNA sample J gave two colonies that were able to be amplified by PCR, with amplicon sizes of approximately 985 bp for J1 and 878 bp for J2. Also, the plasmid and the insert of DNA sample K resulted in an approximate amplicon size of 556 bp, 595 bp, 623 bp, 595 bp and 623 bp.

Due to the ligation and transformation of E, F and G being unsuccessful, this reaction was repeated. It is likely that the ligation and transformation reaction failed due to the molar ratio of sample to vector. The recommended protocol of the CloneJET PCR Cloning Kit (Thermo Scientific™) was to use a molar ratio of 3:1; however, for many of the samples used, the concentration of the vector was higher than the sample (Table 7.3). Furthermore, it is also evident that the reaction wasn't successful due to the low total colony count of the control DNA insert (Table 7.5). Due to the low colony count of the control, it is possible that the transformation efficiency was suboptimal; to control for this, plasmid pUC 19 was used as an additional control for transformation efficiency in subsequent experiments.

The ligation and transformation experiment was repeated. DNA extracted and purified from the agarose gel presented in Figure 7.5 was used for ligation and transformation reactions. Also, purified DNA from band E in Figure 7.4 was again used. Total colony counts of transformed DH10B *E. coli* were calculated for each sample (Table 7.6). All samples produced countable colonies on the lysogeny broth agar; however, sample E only produced three colonies, whereas both samples H and I had colony forming unit counts of greater than 200. The control DNA was used to confirm that the ligation and transformation reaction was successful. The plasmid pUC 19 was used as a control to confirm that the transformation reaction had taken place. As both the pUC 19 control and control DNA produced colony forming units the reaction was successful. The presence of the plasmid and ligated DNA product were confirmed by colony PCR (Figure 7.8).

Table 7.6: *Heterodera schachtii* genomic DNA was ligated into a plasmid vector (pJET 1.2 blunt). Transformed DH10B *Escherichia coli* competent cells were inoculated onto lysogeny broth agar supplemented with ampicillin, using different volumes (200 μ L and 20 μ L). Total number of bacterial colony forming units were counted. DNA samples were obtained by purifying DNA from gel bands in Figure 7.3 and Figure 7.4.

DNA sample	Volume inoculated (μ L)	Total colony count
H	200	269
H	20	5
I	200	230
I	20	0
E	200	3
E	20	0
pUC 19 Control	200	>300
pUC 19 Control	20	24
Control DNA	200	>300
Control DNA	20	84

Figure 7.8 presents an agarose gel of a colony PCR of several colonies selected from samples of transformed *E. coli*. Primers used were specific to the pJET 1.2 blunt plasmid to determine the presence of the plasmid in the bacterial colony. Samples H1-H5 all contained the plasmid but had varying length of DNA insert. The amplicon sizes of the plasmid-insert construct were 5.19 kb, 5.53 kb, 5.29 kb, 5.72 kb and 5.53 kb for samples H1, H2, H3, H4 and H5, respectively. The plasmid-insert construct DNA was also amplified from samples I1, I2, I3 and I5, resulting in amplicons that were 5.56 kb, 5.39 kb, 5.34 kb and 5.55 kb, respectively. Sample I4 was not detected, therefore did not contain the plasmid DNA. The plasmid was also detected in samples E1 and E2; however, E1 only had an amplicon size of 0.37 kb, whereas sample E2 amplicon size was 5.92 kb. Finally, the control reaction band size was 1.12 kb. The plasmid pUC 19 control was not amplified because it does not contain the complementary DNA sequence to the pJET 1.2 primers.

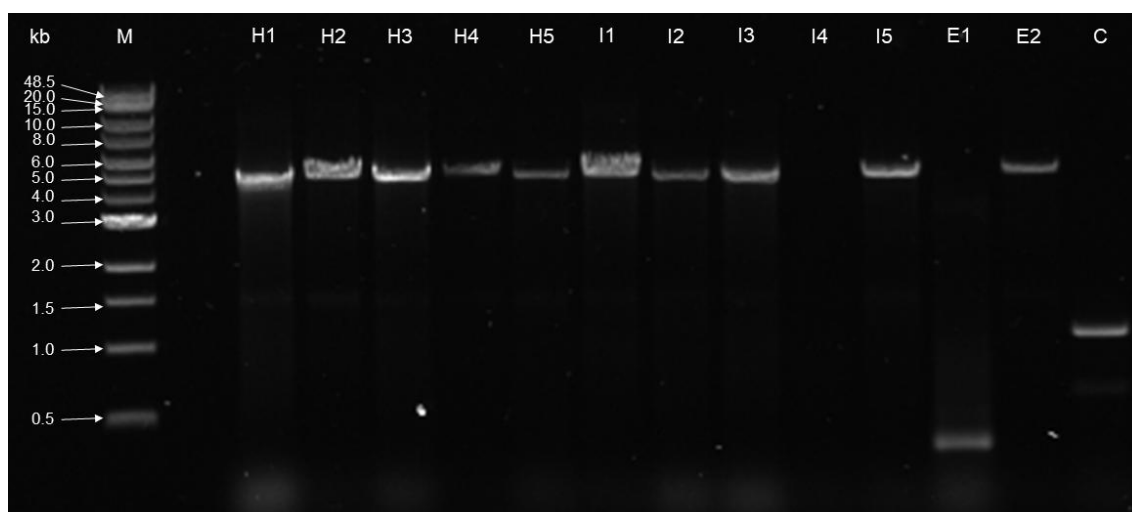


Figure 7.8: Colony PCR to determine the successful ligation and transformation of *Heterodera schachtii* genomic DNA into the pJET 1.2 plasmid construct. Primers used were provided with the CloneJET PCR cloning kit. M: DNA ladder (kilobase pairs). Samples H1- H5 and I1- I5 were ten selected colonies from bacterial transformation using the purified DNA product of bands H and I in Figure 7.5, whereas E1- E2 was obtained from a purified DNA product of band E, in Figure 7.4. Sample C: Control DNA product ligated and transformed into plasmid construct.

7.3.4 Sequencing of Genomic DNA

After the *E. coli* bacterial colonies were transformed and confirmed to contain the plasmid, plasmids were purified, and several samples were selected at various sizes for DNA sequencing. Samples selected from the colony PCR in Figure 7.7 were J1, J2, K1, K2 and K3, whereas from Figure 7.8 samples H1, H2, H3, I1, I2, I3 and E2 were selected for sequencing. These samples were sequenced by Sanger sequencing generating a forward and reverse sequence. All sequences were subject to BLAST searches with the target genome set as *H. schachtii*, to confirm that the target genomic region had been sequenced. Only five samples (H1, H2, H3, I2 and I5) were shown to have sequence homology to the *vit* gene when an expect value of 10 was used in the BLAST search. The resulting E-values were relatively high and may indicate that these alignments were obtained by chance. The sequences were trimmed to remove most of the plasmid sequence and ambiguous regions determined from the chromatograms. The forward and reverse sequences obtained by Sanger sequencing were then aligned to the genomic region of *H. schachtii* containing the *vit* gene Hsc_gene_6726, used as a reference sequence (Supplementary material 7.1 a-f). The alignment position was varied along the length of the reference sequence, for all samples, from both the forward and reverse sequencing.

A limitation of Sanger sequencing was that it will only sequence samples up to approximately 1000 bp. Therefore, to further determine the sequence of the genomic region containing the

vit gene in populations of *H. schachtii*, next generation sequencing was used for sequencing of the whole plasmid construct.

Prior to sequencing samples by next generation sequencing, several of the original colonies selected for colony PCR (H1, H2, H3, I2 and I5) from the ligation and transformation reaction in Table 7.6 were inoculated into lysogeny broth and incubated. Plasmids were again purified from the broth to obtain a greater volume of plasmid DNA from each sample. Several of DNA samples that were sent for Sangar sequencing were also used for next generation sequencing (H2, H3, I2 and I5). Therefore, the originally sequenced samples are now H2a, H3a, I2a and I5a and the newly purified DNA samples are H1, H2b, H3b, I2b and I5b.

After next generation sequencing was completed, sequences were aligned to determine whether there was any variation between the original and newly purified sequences. Each sequence was identical between the a and b counterparts, with a percentage identity of 100%. Therefore, each sample was labelled H1, H2, H3, I2 and I5 to represent the original DNA band purified from the colony PCR in Figure 7.8, in subsequent sequence alignments. Sequences were aligned to the reference sequence of the plasmid and *H. schachtii* DNA insert and pairwise alignment scores were calculated (Table 7.8).

As shown in Table 7.7, samples H1, I2 and I5 contain the same sequence, whereas H2 and H3 also contain the same sequence. All sample sequences only have a percentage identity of 47.85- 48.23% to the reference sequence. Therefore, only half of the sequence from each sample is comparable to the sequence expected to be cloned from *H. schachtii*. A consensus sequence was generated and aligned to the reference sequence (Supplementary material 7.2), to determine whether this gave greater coverage of the sequence. This alignment resulted in a percentage identity of 48.84%. The alignment between the consensus sequence and the reference sequence showed there were no large regions that had not been cloned, instead there were both insertions and deletions along the full length of the generated consensus sequence. A BLAST search of the consensus sequence against the *H. schachtii* genome resulted in two alignments with the Hsc_gene_6726 when the expect value was 100. The genomic location of this alignment was Hsc_scaff007:2222536-2222552 and Hsc_scaff007:2222900-2222920, each with a length of 17 codons.

Table 7.7: Pairwise alignment and percentage identity (%) of samples H1, H2, H3, I2 and I5 compared to each other and the reference sequence of the expected cloned region from *Heterodera schachtii*.

	Reference	H1	H2	H3	I2	I5
Reference	X					
H1	48.23	X				
H2	47.85	49.88	X			
H3	47.86	49.88	99.98	X		
I2	48.23	100	49.86	49.86	X	
I5	48.23	100	49.86	49.86	100	X

7.4 Discussion

The methodology for DNA extraction from *H. schachtii* J2 was compared using either WLB or 10X PCR buffer, with varying concentrations of proteinase K. It was found that there was no great difference in the performance of the buffers but the effect of an increased concentration of proteinase K was more prominent. It has been shown that high concentrations (> 100 µg/mL) of proteinase K can lead to the degradation of DNA (Merck, 2025). The high concentration may have resulted in the degradation of DNA samples in the comparison of extraction methods and the loss of samples with a larger number of base pairs. In initial PCR reactions and gel electrophoresis, there were the presence of bands not at the expected amplicon size. This may have been due to different factors such as annealing temperature or the specificity of the primers. This was improved by isolating and purifying DNA from bands closer to the target amplicon size.

The cloning of the *H. schachtii* J2 DNA into a plasmid construct was successful, with several of the plasmid constructs containing a portion of the target region for amplification. The presence of these inserts was confirmed by colony PCR and once sequenced indicated the presence of *H. schachtii* genomic DNA containing the Hsc_gene_6726 gene, through BLAST searches. The plasmid constructs were able to be pooled into two sets that contained the same DNA sequences. These sets of plasmids were only 50% homologous to each other and the reference sequence; therefore, each sequenced sample only represented part of the target sequence. It is possible that there may have been some variation in the DNA sequence of populations used in this study when compared to the reference sequence. Furthermore, the amplified sequences in PCR reactions were smaller than the anticipated amplicon size, therefore it can be anticipated that the sequenced product would not be the same as the reference sequence.

Two different sequencing methods were used to determine the sequence of the *H. schachtii* DNA and plasmid construct. Sanger sequencing was initially used due to its high accuracy, reliability and cost-effectiveness (Al-Shuhaib and Hashim, 2023); however, due to the short sequencing length of a maximum of 1000 bp (Crossley *et al.*, 2020), other sequencing methodologies were considered. Therefore, the *H. schachtii* DNA- plasmid construct was instead sequenced by a whole plasmid sequencing service using Oxford Nanopore methodology, which offered longer reads of greater than 500 bp (Petersen *et al.*, 2019). Several DNA sequences derived from both methodologies were homologous to a reference sequence for the genomic DNA and plasmid construct. In contrast to Sanger sequencing, the use of Oxford Nanopore methodology resulted in a longer sequencing length and improved sequence analysis.

Multiple bands were observed in the agarose gels when using primer pairs, A, B and C. It is possible that this is due to the initial design of the primers. When using Oligo-Evaluator™, the A reverse primer was a self-dimer, whereas primer pair B and C forward primer were predicted to produce secondary structures. Therefore, these features may have influenced the output of the PCR reaction and resulting agarose gel. Furthermore, by utilising a semi-nested PCR approach, any non-specific amplicons would have been carried forward to subsequent reactions by using the same primer pairs in each reaction. Also, the potential non-specific binding and primer dimers may have been the cause of low yield of DNA from extracted bands, due to less primer availability to bind to the target sequence. Additionally, the Sanger sequencing results may have been influenced by the PCR reaction; if the primers were amplifying non-specific regions or forming dimers, this may have resulted in high background noise or a mixed sequence trace. Both sequencing methods, Sanger sequencing and whole plasmid sequencing, resulted in some regions of the target DNA being amplified that corresponded to the reference sequence; however, with further optimisation the full DNA sequence could be derived.

The region of the *H. schachtii* genome amplified may have been complex, making it more difficult to amplify this region in PCR reactions. The use of a positive control may have assisted in the optimisation of the PCR reactions. One possible positive control could be generated by designing a primer pair to amplify a larger region of *H. schachtii* genomic DNA, containing the original target DNA that includes the Hsc_gene_6726 gene. This amplified region could then be cloned to produce a recombinant plasmid containing the target DNA. The purified plasmid DNA could then be utilised as a positive control for subsequent reactions.

The amplification and cloning of *H. schachtii* genomic DNA containing the Hsc_gene_6726 gene provides an opportunity for the further investigation of the *vit* gene in *H. schachtii*. Future

research could utilise these DNA sequences and clone them into expression vectors for the amplification of the encoding gene. The expressed product could be used to develop assays to further understand the gene function in *H. schachtii*. Furthermore, the specificity of these primers could be tested against other species of *Heterodera* spp. or cyst nematodes. Finally, further research could also utilise these primers against geographically diverse populations of *H. schachtii* to determine variance in vitellogenin DNA across different populations.

In conclusion, primers were developed and used for the successful amplification of *H. schachtii* genomic DNA. Subsequently, the genomic DNA containing the Hsc_gene_6726 gene was cloned into a plasmid construct and sequenced. The analysis of these DNA sequences gave confirmation that the target DNA sequence had been amplified and cloned. Therefore, the work presented here provides a molecular basis for further investigation of vitellogenin in *H. schachtii* and other cyst nematodes.

8 General Discussion

Plant-parasitic nematodes (PPN) are oviparous animals and the developing nematode will need to acquire nutrients from within the egg during embryogenesis. Understanding the proteins present during this important stage of nematode development could provide further insights into their biology and a potential route for control prior to nematode hatch. It has been shown that vitellogenin is an important protein for juvenile development within the egg, with yolk provision, lipid transportation (Hayward *et al.*, 2010; Winter *et al.*, 1996) and providing protection to the nematode (Fischer *et al.*, 2015; Ishii *et al.*, 2002). Whilst a substantial amount of vitellogenin research has been conducted in other animals and insects (Carducci *et al.*, 2019; Hagedorn and Kunkel, 1979), limited research has focused on the identification and characterisation of vitellogenin in many species of cyst, root-knot (RKN) and root-lesion nematodes (RLN).

The aim of the present research was to investigate the presence of, and to characterise, vitellogenin in the beet cyst nematode, *Heterodera schachtii*, with some focus on other economically important cyst nematodes, RKN and RLN. This research initially took a bioinformatic approach and determined the presence of *vit* genes within species of cyst, RKN and RLN, and examined the phylogenetic relationship between them (Chapter 3). Next, again utilising bioinformatic methodology, PPN *vit* gene expression and the life-stage specificity of the expression of these genes within PPN was examined (Chapter 4). Furthermore, antibodies against a vitellogenin sequence were developed, validated and tested against *H. schachtii*, *G. pallida* and *G. rostochiensis* to determine the molecular weight of the protein targets in these species, that may be vitellogenin (Chapter 5). Additionally, to immunolocalise the targets of these antibodies in cyst nematode species, preliminary immunofluorescent investigations were performed (Chapter 6). Also, to advance future molecular study of vitellogenin in cyst nematodes, novel primers were developed and the vitellogenin DNA sequence of populations of *H. schachtii* was determined (Chapter 7). Finally, diagnostic antibodies (MAC 356 and PC320) to egg protein targets were investigated and their immunolocalisation in PCN species were compared in an initial study (Chapter 6).

8.1 PPN Genomes Contain Homologous *vit* Genes

The bioinformatic searches conducted highlighted the presence of *vit* genes in the genomes of cyst nematodes, RKN and RLN; as well as predicting the gene sequences in both *Globodera ellingtonae* and *Pratylenchus scribneri*, that had not been previously determined. It was shown that there were conserved sequences between the PPN species and both *Caenorhabditis elegans* and *C. briggsae*. The sequence conservation between these groups

supports the hypothesis that these genes and proteins are providing a similar function in PPN species. Also, due to the moderate level of sequence conservation between the *vit* genes in these species, there may be a stronger selection pressure on these genes and encoding proteins.

In phylogenetic study, the *vit* genes were largely clustered by their groups (cyst nematode, RKN, RLN or *Caenorhabditis* spp.) but were not clustered by species within these groups. It can be hypothesised that these divergent genes are required to encode different subunits for the formation of protein complexes, as observed in *C. elegans* (Sharrock *et al.*, 1990). Additionally, the PPN *vit* genes contained the same protein domains with similar predicted functions as *C. elegans*, these were lipid transporter activity and lipid transport. This function may be derived from the activity of the Vitellogenin_N domain, which has been predicted to bind lipids (Zhai *et al.*, 2024). Many of the PPN genes had a gene ontology annotation of nutrient reserve; this role has also been hypothesised for *C. elegans vit* genes by Winter *et al.* (1996) and Perez and Lehner (2019).

The work presented here has predicted and examined the *vit* genes of several economically important PPN and illustrated the phylogenetic relationship between them. The relative position of each of the protein domains present in each gene has also been understood and compared. Further research could utilise the predicted gene sequences as a basis for molecular study of vitellogenin. Furthermore, the number of sequenced PPN genomes is increasing, which provides a great opportunity for further investigations of *vit* genes in a wider group of species.

8.2 Vitellogenin Genes are Expressed in PPN Species

Searches were conducted to obtain the *vit* gene expression data for multiple PPN species, *C. elegans* and *C. briggsae*. It was shown that *vit* genes were most highly expressed at the adult female stage of the nematode's lifecycle and was significantly greater than earlier stages, such as the second-stage juvenile (J2) and the egg. It has been previously shown that *vit* gene expression is high in female life stages of *C. elegans* (van Nostrand *et al.*, 2013) and *G. pallida* (Cotton *et al.*, 2014) but has now been illustrated here for other species of PPN. A great importance is placed on yolk provision during the adult female stage of the lifecycle. In *C. elegans* young adult hermaphrodites, synthesis of the vitellogenin proteins comprises 25% of all *de novo* protein synthesis in the intestine (Kimble and Sharrock, 1983). In contrast to previous research, it was shown that *vit* genes are also expressed at low levels in J2 and male stages. This was supported by the recognition of vitellogenin proteins in *H. schachtii* J2 by western blot analysis in Chapter 5.

Whilst the *vit* gene expression from several PPN species was able to be investigated, the study was reliant on expression data available in the public domain. Therefore, there was some variation in how the data was collected and analysed. In future studies, *vit* gene expression across several species could be completed as part of the same study under the same conditions, to allow better comparison of results. These investigations of *vit* gene expression could also include the species *G. ellingtonae*, *Meloidogyne hapla*, *M. javanica* and *P. scribneri* or other PPN species not investigated here, as these data are currently not available in the public domain.

8.3 Vitellogenin Proteins are Likely Present in Females and Eggs of Cyst Nematodes

The molecular weights of proteins from cyst nematodes were obtained by western blot analysis, using an antibody that was designed to target vitellogenin. Initially, to investigate these proteins in *H. schachtii*, *G. pallida* and *G. rostochiensis* two antibodies targeting a sequence within a vitellogenin gene were developed and tested. One of these antibodies was successful in the recognition of the cyst nematode homogenates. The calculated molecular weight of *H. schachtii* female cyst proteins recognised by the anti-vitellogenin antibody were 146 kDa, 94 kDa, 89 kDa, 47 kDa, 34 kDa and 19 kDa. Also, the molecular weight of the *G. pallida* proteins recognised were 157 kDa, 107 kDa, 81 kDa, 55 kDa, 52 kDa and 46 kDa. Finally, the five proteins recognised from *G. rostochiensis* cysts had molecular weights of 110 kDa, 82 kDa, 77 kDa, 47 kDa and 13 kDa. These molecular weights were comparable to those found in other species, including *Heterodera glycines* (Masler, 1999), *C. elegans* (Perez and Lehner, 2019) and *Dolichorhadditis* spp. (Winter, 1992). The antibody recognised proteins in the females, eggs and J2 of *H. schachtii*, with proteins recognised that were specific to each stage. These results agree with the findings of gene expression of *H. schachtii* in Chapter 4, where *vit* genes were shown to be expressed in all three of these life stages. In previous studies, the presence of vitellogenin proteins in J2 has provided contradictory results. In mass spectrometry analysis of *M. hapla* (Mbeunkui *et al.*, 2010) vitellogenin proteins were detected in J2 samples, whereas when anti-vitellogenin antibodies were used against *H. glycines* J2 samples, no protein was recognised (Masler, 2002).

Future investigations could utilise these anti-vitellogenin antibodies to further study egg proteins in cyst nematodes. It may be possible to isolate and purify this protein to understand its structure and continue to determine its function. Furthermore, this work could be expanded to investigate the vitellogenin protein in other species, such as RKN or RLN. Whilst the results of the western blot analysis were convincing, further work may be required to confirm the identity of all the proteins recognised by the anti-vitellogenin antibody.

8.4 Preliminary Immunolocalisation of Vitellogenin Proteins in Cyst Nematodes

In preliminary investigations of the immunolocalisation of vitellogenin proteins in *H. schachtii*, *G. pallida* and *G. rostochiensis*, the anti-vitellogenin antibody was able to detect proteins in different stages of the nematode including the female, egg and J2. This is consistent with western blot analysis results that were presented in Chapter 5. Fluorescence was observed in the eggs of *H. schachtii* and *G. rostochiensis*, which is comparable to findings in *C. elegans* where fluorescence was measured within the eggs using an anti-YP170 antibody (Sharrock, 1983). Surprisingly, fluorescence of proteins were visualised in alternate segments along the head of the *H. schachtii* J2. The anti-vitellogenin antibody may have recognised seam cells present along the nematode cuticle; however, this requires further investigation. Furthermore, this work has provided a methodology that does not rely on a formaldehyde fixative; therefore, it is a safer and useful basis for future immunofluorescent study of antibody targets in PPN.

These results provide a preliminary investigation of the immunolocalisation of the vitellogenin proteins in *H. schachtii*, *G. pallida* and *G. rostochiensis*; however, further work is required to fully determine where these proteins are situated within the tissues of these nematodes, at different stages of the life cycle. This work could also be developed by utilising higher magnifications to determine localisation of vitellogenin proteins at a tissue and cellular level.

8.5 Amplification of *Heterodera schachtii* Genomic DNA containing the *vit* Gene

Molecular methods were utilised to determine the genomic DNA sequence of a region of the *H. schachtii* genome containing the *vit* gene, Hsc_gene_6726. Initially, novel primers were designed to this region and were used to amplify DNA from samples of *H. schachtii* J2. The cloning and sequencing of the amplified PCR product enabled a DNA sequence of the genomic region to be generated. The assembled plasmid constructs could be used for the development of an expression vector to further elucidate the function of *vit* genes in cyst nematodes; however, further work is required to ensure the specificity of these primers. In future work, newly developed primers could be utilised in research of this region of the *H. schachtii* genome. The specificity of these primers could be tested to determine whether they are species-specific and used to investigate vitellogenin in different populations of *H. schachtii*. Alternatively, if these primers lack species-specificity, they could also be used to investigate vitellogenin in other species.

8.6 Antibodies MAC 356 and PC320 Target the Same Protein, Annexin

The diagnostic antibodies MAC 356 and PC320 have previously been independently studied and were shown to recognise proteins within the eggs of *G. pallida* and *G. rostochiensis*. The original target of MAC 356 was a 34 kDa protein from *G. pallida* (Robinson, 1993). This protein was also found in *G. rostochiensis* but had a differential isoelectric focusing point (pI) of 5.7, whereas the *G. pallida* pI was 5.9, as recognised by Fleming and Marks (1983). Separately, the polyclonal PC320 antibody was raised to detect the same target (Curtis *et al.*, 1996). Work by Fioretti *et al.* (2001) was able to use PC320 to identify the cDNA sequence of the protein target as annexin. Although these antibodies were generated against the same target, the identity of the protein recognised by the monoclonal antibody MAC 356 had not been confirmed.

The work presented here was able to confirm using western blot analysis that both antibodies were recognising a protein of the same size, when using the same samples. Furthermore, using mass spectrometry analysis of these protein bands, it was confirmed that the protein recognised in both *G. pallida* and *G. rostochiensis* was likely to be annexin. Intriguingly, a protein of the same size in *H. schachtii* was also detected in western blot analysis. Unfortunately, the protein identity was not able to be confirmed by mass spectrometry; however, it can be hypothesised that this protein was also annexin. This finding is contrary to previous studies, which were not able to show the recognition of *H. schachtii* samples by PC320 (Curtis *et al.*, 1996).

The monoclonal antibody, MAC 356 was shown to be a promising candidate as a diagnostic marker for the differentiation of potato cyst nematode (PCN) species; however, as shown by Robinson *et al.* (1993), there was some cross-reactivity of the antibody between these species. The cross-reactivity of this antibody was also demonstrated in findings presented in Chapter 6. By understanding the identity of the protein targeted by the antibody, differences in the amino acid sequences of *G. pallida* and *G. rostochiensis* annexin could be exploited to further the development of a diagnostic marker.

8.7 Future Study Recommendations

Progress has been made towards the characterisation and understanding of functionality of vitellogenin in PPN, particularly the cyst nematodes. To develop a full understanding of vitellogenin in cyst nematodes, a study of the functionality of these proteins would be required. These proteins could be further examined by utilising structural analysis techniques; for example, the use of X-ray crystallography or nuclear magnetic resonance spectroscopy to determine the three-dimensional structure of these proteins (Alberts *et al.*, 2002). Alternatively, with advances in structural modelling using programs such as AlphaFold (Jumper *et al.*, 2021),

using the gene sequences obtained here, the three-dimensional structure of these proteins could be predicted. By understanding the structure of the protein, the functional sites can be evaluated and further infer the biological functionality of the protein within the nematode. Additionally, one of the hypothesised functions of vitellogenin is lipid transport. This could be assessed by performing a protein-lipid binding study.

Advances in vitellogenin research of cyst nematodes could be made by performing gene knockdown or knockout experiments, by using RNA interference (RNAi). The inhibition of *vit* gene expression could demonstrate phenotypes that further elucidate the function of the proteins. In *C. elegans* studies, RNAi has been used to investigate the role of different *vit* genes and the impact of their knockdown. It was shown that the use of *vit-5* and *vit-6* RNAi prevented the accumulation of vitellogenin. The use of *vit-5* RNAi stopped the accumulation of YP170, whereas *vit-6* RNAi inhibited YP115 and YP88 accumulation. Also, *vit-5* RNAi was shown to decrease intestinal atrophy and increased lifespan (Sornda *et al.*, 2019). Furthermore, when the gene encoding the vitellogenin receptor, *rme-2*, was knocked down, vitellogenin accumulated in the intestine (Ezcurra *et al.*, 2018). Also, it has been suggested that vitellogenin is required for normal development. In a study that used *vit-3* RNAi, there was a slower development rate of the progeny and increased eggshell permeability (Min *et al.*, 2020). Finally, RNAi was used to investigate the role of vitellogenin in response to bacterial infections of *C. elegans*. In this study, the knockdown of *vit-2*, *vit-3*, *vit-5* and *vit-6* reduced the lifespan of the nematode. Also, when *vit-6* RNAi was used, there was an increase in reactive oxygen species levels (Fischer *et al.*, 2013). Similar studies could be conducted in PPN to further investigate the functionality of vitellogenin.

RNAi is already a powerful tool in PPN research. Some studies are deploying this technique to advance the understanding of functionality of different proteins and mediate resistance in the host plant. Many studies have focused on the knockdown of multiple gene targets in *M. incognita* (Niu *et al.*, 2012; Iqbal *et al.*, 2020; Hada *et al.*, 2021), with some of the silenced genes involved in parasitism (Rosso *et al.*, 2007; Dubreuil *et al.*, 2007; Shivakumara *et al.*, 2016, 2019). Furthermore, RNAi was used to understand the gene functionality of four chemosensory genes in *M. graminicola* (Wang *et al.*, 2024). The application of RNAi as a control method of RLN species has also been investigated (Tan *et al.*, 2013). Finally, studies have also utilised RNAi techniques for the application of host-induced gene silencing (Papolu *et al.*, 2013; Chaudhary *et al.*, 2019; Zhang *et al.*, 2022).

8.8 Conclusion

This work has established the presence of vitellogenin in *H. schachtii* and characterised aspects of the encoding genes and proteins, with some focus on other economically important cyst nematodes. This work was able to show the phylogenetic relationship between PPN *vit* genes and compared those to *Caenorhabditis* spp., as well as determining the expression of *vit* genes in these species. The development of an antibody to a vitellogenin sequence enabled the determination of the molecular weight of proteins in *H. schachtii*, *G. pallida* and *G. rostochiensis*, that were likely to be vitellogenin; however, this requires confirmation. This antibody was also used in a preliminary investigation of the immunolocalisation of the target protein, demonstrating that the protein was present within the eggs of *H. schachtii* and *G. rostochiensis*, whilst providing a basis for further immunofluorescent study. Additionally, the DNA sequence of the *H. schachtii vit* gene Hsc_gene_6726 was also assessed, to aid in future molecular study of this gene. Finally, the identity of the protein target of the monoclonal antibody MAC 356 was confirmed to be annexin, the same target as antibody PC320. Therefore, this work provides a novel insight into the vitellogenin protein and *vit* genes of *H. schachtii* and other PPN species and forms a basis for further study of the function of vitellogenin in PPN.

References

- Abad, P., & Williamson, V. M. (2010). Plant nematode Interaction: a sophisticated dialogue. In *Advances in botanical research* (pp. 147–192). [https://doi.org/10.1016/s0065-2296\(10\)53005-2](https://doi.org/10.1016/s0065-2296(10)53005-2)
- Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M. C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., ... Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, 26(8), 909–915. <https://doi.org/10.1038/nbt.1482>
- Agrios, G. N. (2005). Plant diseases caused by nematodes. In *Plant pathology* (pp. 825–874). Elsevier. <https://doi.org/10.1016/b978-0-08-047378-9.50021-x>
- AHDB. (n.d.). *Weeds and volunteer control in potatoes*. AHDB Horticulture & Potatoes. Retrieved July 20, 2025, from <https://horticulture.ahdb.org.uk/knowledge-library/weeds-and-volunteer-control-in-potatoes>
- AHDB. (2019). *Biofumigation for management of potato cyst nematodes*. <https://ahdb.org.uk/knowledge-library/biofumigation-for-management-of-potato-cyst-nematodes-pcn>
- AHDB. (2024). *Potato cyst nematodes (PCN) and their management in potatoes*. <https://archive.ahdb.org.uk/knowledge-library/potato-cyst-nematodes-pcn-and-their-management-in-potatoes>
- Ahuja, A., Joshi, V., Singh, G., Kundu, A., Bhat, C. G., Kumar, S., Rao, U., & Somvanshi, V. S. (2021). Rapid and sensitive detection of potato cyst nematode *Globodera rostochiensis* by loop-mediated isothermal amplification assay. *3 Biotech*, 11(6), 294. <https://doi.org/10.1007/s13205-021-02830-8>
- Ahuja, A., & Somvanshi, V. S. (2021). Diagnosis of plant-parasitic nematodes using loop-mediated isothermal amplification (LAMP): A review. *Crop Protection*, 147, 105459.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Analyzing protein structure and function*. Molecular Biology of the Cell - NCBI Bookshelf. <https://www.ncbi.nlm.nih.gov/books/NBK26820/>
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *B-cells and antibodies*. Molecular Biology of the Cell - NCBI Bookshelf. <https://www.ncbi.nlm.nih.gov/books/NBK26820/>
- Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Van Auken, K., ... Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1). <https://doi.org/10.1093/genetics/iyad031>

- 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(1), e53460. <https://doi.org/10.1371/journal.pone.0053460>
- Al-Shuhaib, M. B. S., & Hashim, H. O. (2023). Mastering DNA chromatogram analysis in Sanger sequencing for reliable clinical analysis. *Journal of Genetic Engineering & Biotechnology*, 21(1), 115. <https://doi.org/10.1186/s43141-023-00587-6>
- Álvarez-Ortega, S., Brito, J. A., & Subbotin, S. A. (2019). Multigene phylogeny of root-knot nematodes and molecular characterization of *Meloidogyne nataliei* Golden, Rose & Bird, 1981 (Nematoda: Tylenchida). *Scientific Reports*, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-48195-0>
- Amiri, S., Subbotin, S. A., & Moens, M. (2002). Identification of the beet cyst nematode *Heterodera schachtii* by PCR. *European Journal of Plant Pathology*, 108(6), 497–506. <https://doi.org/10.1023/A:1019974101225>
- Animal & Plant Health Agency. (2025). *Explanatory guide to the Seed Potato Classification Scheme and Approved Stock Scheme 2025 to 2026* (SPCS & AS Explanatory Guide). GOV.UK. https://assets.publishing.service.gov.uk/media/681386c4b0ef2c9850525450/SPCS_and_AS_explanatory_guide_2025_to_2026.pdf#page=11.87
- Archidona-Yuste, A., Castillo, P., & Palomares-Rius, J. E. (2024). Migratory endoparasitic nematodes. In R. N. Perry, M. Moens, & J. T. Jones (Eds.), *Plant nematology* (3rd ed., pp. 176–214). CAB International.
- Arora, D., Hernandez, A. G., Walden, K. K. O., Fields, C. J., & Yan, G. (2023). First draft genome assembly of root-lesion nematode *Pratylenchus scribneri* generated using long-read sequencing. *International Journal of Molecular Sciences*, 24(8), 7311. <https://doi.org/10.3390/ijms24087311>
- Arora, D., Yan, G., & Baidoo, R. (2020). Developing a real-time PCR assay for direct detection and quantification of *Pratylenchus scribneri* in field soil. *Nematology*, 22(7), 733–744. <https://doi.org/10.1163/15685411-00003336>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556>
- Back, M. A., Cortada, L., Grove, I. G., & Taylor, V. (2018). Field management and control strategies. In R. N. Perry & M. Moens (Eds.), *Cyst nematodes* (pp. 305–327). CAB International.
- Baker, H. J., & Shapiro, D. J. (1978). Rapid accumulation of vitellogenin messenger RNA during secondary estrogen stimulation of *Xenopus laevis*. *Journal of Biological Chemistry*, 253(13), 4521–4524. [https://doi.org/10.1016/S0021-9258\(17\)30418-0](https://doi.org/10.1016/S0021-9258(17)30418-0)
- Baker, M. E. (1988a). Invertebrate vitellogenin is homologous to human von Willebrand factor. *Biochemical Journal*, 256(3), 1059–1061. <https://doi.org/10.1042/BJ2561059>

- Baker, M. E. (1988b). Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase? *Biochemical Journal*, 255(3), 1057–1060.
- Bakhetia, M., Urwin, P. E., & Atkinson, H. J. (2007). qPCR analysis and RNAi define pharyngeal gland cell-expressed genes of *Heterodera glycines* required for initial interactions with the host. *Molecular Plant-Microbe Interactions*, 20(3), 306–312. <https://doi.org/10.1094/MPMI-20-3-0306>
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C., & Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Research*, 32(suppl_1), D138–D141. <https://doi.org/10.1093/nar/gkh121>
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., ... Zhang, J. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>
- Bélair, G., Dauphinais, N., Benoit, D. L., & Fournier, Y. (2007). Reproduction of *Pratylenchus penetrans* on 24 common weeds in potato fields in Québec. *Journal of Nematology*, 39(4), 321–326.
- Bellaïf, S., Shen, Z., Rosso, M. N., Abad, P., Shih, P., & Briggs, S. P. (2008). Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLOS Pathogens*, 4(10), e1000192. <https://doi.org/10.1371/journal.ppat.1000192>
- Bergink, E. W., & Wallace, R. A. (1974). Precursor-product relationship between amphibian vitellogenin and the yolk proteins, lipovitellin and phosvitin. *The Journal of Biological Chemistry*, 249(9), 2897–2903. [https://doi.org/10.1016/S0021-9258\(19\)42715-4](https://doi.org/10.1016/S0021-9258(19)42715-4)
- Berry, S., Fargette, M., Morand, S., & Cadet, P. (2007). Reliability of PCR-based techniques for detection and discrimination of plant-parasitic nematodes of sugarcane. *Nematology*, 9(4), 499–514. <https://doi.org/10.1163/156854107781487297>
- Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., & Apweiler, R. (2009). QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics*, 25(22), 3045–3046. <https://doi.org/10.1093/bioinformatics/btp536>
- Bird, A., & Bird, J. (2012). *The structure of nematodes* (2nd ed.). Academic Press.
- Bird, D. M. K., Opperman, C. H., & Davies, K. G. (2003). Interactions between bacteria and plant-parasitic nematodes: now and then. *International Journal for Parasitology*, 33(11), 1269–1276. [https://doi.org/10.1016/S0020-7519\(03\)00160-7](https://doi.org/10.1016/S0020-7519(03)00160-7)
- Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J., Da Rocha, M., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J., Kozłowski, D. K., Cazareth, J., Couloux, A., Da Silva, C., Guy, J., Kim-Jo, Y., Rancurel, C., Schiex, T., Abad, P., ... Danchin, E. G. J. (2017). Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genetics*, 13(6), e1006777. <https://doi.org/10.1371/journal.pgen.1006777>

- Blaxter, M. (2011). Nematodes: The worm and its relatives. *PLOS Biology*, 9(4) , e1001050. <https://doi.org/10.1371/journal.pbio.1001050>
- Blaxter, M., & Koutsovoulos, G. (2015). The evolution of parasitism in Nematoda. *Parasitology*, 142(S1), 26–39. <https://doi.org/10.1017/S0031182014000791>
- Blok, V. C., Tylka, G., Smiley, R., de Jong, W., & Daub, M. (2018). Resistance breeding. In R. N. Perry & M. Moens (Eds.), *Cyst nematodes* (pp. 174–177). CAB International.
- Blumenthal, T., Squire, M., Kirtland, S., Cane, J., Donegan, M., Spieth, J., & Sharrock, W. (1984). Cloning of a yolk protein gene family from *Caenorhabditis elegans*. *Journal of Molecular Biology*, 174(1), 1–18. [https://doi.org/10.1016/0022-2836\(84\)90361-9](https://doi.org/10.1016/0022-2836(84)90361-9)
- Borovsky, D., & Whitney, P. L. (1987). Biosynthesis, purification, and characterization of *Aedes aegypti* vitellin and vitellogenin. *Archives of Insect Biochemistry and Physiology*, 4(2), 81–99. <https://doi.org/10.1002/ARCH.940040202>
- Brennan, R. J. B., Glaze-Corcoran, S., Wick, R. & Hashemi, M. (2020). Biofumigation: An alternative strategy for the control of plant parasitic nematodes. *Journal of Integrative Agriculture*, 19(7), 1680-1690. [https://doi.org/10.1016/S2095-3119\(19\)62817-0](https://doi.org/10.1016/S2095-3119(19)62817-0)
- Bürglin, T. R., Lobos, E., & Blaxter, M. L. (1998). *Caenorhabditis elegans* as a model for parasitic nematodes. *International Journal for Parasitology*, 28(3), 395–411. [https://doi.org/10.1016/S0020-7519\(97\)00208-7](https://doi.org/10.1016/S0020-7519(97)00208-7)
- Camacho, M. J., Inácio, M. L., Mota, M., & de Andrade, E. (2021). Development and validation of a loop-mediated isothermal amplification diagnostic method to detect the quarantine potato pale cyst nematode, *Globodera pallida*. *Pathogens*, 10(6), 744. <https://doi.org/10.3390/pathogens10060744>
- Carducci, F., Biscotti, M. A., & Canapa, A. (2019). Vitellogenin gene family in vertebrates: Evolution and functions. *The European Zoological Journal*, 86(1), 233–240. <https://doi.org/10.1080/24750263.2019.1631398>
- Carneiro, R. M. D. G., Lima, F. S. de O., & Correia, V. R. (2017). Methods and tools currently used for the identification of plant parasitic nematodes. In *Nematology - Concepts, Diagnosis and Control*. In *InTech ebooks*. <https://doi.org/10.5772/intechopen.69403>
- Chaudhary, S., Dutta, T. K., Tyagi, N., Shivakumara, T. N., Papolu, P. K., Chobhe, K. A., & Rao, U. (2019). Host-induced silencing of Mi-msp-1 confers resistance to root-knot nematode *Meloidogyne incognita* in eggplant. *Transgenic Research*, 28(3–4), 327–340. <https://doi.org/10.1007/S11248-019-00126-5/metrics>
- Choi, I., Subramanian, P., Shim, D., Oh, B. J., & Hahn, B. S. (2017). RNA-seq of plant-parasitic nematode *Meloidogyne incognita* at various stages of its development. *Frontiers in Genetics*, 8. <https://doi.org/10.3389/fgene.2017.00190/pdf>
- Cobb, N. A. (1918). Estimating the nema population of soil, with special reference to the sugar-beet and root-gall nemas, *Heterodera schachtii* Schmidt and *Heterodera radicola* (Greef) Müller: And with a description of *Tylencholaimus aequalis* n. sp (Vol. 1). US Government Printing Office.

- Collins, H. P., Alva, A., Boydston, R. A., Cochran, R. L., Hamm, P. B., McGuire, A., & Riga, E. (2006). Soil microbial, fungal, and nematode responses to soil fumigation and cover crops under potato production. *Biology and Fertility of Soils*, 42(3), 247–257. <https://doi.org/10.1007/s00374-005-0022-0>
- Colonello-Frattini, N. A., Guidugli-Lazzarini, K. R., Simões, Z. L. P., & Hartfelder, K. (2010). Mars is close to Venus – Female reproductive proteins are expressed in the fat body and reproductive tract of honey bee (*Apis mellifera* L.) drones. *Journal of Insect Physiology*, 56(11), 1638–1644. <https://doi.org/10.1016/j.jinsphys.2010.06.008>
- Cottee, P. A., Nisbet, A. J., Abs El-Osta, Y. G., Webster, T. L., & Gasser, R. B. (2006). Construction of gender-enriched cDNA archives for adult *Oesophagostomum dentatum* by suppressive-subtractive hybridization and a microarray analysis of expressed sequence tags. *Parasitology*, 132(5), 691–708. <https://doi.org/10.1017/S0031182005009728>
- 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. A., den Akker, S. E. van, Holroyd, N., Hunt, M., Mantelin, S., Naghra, H., Pain, A., Palomares-Rius, J. E., Zarowiecki, M., Berriman, M., ... Urwin, P. E. (2014). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology*, 15(3), 1–17. <https://doi.org/10.1186/gb-2014-15-3-r43/figures/5>
- Creutz, C. E., Snyder, S. L., Daigle, S. N., & Redick, J. (1996). Identification, localization, and functional implications of an abundant nematode annexin. *Journal of Cell Biology*, 132(6), 1079–1092. <https://doi.org/10.1083/jcb.132.6.1079>
- Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., Clement, T., & Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation*, 32(6), 767. <https://doi.org/10.1177/1040638720905833>
- Curtis, R. H. C. (2007). Plant parasitic nematode proteins and the host–parasite interaction. *Briefings in Functional Genomics*, 6(1), 50–58. <https://doi.org/10.1093/BFGP/ELM006>
- Curtis, R. H. C., Dunn, J., Yeung, M., Robinson, M. P., Martins, F., & Evans, K. (1998). Serological identification and quantification of potato cyst nematodes from clean cysts and processed soil samples. *Annals of Applied Biology*, 133(1), 65–79. <https://doi.org/10.1111/J.1744-7348.1998.TB05803.X>
- Curtis, R. H. C., Segers, I., & Evans, K. (1996). Observations on the specificity of *Globodera pallida* and *G. rostochiensis* diagnostic proteins to amphids of *Globodera* species. *Nematologica*, 42(4), 466–480. <https://doi.org/10.1163/004525996X00073>
- Custodia, N., Won, S. J., Novillo, A., Wieland, M., Li, C., & Callard, I. P. (2001). *Caenorhabditis elegans* as an environmental monitor using DNA microarray analysis. *Annals of the New York Academy of Sciences*, 948(1), 32–42. <https://doi.org/10.1111/J.1749-6632.2001.TB03984.X>
- Daub, M. (2021). The beet cyst nematode (*Heterodera schachtii*): an ancient threat to sugar beet crops in Central Europe has become an invisible actor. In *Integrated Nematode Management: State-of-the-art and visions for the future* (pp. 394–399). Wallingford UK: CAB International. <https://doi.org/10.1079/9781789247541.0055>

Davies, K. G., Curtis, R., & Evans, K. (1996). Serologically based diagnostic and quantification tests for nematodes. *Pesticide Science*, 47, 81–87.

Davis, G. M., Hipwell, H., & Boag, P. R. (2023). Oogenesis in *Caenorhabditis elegans*. *Sexual Development*, 17(2-3), 73-83. <https://doi.org/10.1159/000531019>

de Oliveira, C. M. G., Monteiro, A. R., & Blok, V. C. (2011). Morphological and molecular diagnostics for plant-parasitic nematodes: Working together to get the identification done. *Tropical Plant Pathology*, 36(2), 65–73. <https://doi.org/10.1590/S1982-56762011000200001>

De Waele, D., & Elsen, A. (2008). Challenges in tropical plant nematology. *Annual Review of Phytopathology*, 45, 457–485. <https://doi.org/10.1146/annurev.phyto.45.062806.094438>

Decraemer, W., & Hunt, D. (2024). Structure and classification. In R. Perry, M. Moens, & J. T. Jones (Eds.), *Plant Nematology* (3rd ed., pp. 30–66). CAB International.

Deeley, R., Mullinix, D., Wetekam, W., Kronenberg, H., Meyers, M., Eldridge, J., & Goldberger, R. (1975). Vitellogenin synthesis in the avian liver. Vitellogenin is the precursor of the egg yolk phosphoproteins. *Journal of Biological Chemistry*, 250(23), 9060–9066. [https://doi.org/10.1016/S0021-9258\(19\)40693-5](https://doi.org/10.1016/S0021-9258(19)40693-5)

Department for Environment, Food & Rural Affairs, and Animal and Plant Health Agency. (2024, October 11). *Plant health controls*. GOV.UK. <https://www.gov.uk/guidance/plant-health-controls#legislation>

DiLegge, M. J., Manter, D. K., & Vivanco, J. M. (2019). A novel approach to determine generalist nematophagous microbes reveals *Mortierella globalpina* as a new biocontrol agent against *Meloidogyne* spp. nematodes. *Scientific Reports*, 9(1), 1–9. <https://doi.org/10.1038/s41598-019-44010-y>

Dillman, A. R., Macchietto, M., Porter, C. F., Rogers, A., Williams, B., Antoshechkin, I., Lee, M. M., Goodwin, Z., Lu, X., Lewis, E. E., Goodrich-Blair, H., Stock, S. P., Adams, B. J., Sternberg, P. W., & Mortazavi, A. (2015). Comparative genomics of *Steinernema* reveals deeply conserved gene regulatory networks. *Genome Biology*, 16(1), 1–21. <https://doi.org/10.1186/S13059-015-0746-6>

Dubreuil, G., Magliano, M., Deleury, E., Abad, P., & Rosso, M. N. (2007). Transcriptome analysis of root-knot nematode functions induced in the early stages of parasitism. *New Phytologist*, 176(2), 426–436. <https://doi.org/10.1111/J.1469-8137.2007.02181.X>

Duceppe, M. O., Lafond-Lapalme, J., Palomares-Rius, J. E., Sabeh, M., Blok, V., Moffett, P., & Mimee, B. (2017). Analysis of survival and hatching transcriptomes from potato cyst nematodes, *Globodera rostochiensis* and *G. pallida*. *Scientific Reports*, 7(1), 1–13. <https://doi.org/10.1038/s41598-017-03871-x>

Elling, A. A. (2006). *Genomics and molecular biology approaches to plant-parasitic nematode biology and parasitism*. Unpublished doctoral dissertation, Iowa State University, United States. <https://lib.dr.iastate.edu/rtd>

Eren, K., Taktakoğlu, N., & Pirim, I. (2022). DNA sequencing methods: From past to present. *The Eurasian Journal of Medicine*, 54(Suppl 1), S47. <https://doi.org/10.5152/EURASIANJMED.2022.22280>

- Evans, C., Hardin, J., & Stoebe, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, 19(5), 776. <https://doi.org/10.1093/BIB/BBX008>
- Evans, K., & Rowe, J. A. (1998). Distribution and economic importance. In *The cyst nematodes* (pp. 1-30). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-9018-1_1
- Evans, M. I., Silva, R., & Burch, J. B. (1988). Isolation of chicken vitellogenin I and III cDNAs and the developmental regulation of five estrogen-responsive genes in the embryonic liver. *Genes & Development*, 2(1), 116–124. <https://doi.org/10.1101/GAD.2.1.116>
- Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G. J., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry, J., Mimeo, B., Duceppe, M. O., Boyes, I., Marvin, J. M. C., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., ... Jones, J. T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology*, 17(1), 124–124. <https://doi.org/10.1186/S13059-016-0985-1>
- Ezurra, M., Benedetto, A., Sornda, T., Gilliat, A. F., Au, C., Zhang, Q., van Schelt, S., Petrache, A. L., Wang, H., de la Guardia, Y., Bar-Nun, S., Tyler, E., Wakelam, M. J., & Gems, D. (2018). *C. elegans* eats its own intestine to make yolk leading to multiple senescent pathologies. *Current Biology*, 28(16), 2544–2556.e5. <https://doi.org/10.1016/j.cub.2018.06.035>
- Fioretti, L., Warry, A., Porter, A., Haydock, P., & Curtis, R. (2001). Isolation and localisation of an annexin gene (gp-nex) from the potato cyst nematode, *Globodera pallida*. *Nematology*, 3(1), 45–54. <https://doi.org/10.1163/156854101300106883>
- 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(8–9), 569–578. <https://doi.org/10.1016/j.micinf.2013.05.002>
- Flaherty, D. K. (2012). Immunogenicity and antigenicity. *Immunology for pharmacy* (pp. 23–30). Mosby. <https://doi.org/10.1016/B978-0-323-06947-2.10003-3>
- Fleming, C. C., & Marks, R. J. (1983). The identification of the potato cyst nematodes *Globodera rostochiensis* and *G. pallida* by isoelectric focusing of proteins on polyacrylamide gels. *Annals of Applied Biology*, 103(2), 277–281. <https://doi.org/10.1111/j.1744-7348.1983.tb02765.x>
- Floyd, R., Abebe, E., Papert, A., & Blaxter, M. (2002). Molecular barcodes for soil nematode identification. *Molecular Ecology*, 11(4), 839–850. <https://doi.org/10.1046/j.1365-294X.2002.01485.x>
- 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(1), 1–15. <https://doi.org/10.1038/s41598-018-20536-5>

- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Totowa, NJ: Humana press.
- Gerber-Huber, S., Nardelli, D., Haefliger, J. A., Cooper, D. N., Givel, F., Germond, J. E., Engel, J., Green, N. M., & Wahli, W. (1987). Precursor-product relationship between vitellogenin and the yolk proteins as derived from the complete sequence of a *Xenopus* vitellogenin gene. *Nucleic Acids Research*, 15(12), 4737–4760. <https://doi.org/10.1093/nar/15.12.4737>
- Gerke, V., Gavins, F. N. E., Geisow, M., Grewal, T., Jaiswal, J. K., Nylandsted, J., & Rescher, U. (2024). Annexins—a family of proteins with distinctive tastes for cell signaling and membrane dynamics. *Nature Communications*, 15(1), 1–15. <https://doi.org/10.1038/s41467-024-45954-0>
- Gill, H. K., Aujla, I. S., De Bellis, L., & Luvisi, A. (2017). The role of soil solarization in India: How an unnoticed practice could support pest control. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.01515>
- Grant, B., & Hirsh, D. (1999). Receptor-mediated endocytosis in the *Caenorhabditis elegans* oocyte. *Molecular Biology of the Cell*, 10(12), 4311. <https://doi.org/10.1091/mbc.10.12.4311>
- Grün, D., Kirchner, M., Thierfelder, N., Stoeckius, M., Selbach, M., & Rajewsky, N. (2014). Conservation of mRNA and protein expression during development of *C. elegans*. *Cell Reports*, 6(3), 565–577. <https://doi.org/10.1016/j.celrep.2014.01.001>
- Hada, A., Singh, D., Papolu, P. K., Banakar, P., Raj, A., & Rao, U. (2021). Host-mediated RNAi for simultaneous silencing of different functional groups of genes in *Meloidogyne incognita* using fusion cassettes in *Nicotiana tabacum*. *Plant Cell Reports*, 40(12), 2287–2302. <https://doi.org/10.1007/s00299-021-02767-5>
- Hagedorn, H. H., & Kunkel, J. G. (1979). Vitellogenin and vitellin in insects. *Annual Review of Entomology*, 24, 475–505.
- Hall, D. H., Herndon, L. A., & Altun, Z. (2017). *Introduction to C. elegans Embryo Anatomy*. WormAtlas. <https://www.wormatlas.org>
- Hartman, D., Donald, D. R., Nikolaou, S., Savin, K. W., Hasse, D., Presidente, P. J. A., & Newton, S. E. (2001). Analysis of developmentally regulated genes of the parasite *Haemonchus contortus*. *International Journal for Parasitology*, 31(11), 1236–1245. [https://doi.org/10.1016/S0020-7519\(01\)00248-X](https://doi.org/10.1016/S0020-7519(01)00248-X)
- Hashimshony, T., Feder, M., Levin, M., Hall, B. K., & Yanai, I. (2014). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature*, 519(7542), 219–222. <https://doi.org/10.1038/nature13996>
- He, Q., Wang, D., Tang, B., Wang, J., Zhang, D., Liu, Y., & Cheng, F. (2021). Rapid and sensitive detection of *Meloidogyne graminicola* in soil using conventional PCR, loop-mediated isothermal amplification, and real-time PCR methods. *Plant Disease*, 105(2), 456–463. <https://doi.org/10.1094/PDIS-06-20-1291-RE>
- Heine, U., & Blumenthal, T. (1986). Characterization of regions of the *Caenorhabditis elegans* X chromosome containing vitellogenin genes. *Journal of Molecular Biology*, 188(3), 301–312. [https://doi.org/10.1016/0022-2836\(86\)90156-7](https://doi.org/10.1016/0022-2836(86)90156-7)

- Hesse, C. N., Moreno, I., Acevedo Pardo, O., Pacheco Fuentes, H., Grenier, E., Dandurand, L. M., Zasada, I. A., & Digennaro, P. (2021). Characterization of *Globodera ellingtonae* populations from Chile utilizing whole genome sequencing. *Journal of Nematology*, 53, 2021–2088. <https://doi.org/10.21307/jofnem-2021-088>
- Hockland, S., Prior, T., Stanley, J. D., Insera, R. N., & Kohl, L. M. (2024). International plant health – Putting legislation into practice. In R. Perry, M. Moens, & J. T. Jones (Eds.), *Plant Nematology* (3rd ed., pp. 415–422). CAB International.
- Hodge, K., Have, S. Ten, Hutton, L., & Lamond, A. I. (2013). Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *Journal of Proteomics*, 88, 92–103. <https://doi.org/10.1016/j.jprot.2013.02.023>
- Holbein, J., Grundler, F. M. W., & Siddique, S. (2016). Plant basal resistance to nematodes: An update. *Journal of Experimental Botany*, 67(7), 2049–2061. <https://doi.org/10.1093/jxb/erw005>
- Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., & Berriman, M. (2017). WormBase ParaSite – a comprehensive resource for helminth genomics. *Molecular and Biochemical Parasitology*, 215, 2–10. <https://doi.org/10.1016/j.molbiopara.2016.11.005>
- Ishii, N., Goto, S., & Hartman, P. S. (2002). Protein oxidation during aging of the nematode *Caenorhabditis elegans*. *Free Radical Biology and Medicine*, 33(8), 1021–1025.
- Iqbal, S., Fosu-Nyarko, J., & Jones, M. G. K. (2020). Attempt to silence genes of the RNAi pathways of the root-knot nematode, *Meloidogyne incognita*, results in diverse responses including increase and no change in expression of some genes. *Frontiers in Plant Science*, 11, 501032. <https://doi.org/10.3389/fpls.2020.00328>
- Janeway, C. A., Jr., Travers, P., Walport, M., & Shlomchik, M. J. (2001). B-cell activation by armed helper T cells. *Immunobiology: The immune system in health and disease* (5th ed.). Garland Science.
- Jones, F. G. W., & Jones, M. (1964). *Pests of field crops*. St. Martin's Press, Inc.
- Jones, J. T., Haegeman, A., Danchin, E. G. J., Gaur, H. S., Helder, J., Jones, M. G. K., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J. E., Wesemael, W. M. L., & Perry, R. N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, 14(9), 946–961. <https://doi.org/10.1111/mpp.12057>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Jung, C., Cai, D., & Kleine, M. (1998). Engineering nematode resistance in crop species. *Trends in Plant Science*, 3(7), 266–271. [https://doi.org/10.1016/S1360-1385\(98\)01247-3](https://doi.org/10.1016/S1360-1385(98)01247-3)
- Kayani, M. Z., Mukhtar, T., & Hussain, M. A. (2017). Effects of southern root knot nematode population densities and plant age on growth and yield parameters of cucumber. *Crop Protection*, 92, 207–212. <https://doi.org/10.1016/j.cropro.2016.09.007>

- Kennedy, M. J., Sghoelz, J. E., Donald, P. A., & Niblack, T. L. (1997). Unique immunogenic proteins in *Heterodera glycines* eggshells. *Journal of Nematology*, 29(3), 276.
- Kerry, B. (1980). Biocontrol: Fungal parasites of female cyst nematodes. *Journal of Nematology*, 12(4), 253.
- Khan, Z., & Kim, Y. H. (2007). A review on the role of predatory soil nematodes in the biological control of plant parasitic nematodes. *Applied Soil Ecology*, 35(2), 370–379. <https://doi.org/10.1016/j.apsoil.2006.07.007>
- Kikuchi, T., Eves-Van Den Akker, S., & Jones, J. T. (2017). Genome evolution of plant-parasitic nematodes. *Annual Review of Phytopathology*, 55, 333–354. <https://doi.org/10.1146/annurev-phyto-080516-035434>
- Kimble, J., & Sharrock, W. J. (1983). Tissue-specific synthesis of yolk proteins in *Caenorhabditis elegans*. *Developmental Biology*, 96(1), 189–196. [https://doi.org/10.1016/0012-1606\(83\)90322-6](https://doi.org/10.1016/0012-1606(83)90322-6)
- Kleine, M., Cai, D., Elbl, C., Herrmann, R. G., & Jung, C. (1995). Physical mapping and cloning of a translocation in sugar beet (*Beta vulgaris* L.) carrying a gene for nematode (*Heterodera schachtii*) resistance from *B. procumbens*. *Theoretical and Applied Genetics*, 90(3–4), 399–406. <https://doi.org/10.1007/BF00221982>
- Klink, V. P., Kim, K. H., Martins, V., MacDonald, M. H., Beard, H. S., Alkharouf, N. W., Lee, S. K., Park, S. C., & Matthews, B. F. (2009). A correlation between host-mediated expression of parasite genes as tandem inverted repeats and abrogation of development of female *Heterodera glycines* cyst formation during infection of *Glycine max*. *Planta*, 230(1), 53–71. <https://doi.org/10.1007/S00425-009-0926-2>
- Ko, H. R., Kim, E. H., Kim, J., Lee, J. K., & Lee, W. H. (2017). Rapid methods to distinguish *Heterodera schachtii* from *Heterodera glycines* using PCR technique. *Research in Plant Disease*, 23(3), 241–248. <https://doi.org/10.5423/RPD.2017.23.3.241>
- Koenning, S. R. (2004). *Biology and management of soybean cyst nematode*. D. P. Schmitt, R. D. Riggs, & J. A. Wrather, Eds.; 2nd ed.). Schmitt & Associates of Marceline.
- Kohlmeier, P., Feldmeyer, B., & Foitzik, S. (2018). Vitellogenin-like A-associated shifts in social cue responsiveness regulate behavioral task specialization in an ant. *PLOS Biology*, 16(6), e2005747. <https://doi.org/10.1371/JOURNAL.PBIO.2005747>
- Koutsovolos, G. D., Pouillet, M., Elashry, A., Kozłowski, D. K. L., Sallet, E., Da Rocha, M., Perfus-Barbeoch, L., Martin-Jimenez, C., Frey, J. E., Ahrens, C. H., Kiewnick, S., & Danchin, E. G. J. (2020). Genome assembly and annotation of *Meloidogyne enterolobii*, an emerging parthenogenetic root-knot nematode. *Scientific Data*, 7(1), 1–13. <https://doi.org/10.1038/s41597-020-00666-0>
- Krissinel, E. (2007). On the relationship between sequence and structure similarities in proteomics. *Bioinformatics*, 23(6), 717–723.
- Lambert, K., & Bekal, S. (2002). Introduction to plant-parasitic nematodes. *The Plant Health Instructor*. <https://doi.org/10.1094/phi-i-2002-1218-01>

- Le Guellec, K., Lawless, K., Valotaire, Y., Kress, M., & Tenniswood, M. (1988). Vitellogenin gene expression in male rainbow trout (*Salmo gairdneri*). *General and Comparative Endocrinology*, 71(3), 359–371. [https://doi.org/10.1016/0016-6480\(88\)90264-X](https://doi.org/10.1016/0016-6480(88)90264-X)
- Lewis, K. A., Tzilivakis, J., Warner, D. J., & Green, A. (2016). An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment*, 22(4), 1050–1064. <https://doi.org/10.1080/10807039.2015.1133242>
- Li, J., Zou, C., Xu, J., Ji, X., Niu, X., Yang, J., Huang, X., & Zhang, K. Q. (2015). Molecular mechanisms of nematode-nematophagous microbe interactions: Basis for biological control of plant-parasitic nematodes. *Annual Review of Phytopathology*, 53, 67–95. <https://doi.org/10.1146/annurev-phyto-080614-120336>
- Lilley, C. J., Atkinson, H. J., & Urwin, P. E. (2005). Molecular aspects of cyst nematodes. *Molecular Plant Pathology*, 6(6), 577–588. <https://doi.org/10.1111/j.1364-3703.2005.00306.x>
- Linsell, K. J., Riley, I. T., Davies, K. A., & Oldach, K. H. (2014). Characterization of resistance to *Pratylenchus thornei* (Nematoda) in wheat (*Triticum aestivum*): Attraction, penetration, motility, and reproduction. *Phytopathology*, 104(2), 174–187. <https://doi.org/10.1094/PHYTO-12-12-0345-R>
- Logsdon, J. M., Tyshenko, M. G., Dixon, C., D.-Jafari, J., Walker, V. K., & Palmer, J. D. (1995). Seven newly discovered intron positions in the triose-phosphate isomerase gene: Evidence for the introns-late theory. *Proceedings of the National Academy of Sciences of the United States of America*, 92(18), 8507–8511. <https://doi.org/10.1073/PNAS.92.18.8507>
- Maggenti, A. R. (1981). Nematodes: Development as plant parasites. *Annual Review of Microbiology*, 35, 135–154. <https://doi.org/10.1146/annurev.mi.35.100181.001031>
- Maier, T. R., Hewezi, T., Peng, J., & Baum, T. J. (2012). Isolation of whole esophageal gland cells from plant-parasitic nematodes for transcriptome analyses and effector identification. *Molecular Plant-Microbe Interactions*, 26(1), 31–35. <https://doi.org/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(3), 305. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2620382/>
- Masler, E. P. (2002). A polyclonal antiserum that recognizes the major female-specific proteins from the soybean cyst nematode *Heterodera glycines*. *International Journal of Nematology*, 12(2), 119–124.
- Masonbrink, R. E., Maier, T. R., Hudson, M., Severin, A., & Baum, T. (2021). A chromosomal assembly of the soybean cyst nematode genome. *Molecular Ecology Resources*, 21(7), 2407–2422. <https://doi.org/10.1111/1755-0998.13432>
- Matyash, V., Geier, C., Henske, A., Mukherjee, S., Hirsh, D., Thiele, C., Grant, B., Maxfield, F. R., & Kurzchalia, T. V. (2001). Distribution and transport of cholesterol in *Caenorhabditis elegans*. *Molecular Biology of the Cell*, 12(6), 1725–1736. <https://doi.org/10.1091/MBC.12.6.1725>
- Mbeunkui, F., Scholl, E. H., Opperman, C. H., Goshe, M. B., & Bird, D. M. K. (2010). Proteomic and bioinformatic analysis of the root-knot nematode *Meloidogyne hapla*: The basis for plant

parasitism. *Journal of Proteome Research*, 9(10), 5370–5381. <https://doi.org/10.1021/PR1006069>

Merck. (2025). *Guide on how to use Proteinase K in different procedures*. DNA & RNA Purification. <https://www.sigmaaldrich.com/GB/en/technical-documents/technical-article/genomics/dna-and-rna-purification/guide-to-use-proteinase-k-procedures>

Min, H., Youn, E., & Shim, Y. H. (2020). Maternal caffeine intake disrupts eggshell integrity and retards larval development by reducing yolk production in a *Caenorhabditis elegans* model. *Nutrients*, 12(5), 1334. <https://doi.org/10.3390/NU12051334>

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(2), 174–187. <https://doi.org/10.1016/j.pt.2021.08.008>

Moens, M., Perry, R. N., & Jones, J. (2018). Cyst nematodes: Life cycle and economic importance. In R. N. Perry & M. Moens (Eds.). *Cyst nematodes* (pp. 1–26). CAB International.

Montarry, J., Mimee, B., Danchin, E. G. J., Koutsovoulos, G. D., Ste-Croix, D. T., & Grenier, E. (2021). Recent advances in population genomics of plant-parasitic nematodes. *Phytopathology*, 111(1), 40–48. <https://doi.org/10.1094/PHYTO-09-20-0418-RVW>

Morandin, C., Hietala, A., & Helanterä, H. (2019). Vitellogenin and vitellogenin-like gene expression patterns in relation to caste and task in the ant *Formica fusca*. *Insectes Sociaux*, 66(4), 519–531. <https://doi.org/10.1007/S00040-019-00725-9>

Moya, A., Tejedor, D., Manetti, M., Clavijo, A., Pagano, E., Munarriz, E., & Kronberg, M. F. (2022). Reproductive toxicity by exposure to low concentrations of pesticides in *Caenorhabditis elegans*. *Toxicology*, 475, 153229. <https://doi.org/10.1016/J.TOX.2022.153229>

Mulholland, V., Carde, L., O'Donnell, K. L., Fleming, C. C., & Powers, T. O. (1996). Use of the polymerase chain reaction to discriminate potato cyst nematode at the species level. In C. Marshall (Ed.). *Proceedings of Diagnostics in Crop Production Symposium* (pp. 247–252). British Crop Production Council.

Münch, D., Ihle, K. E., Salmela, H., & Amdam, G. V. (2015). Vitellogenin in the honey bee brain: Atypical localization of a reproductive protein that promotes longevity. *Experimental Gerontology*, 71, 103–108. <https://doi.org/10.1016/J.EXGER.2015.08.001>

Nardelli, D., Wahli, W., Gerber-Huber, S., van het Schip, F. D., Gruber, M., & Ab, G. (1987). Vertebrate and nematode genes coding for yolk proteins are derived from a common ancestor. *Biochemistry*, 26(20), 6397–6402. <https://doi.org/10.1021/BI00394A014>

National Center for Biotechnology Information. (2017). *Gene expression*. Bethesda (MD): National Library of Medicine (US).

National Research Council (US) Committee on Methods of Producing Monoclonal Antibodies. (1999). *Monoclonal antibody production*. National Academies Press. <https://www.ncbi.nlm.nih.gov/books/NBK10008/>

Nicol, J. M., Turner, S. J., Coyne, D. L., Nijs, L. den, Hockland, S., & Maafi, Z. T. (2011). Current nematode threats to world agriculture. In *Genomics and Molecular Genetics of Plant-*

Nematode Interactions (pp. 21–43). Springer Netherlands. https://doi.org/10.1007/978-94-007-0434-3_2

Niebur, E., & Erdős, P. (1991). Theory of the locomotion of nematodes: Dynamics of undulatory progression on a surface. *Biophysical Journal*, 60(5), 1132–1146. [https://doi.org/10.1016/S0006-3495\(91\)82149-X](https://doi.org/10.1016/S0006-3495(91)82149-X)

Nisbet, A. J., & Gasser, R. B. (2004). Profiling of gender-specific gene expression for *Trichostrongylus vitrinus* (Nematoda: Strongylida) by microarray analysis of expressed sequence tag libraries constructed by suppressive-subtractive hybridisation. *International Journal for Parasitology*, 34(5), 633–643. <https://doi.org/10.1016/J.IJPARA.2003.12.007>

Niu, J. H., Guo, Q. X., Jian, H., Chen, C. L., Yang, D., Liu, Q., & Guo, Y. D. (2011). Rapid detection of *Meloidogyne* spp. by LAMP assay in soil and roots. *Crop Protection*, 30(8), 1063–1069. <https://doi.org/10.1016/j.cropro.2011.03.028>

Niu, J., Jian, H., Xu, J., Chen, C., Guo, Q., Liu, Q., & Guo, Y. (2012). RNAi silencing of the *Meloidogyne incognita* Rpn7 gene reduces nematode parasitic success. *European Journal of Plant Pathology*, 134(1), 131–144. <https://doi.org/10.1007/S10658-012-9971-Y>

Norshie, P. M., Grove, I. G., & Back, M. A. (2016). Field evaluation of the nematicide fluensulfone for control of the potato cyst nematode *Globodera pallida*. *Pest Management Science*, 72(10), 2001–2007. <https://doi.org/10.1002/ps.4329>

O'Rourke, D., Gravato-Nobre, M. J., Stroud, D., Pritchett, E., Barker, E., Price, R. L., Robinson, S. A., Spiro, S., Kuwabara, P., & Hodgkin, J. (2023). Isolation and molecular identification of nematode surface mutants with resistance to bacterial pathogens. *G3 Genes|Genomes|Genetics*, 13(5), 56. <https://doi.org/10.1093/G3JOURNAL/JKAD056>

Olson, S. K., Greenan, G., Desai, A., Müller-Reichert, T., & Oegema, K. (2012). Hierarchical assembly of the eggshell and permeability barrier in *C. elegans*. *Journal of Cell Biology*, 198(4), 731–748. <https://doi.org/10.1083/JCB.201206008>

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., & Windham, E. (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39), 14802–14807. https://doi.org/10.1073/PNAS.0805946105/SUPPL_FILE/0805946105SI.PDF

Ou, S., Peng, D., Liu, X., Li, Y., & Moens, M. (2008). Identification of *Heterodera glycines* using PCR with sequence characterised amplified region (SCAR) primers. *Nematology*, 10(3), 397–403. <https://doi.org/10.1163/156854108783900212>

Page, A., & Johnstone, I. L. (2007). *The Cuticle*. WormBook. Retrieved: December, 19, 2020, from http://www.wormbook.org/chapters/www_cuticle/cuticle.html

Palomares-Rius, J. E., Hedley, P., Cock, P. J., Morris, J. A., Jones, J. T., & Blok, V. C. (2016). Gene expression changes in diapause or quiescent potato cyst nematode, *Globodera pallida*, eggs after hydration or exposure to tomato root diffusate. *PeerJ*, 4, e1654. <https://doi.org/10.7717/peerj.1654>

- Papolu, P. K., Gantasala, N. P., Kamaraju, D., Banakar, P., Sreevathsa, R., & Rao, U. (2013). Utility of host delivered RNAi of two FMRF amide-like peptides, *flp-14* and *flp-18*, for the management of root knot nematode, *Meloidogyne incognita*. *PLOS ONE*, 8(11), e80603. <https://doi.org/10.1371/JOURNAL.PONE.0080603>
- Pateraki, L. E., & Stratakis, E. (1997). Characterization of vitellogenin and vitellin from land crab *Potamon potamios*: Identification of a precursor polypeptide in the molecule. *Journal of Experimental Zoology*, 279(6), 597-608.
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., . . . Bateman, A. (2022). InterPro in 2022. *Nucleic Acids Research*, 51(D1), D418–D427. <https://doi.org/10.1093/nar/gkac993>
- Perez, M. F., & Lehner, B. (2019). Vitellogenins - yolk gene function and regulation in *Caenorhabditis elegans*. *Frontiers in Physiology*, 10, 1067. <https://doi.org/10.3389/fphys.2019.01067>
- Perez, M. F., Francesconi, M., Hidalgo-Carcedo, C., & Lehner, B. (2017). Maternal age generates phenotypic variation in *Caenorhabditis elegans*. *Nature*, 552(7683), 106–109. <https://doi.org/10.1038/nature25012>
- Perry, R. N. (1996). Chemoreception in plant parasitic nematodes. *Annual Review of Phytopathology*, 34, 181–199. <https://doi.org/10.1146/ANNUREV.PHYTO.34.1.181/CITE/REFWORKS>
- Perry, R. N., & Moens, M. (2011). Survival of parasitic nematodes outside the host. In *Molecular and Physiological Basis of Nematode Survival* (pp. 1–27). Springer Science & Business Media. <https://doi.org/10.1079/9781845936877.0001>
- Perry, R. N., Wright, D. J., Chitwood, D. J., & Price, J. A. (2024). Reproduction, physiology, and biochemistry. In R. N. Perry, M. Moens, & J. T. Jones (Eds.), *Plant Nematology* (3rd ed., pp. 231–261). CAB International.
- Petersen, L. M., Martin, I. W., Moschetti, W. E., Kershaw, C. M., & Tsongalis, G. J. (2019). Third-generation sequencing in the clinical laboratory: Exploring the advantages and challenges of nanopore sequencing. *Journal of Clinical Microbiology*, 58(1). <https://doi.org/10.1128/JCM.01315-19>
- Phan, N. T., Orjuela, J., Danchin, E. G. J., Klopp, C., Perfus-Barbeoch, L., Kozlowski, D. K., Koutsovoulos, G. D., Lopez-Roques, C., Bouchez, O., Zahm, M., Besnard, G., & Bellafiore, S. (2020). Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*). *Ecology and Evolution*, 10(20), 11006–11021. <https://doi.org/10.1002/ECE3.6680>
- Phillips, W. S., Howe, D. K., Brown, A. M. V., Eves-Van Den Akker, S., Dettwyler, L., Peetz, A. B., Denver, D. R., & Zasada, I. A. (2017). The draft genome of *Globodera ellingtonae*. *Journal of Nematology*, 49(2), 127. <https://doi.org/10.21307/JOFNEM-2017-054>
- Phiriyangkul, P., & Utarabhand, P. (2006). Molecular characterization of a cDNA encoding vitellogenin in the banana shrimp, *Penaeus (Litopenaeus) merguensis* and sites of

vitellogenin mRNA expression. *Molecular Reproduction and Development*, 73(4), 410–423. <https://doi.org/10.1002/MRD.20424>

Poveda, J., Abril-Urias, P., & Escobar, C. (2020). Biological control of plant-parasitic nematodes by filamentous fungi inducers of resistance: *Trichoderma*, mycorrhizal and endophytic fungi. *Frontiers in Microbiology*, 11, 992. <https://doi.org/10.3389/fmicb.2020.00992>

Price, J. A., Ali, M. F., Major, L. L., Smith, T. K., & Jones, J. T. (2023). An eggshell-localised annexin plays a key role in the coordination of the life cycle of a plant-parasitic nematode with its host. *PLOS Pathogens*, 19(2), e1011147. <https://doi.org/10.1371/JOURNAL.PPAT.1011147>

Price, J. A., Hunt, K., Nyagah, N., Orage, C., Mwaura, O., Haukeland, S., Coyne, D., & Jones, J. T. (2024). Characterisation of ‘giant’ cysts of the potato cyst nematode, *Globodera rostochiensis*, present in potato fields in Kenya. *Nematology*, 26(7), 795–801. <https://doi.org/10.1163/15685411-BJA10339>

Qing, X., Zhang, Y. M., Sun, S., Ahmed, M., Lo, W.-S., Bert, W., Holovachov, O., & Li, H. (2023). Phylogenomic insights into the evolution and origin of Nematoda. *BioRxiv*. <https://doi.org/10.1101/2023.12.13.571554>

Quintero-Ronderos, P., Arango, M. T., Castiblanco, J., Rojas-Villarraga, A., Shoenfeld, Y., & Anaya, J. M. (2013). Analysis of proteins and antibodies. In J. M. Anaya, Y. Shoenfeld, A. Rojas-Villarraga, G. A. Levy, & R. Cervera (Eds.), *Autoimmunity: From bench to bedside*. El Rosario University Press.

Rau, A., Marot, G., & Jaffrézic, F. (2014). Differential meta-analysis of RNA-seq data from multiple studies. *BMC bioinformatics*, 15(1), 91. <https://doi.org/10.1186/1471-2105-15-91>

Ramírez-Clavijo, S., & Montoya-Ortíz, G. (2013). Gene expression and regulation. In J. M. Anaya, Y. Shoenfeld, A. Rojas-Villarraga, R. A. Levy, & R. Cervera (Eds.), *Autoimmunity: From Bench to Bedside*. El Rosario University Press.

Reuther, M., Lang, C., & Grundler, F. M. (2017). Nematode-tolerant sugar beet varieties—resistant or susceptible to the beet cyst nematode *Heterodera schachtii*. *Sugar Industry*, 142(5), 277–284.

Robinson, M. P., Atkinson, H. J., & Perry, R. N. (1987). The influence of soil moisture and storage time on the motility, infectivity and lipid utilization of second stage juveniles of the potato cyst nematodes *Globodera rostochiensis* and *G. pallida*. *Revue de Nématologie*, 10(3), 343–348.

Robinson, M. P., Butcher, G., Curtis, R. H., Davies, K. G., & Evans, K. (1993). Characterisation of a 34 kD protein from potato cyst nematodes, using monoclonal antibodies with potential for species diagnosis. *Annals of Applied Biology*, 123(2), 337–347. <https://doi.org/10.1111/j.1744-7348.1993.tb04096.x>

Rosso, M. N., Dubrana, M. P., Cimbolini, N., Jaubert, S., & Abad, P. (2005). Application of RNA interference to root-knot nematode genes encoding esophageal gland proteins. *Molecular Plant-Microbe Interactions*, 18(7), 615–620. <https://doi.org/10.1094/MPMI-18-0615>

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467. <https://doi.org/10.1073/PNAS.74.12.5463>

Sappington, T. W., & Raikhel, A. S. (1998). Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochemistry and Molecular Biology*, 28(5–6), 277–300. [https://doi.org/10.1016/S0965-1748\(97\)00110-0](https://doi.org/10.1016/S0965-1748(97)00110-0)

Schots, A., Gommers, F. J., Bakker, J., & Egberts, E. (1990). Serological differentiation of plant-parasitic nematode species with polyclonal and monoclonal antibodies. *Journal of Nematology*, 22(1), 16.

Schots, A., Bakker, J., Gommers, F. J., Bouwman-Smits, L., & Egberts, E. (1987). Serological differentiation of the potato-cyst nematodes *Globodera pallida* and *G. rostochiensis*: Partial purification of species-specific proteins. *Parasitology*, 95(2), 421–428. <https://doi.org/10.1017/S003118200005784X>

Schots, A., Gommers, F. J., & Egberts, E. (1992). Quantitative ELISA for the detection of potato cyst nematodes in soil samples. *Fundamental and Applied Nematology*, 15, 55–61.

Sharrock, W. J. (1983). Yolk proteins of *Caenorhabditis elegans*. *Developmental Biology*, 96(1), 182–188. [https://doi.org/10.1016/0012-1606\(83\)90321-4](https://doi.org/10.1016/0012-1606(83)90321-4)

Sharrock, W. J., Sutherlin, M. E., Leske, K., Cheng, T. K., & Kim, T. Y. (1990). Two distinct yolk lipoprotein complexes from *Caenorhabditis elegans*. *Journal of Biological Chemistry*, 265(24), 14422–14431. [https://doi.org/10.1016/S0021-9258\(18\)77319-5](https://doi.org/10.1016/S0021-9258(18)77319-5)

Shen, M. M., & Hodgkin, J. (1988). *mab-3*, a gene required for sex-specific yolk protein expression and a male-specific lineage in *C. elegans*. *Cell*, 54(7), 1019–1031. [https://doi.org/10.1016/0092-8674\(88\)90117-1](https://doi.org/10.1016/0092-8674(88)90117-1)

Shivakumara, T. N., Papolu, P. K., Dutta, T. K., Kamaraju, D., Chaudhary, S., & Rao, U. (2016). RNAi-induced silencing of an effector confers transcriptional oscillation in another group of effectors in the root-knot nematode, *Meloidogyne incognita*. *Nematology*, 18(7), 857–870. <https://doi.org/10.1163/15685411-00003003>

Shivakumara, T. N., Somvanshi, V. S., Phani, V., Chaudhary, S., Hada, A., Budhwar, R., Shukla, R. N., & Rao, U. (2019). *Meloidogyne incognita* (Nematoda: Meloidogynidae) sterol-binding protein Mi-SBP-1 as a target for its management. *International Journal for Parasitology*, 49(13–14), 1061–1073. <https://doi.org/10.1016/J.IJPARA.2019.09.002>

Shyu, A. B., Raff, R. A., & Blumenthal, T. (1986). Expression of the vitellogenin gene in female and male sea urchin. *Proceedings of the National Academy of Sciences*, 83(11), 3865–3869. <https://doi.org/10.1073/PNAS.83.11.3865>

Siddique, S., & Grundler, F. M. (2018). Parasitic nematodes manipulate plant development to establish feeding sites. *Current Opinion in Microbiology*, 46, 102–108. <https://doi.org/10.1016/j.mib.2018.09.004>

Siddique, S., Radakovic, Z. S., Hiltl, C., Pellegrin, C., Baum, T. J., Beasley, H., Bent, A. F., Chitambo, O., Chopra, D., Danchin, E. G. J., Grenier, E., Habash, S. S., Hasan, M. S., Helder, J., Hewezi, T., Holbein, J., Holterman, M., Janakowski, S., Koutsovoulos, G. D., ... Eves-van den Akker, S. (2022). 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 B5. *Nature Communications*, 13(1), 1–19. <https://doi.org/10.1038/s41467-022-33769-w>

- Sigma-Aldrich Co. (2014). *Oligo-Evaluator*TM. <https://www.oligoevaluator.com/LoginServlet>
- Sikder, M. M., & Vestergård, M. (2020). Impacts of root metabolites on soil nematodes. *Frontiers in Plant Science*, 10, 1792. <https://doi.org/10.3389/fpls.2019.01792>
- Singh, K. P., Miaskowski, C., Dhruva, A. A., Flowers, E., & Kober, K. M. (2018). Mechanisms and measurement of changes in gene expression. *Biological Research for Nursing*, 20(4), 369. <https://doi.org/10.1177/1099800418772161>
- Singh, P. R., Karssen, G., Couvreur, M., & Bert, W. (2020). Morphological and molecular characterization of *Heterodera dunensis* n. sp. (Nematoda: Heteroderidae) from Gran Canaria, Canary Islands. *Journal of Nematology*, 52, 1–14. <https://doi.org/10.21307/jofnem-2020-098>
- Smolenaars, M. M. W., Madsen, O., Rodenburg, K. W., & Van Der Horst, D. J. (2007). Molecular diversity and evolution of the large lipid transfer protein superfamily. *Journal of Lipid Research*, 48(3), 489–502.
- Sornda, T., Ezcurra, M., Kern, C., Galimov, E. R., Au, C., De La Guardia, Y., & Gems, D. (2019). Production of YP170 vitellogenins promotes intestinal senescence in *Caenorhabditis elegans*. *The Journals of Gerontology: Series A*, 74(8), 1180–1188. <https://doi.org/10.1093/GERONA/GLZ067>
- Sperling, A. L., & Eves-van den Akker, S. (2023). Whole mount multiplexed visualization of DNA, mRNA, and protein in plant-parasitic nematodes. *Plant Methods*, 19(1). <https://doi.org/10.1186/S13007-023-01112-Z>
- Spieth, J., & Blumenthal, T. (1985). The *Caenorhabditis elegans* vitellogenin gene family includes a gene encoding a distantly related protein. *Molecular and Cellular Biology*, 5(10), 2495–2501. <https://doi.org/10.1128/MCB.5.10.2495-2501.1985>
- Spieth, J., Denison, K., Kirtland, S., Cane, J., & Blumenthal, T. (1985). The *C. elegans* vitellogenin genes: Short sequence repeats in the promoter regions and homology to the vertebrate genes. *Nucleic Acids Research*, 13(14), 5283–5295. <https://doi.org/10.1093/nar/13.14.5283>
- Spieth, J., Denison, K., Zucker, E., & Blumenthal, T. (1985). The nucleotide sequence of a nematode vitellogenin gene. *Nucleic Acids Research*, 13(19), 7129–7138. <https://doi.org/10.1093/nar/13.19.7129>
- Starr, J. L., Koenning, S. R., Kirkpatrick, T. L., Robinson, A. F., Roberts, P. A., & Nichols, R. L. (2007). The future of nematode management in cotton. *Journal of Nematology*, 39(4), 283–294.
- Stevanato, P., Trebbi, D., Panella, L., Richardson, K., Broccanello, C., Pakish, L., Fenwick, A. L., & Saccomani, M. (2015). Identification and validation of a SNP marker linked to the gene HsBvm-1 for nematode resistance in sugar beet. *Plant Molecular Biology Reporter*, 33(3), 474–479. <https://doi.org/10.1007/S11105-014-0763-8/METRICS>
- Subbotin, S. A. (2021). Molecular identification of nematodes using polymerase chain reaction (PCR). In R. N. Perry, D. J. Hunt, & S. A. Subbotin (Eds.), *Techniques for Work with Plant and Soil Nematodes* (pp. 218–239). CAB International.

- Subbotin, S. A., Toumi, F., Elekçioğlu, I. H., Waeyenberge, L., & Tanha Maafi, Z. (2018). DNA barcoding, phylogeny and phylogeography of the cyst nematode species of the Avenae group from the genus *Heterodera* (Tylenchida: Heteroderidae). *Nematology*, 20(7), 671–702. <https://doi.org/10.1163/15685411-00003170>
- Subbotin, S. A., Vierstraete, A., De Ley, P., Rowe, J., Waeyenberge, L., Moens, M., & Vanfleteren, J. R. (2001). Phylogenetic relationships within the cyst-forming nematodes (Nematoda, Heteroderidae) based on analysis of sequences from the ITS regions of ribosomal DNA. *Molecular Phylogenetics and Evolution*, 21(1), 1–16. <https://doi.org/10.1006/MPEV.2001.0998>
- Subbotin, S., & Hallmann, J. (2024). Cyst Nematodes. In *Plant Nematology* (3rd ed., pp. 138–175). CAB International.
- Subedi, S., Thapa, B., & Shrestha, J. (2020). Overview of root-knot nematode (*Meloidogyne incognita*) and control management. *Journal of Agriculture and Natural Resources*, 3(2), 21–31. <https://doi.org/10.3126/janr.v3i2.32298>
- Tan, J. A. C. H., Jones, M. G. K., & Fosu-Nyarko, J. (2013). Gene silencing in root lesion nematodes (*Pratylenchus* spp.) significantly reduces reproduction in a plant host. *Experimental Parasitology*, 133(2), 166–178. <https://doi.org/10.1016/J.EXPPARA.2012.11.011>
- The C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, 282(5396), 2012–2018. https://doi.org/10.1126/SCIENCE.282.5396.2012/SUPPL_FILE/C-ELEGANS.XHTML
- Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), 418–426. <https://doi.org/10.1016/J.TIG.2014.07.001>
- Thermo Fisher Scientific. (2025). *Custom peptide synthesis and antigen preparation*. Thermo Fisher Scientific. <https://www.thermofisher.com>
- Thomas, C. G., Li, R., Smith, H. E., Woodruff, G. C., Oliver, B., & Haag, E. S. (2012). Simplification and desexualization of gene expression in self-fertile nematodes. *Current Biology*, 22(22), 2167–2172. <https://doi.org/10.1016/J.CUB.2012.09.038>
- Tian, B., Yang, J., & Zhang, K.-Q. (2007). Bacteria used in the biological control of plant-parasitic nematodes: Populations, mechanisms of action, and future prospects. *FEMS Microbiology Ecology*, 61(2), 197–213. <https://doi.org/10.1111/j.1574-6941.2007.00349.x>
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic acids research*, 40(15), e115–e115.
- van de Walle, P., Geens, E., Baggerman, G., Naranjo-Galindo, F. J., Askjaer, P., Schoofs, L., & Temmerman, L. (2019). CEH-60/PBX regulates vitellogenesis and cuticle permeability through intestinal interaction with UNC-62/MEIS in *Caenorhabditis elegans*. *PLOS Biology*, 17(11), e3000499. <https://doi.org/10.1371/JOURNAL.PBIO.3000499>
- Van Nostrand, E. L., Sánchez-Blanco, A., Wu, B., Nguyen, A., & Kim, S. K. (2013). Roles of the Developmental Regulator unc-62/Homothorax in Limiting Longevity in *Caenorhabditis elegans*. *PLoS Genetics*, 9(2), e1003325. <https://doi.org/10.1371/journal.pgen.1003325>

- 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(12), e0144674. <https://doi.org/10.1371/journal.pone.0144674>
- Vieira, P., Maier, T. R., Eves-Van Den Akker, S., Howe, D. K., Zasada, I., Baum, T. J., Eisenback, J. D., & Kamo, K. (2018). Identification of candidate effector genes of *Pratylenchus penetrans*. *Molecular Plant Pathology*, 19(8), 1887–1907. <https://doi.org/10.1111/mpp.12666>
- Vrablik, T. L., Petyuk, V. A., Larson, E. M., Smith, R. D., & Watts, J. L. (2015). Lipidomic and proteomic analysis of *Caenorhabditis elegans* lipid droplets and identification of ACS-4 as a lipid droplet-associated protein. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1851(10), 1337–1345. <https://doi.org/10.1016/j.bbalip.2015.06.004>
- Waeyenberge, L., Ryss, A., Moens, M., Pinochet, J., & Vrain, T. C. (2000). Molecular characterisation of 18 *Pratylenchus* species using rDNA Restriction Fragment Length Polymorphism. *Nematology*, 2(2), 135–142. <https://doi.org/10.1163/156854100509024>
- Wahli, W., Dawid, I. B., Ryffel, G. U., & Weber, R. (1981). Vitellogenesis and the vitellogenin gene family. *Science*, 212(4492), 298–304. <https://doi.org/10.1126/science.7209528>
- Wallace, R. A. (1985). Vitellogenesis and oocyte growth in nonmammalian vertebrates. *Oogenesis*, 127–177. https://doi.org/10.1007/978-1-4615-6814-8_3
- Wang, D., Shi, X., Long, X., Zhang, Z., Ye, S., Ding, Z., & Yang, Z. (2024). Effect of RNA interference targeting chemosensory genes on chemotactic host-seeking in *Meloidogyne graminicola*. *European Journal of Plant Pathology*, 169(1), 187–200. <https://doi.org/10.1007/s10658-024-02818-z>
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Webster, J. M., & Hooper, D. J. (1968). Serological and morphological studies on the inter- and intraspecific differences of the plant-parasitic nematodes *Heterodera* and *Ditylenchus*. *Parasitology*, 58(4), 879–891. <https://doi.org/10.1017/s0031182000069651>
- Wesemael, W., Moens, M., & Bellafiore, S. (2024). Root-knot nematodes. In R. N. Perry, M. Moens, & J. T. Jones (Eds.), *Plant Nematology* (3rd ed., pp. 101–137). CAB International.
- Wharton, D. (1980). Nematode egg shells. *Parasitology*, 81(2), 447–463.
- Wheeler, D. E., & Kawooya, J. K. (1990). Purification and characterization of honey bee vitellogenin. *Archives of Insect Biochemistry and Physiology*, 14(4), 253–267. <https://doi.org/10.1002/arch.940140405>
- Wileman, H. J., Perry, R. N., & Davies, K. G. (2023). Comparative phylogenetic analysis of vitellogenin in species of cyst and root-knot nematodes. *Nematology*, 25(4), 467–476. <https://doi.org/10.1163/15685411-bja10232>
- Williamson, V. M., & Gleason, C. A. (2003). Plant-nematode interactions. *Current Opinion in Plant Biology*, 6(4), 327–333. [https://doi.org/10.1016/s1369-5266\(03\)00059-1](https://doi.org/10.1016/s1369-5266(03)00059-1)

- Williamson, V. M., & Kumar, A. (2006). Nematode resistance in plants: The battle underground. *Trends in Genetics*, 22(7), 396–403. <https://doi.org/10.1016/j.tig.2006.05.003>
- Winter, C. E. (1992). The yolk polypeptides of a free-living rhabditid nematode. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 103(1), 189–196. [https://doi.org/10.1016/0305-0491\(92\)90430-y](https://doi.org/10.1016/0305-0491(92)90430-y)
- 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(5), 674–684. <https://doi.org/10.1093/oxfordjournals.molbev.a025628>
- Wright, A. J. D., Stevens, M., Back, M. A., & Sparkes, D. L. (2022). A new method to validate and compare varietal resistance and yield tolerance of sugar beet (*Beta vulgaris*) against the beet cyst nematode, *Heterodera schachtii* Schmidt. *Pest Management Science*, 78(7), 2767–2778. <https://doi.org/10.1002/ps.6885>
- Wright, A., Stevens, M., Back, M., & Sparkes, D. (2017). Interactions between *Heterodera schachtii* and sugar beet. In R. Hull, G. Jellis, M. May, P. Miller, S. Moss, C. Nicholls, & J. Orson (Eds.), *Crop Production in Southern Britain* (pp. 81–88). Association of Applied Biologists.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, 13(1), 134.
- Ye, W., Zeng, Y., & Kerns, J. (2015). Molecular characterisation and diagnosis of root-knot nematodes (*Meloidogyne* spp.) from turfgrasses in North Carolina, USA. *PLOS ONE*, 10(11). <https://doi.org/10.1371/journal.pone.0143556>
- Yi, W., & Zarkower, D. (1999). Similarity of DNA binding and transcriptional regulation by *Caenorhabditis elegans* MAB-3 and *Drosophila melanogaster* DSX suggests conservation of sex determining mechanisms. *Development*, 126(5), 873–881. <https://doi.org/10.1242/dev.126.5.873>
- Zhai, C., Zhang, N., Li, X. X., Tan, X. K., Sun, F., & Dong, M. Q. (2024). Immuno-electron microscopy localizes *Caenorhabditis elegans* vitellogenins along the classic exocytosis route. *Life Metabolism*, 3(6). <https://doi.org/10.1093/lifemeta/loae025>
- Zhang, T., Zhang, G., Zeng, F., Mao, J., Liang, H., & Liu, F. (2017). Molecular cloning of the vitellogenin gene and the effects of vitellogenin protein expression on the physiology of *Harmonia axyridis* (Coleoptera: Coccinellidae). *Scientific Reports* 2017, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-14339-3>
- Zhang, Y., Zhao, Q., Zhang, J., Niu, L., Yang, J., Liu, X., Xing, G., Zhong, X., & Yang, X. (2022). Enhanced resistance to soybean cyst nematode in transgenic soybean via host-induced silencing of vital *Heterodera glycines* genes. *Transgenic Research*, 31(2), 239–248. <https://doi.org/10.1007/s11248-022-00298-7>
- Zhao, J., Sun, Y., Xiao, L., Tan, Y., Jiang, Y., & Bai, L. (2018). Vitellogenin and vitellogenin receptor gene expression profiles in *Spodoptera exigua* are related to host plant suitability. *Pest Management Science*, 74(4), 950–958. <https://doi.org/10.1002/ps.4794>

Zhao, S., Ye, Z., & Stanton, R. (2020). Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA*, 26(8), 903. <https://doi.org/10.1261/rna.074922.120>

Zucker-Aprison, E., & Blumenthal, T. (1989). Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. *Journal of Molecular Evolution*, 28(6), 487–496. <https://doi.org/10.1007/bf02602929>

Zunke, U. (1990). Observations on the invasion and endoparasitic behavior of the root lesion nematode *Pratylenchus penetrans*. *Journal of Nematology*, 22(3), 309–320.

Appendices

Supplementary material 3.1: BLAST results using *Caenorhabditis elegans vit* genes as query sequences.

Table S3.1.1: BLAST result. Query sequence: *Caenorhabditis elegans vit* 1-6 gene sequences. Subject: *C. briggsae* genome.

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans vit-1</i>					
X:8485127-8486572	CBG14203	919	1400	0	58.3
X:8483534-8484340	CBG14203	387	655	0	60.6
X:8484578-8485132	CBG14203	737	921	0	44.3
X:8482337-8482789	CBG14203	18	168	0	56.3
X:8486605-8486829	CBG14203	1397	1471	0	69.3
X:8484320-8484571	CBG14203	650	733	0	52.4
X:8483197-8483484	CBG14203	292	387	0	42.7
X:8482772-8483179	CBG14203	166	301	1.20E-72	54.4
X:8487085-8487159	CBG14203	1533	1557	1.20E-72	92
X:8487154-8487267	CBG14203	1557	1594	1.80E-07	55.3
X:8685854-8687185	CBG14234	957	1400	0	59.7
X:8688079-8688900	CBG14234	382	655	0	59.5
X:8687371-8687841	CBG14234	737	893	0	44.6
X:8685597-8685821	CBG14234	1397	1471	0	68
X:8687848-8688099	CBG14234	650	733	0	52.4
X:8687288-8687377	CBG14234	892	921	0	43.3
X:8689623-8690075	CBG14234	18	168	4.50E-134	57
X:8689260-8689640	CBG14234	166	292	4.50E-134	56.7
X:8688904-8689233	CBG14234	286	395	4.50E-134	39.1
X:8685148-8685261	CBG14234	1557	1594	4.50E-134	60.5
X:8685256-8685330	CBG14234	1533	1557	4.50E-134	88
X:8687135-8687293	CBG14234	919	971	4.50E-134	37.7
X:13323465-13324649	CBG16767	650	1044	0	71.6
X:13324887-13325867	CBG16767	1122	1448	0	79.5
X:13322679-13323494	CBG16767	387	658	0	75
X:13321913-13322326	CBG16767	161	298	0	76.8
X:13324650-13324898	CBG16767	1044	1126	0	72.3
X:13326272-13326451	CBG16767	1557	1616	0	61.7
X:13326197-13326277	CBG16767	1531	1557	0	88.9
X:13325912-13325983	CBG16767	1448	1471	0	75
X:13322341-13322664	CBG16767	288	395	9.50E-91	63.9
X:13321631-13321903	CBG16767	64	154	9.50E-91	69.2
X:13321496-13321636	CBG16767	18	64	9.50E-91	76.6
X:10429983-10431950	Cbr-vit-2	387	1042	0	79
X:10432194-10433174	Cbr-vit-2	1122	1448	0	82.3
X:10428781-10429605	Cbr-vit-2	18	292	0	81.1
X:10429648-10429947	Cbr-vit-2	291	390	0	82
X:10431957-10432205	Cbr-vit-2	1044	1126	0	71.1
X:10433207-10433290	Cbr-vit-2	1444	1471	0	71.4
X:10433588-10433767	Cbr-vit-2	1557	1616	1.20E-72	61.7
X:10433513-10433626	Cbr-vit-2	1531	1568	1.60E-09	71.1
IV:11241575-11242066	Cbr-vit-6	487	650	6.20E-105	42.1
IV:11240927-11241226	Cbr-vit-6	766	865	6.20E-105	38
IV:11241314-11241559	Cbr-vit-6	655	736	6.20E-105	34.1
IV:11240567-11240899	Cbr-vit-6	869	979	6.20E-105	30.6
IV:11242121-11242297	Cbr-vit-6	414	472	6.20E-105	37.3
IV:11242301-11242396	Cbr-vit-6	382	413	6.20E-105	31.2
IV:11239853-11240041	Cbr-vit-6	1176	1238	5.10E-21	36.5

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
IV:11243184-11243369	Cbr-vit-6	86	147	5.10E-21	32.3
IV:11239478-11239579	Cbr-vit-6	1312	1345	5.10E-21	50
IV:11243370-11243567	Cbr-vit-6	22	87	5.10E-21	30.3
IV:11239112-11239189	Cbr-vit-6	1435	1460	5.10E-21	53.8
IV:11242991-11243050	Cbr-vit-6	178	197	0.0086	55
IV:11242931-11242972	Cbr-vit-6	202	215	0.0086	78.6
X:13335752-13335976		187	261	1.30E-107	77.3
X:13341795-13342067		64	154	1.30E-107	65.9
X:13334602-13334850		831	913	1.30E-107	50.6
X:13333865-13333987		1133	1173	1.30E-107	75.6
X:13334445-13334531		946	974	1.30E-107	58.6
X:13341702-13341785		161	188	1.30E-107	67.9
X:13333773-13333838		1427	1448	1.30E-107	72.7
X:13334529-13334603		921	945	1.30E-107	52
X:13334302-13334448		896	944	2.00E-14	44.9
X:13334253-13334318		944	965	2.00E-14	59.1
X:13334140-13334184		994	1008	2.00E-14	60
Query Sequence: <i>C. elegans vit-2</i>					
X:8485127-8486572	CBG14203	917	1398	0.00E+00	59.5
X:8483642-8484340	CBG14203	421	653	0.00E+00	62.2
X:8482337-8482789	CBG14203	18	168	0.00E+00	59.6
X:8484578-8485132	CBG14203	735	919	0.00E+00	45.4
X:8486605-8486829	CBG14203	1395	1469	0.00E+00	70.7
X:8484320-8484538	CBG14203	648	720	0.00E+00	54.8
X:8483197-8483484	CBG14203	292	387	0.00E+00	42.7
X:8483534-8483632	CBG14203	387	419	0.00E+00	51.5
X:8482772-8483179	CBG14203	166	301	1.10E-93	55.1
X:8487085-8487261	CBG14203	1531	1589	4.60E-28	69.5
X:8487241-8487330	CBG14203	1582	1611	4.60E-28	56.7
X:8685854-8687185	CBG14234	955	1398	0.00E+00	60.8
X:8688079-8688777	CBG14234	421	653	0.00E+00	62.2
X:8687371-8687841	CBG14234	735	891	0.00E+00	46.5
X:8685597-8685821	CBG14234	1395	1469	0.00E+00	69.3
X:8687881-8688099	CBG14234	648	720	0.00E+00	54.8
X:8688787-8688900	CBG14234	382	419	0.00E+00	42.1
X:8687288-8687377	CBG14234	890	919	0.00E+00	40
X:8689623-8690075	CBG14234	18	168	1.80E-145	58.3
X:8689260-8689640	CBG14234	166	292	1.80E-145	57.5
X:8685154-8685330	CBG14234	1531	1589	1.80E-145	66.1
X:8688904-8689233	CBG14234	286	395	1.80E-145	39.1
X:8687135-8687293	CBG14234	917	969	1.80E-145	39.6
X:8685079-8685174	CBG14234	1582	1613	1.80E-145	53.1
X:13323465-13324649	CBG16767	648	1042	0.00E+00	73.2
X:13324887-13325867	CBG16767	1120	1446	0.00E+00	80.7
X:13322781-13323494	CBG16767	419	656	0.00E+00	78.2
X:13321913-13322326	CBG16767	161	298	0.00E+00	79
X:13326197-13326451	CBG16767	1529	1613	0.00E+00	75.3
X:13324650-13324898	CBG16767	1042	1124	0.00E+00	77.1
X:13325912-13325983	CBG16767	1446	1469	0.00E+00	79.2
X:13322679-13322777	CBG16767	387	419	0.00E+00	51.5
X:13322341-13322664	CBG16767	288	395	2.80E-93	63.9
X:13321631-13321903	CBG16767	64	154	2.80E-93	71.4
X:13321496-13321636	CBG16767	18	64	2.80E-93	83
X:10430091-10431950	Cbr-vit-2	421	1040	0.00E+00	81.1
X:10432194-10433174	Cbr-vit-2	1120	1446	0.00E+00	83.5
X:10428781-10429605	Cbr-vit-2	18	292	0.00E+00	84.7
X:10429648-10429947	Cbr-vit-2	291	390	0.00E+00	83

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
X:10431957-10432205	Cbr-vit-2	1042	1124	0.00E+00	75.9
X:10429983-10430081	Cbr-vit-2	387	419	0.00E+00	72.7
X:10433207-10433290	Cbr-vit-2	1442	1469	0.00E+00	75
X:10433513-10433767	Cbr-vit-2	1529	1613	1.10E-93	89.4
IV:11241575-11242066	Cbr-vit-6	485	648	6.70E-105	42.7
IV:11240927-11241226	Cbr-vit-6	764	863	6.70E-105	39
IV:11241362-11241559	Cbr-vit-6	653	718	6.70E-105	39.4
IV:11240567-11240899	Cbr-vit-6	867	977	6.70E-105	30.6
IV:11239853-11240041	Cbr-vit-6	1174	1236	6.70E-105	36.5
IV:11243184-11243369	Cbr-vit-6	86	147	6.70E-105	32.3
IV:11239478-11239579	Cbr-vit-6	1310	1343	6.70E-105	50
IV:11242121-11242264	Cbr-vit-6	423	470	6.70E-105	35.4
IV:11243370-11243567	Cbr-vit-6	22	87	6.70E-105	30.3
IV:11239112-11239189	Cbr-vit-6	1433	1458	6.70E-105	53.8
IV:11242991-11243050	Cbr-vit-6	178	197	8.70E-03	55
IV:11242931-11242972	Cbr-vit-6	202	215	8.70E-03	78.6
X:13335752-13335976		187	261	6.30E-112	80
X:13341795-13342067		64	154	6.30E-112	69.2
X:13334593-13334850		829	914	6.30E-112	52.3
X:13333865-13333987		1131	1171	6.30E-112	78
X:13334445-13334531		944	972	6.30E-112	58.6
X:13341702-13341785		161	188	6.30E-112	67.9
X:13333773-13333838		1425	1446	6.30E-112	72.7
X:13334529-13334603		919	943	6.30E-112	56
X:13334302-13334448		894	942	1.50E-14	46.9
X:13334253-13334318		942	963	1.50E-14	59.1
X:13334140-13334184		992	1006	1.50E-14	60
Query Sequence: <i>C. elegans vit-3</i>					
X:8483504-8486449	CBG14203	374	1355	0.00E+00	82
X:8482346-8483209	CBG14203	21	308	0.00E+00	77.1
X:8486602-8486844	CBG14203	1392	1472	0.00E+00	86.4
X:8487088-8487330	CBG14203	1521	1601	0.00E+00	75.3
X:8486891-8487031	CBG14203	1472	1518	0.00E+00	63.8
X:8486504-8486593	CBG14203	1374	1403	0.00E+00	46.7
X:8483197-8483514	CBG14203	289	394	2.60E-40	66
X:8687371-8688900	CBG14234	379	888	0.00E+00	82
X:8685977-8687194	CBG14234	950	1355	0.00E+00	81.3
X:8689260-8690066	CBG14234	21	289	0.00E+00	82.5
X:8685582-8685824	CBG14234	1392	1472	0.00E+00	84
X:8688928-8689212	CBG14234	290	384	0.00E+00	67.4
X:8685079-8685327	CBG14234	1521	1603	0.00E+00	74.7
X:8687135-8687386	CBG14234	884	967	0.00E+00	64.3
X:8685407-8685535	CBG14234	1472	1514	0.00E+00	69.8
X:13322679-13323716	CBG16767	384	729	0.00E+00	58.7
X:13324890-13325600	CBG16767	1119	1355	0.00E+00	66.2
X:13323699-13324280	CBG16767	723	916	0.00E+00	47.4
X:13324272-13324649	CBG16767	915	1040	0.00E+00	44.4
X:13322335-13322640	CBG16767	283	384	0.00E+00	44.1
X:13324656-13324898	CBG16767	1042	1122	0.00E+00	54.3
X:13325658-13325867	CBG16767	1375	1444	0.00E+00	58.6
X:13326197-13326379	CBG16767	1518	1578	0.00E+00	63.9
X:13325921-13326097	CBG16767	1447	1505	0.00E+00	57.6
X:13321631-13322323	CBG16767	64	294	8.60E-91	55.8
X:13321505-13321636	CBG16767	21	64	8.60E-91	56.8
X:13326350-13326451	CBG16767	1570	1603	8.20E-50	55.9
V:18570755-18570988	CBG20715	1453	1530	6.80E-04	24.4
X:10429983-10430789	Cbr-vit-2	384	652	0.00E+00	62.1

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
X:10432197-10432907	Cbr-vit-2	1119	1355	0.00E+00	67.1
X:10431006-10431587	Cbr-vit-2	723	916	0.00E+00	48.5
X:10431579-10431950	Cbr-vit-2	915	1038	0.00E+00	46.8
X:10430772-10431023	Cbr-vit-2	646	729	0.00E+00	60.7
X:10429582-10429938	Cbr-vit-2	266	384	0.00E+00	40.3
X:10431957-10432205	Cbr-vit-2	1040	1122	0.00E+00	55.4
X:10432965-10433174	Cbr-vit-2	1375	1444	0.00E+00	58.6
X:10433513-10433695	Cbr-vit-2	1518	1578	0.00E+00	60.7
X:10433207-10433455	Cbr-vit-2	1440	1522	0.00E+00	50.6
X:10433672-10433767	Cbr-vit-2	1572	1603	0.00E+00	56.2
X:10429138-10429605	Cbr-vit-2	134	289	8.20E-50	52.6
X:10428790-10429200	Cbr-vit-2	21	157	1.00E-44	58.4
IV:11241575-11242081	Cbr-vit-6	479	647	3.80E-114	42
IV:11240927-11241235	Cbr-vit-6	758	860	3.80E-114	42.7
IV:11241302-11241580	Cbr-vit-6	644	736	3.80E-114	36.6
IV:11242109-11242300	Cbr-vit-6	410	473	3.80E-114	39.1
IV:11240765-11240902	Cbr-vit-6	863	908	3.80E-114	43.5
IV:11240549-11240749	Cbr-vit-6	915	981	3.80E-114	23.9
IV:11242298-11242396	Cbr-vit-6	379	411	3.80E-114	30.3
IV:11242403-11242444	Cbr-vit-6	331	344	3.80E-114	50
IV:11239853-11240041	Cbr-vit-6	1172	1234	1.70E-18	31.7
IV:11239478-11239585	Cbr-vit-6	1306	1341	1.70E-18	44.4
IV:11239685-11239771	Cbr-vit-6	1256	1284	1.70E-18	44.8
IV:11239112-11239201	Cbr-vit-6	1427	1456	1.40E-08	50
IV:11238873-11239049	Cbr-vit-6	1461	1519	1.40E-08	20.3
IV:11239178-11239336	Cbr-vit-6	1381	1433	1.40E-08	18.9
IV:11243187-11243297	Cbr-vit-6	110	146	2.80E-07	35.1
IV:11243370-11243567	Cbr-vit-6	22	87	2.80E-07	27.3
X:13341702-13342067		64	185	0.00E+00	52.5
X:13335752-13335976		184	258	0.00E+00	60
X:13334587-13334850		826	913	0.00E+00	35.2
X:13333871-13333987		1129	1167	3.60E-11	74.4
X:13333649-13333708		1467	1486	1.40E-05	65
X:13333761-13333838		1423	1448	1.40E-05	46.2
Query Sequence: <i>C. elegans vit-4</i>					
X:8483504-8486449	CBG14203	374	1355	0.00E+00	82.2
X:8482346-8483209	CBG14203	21	308	0.00E+00	75.7
X:8486602-8486844	CBG14203	1392	1472	0.00E+00	86.4
X:8487109-8487330	CBG14203	1528	1601	0.00E+00	77
X:8486891-8486956	CBG14203	1472	1493	0.00E+00	77.3
X:8486504-8486593	CBG14203	1374	1403	0.00E+00	50
X:8483197-8483514	CBG14203	289	394	1.60E-40	66
X:8687371-8688900	CBG14234	379	888	0.00E+00	82
X:8685977-8687194	CBG14234	950	1355	0.00E+00	82
X:8689260-8690066	CBG14234	21	289	0.00E+00	81
X:8685582-8685824	CBG14234	1392	1472	0.00E+00	84
X:8688928-8689212	CBG14234	290	384	0.00E+00	68.4
X:8685079-8685306	CBG14234	1528	1603	0.00E+00	77.6
X:8687135-8687386	CBG14234	884	967	0.00E+00	64.3
X:8685470-8685535	CBG14234	1472	1493	0.00E+00	77.3
X:13322679-13323716	CBG16767	384	729	0.00E+00	58.7
X:13324890-13325600	CBG16767	1119	1355	0.00E+00	66.2
X:13323699-13324280	CBG16767	723	916	0.00E+00	47.4
X:13324272-13324649	CBG16767	915	1040	0.00E+00	44.4
X:13322335-13322640	CBG16767	283	384	0.00E+00	46.1
X:13324656-13324898	CBG16767	1042	1122	0.00E+00	55.6
X:13325658-13325867	CBG16767	1375	1444	0.00E+00	60

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
X:13325921-13326046	CBG16767	1447	1488	0.00E+00	81
X:13321631-13322323	CBG16767	64	294	1.00E-99	55.4
X:13326230-13326379	CBG16767	1529	1578	1.00E-99	68
X:13321505-13321636	CBG16767	21	64	1.00E-99	56.8
X:13326350-13326451	CBG16767	1570	1603	1.50E-45	55.9
X:10429983-10430789	Cbr-vit-2	384	652	0.00E+00	62.1
X:10432197-10432907	Cbr-vit-2	1119	1355	0.00E+00	67.1
X:10431006-10431587	Cbr-vit-2	723	916	0.00E+00	48.5
X:10431579-10431950	Cbr-vit-2	915	1038	0.00E+00	47.6
X:10430772-10431023	Cbr-vit-2	646	729	0.00E+00	60.7
X:10429582-10429938	Cbr-vit-2	266	384	0.00E+00	42
X:10431957-10432205	Cbr-vit-2	1040	1122	0.00E+00	56.6
X:10432965-10433174	Cbr-vit-2	1375	1444	0.00E+00	60
X:10433207-10433353	Cbr-vit-2	1440	1488	0.00E+00	71.4
X:10429138-10429605	Cbr-vit-2	134	289	2.40E-66	51.9
X:10433543-10433695	Cbr-vit-2	1528	1578	2.40E-66	62.7
X:10433672-10433767	Cbr-vit-2	1572	1603	2.40E-66	56.2
X:10428790-10429200	Cbr-vit-2	21	157	1.50E-45	58.4
IV:11241575-11242081	Cbr-vit-6	479	647	9.00E-115	42
IV:11240927-11241235	Cbr-vit-6	758	860	9.00E-115	42.7
IV:11241302-11241580	Cbr-vit-6	644	736	9.00E-115	36.6
IV:11242109-11242300	Cbr-vit-6	410	473	9.00E-115	39.1
IV:11240765-11240902	Cbr-vit-6	863	908	9.00E-115	43.5
IV:11240549-11240749	Cbr-vit-6	915	981	9.00E-115	23.9
IV:11242298-11242396	Cbr-vit-6	379	411	9.00E-115	30.3
IV:11242403-11242444	Cbr-vit-6	331	344	9.00E-115	57.1
IV:11239853-11240041	Cbr-vit-6	1172	1234	1.10E-18	31.7
IV:11239478-11239585	Cbr-vit-6	1306	1341	1.10E-18	44.4
IV:11239685-11239771	Cbr-vit-6	1256	1284	1.10E-18	44.8
IV:11243187-11243297	Cbr-vit-6	110	146	2.60E-07	35.1
IV:11243370-11243567	Cbr-vit-6	22	87	2.60E-07	27.3
IV:11239112-11239201	Cbr-vit-6	1427	1456	8.20E-04	50
IV:11239178-11239306	Cbr-vit-6	1391	1433	8.20E-04	23.3
X:13341702-13342067		64	185	0.00E+00	52.5
X:13335752-13335976		184	258	0.00E+00	60
X:13334587-13334850		826	913	0.00E+00	35.2
X:13333871-13333987		1129	1167	3.30E-11	74.4
X:13333649-13333708		1467	1486	2.30E-05	65
X:13333761-13333838		1423	1448	2.30E-05	46.2
Query Sequence: <i>C. elegans vit-5</i>					
X:8483504-8486449	CBG14203	374	1355	0.00E+00	81.9
X:8482346-8483209	CBG14203	21	308	0.00E+00	77.8
X:8486602-8486844	CBG14203	1392	1472	0.00E+00	87.7
X:8487109-8487330	CBG14203	1528	1601	0.00E+00	75.7
X:8486891-8486956	CBG14203	1472	1493	0.00E+00	86.4
X:8486504-8486593	CBG14203	1374	1403	0.00E+00	50
X:8483197-8483514	CBG14203	289	394	2.60E-41	67
X:8687371-8688900	CBG14234	379	888	0.00E+00	81.6
X:8685977-8687194	CBG14234	950	1355	0.00E+00	81.8
X:8689260-8690066	CBG14234	21	289	0.00E+00	82.2
X:8685582-8685824	CBG14234	1392	1472	0.00E+00	85.2
X:8688928-8689212	CBG14234	290	384	0.00E+00	68.4
X:8685079-8685306	CBG14234	1528	1603	0.00E+00	76.3
X:8687135-8687386	CBG14234	884	967	0.00E+00	64.3
X:8685470-8685535	CBG14234	1472	1493	0.00E+00	86.4
X:13322679-13323716	CBG16767	384	729	0.00E+00	59
X:13324890-13325600	CBG16767	1119	1355	0.00E+00	65

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
X:13323699-13324280	CBG16767	723	916	0.00E+00	47.9
X:13324272-13324649	CBG16767	915	1040	0.00E+00	44.4
X:13322335-13322640	CBG16767	283	384	0.00E+00	45.1
X:13325658-13325867	CBG16767	1375	1444	0.00E+00	60
X:13324656-13324898	CBG16767	1042	1122	0.00E+00	54.3
X:13325921-13326046	CBG16767	1447	1488	0.00E+00	78.6
X:13321631-13322323	CBG16767	64	294	5.10E-107	55.8
X:13326230-13326379	CBG16767	1529	1578	5.10E-107	68
X:13321505-13321636	CBG16767	21	64	5.10E-107	59.1
X:13326350-13326451	CBG16767	1570	1603	9.60E-50	55.9
X:10429983-10430789	Cbr-vit-2	384	652	0.00E+00	61.7
X:10432197-10432907	Cbr-vit-2	1119	1355	0.00E+00	65.8
X:10431006-10431587	Cbr-vit-2	723	916	0.00E+00	47.9
X:10431579-10431950	Cbr-vit-2	915	1038	0.00E+00	47.6
X:10430772-10431023	Cbr-vit-2	646	729	0.00E+00	60.7
X:10429582-10429938	Cbr-vit-2	266	384	0.00E+00	41.2
X:10431957-10432205	Cbr-vit-2	1040	1122	0.00E+00	55.4
X:10432965-10433174	Cbr-vit-2	1375	1444	0.00E+00	60
X:10433207-10433353	Cbr-vit-2	1440	1488	0.00E+00	73.5
X:10428790-10429200	Cbr-vit-2	21	157	1.10E-69	59.9
X:10433543-10433695	Cbr-vit-2	1528	1578	1.10E-69	62.7
X:10433672-10433767	Cbr-vit-2	1572	1603	1.10E-69	59.4
X:10429138-10429605	Cbr-vit-2	134	289	9.60E-50	51.9
IV:11241575-11242081	Cbr-vit-6	479	647	9.40E-114	42.6
IV:11240927-11241235	Cbr-vit-6	758	860	9.40E-114	41.7
IV:11241302-11241580	Cbr-vit-6	644	736	9.40E-114	36.6
IV:11242109-11242300	Cbr-vit-6	410	473	9.40E-114	39.1
IV:11240765-11240902	Cbr-vit-6	863	908	9.40E-114	43.5
IV:11240549-11240749	Cbr-vit-6	915	981	9.40E-114	23.9
IV:11242298-11242396	Cbr-vit-6	379	411	9.40E-114	30.3
IV:11242403-11242444	Cbr-vit-6	331	344	9.40E-114	50
IV:11239853-11240041	Cbr-vit-6	1172	1234	1.60E-18	31.7
IV:11239478-11239585	Cbr-vit-6	1306	1341	1.60E-18	44.4
IV:11239685-11239771	Cbr-vit-6	1256	1284	1.60E-18	44.8
IV:11243187-11243342	Cbr-vit-6	95	146	3.40E-10	30.8
IV:11243370-11243567	Cbr-vit-6	22	87	3.40E-10	28.8
IV:11239112-11239201	Cbr-vit-6	1427	1456	9.90E-08	50
IV:11238966-11239049	Cbr-vit-6	1461	1488	9.90E-08	32.1
IV:11239178-11239306	Cbr-vit-6	1391	1433	9.90E-08	23.3
X:13341702-13342067		64	185	0.00E+00	53.3
X:13335752-13335976		184	258	0.00E+00	60
X:13334587-13334850		826	913	0.00E+00	35.2
X:13333865-13333987		1129	1169	8.80E-11	68.3
X:13333649-13333708		1467	1486	1.30E-05	65
X:13333761-13333838		1423	1448	1.30E-05	46.2
Query Sequence: <i>C. elegans vit-6</i>					
X:15233032-15233181	CBG00277	1509	1558	1.50E-98	38
I:4189092-4189244	CBG04160, Cbr-kcnl-2	1517	1567	8.50E-03	35.3
X:8483834-8484325	CBG14203	513	676	1.30E-100	42.1
X:8484659-8484964	CBG14203	791	892	1.30E-100	31.4
X:8484314-8484610	CBG14203	675	773	1.30E-100	32.3
X:8483612-8483803	CBG14203	435	498	1.30E-100	34.4
X:8485154-8485318	CBG14203	961	1015	1.30E-100	32.7
X:8484971-8485108	CBG14203	901	946	1.30E-100	41.3
X:8486297-8486407	CBG14203	1339	1375	5.30E-18	43.2
X:8485892-8486086	CBG14203	1186	1250	5.30E-18	29.2
X:8486099-8486236	CBG14203	1261	1306	5.30E-18	26.1

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
X:8482355-8482525	CBG14203	34	90	1.00E-15	29.8
X:8482487-8482723	CBG14203	80	158	3.10E-15	31.6
X:8482796-8482855	CBG14203	181	200	3.10E-15	60
X:8482850-8482915	CBG14203	201	222	3.10E-15	40.9
X:8688094-8688585	CBG14234	513	676	9.50E-88	42.7
X:8687455-8687763	CBG14234	790	892	9.50E-88	30.1
X:8687809-8688105	CBG14234	675	773	9.50E-88	32.3
X:8686019-8686129	CBG14234	1339	1375	9.50E-88	43.2
X:8688616-8688807	CBG14234	435	498	9.50E-88	31.2
X:8686340-8686534	CBG14234	1186	1250	9.50E-88	29.2
X:8685630-8685719	CBG14234	1468	1497	9.50E-88	50
X:8689557-8689667	CBG14234	164	200	9.50E-88	37.8
X:8686190-8686327	CBG14234	1261	1306	9.50E-88	26.1
X:8689689-8689925	CBG14234	80	158	5.30E-05	30.4
X:13322979-13323470	CBG16767	513	676	1.50E-98	39.6
X:13323813-13324112	CBG16767	793	892	1.50E-98	40
X:13323504-13323698	CBG16767	690	754	1.50E-98	41.5
X:13321649-13321882	CBG16767	82	159	1.50E-98	29.5
X:13325424-13325558	CBG16767	1331	1375	1.50E-98	44.4
X:13322766-13322936	CBG16767	438	494	1.50E-98	31.6
X:13325067-13325237	CBG16767	1194	1250	1.50E-98	29.8
X:13324122-13324259	CBG16767	902	947	1.50E-98	37
X:13322476-13322640	CBG16767	350	404	1.50E-98	32.7
X:13325250-13325387	CBG16767	1261	1306	1.50E-98	30.4
X:13321895-13322023	CBG16767	162	204	1.50E-98	37.2
I:7774747-7774923	Cbr-unc-73	1509	1567	1.70E-03	33.9
X:10430283-10430774	Cbr-vit-2	513	676	2.40E-102	40.2
X:10431120-10431419	Cbr-vit-2	793	892	2.40E-102	41
X:10430787-10431032	Cbr-vit-2	682	763	2.40E-102	37.8
X:10430064-10430240	Cbr-vit-2	436	494	2.40E-102	35.6
X:10431429-10431554	Cbr-vit-2	902	943	2.40E-102	45.2
X:10429983-10430063	Cbr-vit-2	408	434	2.40E-102	29.6
X:10428790-10428969	Cbr-vit-2	31	90	1.50E-98	30
X:10432731-10432865	Cbr-vit-2	1331	1375	6.30E-19	42.2
X:10432374-10432544	Cbr-vit-2	1194	1250	6.30E-19	28.1
X:10432557-10432694	Cbr-vit-2	1261	1306	6.30E-19	28.3
X:10428985-10429170	Cbr-vit-2	98	159	1.00E-15	33.9
X:10429216-10429320	Cbr-vit-2	170	204	1.00E-15	42.9
IV:11239100-11243068	Cbr-vit-6	179	1501	0.00E+00	89
IV:11243106-11243576	Cbr-vit-6	29	185	0.00E+00	86.6
IV:11238780-11239046	Cbr-vit-6	1502	1590	0.00E+00	75.3
IV:11238555-11238734	Cbr-vit-6	1591	1650	0.00E+00	81.7
X:13341816-13342049		82	159	9.50E-88	26.9
X:7247588-7247734		1522	1570	9.50E-88	26.5
X:21468486-21468572		1313	1341	1.00E-15	62.1
X:21468377-21468475		1278	1310	1.00E-15	51.5
X:21468115-21468180		1100	1121	1.00E-15	68.2

Table S3.1.2: BLAST result. Query sequence: *Caenorhabditis elegans* vit 1-6 gene sequences. Subject: *Globodera pallida* genome.

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans</i> vit-1					
scaffold_2:257050-257448	GPLIN_000012800	501	633	8.00E-34	32.3
scaffold_2:259164-259316	GPLIN_000012800	94	144	8.00E-34	39.2
scaffold_2:255139-255393	GPLIN_000012800	1055	1139	8.00E-34	24.7
scaffold_2:256249-256416	GPLIN_000012800	814	869	8.00E-34	28.6
scaffold_2:254597-254686	GPLIN_000012800	1259	1288	8.00E-34	23.3
scaffold_2:253030-253131	GPLIN_000012800	1564	1597	8.00E-34	35.3
scaffold_152:174962-175093	GPLIN_000564000	384	427	1.30E-11	43.2
scaffold_152:173967-174050	GPLIN_000564000	200	227	1.30E-11	50
scaffold_152:181240-181320	GPLIN_000564000	1566	1592	1.30E-11	51.9
scaffold_152:180726-180821	GPLIN_000564000	1440	1471	1.30E-11	37.5
scaffold_152:180083-180136	GPLIN_000564000	1328	1345	1.30E-11	50
scaffold_221:78608-79051	GPLIN_000702400	486	633	1.40E-31	30.4
scaffold_221:74514-74600	GPLIN_000702400	1564	1592	1.40E-31	48.3
scaffold_221:80777-80839	GPLIN_000702400	127	147	1.40E-31	57.1
scaffold_221:74673-74750	GPLIN_000702400	1535	1560	1.40E-31	42.3
scaffold_221:77755-77922	GPLIN_000702400	814	869	1.40E-31	28.6
scaffold_221:76246-76350	GPLIN_000702400	1254	1288	1.40E-31	22.9
scaffold_221:80434-80541	GPLIN_000702400	193	228	1.40E-31	30.6
scaffold_377:31179-31541	GPLIN_000945900	508	628	1.90E-42	36.4
scaffold_377:31759-31947	GPLIN_000945900	414	476	1.90E-42	28.6
scaffold_377:33358-33522	GPLIN_000945900	96	150	1.90E-42	34.5
scaffold_377:27250-27318	GPLIN_000945900	1535	1557	1.90E-42	52.2
scaffold_377:32966-33073	GPLIN_000945900	195	230	1.90E-42	33.3
scaffold_377:27082-27165	GPLIN_000945900	1565	1592	1.90E-42	42.9
scaffold_377:30378-30461	GPLIN_000945900	835	862	1.90E-42	35.7
scaffold_377:28022-28075	GPLIN_000945900	1435	1452	1.90E-42	55.6
Query Sequence: <i>C. elegans</i> vit-2					
scaffold_2:257050-257448	GPLIN_000012800	499	631	1.50E-32	31.6
scaffold_2:259164-259316	GPLIN_000012800	94	144	1.50E-32	39.2
scaffold_2:255139-255393	GPLIN_000012800	1053	1137	1.50E-32	25.9
scaffold_2:256249-256416	GPLIN_000012800	812	867	1.50E-32	30.4
scaffold_2:254597-254686	GPLIN_000012800	1257	1286	1.50E-32	20
scaffold_152:173967-174050	GPLIN_000564000	200	227	3.90E-06	50
scaffold_152:180726-180821	GPLIN_000564000	1438	1469	3.90E-06	40.6
scaffold_152:181207-181317	GPLIN_000564000	1552	1588	3.90E-06	32.4
scaffold_152:180083-180136	GPLIN_000564000	1326	1343	3.90E-06	50
scaffold_221:78608-79051	GPLIN_000702400	484	631	4.90E-31	30.4
scaffold_221:74679-74750	GPLIN_000702400	1533	1556	4.90E-31	45.8
scaffold_221:80777-80839	GPLIN_000702400	127	147	4.90E-31	57.1
scaffold_221:77755-77922	GPLIN_000702400	812	867	4.90E-31	30.4
scaffold_221:76246-76350	GPLIN_000702400	1252	1286	4.90E-31	20
scaffold_221:80434-80541	GPLIN_000702400	193	228	4.90E-31	30.6

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
scaffold_221:74517-74600	GPLIN_000702400	1561	1588	4.90E-31	35.7
scaffold_377:31179-31541	GPLIN_000945900	506	626	6.90E-40	35.5
scaffold_377:33358-33522	GPLIN_000945900	96	150	6.90E-40	34.5
scaffold_377:31759-31920	GPLIN_000945900	421	474	6.90E-40	29.6
scaffold_377:27244-27318	GPLIN_000945900	1533	1557	6.90E-40	52
scaffold_377:32966-33073	GPLIN_000945900	195	230	6.90E-40	33.3
scaffold_377:30369-30461	GPLIN_000945900	833	863	6.90E-40	35.5
scaffold_377:28022-28075	GPLIN_000945900	1433	1450	6.90E-40	55.6
Query Sequence: <i>C. elegans vit-3</i>					
scaffold_2:257050-257448	GPLIN_000012800	498	630	1.20E-31	30.8
scaffold_2:255139-255267	GPLIN_000012800	1093	1135	1.20E-31	32.6
scaffold_2:256261-256365	GPLIN_000012800	826	860	1.20E-31	40
scaffold_2:259158-259268	GPLIN_000012800	110	146	1.20E-31	35.1
scaffold_2:257639-257815	GPLIN_000012800	408	466	1.20E-31	27.1
scaffold_152:174995-175093	GPLIN_000564000	392	424	3.80E-09	45.5
scaffold_152:180726-180845	GPLIN_000564000	1436	1475	3.80E-09	37.5
scaffold_152:173970-174050	GPLIN_000564000	198	224	3.80E-09	44.4
scaffold_152:181225-181341	GPLIN_000564000	1547	1585	3.80E-09	35.9
scaffold_152:175121-175210	GPLIN_000564000	431	460	3.80E-09	43.3
scaffold_152:175266-175364	GPLIN_000564000	392	424	2.70E-05	45.5
scaffold_152:175392-175481	GPLIN_000564000	431	460	2.70E-05	43.3
scaffold_221:78623-79027	GPLIN_000702400	491	625	4.20E-27	31.9
scaffold_221:77767-77871	GPLIN_000702400	826	860	4.20E-27	42.9
scaffold_221:74670-74750	GPLIN_000702400	1522	1548	4.20E-27	44.4
scaffold_221:80771-80836	GPLIN_000702400	128	149	4.20E-27	40.9
scaffold_221:79258-79398	GPLIN_000702400	420	466	4.20E-27	31.9
scaffold_377:31164-31541	GPLIN_000945900	505	630	3.00E-37	35.7
scaffold_377:31780-31947	GPLIN_000945900	411	466	3.00E-37	28.6
scaffold_377:27247-27318	GPLIN_000945900	1522	1545	3.00E-37	54.2
scaffold_377:30369-30473	GPLIN_000945900	826	860	3.00E-37	40
scaffold_377:33370-33480	GPLIN_000945900	110	146	3.00E-37	29.7
scaffold_377:28700-28804	GPLIN_000945900	1255	1289	3.00E-37	31.4
Query Sequence: <i>C. elegans vit-4</i>					
scaffold_2:257050-257448	GPLIN_000012800	498	630	4.00E-32	30.8
scaffold_2:255139-255267	GPLIN_000012800	1093	1135	4.00E-32	34.9
scaffold_2:256261-256365	GPLIN_000012800	826	860	4.00E-32	40
scaffold_2:259158-259268	GPLIN_000012800	110	146	4.00E-32	35.1
scaffold_2:257639-257815	GPLIN_000012800	408	466	4.00E-32	27.1
scaffold_152:174995-175093	GPLIN_000564000	392	424	4.40E-09	45.5
scaffold_152:180726-180845	GPLIN_000564000	1436	1475	4.40E-09	37.5
scaffold_152:173970-174050	GPLIN_000564000	198	224	4.40E-09	44.4
scaffold_152:181225-181341	GPLIN_000564000	1547	1585	4.40E-09	35.9
scaffold_152:175121-175210	GPLIN_000564000	431	460	4.40E-09	43.3
scaffold_152:175266-175364	GPLIN_000564000	392	424	2.50E-05	45.5
scaffold_152:175392-175481	GPLIN_000564000	431	460	2.50E-05	43.3

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
scaffold_221:78623-79027	GPLIN_000702400	491	625	2.40E-27	31.9
scaffold_221:77767-77871	GPLIN_000702400	826	860	2.40E-27	42.9
scaffold_221:80771-80836	GPLIN_000702400	128	149	2.40E-27	40.9
scaffold_221:79258-79398	GPLIN_000702400	420	466	2.40E-27	31.9
scaffold_221:76852-76926	GPLIN_000702400	1098	1122	2.40E-27	40
scaffold_221:74517-74609	GPLIN_000702400	1547	1577	2.40E-27	32.3
scaffold_377:31164-31541	GPLIN_000945900	505	630	5.30E-35	35.7
scaffold_377:31780-31947	GPLIN_000945900	411	466	5.30E-35	28.6
scaffold_377:30369-30473	GPLIN_000945900	826	860	5.30E-35	40
scaffold_377:33370-33480	GPLIN_000945900	110	146	5.30E-35	29.7
scaffold_377:28700-28804	GPLIN_000945900	1255	1289	5.30E-35	31.4
scaffold_377:27247-27297	GPLIN_000945900	1529	1545	5.30E-35	58.8
Query Sequence: <i>C. elegans vit-5</i>					
scaffold_2:257050-257448	GPLIN_000012800	498	630	5.20E-33	31.6
scaffold_2:259158-259301	GPLIN_000012800	99	146	5.20E-33	33.3
scaffold_2:255139-255267	GPLIN_000012800	1093	1135	5.20E-33	32.6
scaffold_2:256261-256365	GPLIN_000012800	826	860	5.20E-33	40
scaffold_2:257639-257815	GPLIN_000012800	408	466	5.20E-33	27.1
scaffold_152:174995-175093	GPLIN_000564000	392	424	2.60E-09	45.5
scaffold_152:180726-180845	GPLIN_000564000	1436	1475	2.60E-09	37.5
scaffold_152:181225-181341	GPLIN_000564000	1547	1585	2.60E-09	35.9
scaffold_152:173970-174050	GPLIN_000564000	198	224	2.60E-09	44.4
scaffold_152:175121-175210	GPLIN_000564000	431	460	2.60E-09	43.3
scaffold_152:175266-175364	GPLIN_000564000	392	424	2.60E-05	45.5
scaffold_152:175392-175481	GPLIN_000564000	431	460	2.60E-05	43.3
scaffold_221:78623-79027	GPLIN_000702400	491	625	4.00E-26	32.6
scaffold_221:77767-77871	GPLIN_000702400	826	860	4.00E-26	42.9
scaffold_221:80771-80836	GPLIN_000702400	128	149	4.00E-26	40.9
scaffold_221:79258-79398	GPLIN_000702400	420	466	4.00E-26	31.9
scaffold_221:74517-74609	GPLIN_000702400	1547	1577	4.00E-26	32.3
scaffold_377:31164-31541	GPLIN_000945900	505	630	2.90E-36	36.5
scaffold_377:33370-33522	GPLIN_000945900	96	146	2.90E-36	27.5
scaffold_377:31780-31947	GPLIN_000945900	411	466	2.90E-36	28.6
scaffold_377:30369-30473	GPLIN_000945900	826	860	2.90E-36	40
scaffold_377:28700-28804	GPLIN_000945900	1255	1289	2.90E-36	31.4
scaffold_377:27247-27297	GPLIN_000945900	1529	1545	2.90E-36	58.8
Query Sequence: <i>C. elegans vit-6</i>					
scaffold_2:257056-257460	GPLIN_000012800	523	657	2.70E-52	31.1
scaffold_2:257627-257776	GPLIN_000012800	446	495	2.70E-52	44
scaffold_2:253878-254015	GPLIN_000012800	1444	1489	2.70E-52	39.1
scaffold_2:259152-259286	GPLIN_000012800	116	160	2.70E-52	37.8
scaffold_2:256213-256416	GPLIN_000012800	841	908	2.70E-52	27.9
scaffold_2:258820-258921	GPLIN_000012800	187	220	2.70E-52	41.2
scaffold_2:254588-254692	GPLIN_000012800	1275	1309	2.70E-52	34.3
scaffold_2:253180-253260	GPLIN_000012800	1555	1581	2.70E-52	40.7

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
scaffold_2:251809-251946	GPLIN_000012800	1444	1489	6.50E-08	39.1
scaffold_2:252513-252617	GPLIN_000012800	1275	1309	6.50E-08	34.3
scaffold_2:251113-251193	GPLIN_000012800	1555	1581	6.50E-08	40.7
scaffold_377:31173-31541	GPLIN_000945900	534	656	4.20E-46	33.3
scaffold_377:31768-31947	GPLIN_000945900	436	495	4.20E-46	33.3
scaffold_377:28022-28138	GPLIN_000945900	1451	1489	4.20E-46	41
scaffold_377:33328-33489	GPLIN_000945900	119	172	4.20E-46	31.5
scaffold_377:30321-30473	GPLIN_000945900	858	908	4.20E-46	27.5
scaffold_377:28442-28486	GPLIN_000945900	1364	1378	4.20E-46	73.3
scaffold_377:33062-33130	GPLIN_000945900	187	209	4.20E-46	34.8
scaffold_221:78614-79030	GPLIN_000702400	519	657	7.10E-49	29.5
scaffold_221:79246-79395	GPLIN_000702400	446	495	7.10E-49	44
scaffold_221:75503-75640	GPLIN_000702400	1444	1489	7.10E-49	39.1
scaffold_221:80485-80586	GPLIN_000702400	187	220	7.10E-49	44.1
scaffold_221:77719-77922	GPLIN_000702400	841	908	7.10E-49	26.5
scaffold_221:80777-80839	GPLIN_000702400	139	159	7.10E-49	57.1
scaffold_221:76237-76347	GPLIN_000702400	1273	1309	7.10E-49	27
scaffold_221:75934-75978	GPLIN_000702400	1364	1378	7.10E-49	73.3

Table S3.1.3: BLAST result. Query sequence: *Caenorhabditis elegans vit* 1-6 gene sequences. Subject: *Globodera rostochiensis* genome.

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans vit</i> -1					
GROS_00031:120953-121336	GROS_g02537	501	628	2.10E-41	33.6
GROS_00031:123265-123456	GROS_g02537	1090	1153	2.10E-41	31.2
GROS_00031:119060-119212	GROS_g02537	94	144	2.10E-41	39.2
GROS_00031:125401-125505	GROS_g02537	1563	1597	2.10E-41	42.9
GROS_00031:121989-122156	GROS_g02537	814	869	2.10E-41	28.6
GROS_00031:119520-119606	GROS_g02537	200	228	2.10E-41	37.9
GROS_00031:123866-123970	GROS_g02537	1259	1293	2.10E-41	22.9
GROS_00034:70329-70751	GROS_g02714	488	628	5.30E-37	31.9
GROS_00034:70907-71095	GROS_g02714	414	476	5.30E-37	30.2
GROS_00034:72490-72636	GROS_g02714	96	144	5.30E-37	34.7
GROS_00034:66353-66421	GROS_g02714	1535	1557	5.30E-37	52.2
GROS_00034:66179-66262	GROS_g02714	1565	1592	5.30E-37	42.9
GROS_00034:72094-72201	GROS_g02714	195	230	5.30E-37	33.3
GROS_00039:12019-12099	GROS_g03015	1566	1592	8.50E-11	51.9
GROS_00039:12537-12632	GROS_g03015	1440	1471	8.50E-11	37.5
GROS_00039:13218-13271	GROS_g03015	1328	1345	8.50E-11	50
GROS_00039:20230-20361	GROS_g03016	384	427	8.50E-11	43.2
GROS_00039:22732-22815	GROS_g03016	200	227	8.50E-11	46.4
GROS_00039:17973-18104	GROS_g03016	384	427	6.50E-05	43.2
GROS_00039:21882-21965	GROS_g03016	200	227	6.50E-05	46.4
GROS_00136:83588-84031	GROS_g06238	486	633	3.00E-38	31.1

GROS_00136:81585-81881	GROS_g06238	1055	1153	3.00E-38	26.3
GROS_00136:79438-79515	GROS_g06238	1535	1560	3.00E-38	46.2
GROS_00136:85745-85807	GROS_g06238	127	147	3.00E-38	57.1
GROS_00136:79279-79365	GROS_g06238	1564	1592	3.00E-38	44.8
GROS_00136:82728-82895	GROS_g06238	814	869	3.00E-38	28.6
GROS_00136:85402-85488	GROS_g06238	200	228	3.00E-38	34.5
Query Sequence: <i>C. elegans vit-2</i>					
GROS_00031:120953-121336	GROS_g02537	499	626	1.50E-38	32.8
GROS_00031:123265-123456	GROS_g02537	1088	1151	1.50E-38	31.2
GROS_00031:119060-119212	GROS_g02537	94	144	1.50E-38	39.2
GROS_00031:121989-122156	GROS_g02537	812	867	1.50E-38	30.4
GROS_00031:119520-119606	GROS_g02537	200	228	1.50E-38	37.9
GROS_00031:123866-123970	GROS_g02537	1257	1291	1.50E-38	20
GROS_00034:70329-70751	GROS_g02714	486	626	7.20E-35	31.2
GROS_00034:72490-72636	GROS_g02714	96	144	7.20E-35	34.7
GROS_00034:70907-71068	GROS_g02714	421	474	7.20E-35	31.5
GROS_00034:66350-66421	GROS_g02714	1533	1556	7.20E-35	54.2
GROS_00034:72094-72201	GROS_g02714	195	230	7.20E-35	33.3
GROS_00034:69522-69614	GROS_g02714	833	863	7.20E-35	35.5
GROS_00039:12022-12132	GROS_g03015	1552	1588	1.20E-06	37.8
GROS_00039:12537-12632	GROS_g03015	1438	1469	1.20E-06	40.6
GROS_00039:13218-13271	GROS_g03015	1326	1343	1.20E-06	50
GROS_00039:22732-22815	GROS_g03016	200	227	1.20E-06	46.4
GROS_00039:21882-21965	GROS_g03016	200	227	0.006	46.4
GROS_00039:21969-22028	GROS_g03016	180	199	0.006	30
GROS_00136:83588-84031	GROS_g06238	484	631	2.90E-36	31.1
GROS_00136:81585-81776	GROS_g06238	1088	1151	2.90E-36	32.8
GROS_00136:79444-79515	GROS_g06238	1533	1556	2.90E-36	50
GROS_00136:85745-85807	GROS_g06238	127	147	2.90E-36	57.1
GROS_00136:82728-82895	GROS_g06238	812	867	2.90E-36	30.4
GROS_00136:85402-85488	GROS_g06238	200	228	2.90E-36	34.5
Query Sequence: <i>C. elegans vit-3</i>					
GROS_00031:120953-121336	GROS_g02537	498	625	4.00E-38	32
GROS_00031:123286-123483	GROS_g02537	1093	1158	4.00E-38	30.3
GROS_00031:122040-122144	GROS_g02537	826	860	4.00E-38	42.9
GROS_00031:119108-119227	GROS_g02537	110	149	4.00E-38	35
GROS_00031:120542-120763	GROS_g02537	393	466	4.00E-38	24.3
GROS_00034:70314-70691	GROS_g02714	505	630	2.40E-34	34.9
GROS_00034:66350-66421	GROS_g02714	1522	1545	2.40E-34	54.2
GROS_00034:69522-69626	GROS_g02714	826	860	2.40E-34	40
GROS_00034:70928-71095	GROS_g02714	411	466	2.40E-34	28.6
GROS_00034:72484-72594	GROS_g02714	110	146	2.40E-34	29.7
GROS_00034:67854-67958	GROS_g02714	1255	1289	2.40E-34	34.3
GROS_00039:11998-12114	GROS_g03015	1547	1585	3.70E-10	38.5
GROS_00039:12513-12632	GROS_g03015	1436	1475	3.70E-10	37.5
GROS_00039:22732-22812	GROS_g03016	198	224	3.70E-10	48.1

GROS_00039:20230-20328	GROS_g03016	392	424	3.70E-10	45.5
GROS_00039:17856-17945	GROS_g03016	431	460	3.70E-10	43.3
GROS_00039:20717-20815	GROS_g03016	392	424	2.10E-05	45.5
GROS_00039:20843-20932	GROS_g03016	431	460	2.10E-05	43.3
GROS_00039:21882-21962	GROS_g03016	198	224	0.0043	48.1
GROS_00039:17973-18071	GROS_g03016	392	424	0.0043	45.5
GROS_00136:83603-84007	GROS_g06238	491	625	3.20E-35	31.1
GROS_00136:81558-81755	GROS_g06238	1093	1158	3.20E-35	31.8
GROS_00136:82740-82844	GROS_g06238	826	860	3.20E-35	42.9
GROS_00136:79435-79515	GROS_g06238	1522	1548	3.20E-35	48.1
GROS_00136:85739-85804	GROS_g06238	128	149	3.20E-35	40.9
GROS_00136:84230-84370	GROS_g06238	420	466	3.20E-35	31.9
Query Sequence: <i>C. elegans vit-4</i>					
GROS_00031:120953-121336	GROS_g02537	498	625	1.20E-38	32
GROS_00031:123286-123483	GROS_g02537	1093	1158	1.20E-38	31.8
GROS_00031:122040-122144	GROS_g02537	826	860	1.20E-38	42.9
GROS_00031:119108-119227	GROS_g02537	110	149	1.20E-38	35
GROS_00031:120542-120763	GROS_g02537	393	466	1.20E-38	24.3
GROS_00034:70314-70691	GROS_g02714	505	630	2.50E-31	34.9
GROS_00034:69522-69626	GROS_g02714	826	860	2.50E-31	40
GROS_00034:70928-71095	GROS_g02714	411	466	2.50E-31	28.6
GROS_00034:72484-72594	GROS_g02714	110	146	2.50E-31	29.7
GROS_00034:67854-67958	GROS_g02714	1255	1289	2.50E-31	34.3
GROS_00039:12513-12632	GROS_g03015	1436	1475	4.20E-10	37.5
GROS_00039:11998-12114	GROS_g03015	1547	1585	4.20E-10	38.5
GROS_00039:22732-22812	GROS_g03016	198	224	4.20E-10	48.1
GROS_00039:20230-20328	GROS_g03016	392	424	4.20E-10	45.5
GROS_00039:17856-17945	GROS_g03016	431	460	4.20E-10	43.3
GROS_00039:20717-20815	GROS_g03016	392	424	1.90E-05	45.5
GROS_00039:20843-20932	GROS_g03016	431	460	1.90E-05	43.3
GROS_00039:21882-21962	GROS_g03016	198	224	0.004	48.1
GROS_00039:17973-18071	GROS_g03016	392	424	0.004	45.5
GROS_00136:83603-84007	GROS_g06238	491	625	8.40E-34	31.1
GROS_00136:81558-81755	GROS_g06238	1093	1158	8.40E-34	33.3
GROS_00136:82740-82844	GROS_g06238	826	860	8.40E-34	42.9
GROS_00136:85739-85804	GROS_g06238	128	149	8.40E-34	40.9
GROS_00136:84230-84370	GROS_g06238	420	466	8.40E-34	31.9
Query Sequence: <i>C. elegans vit-5</i>					
GROS_00031:120953-121336	GROS_g02537	498	625	2.20E-39	32.8
GROS_00031:123286-123483	GROS_g02537	1093	1158	2.20E-39	30.3
GROS_00031:119075-119227	GROS_g02537	99	149	2.20E-39	33.3
GROS_00031:122040-122144	GROS_g02537	826	860	2.20E-39	42.9
GROS_00031:120542-120763	GROS_g02537	393	466	2.20E-39	24.3
GROS_00034:70314-70691	GROS_g02714	505	630	3.90E-33	35.7
GROS_00034:72484-72636	GROS_g02714	96	146	3.90E-33	29.4
GROS_00034:69522-69626	GROS_g02714	826	860	3.90E-33	40

GROS_00034:70928-71095	GROS_g02714	411	466	3.90E-33	28.6
GROS_00034:67854-67958	GROS_g02714	1255	1289	3.90E-33	34.3
GROS_00039:11998-12114	GROS_g03015	1547	1585	2.50E-10	38.5
GROS_00039:12513-12632	GROS_g03015	1436	1475	2.50E-10	37.5
GROS_00039:22732-22812	GROS_g03016	198	224	2.50E-10	48.1
GROS_00039:20230-20328	GROS_g03016	392	424	2.50E-10	45.5
GROS_00039:17856-17945	GROS_g03016	431	460	2.50E-10	43.3
GROS_00039:20717-20815	GROS_g03016	392	424	2.00E-05	45.5
GROS_00039:20843-20932	GROS_g03016	431	460	2.00E-05	43.3
GROS_00039:21882-21962	GROS_g03016	198	224	0.0041	48.1
GROS_00039:17973-18071	GROS_g03016	392	424	0.0041	45.5
GROS_00136:83603-84007	GROS_g06238	491	625	4.20E-33	31.9
GROS_00136:81558-81755	GROS_g06238	1093	1158	4.20E-33	31.8
GROS_00136:82740-82844	GROS_g06238	826	860	4.20E-33	42.9
GROS_00136:85739-85804	GROS_g06238	128	149	4.20E-33	40.9
GROS_00136:84230-84370	GROS_g06238	420	466	4.20E-33	31.9
Query Sequence: <i>C. elegans vit-6</i>					
GROS_00007:218919-219152	GROS_g00996	1481	1558	0.0015	28.2
GROS_00007:218955-219164	GROS_g00996	1485	1554	0.0054	28.6
GROS_00031:120941-121342	GROS_g02537	523	656	9.60E-55	32.1
GROS_00031:120626-120775	GROS_g02537	446	495	9.60E-55	44
GROS_00031:124531-124668	GROS_g02537	1444	1489	9.60E-55	41.3
GROS_00031:119090-119218	GROS_g02537	116	158	9.60E-55	37.2
GROS_00031:121989-122192	GROS_g02537	841	908	9.60E-55	27.9
GROS_00031:119454-119555	GROS_g02537	187	220	9.60E-55	38.2
GROS_00031:123860-123964	GROS_g02537	1275	1309	9.60E-55	34.3
GROS_00031:124185-124229	GROS_g02537	1364	1378	9.60E-55	66.7
GROS_00031:167515-167619	GROS_g02549	1517	1551	9.60E-55	42.9
GROS_00034:70323-70691	GROS_g02714	534	656	8.20E-43	33.3
GROS_00034:70916-71095	GROS_g02714	436	495	8.20E-43	33.3
GROS_00034:67151-67267	GROS_g02714	1451	1489	8.20E-43	41
GROS_00034:72484-72603	GROS_g02714	119	158	8.20E-43	37.5
GROS_00034:69474-69626	GROS_g02714	858	908	8.20E-43	29.4
GROS_00034:67592-67636	GROS_g02714	1364	1378	8.20E-43	73.3
GROS_00136:83594-84010	GROS_g06238	519	657	9.90E-46	29.5
GROS_00136:84218-84367	GROS_g06238	446	495	9.90E-46	44
GROS_00136:80316-80432	GROS_g06238	1451	1489	9.90E-46	43.6
GROS_00136:82692-82895	GROS_g06238	841	908	9.90E-46	27.9
GROS_00136:85453-85554	GROS_g06238	187	220	9.90E-46	41.2
GROS_00136:85745-85807	GROS_g06238	139	159	9.90E-46	57.1
GROS_00136:80742-80786	GROS_g06238	1364	1378	9.90E-46	73.3

Table S3.1.4: BLAST result. Query sequence: *Caenorhabditis elegans vit* 1-6 gene sequences. Subject: *Heterodera glycines* genome.

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans vit-1</i>					
Chromosome_3:15 446105-15446251	Hetgly05118	93	141	0.0091	34.7
Chromosome_3:15 509433-15509579	Hetgly05124, Hetgly05126, Hetgly05131	93	141	0.0091	34.7
Chromosome_7:87 49068-8749160	Hetgly17961 (Hetgly.G000013367)	1564	1594	0.00016	51.6
Chromosome_7:87 54562-8754675	Hetgly17961 (Hetgly.G000013367)	390	427	0.00016	42.1
Chromosome_7:87 55852-8755935	Hetgly17961 (Hetgly.G000013367)	200	227	0.00016	39.3
Chromosome_7:87 50185-8750268	Hetgly17961 (Hetgly.G000013367)	1318	1345	0.00016	42.9
Chromosome_8:77 25410-7725646	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	510	588	1.30E-15	39.2
Chromosome_8:77 27399-7727692	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	1054	1151	1.30E-15	26.5
Chromosome_8:10 251903-10251986	Hetgly22125	1565	1592	6.70E-05	42.9
Chromosome_8:10 252035-10252100	Hetgly22125	1536	1557	6.70E-05	50
Chromosome_8:10 222642-10223028		500	628	3.80E-24	33.3
Chromosome_8:10 220441-10220602		96	149	3.80E-24	38.9
Chromosome_8:10 228183-10228266		1565	1592	3.80E-24	42.9
Chromosome_8:10 225494-10225652		1194	1246	3.80E-24	26.4
Chromosome_8:77 59507-7759770		501	588	8.30E-20	40.9
Chromosome_8:77 61523-7761816		1054	1151	8.30E-20	26.5
Chromosome_8:77 62196-7762306		1252	1288	8.30E-20	21.6
Query Sequence: <i>C. elegans vit-2</i>					
Chromosome_3:15 446105-15446251	Hetgly05118	93	141	0.0092	34.7
Chromosome_3:15 509433-15509579	Hetgly05124, Hetgly05126, Hetgly05131	93	141	0.0092	34.7
Chromosome_8:77 25410-7725646	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	508	586	2.20E-16	40.5
Chromosome_8:77 27507-7727692	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	1088	1149	2.20E-16	33.9
Chromosome_8:10 222642-10223028		498	626	7.50E-24	33.3
Chromosome_8:10 220441-10220602		96	149	7.50E-24	38.9
Chromosome_8:10 228069-10228137		1534	1556	7.50E-24	52.2
Chromosome_8:10 225494-10225652		1192	1244	7.50E-24	26.4
Chromosome_8:77 59507-7759770		499	586	5.60E-21	42
Chromosome_8:77 61631-7761816		1088	1149	5.60E-21	33.9
Chromosome_8:77 62196-7762363		1250	1305	5.60E-21	21.4
Query Sequence: <i>C. elegans vit-3</i>					
Chromosome_5:29 32390-2932518	Hetgly00448	1475	1517	0.003	27.9
Chromosome_8:77 25410-7725646	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	507	585	6.30E-17	41.8
Chromosome_8:77 27528-7727725	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	1093	1158	6.30E-17	30.3
Chromosome_8:77 57727-7757816	Hetgly21550	117	146	2.80E-22	33.3

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Chromosome_8:10 090144-10090257	Hetgly22088	1250	1287	2.80E-22	28.9
Chromosome_8:77 59507-7759770		498	585	2.80E-22	42
Chromosome_8:77 61652-7761849		1093	1158	2.80E-22	30.3
Chromosome_8:10 222666-10223043		505	630	6.20E-20	31.7
Chromosome_8:10 222183-10222389		405	473	6.20E-20	30.4
Chromosome_8:10 220483-10220593		110	146	6.20E-20	29.7
Chromosome_8:10 225988-10226101		1250	1287	6.20E-20	28.9
Chromosome_5:29 32291-2932344		1461	1478	0.003	44.4
Query Sequence: <i>C. elegans vit-4</i>					
Chromosome_8:77 25410-7725646	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	507	585	2.20E-17	41.8
Chromosome_8:77 27528-7727725	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	1093	1158	2.20E-17	31.8
Chromosome_8:77 57727-7757816	Hetgly21550	117	146	9.30E-23	33.3
Chromosome_8:10 090144-10090257	Hetgly22088	1250	1287	9.30E-23	28.9
Chromosome_8:77 59507-7759770		498	585	9.30E-23	42
Chromosome_8:77 61652-7761849		1093	1158	9.30E-23	31.8
Chromosome_8:10 222666-10223043		505	630	5.20E-20	31.7
Chromosome_8:10 222183-10222389		405	473	5.20E-20	30.4
Chromosome_8:10 220483-10220593		110	146	5.20E-20	29.7
Chromosome_8:10 225988-10226101		1250	1287	5.20E-20	28.9
Query Sequence: <i>C. elegans vit-5</i>					
Chromosome_8:77 25410-7725646	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	507	585	4.20E-17	43
Chromosome_8:77 27528-7727725	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	1093	1158	4.20E-17	30.3
Chromosome_8:77 57673-7757816	Hetgly21550	99	146	5.40E-23	29.2
Chromosome_8:10 090144-10090257	Hetgly22088	1250	1287	5.40E-23	28.9
Chromosome_8:77 59507-7759770		498	585	5.40E-23	43.2
Chromosome_8:77 61652-7761849		1093	1158	5.40E-23	30.3
Chromosome_8:10 222666-10223043		505	630	1.50E-21	32.5
Chromosome_8:10 220441-10220593		96	146	1.50E-21	29.4
Chromosome_8:10 222183-10222389		405	473	1.50E-21	30.4
Chromosome_8:10 225988-10226101		1250	1287	1.50E-21	28.9
Query Sequence: <i>C. elegans vit-6</i>					
Chromosome_5:75 11395-7511649	Hetgly01116	1483	1567	0.0013	29.4
Chromosome_1:13 822006-13822218	Hetgly07363	1483	1553	0.0066	28.2
Chromosome_6:72 30545-7230835	Hetgly15212 (Hetgly.G000006510)	575	671	4.60E-17	29.9
Chromosome_6:72 30184-7230291	Hetgly15212 (Hetgly.G000006510)	1454	1489	4.60E-17	50
Chromosome_6:72 31108-7231209	Hetgly15212 (Hetgly.G000006510)	456	489	4.60E-17	47.1
Chromosome_6:72 62580-7262870	Hetgly15218	575	671	2.40E-11	29.9

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Chromosome_6:72 63143-7263244	Hetgly15218	456	489	2.40E-11	44.1
Chromosome_8:77 23855-7724019	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	166	220	1.70E-26	29.1
Chromosome_8:77 25410-7725643	Hetgly21536, Hetgly21542 (Hetgly.G000023424)	536	613	1.20E-09	35.9
Chromosome_8:92 35522-9235659	Hetgly21883, Hetgly21884	1513	1558	2.80E-18	34.8
Chromosome_8:10 086415-10086543	Hetgly22088	456	498	4.50E-06	46.5
Chromosome_8:10 154825-10155115	Hetgly22100	575	671	2.80E-18	29.9
Chromosome_8:10 154464-10154571	Hetgly22100	1454	1489	2.80E-18	50
Chromosome_8:10 155388-10155489	Hetgly22100	456	489	2.80E-18	47.1
Chromosome_8:10 222666-10223034		534	656	6.80E-35	35.8
Chromosome_8:10 222261-10222389		456	498	6.80E-35	46.5
Chromosome_8:10 226765-10226872		1454	1489	6.80E-35	50
Chromosome_8:10 220459-10220596		114	159	6.80E-35	30.4
Chromosome_8:77 59507-7759767		527	613	1.70E-26	35.6
Chromosome_8:77 59143-7759322		436	495	1.70E-26	40
Chromosome_8:77 62916-7763047		1444	1487	1.70E-26	36.4
Chromosome_8:77 62178-7762315		1264	1309	1.70E-26	32.6
Chromosome_8:70 74593-7074916		1483	1590	0.0035	23.1

Table S3.1.5: BLAST result. Query sequence: *Caenorhabditis elegans* vit 1-6 gene sequences. Subject: *Heterodera schachtii* genome.

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans</i> vit-1					
004:241426-241539	Hsc_gene_19700	390	427	3.90E-09	44.7
004:235911-236003	Hsc_gene_19700	1564	1594	3.90E-09	51.6
004:242709-242792	Hsc_gene_19700	200	227	3.90E-09	39.3
004:237041-237124	Hsc_gene_19700	1318	1345	3.90E-09	42.9
004:237800-237949	Hsc_gene_19700	1098	1147	3.90E-09	24
004:236387-236509	Hsc_gene_19700	1431	1471	3.90E-09	34.1
011:2145527-2145913	Hsc_gene_5564	500	628	3.10E-26	32.6
011:2143356-2143517	Hsc_gene_5564	96	149	3.10E-26	38.9
011:2148422-2148628	Hsc_gene_5566	1184	1252	3.10E-26	23.2
011:2150878-2150943	Hsc_gene_5566	1536	1557	3.10E-26	50
011:2150992-2151072	Hsc_gene_5566	1565	1591	3.10E-26	40.7
007:2226848-2227135	Hsc_gene_6726	493	588	3.70E-22	34.4
007:2224804-2225097	Hsc_gene_6726	1054	1151	3.70E-22	27.6
007:2222607-2222723	Hsc_gene_6726	1559	1597	3.70E-22	38.5
007:2224314-2224424	Hsc_gene_6726	1252	1288	3.70E-22	21.6
007:2256263-2256526	Hsc_gene_6737	501	588	5.00E-13	37.5
Query Sequence: <i>C. elegans</i> vit-2					
004:235920-236006	Hsc_gene_19700	1560	1588	4.40E-05	41.4

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
004:242709-242792	Hsc_gene_19700	200	227	4.40E-05	39.3
004:236387-236509	Hsc_gene_19700	1429	1469	4.40E-05	36.6
004:237800-237949	Hsc_gene_19700	1096	1145	4.40E-05	26
004:237041-237124	Hsc_gene_19700	1316	1343	4.40E-05	42.9
011:2145527-2145913	Hsc_gene_5564	498	626	2.90E-26	32.6
011:2143356-2143517	Hsc_gene_5564	96	149	2.90E-26	38.9
011:2148422-2148628	Hsc_gene_5566	1182	1250	2.90E-26	23.2
011:2150878-2150946	Hsc_gene_5566	1534	1556	2.90E-26	52.2
007:2226848-2227135	Hsc_gene_6726	491	586	2.20E-22	35.4
007:2224804-2224989	Hsc_gene_6726	1088	1149	2.20E-22	35.5
007:2224257-2224424	Hsc_gene_6726	1250	1305	2.20E-22	21.4
007:2222625-2222723	Hsc_gene_6726	1556	1588	2.20E-22	36.4
007:2256263-2256526	Hsc_gene_6737	499	586	5.00E-17	38.6
007:2256536-2256652	Hsc_gene_6737	588	626	5.00E-17	28.2
Query Sequence: <i>C. elegans vit-3</i>					
004:241426-241527	Hsc_gene_19700	391	424	3.40E-05	47.1
004:241309-241398	Hsc_gene_19700	431	460	3.40E-05	43.3
284:16885-17055	Hsc_gene_25698	329	385	2.20E-05	29.8
284:16673-16801	Hsc_gene_25698	431	473	2.20E-05	39.5
284:18245-18355	Hsc_gene_25699	110	146	2.20E-05	29.7
011:2145551-2145928	Hsc_gene_5564	505	630	5.30E-21	31
011:2145062-2145271	Hsc_gene_5564	404	473	5.30E-21	30
011:2143398-2143508	Hsc_gene_5564	110	146	5.30E-21	29.7
011:2148931-2149044	Hsc_gene_5566	1250	1287	5.30E-21	28.9
007:2226848-2227135	Hsc_gene_6726	490	585	3.30E-23	38.5
007:2224771-2224968	Hsc_gene_6726	1093	1158	3.30E-23	31.8
007:2228806-2228895	Hsc_gene_6726	117	146	3.30E-23	33.3
007:2256263-2256526	Hsc_gene_6737	498	585	4.80E-16	40.9
Query Sequence: <i>C. elegans vit-4</i>					
004:241426-241527	Hsc_gene_19700	391	424	3.20E-05	47.1
004:241309-241398	Hsc_gene_19700	431	460	3.20E-05	43.3
284:16885-17055	Hsc_gene_25698	329	385	1.10E-05	29.8
284:16673-16801	Hsc_gene_25698	431	473	1.10E-05	39.5
284:18245-18355	Hsc_gene_25699	110	146	1.10E-05	29.7
011:2145551-2145928	Hsc_gene_5564	505	630	4.40E-21	31
011:2145062-2145271	Hsc_gene_5564	404	473	4.40E-21	30
011:2143398-2143508	Hsc_gene_5564	110	146	4.40E-21	29.7
011:2148931-2149044	Hsc_gene_5566	1250	1287	4.40E-21	28.9
007:2226848-2227135	Hsc_gene_6726	490	585	1.10E-23	38.5
007:2224771-2224968	Hsc_gene_6726	1093	1158	1.10E-23	33.3
007:2228806-2228895	Hsc_gene_6726	117	146	1.10E-23	33.3
007:2256263-2256526	Hsc_gene_6737	498	585	4.40E-16	40.9
Query Sequence: <i>C. elegans vit-5</i>					
004:241426-241527	Hsc_gene_19700	391	424	3.20E-05	47.1
004:241309-241398	Hsc_gene_19700	431	460	3.20E-05	43.3

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
284:16885-17055	Hsc_gene_25698	329	385	5.60E-08	29.8
284:16673-16801	Hsc_gene_25698	431	473	5.60E-08	39.5
284:18245-18406	Hsc_gene_25699	93	146	5.60E-08	27.8
011:2145551-2145928	Hsc_gene_5564	505	630	2.00E-23	31.7
011:2143356-2143508	Hsc_gene_5564	96	146	2.00E-23	29.4
011:2145062-2145271	Hsc_gene_5564	404	473	2.00E-23	30
011:2148931-2149044	Hsc_gene_5566	1250	1287	2.00E-23	28.9
007:2226848-2227135	Hsc_gene_6726	490	585	6.40E-24	39.6
007:2224771-2224968	Hsc_gene_6726	1093	1158	6.40E-24	31.8
007:2228806-2228949	Hsc_gene_6726	99	146	6.40E-24	29.2
007:2256263-2256526	Hsc_gene_6737	498	585	2.40E-16	42
Query Sequence: <i>C. elegans vit-6</i>					
043:1129736-1129882	Hsc_gene_15837	113	161	1.50E-03	30.6
132:21918-22145	Hsc_gene_22368	1483	1558	1.10E-03	28.9
284:16673-16801	Hsc_gene_25698	456	498	7.30E-11	48.8
284:18245-18373	Hsc_gene_25699	116	158	7.30E-11	32.6
011:2145551-2145919	Hsc_gene_5564	534	656	1.80E-37	35.8
011:2145083-2145271	Hsc_gene_5564	436	498	1.80E-37	36.5
011:2143380-2143511	Hsc_gene_5564	116	159	1.80E-37	31.8
011:2149714-2149821	Hsc_gene_5566	1454	1489	1.80E-37	50
007:856681-856818	Hsc_gene_6329	1513	1558	1.10E-30	34.8
007:2226851-2227111	Hsc_gene_6726	527	613	1.10E-30	37.9
007:2227299-2227493	Hsc_gene_6726	431	495	1.10E-30	38.5
007:2223565-2223696	Hsc_gene_6726	1444	1487	1.10E-30	36.4
007:2224305-2224442	Hsc_gene_6726	1264	1309	1.10E-30	32.6
007:2228479-2228580	Hsc_gene_6726	187	220	1.10E-30	38.2
007:2256263-2256523	Hsc_gene_6737	527	613	5.90E-14	36.8
007:2256034-2256078	Hsc_gene_6737	481	495	5.90E-14	80

Table S3.1.6: BLAST result. Query sequence: *Caenorhabditis elegans vit* 1-6 gene sequences. Subject: *Meloidogyne hapla* genome.

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans vit-1</i>					
Contig1717:11721-12128	Contig1717.frz3.gene3	493	628	8.90E-61	44.1
Contig1717:10043-10195	Contig1717.frz3.gene3	97	147	8.90E-61	45.1
Contig1717:10445-10543	Contig1717.frz3.gene3	200	232	8.90E-61	36.4
Contig1717:10304-10360	Contig1717.frz3.gene3	173	191	8.90E-61	52.6
Contig1717:12908-12985	Contig1717.frz3.gene4	837	862	8.90E-61	38.5
Contig1717:13094-13180	Contig1717.frz3.gene4	884	912	8.90E-61	34.5
Contig1717:13950-14177	Contig1717.frz3.gene5	1090	1165	8.90E-61	28.9
Contig1717:15319-15372	Contig1717.frz3.gene5	1435	1452	8.90E-61	61.1
Contig1717:15992-16063	Contig1717.frz3.gene6	1568	1591	8.90E-61	50
Contig2399:8649-8720	Contig2399.frz3.fgene1	1568	1591	8.20E-64	50
Contig2399:9340-9393	Contig2399.frz3.fgene1	1435	1452	8.20E-64	61.1

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Contig2399:9382-9477	Contig2399.frz3.fgene1	1406	1437	8.20E-64	31.2
Contig2399:9747-9779	Contig2399.frz3.fgene1	1332	1342	8.20E-64	72.7
Contig2399:10615-10761	Contig2399.frz3.fgene2	1090	1138	8.20E-64	34.7
Contig2399:11727-11804	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	837	862	8.20E-64	38.5
Contig2399:11532-11618	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	884	912	8.20E-64	34.5
Contig2399:12584-12991	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	493	628	8.20E-64	44.1
Contig2399:14519-14671	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	97	147	8.20E-64	45.1
Contig2399:14171-14269	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	200	232	8.20E-64	36.4
Contig2399:14354-14410	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	173	191	8.20E-64	52.6
Contig2584:1454-1606	Contig2584.frz3.gene1	97	147	7.50E-15	45.1
Contig2584:1106-1204	Contig2584.frz3.gene1	200	232	7.50E-15	36.4
Contig2584:1289-1345	Contig2584.frz3.gene1	173	191	7.50E-15	52.6
Contig363:36651-36821	Contig363.frz3.fgene1	1090	1146	8.70E-49	24.6
Contig363:37977-38030	Contig363.frz3.fgene1	1435	1452	8.70E-49	50
Contig363:35689-35856	Contig363.frz3.fgene1, Contig363.frz3.gene10	826	881	8.70E-49	26.8
Contig363:34490-34933	Contig363.frz3.fgene1, Contig363.frz3.gene9	486	633	8.70E-49	40.5
Contig363:32909-33064	Contig363.frz3.fgene1, Contig363.frz3.gene9	96	147	8.70E-49	38.5
Contig363:34210-34383	Contig363.frz3.fgene1, Contig363.frz3.gene9	420	477	8.70E-49	27.6
Contig363:33239-33331	Contig363.frz3.fgene1, Contig363.frz3.gene9	200	230	8.70E-49	35.5
Query Sequence: <i>C. elegans vit-2</i>					
Contig1717:11721-12128	Contig1717.frz3.gene3	491	626	4.50E-60	42.6
Contig1717:10043-10195	Contig1717.frz3.gene3	97	147	4.50E-60	47.1
Contig1717:10445-10543	Contig1717.frz3.gene3	200	232	4.50E-60	36.4
Contig1717:10304-10360	Contig1717.frz3.gene3	173	191	4.50E-60	52.6
Contig1717:12908-12988	Contig1717.frz3.gene4	835	861	4.50E-60	40.7
Contig1717:13094-13180	Contig1717.frz3.gene4	882	910	4.50E-60	34.5
Contig1717:13950-14177	Contig1717.frz3.gene5	1088	1163	4.50E-60	30.3
Contig1717:15319-15372	Contig1717.frz3.gene5	1433	1450	4.50E-60	61.1
Contig1717:15992-16075	Contig1717.frz3.gene6	1565	1592	4.50E-60	39.3
Contig2399:9340-9393	Contig2399.frz3.fgene1	1433	1450	1.50E-61	61.1
Contig2399:8637-8720	Contig2399.frz3.fgene1	1565	1592	1.50E-61	39.3
Contig2399:9747-9779	Contig2399.frz3.fgene1	1330	1340	1.50E-61	72.7
Contig2399:10615-10761	Contig2399.frz3.fgene2	1088	1136	1.50E-61	36.7
Contig2399:11724-11804	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	835	861	1.50E-61	40.7
Contig2399:11532-11618	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	882	910	1.50E-61	34.5
Contig2399:12584-12991	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	491	626	1.50E-61	42.6
Contig2399:14519-14671	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	97	147	1.50E-61	47.1
Contig2399:14171-14269	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	200	232	1.50E-61	36.4
Contig2399:14354-14410	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	173	191	1.50E-61	52.6
Contig2584:1454-1606	Contig2584.frz3.gene1	97	147	4.20E-15	47.1

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Contig2584:1106-1204	Contig2584.frz3.gene1	200	232	4.20E-15	36.4
Contig2584:1289-1345	Contig2584.frz3.gene1	173	191	4.20E-15	52.6
Contig363:36651-36821	Contig363.frz3.fgene1	1088	1144	2.10E-49	26.3
Contig363:37977-38030	Contig363.frz3.fgene1	1433	1450	2.10E-49	50
Contig363:35689-35802	Contig363.frz3.fgene1, Contig363.frz3.gene10	824	861	2.10E-49	28.9
Contig363:34490-34933	Contig363.frz3.fgene1, Contig363.frz3.gene9	484	631	2.10E-49	40.5
Contig363:32909-33064	Contig363.frz3.fgene1, Contig363.frz3.gene9	96	147	2.10E-49	40.4
Contig363:34225-34383	Contig363.frz3.fgene1, Contig363.frz3.gene9	423	475	2.10E-49	30.2
Contig363:33239-33331	Contig363.frz3.fgene1, Contig363.frz3.gene9	200	230	2.10E-49	35.5
Query Sequence: <i>C. elegans vit-3</i>					
Contig1717:11721-12143	Contig1717.frz3.gene3	490	630	2.00E-51	37.6
Contig1717:10103-10192	Contig1717.frz3.gene3	117	146	2.00E-51	40
Contig1717:10259-10360	Contig1717.frz3.gene3	155	188	2.00E-51	32.4
Contig1717:12890-12988	Contig1717.frz3.gene4	826	858	2.00E-51	42.4
Contig1717:13106-13183	Contig1717.frz3.gene4	883	908	2.00E-51	34.6
Contig1717:13950-14177	Contig1717.frz3.gene5	1086	1161	2.00E-51	26.3
Contig1717:14620-14730	Contig1717.frz3.gene5	1253	1289	2.00E-51	29.7
Contig1717:15986-16066	Contig1717.frz3.gene6	1552	1578	2.00E-51	44.4
Contig2399:9982-10092	Contig2399.frz3.fgene1	1253	1289	5.20E-52	29.7
Contig2399:8646-8726	Contig2399.frz3.fgene1	1552	1578	5.20E-52	44.4
Contig2399:10615-10761	Contig2399.frz3.fgene2	1086	1134	5.20E-52	28.6
Contig2399:11724-11822	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	826	858	5.20E-52	42.4
Contig2399:11529-11606	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	883	908	5.20E-52	34.6
Contig2399:12569-12991	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	490	630	5.20E-52	37.6
Contig2399:14522-14611	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	117	146	5.20E-52	40
Contig2399:14354-14455	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	155	188	5.20E-52	32.4
Contig2399:13351-13506	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	354	405	5.20E-52	19.2
Contig2584:1457-1546	Contig2584.frz3.gene1	117	146	1.30E-05	40
Contig2584:1289-1390	Contig2584.frz3.gene1	155	188	1.30E-05	32.4
Contig363:36666-36821	Contig363.frz3.fgene1	1091	1142	2.90E-42	28.8
Contig363:35698-35802	Contig363.frz3.fgene1, Contig363.frz3.gene10	824	858	2.90E-42	31.4
Contig363:34502-34933	Contig363.frz3.fgene1, Contig363.frz3.gene9	487	630	2.90E-42	38.2
Contig363:32972-33187	Contig363.frz3.fgene1, Contig363.frz3.gene9	117	188	2.90E-42	34.7
Query Sequence: <i>C. elegans vit-4</i>					
Contig1717:11721-12143	Contig1717.frz3.gene3	490	630	1.70E-52	37.6
Contig1717:10103-10192	Contig1717.frz3.gene3	117	146	1.70E-52	40
Contig1717:10259-10360	Contig1717.frz3.gene3	155	188	1.70E-52	32.4
Contig1717:12890-12988	Contig1717.frz3.gene4	826	858	1.70E-52	42.4
Contig1717:13106-13183	Contig1717.frz3.gene4	883	908	1.70E-52	34.6
Contig1717:13950-14177	Contig1717.frz3.gene5	1086	1161	1.70E-52	27.6
Contig1717:14620-14730	Contig1717.frz3.gene5	1253	1289	1.70E-52	29.7

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Contig1717:15986-16066	Contig1717.frz3.gene6	1552	1578	1.70E-52	48.1
Contig2399:9982-10092	Contig2399.frz3.fgene1	1253	1289	4.50E-53	29.7
Contig2399:8646-8726	Contig2399.frz3.fgene1	1552	1578	4.50E-53	48.1
Contig2399:10615-10761	Contig2399.frz3.fgene2	1086	1134	4.50E-53	30.6
Contig2399:11724-11822	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	826	858	4.50E-53	42.4
Contig2399:11529-11606	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	883	908	4.50E-53	34.6
Contig2399:12569-12991	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	490	630	4.50E-53	37.6
Contig2399:14522-14611	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	117	146	4.50E-53	40
Contig2399:14354-14455	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	155	188	4.50E-53	32.4
Contig2399:13351-13506	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	354	405	4.50E-53	19.2
Contig2584:1457-1546	Contig2584.frz3.gene1	117	146	1.20E-05	40
Contig2584:1289-1390	Contig2584.frz3.gene1	155	188	1.20E-05	32.4
Contig363:36666-36821	Contig363.frz3.fgene1	1091	1142	7.90E-42	26.9
Contig363:35698-35802	Contig363.frz3.fgene1, Contig363.frz3.gene10	824	858	7.90E-42	31.4
Contig363:34502-34933	Contig363.frz3.fgene1, Contig363.frz3.gene9	487	630	7.90E-42	38.2
Contig363:32972-33187	Contig363.frz3.fgene1, Contig363.frz3.gene9	117	188	7.90E-42	34.7
Query Sequence: <i>C. elegans</i> vit-5					
Contig1717:11721-12143	Contig1717.frz3.gene3	490	630	4.30E-53	38.3
Contig1717:10043-10192	Contig1717.frz3.gene3	97	146	4.30E-53	34
Contig1717:10259-10360	Contig1717.frz3.gene3	155	188	4.30E-53	32.4
Contig1717:12890-12988	Contig1717.frz3.gene4	826	858	4.30E-53	42.4
Contig1717:13106-13183	Contig1717.frz3.gene4	883	908	4.30E-53	34.6
Contig1717:13950-14177	Contig1717.frz3.gene5	1086	1161	4.30E-53	25
Contig1717:14620-14730	Contig1717.frz3.gene5	1253	1289	4.30E-53	29.7
Contig1717:15986-16066	Contig1717.frz3.gene6	1552	1578	4.30E-53	48.1
Contig2399:9982-10092	Contig2399.frz3.fgene1	1253	1289	3.40E-54	29.7
Contig2399:8646-8726	Contig2399.frz3.fgene1	1552	1578	3.40E-54	48.1
Contig2399:10615-10761	Contig2399.frz3.fgene2	1086	1134	3.40E-54	28.6
Contig2399:11724-11822	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	826	858	3.40E-54	42.4
Contig2399:11529-11606	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	883	908	3.40E-54	34.6
Contig2399:12569-12991	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	490	630	3.40E-54	38.3
Contig2399:14522-14671	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	97	146	3.40E-54	34
Contig2399:14354-14455	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	155	188	3.40E-54	32.4
Contig2399:13351-13506	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	354	405	3.40E-54	19.2
Contig2584:1457-1606	Contig2584.frz3.gene1	97	146	2.40E-07	34
Contig2584:1289-1390	Contig2584.frz3.gene1	155	188	2.40E-07	32.4
Contig363:36666-36821	Contig363.frz3.fgene1	1091	1142	8.40E-44	28.8
Contig363:35698-35802	Contig363.frz3.fgene1, Contig363.frz3.gene10	824	858	8.40E-44	31.4
Contig363:34502-34933	Contig363.frz3.fgene1, Contig363.frz3.gene9	487	630	8.40E-44	38.9
Contig363:32909-33187	Contig363.frz3.fgene1, Contig363.frz3.gene9	96	188	8.40E-44	31.2

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans vit-6</i>					
Contig1717:11733-12137	Contig1717.frz3.gene3	523	657	1.90E-60	38.5
Contig1717:10043-10219	Contig1717.frz3.gene3	109	167	1.90E-60	33.9
Contig1717:11424-11603	Contig1717.frz3.gene3	438	497	1.90E-60	35
Contig1717:10304-10360	Contig1717.frz3.gene3	180	198	1.90E-60	57.9
Contig1717:12875-13006	Contig1717.frz3.gene4	853	896	1.90E-60	29.5
Contig1717:15235-15372	Contig1717.frz3.gene5	1444	1489	1.90E-60	45.7
Contig1717:14933-14971	Contig1717.frz3.gene5	1362	1374	1.90E-60	84.6
Contig1717:14626-14724	Contig1717.frz3.gene5	1277	1309	1.90E-60	30.3
Contig2399:9340-9477	Contig2399.frz3.fgene1	1444	1489	1.60E-62	45.7
Contig2399:9741-9779	Contig2399.frz3.fgene1	1362	1374	1.60E-62	84.6
Contig2399:9988-10086	Contig2399.frz3.fgene1	1277	1309	1.60E-62	30.3
Contig2399:11706-11837	Contig2399.frz3.fgene2, Contig2399.frz3.gene6	853	896	1.60E-62	29.5
Contig2399:12575-12979	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	523	657	1.60E-62	38.5
Contig2399:14495-14671	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	109	167	1.60E-62	33.9
Contig2399:13108-13287	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	438	497	1.60E-62	35
Contig2399:14354-14410	Contig2399.frz3.fgene2, Contig2399.frz3.gene7	180	198	1.60E-62	57.9
Contig2584:1430-1606	Contig2584.frz3.gene1	109	167	3.60E-18	33.9
Contig2584:1289-1345	Contig2584.frz3.gene1	180	198	3.60E-18	57.9
Contig2800:3912-3998	Contig2800.frz3.gene1	1523	1551	7.30E-03	37.9
Contig2800:3876-3929	Contig2800.frz3.gene1	1480	1497	7.30E-03	61.1
Contig363:37893-38030	Contig363.frz3.fgene1	1444	1489	1.40E-57	50
Contig363:37591-37623	Contig363.frz3.fgene1	1364	1374	1.40E-57	90.9
Contig363:35719-35859	Contig363.frz3.fgene1, Contig363.frz3.gene10	863	909	1.40E-57	29.8
Contig363:34523-34927	Contig363.frz3.fgene1, Contig363.frz3.gene9	523	657	1.40E-57	38.5
Contig363:32915-33067	Contig363.frz3.fgene1, Contig363.frz3.gene9	110	160	1.40E-57	45.1
Contig363:34192-34398	Contig363.frz3.fgene1, Contig363.frz3.gene9	436	504	1.40E-57	33.3
Contig2584:43-222		438	497	3.60E-18	35

Table S3.1.7: BLAST result. Query sequence: *Caenorhabditis elegans vit* 1-6 gene sequences. Subject: *Meloidogyne incognita* genome.

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans vit-1</i>					
FXSY01000180.1:1963-2115	Minc3s00180g06901	97	147	8.00E-49	43.1
FXSY01000180.1:5013-5417	Minc3s00180g06902	494	628	8.00E-49	44.4
FXSY01000180.1:7198-7383	Minc3s00180g06902	1089	1150	8.00E-49	30.6
FXSY01000180.1:6220-6297	Minc3s00180g06902	837	862	8.00E-49	38.5
FXSY01000180.1:9292-9363	Minc3s00180g06902	1568	1591	8.00E-49	45.8
FXSY01000180.1:8584-8637	Minc3s00180g06902	1435	1452	8.00E-49	55.6
FXSY01000286.1:9228-9632	Minc3s00286g09435	494	628	9.10E-52	43.7
FXSY01000286.1:7550-7702	Minc3s00286g09435	97	147	9.10E-52	45.1
FXSY01000286.1:11483-11674	Minc3s00286g09435	1090	1153	9.10E-52	26.6

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01000286.1:10435-10512	Minc3s00286g09435	837	862	9.10E-52	42.3
FXSY01000286.1:13603-13674	Minc3s00286g09435	1568	1591	9.10E-52	45.8
FXSY01000286.1:10645-10719	Minc3s00286g09435	888	912	9.10E-52	44
FXSY01000286.1:7766-7867	Minc3s00286g09435	158	191	9.10E-52	35.3
FXSY01000286.1:12840-12893	Minc3s00286g09435	1435	1452	9.10E-52	55.6
FXSY01000369.1:11838-12257	Minc3s00369g11129	495	634	1.50E-48	43.6
FXSY01000369.1:9915-10070	Minc3s00369g11129	96	147	1.50E-48	38.5
FXSY01000369.1:11388-11558	Minc3s00369g11129	420	476	1.50E-48	29.8
FXSY01000369.1:14280-14432	Minc3s00369g11129	1088	1138	1.50E-48	23.5
FXSY01000369.1:10137-10211	Minc3s00369g11129	173	197	1.50E-48	44
FXSY01000369.1:13022-13135	Minc3s00369g11129	825	862	1.50E-48	28.9
FXSY01000369.1:16654-16734	Minc3s00369g11131	1534	1560	1.50E-48	37
FXSY01000369.1:16813-16905	Minc3s00369g11131	1568	1598	1.50E-48	35.5
FXSY01000805.1:30425-30829	Minc3s00805g17578	494	628	4.00E-52	45.2
FXSY01000805.1:28466-28651	Minc3s00805g17578	1089	1150	4.00E-52	30.6
FXSY01000805.1:29548-29625	Minc3s00805g17578	837	862	4.00E-52	42.3
FXSY01000805.1:26445-26516	Minc3s00805g17578	1568	1591	4.00E-52	45.8
FXSY01000805.1:27217-27270	Minc3s00805g17578	1435	1452	4.00E-52	55.6
FXSY01000805.1:32573-32725	Minc3s00805g17579	97	147	4.00E-52	43.1
FXSY01000805.1:32408-32509	Minc3s00805g17579	158	191	4.00E-52	35.3
FXSY01001515.1:30673-31077	Minc3s01514g24409	494	628	1.90E-41	43.7
FXSY01001515.1:28644-28826	Minc3s01514g24409	1090	1150	1.90E-41	27.9
FXSY01001515.1:26644-26715	Minc3s01514g24409	1568	1591	1.90E-41	45.8
FXSY01001515.1:29586-29660	Minc3s01514g24409	888	912	1.90E-41	44
FXSY01001515.1:27421-27474	Minc3s01514g24409	1435	1452	1.90E-41	55.6
FXSY01001515.1:29793-29870	Minc3s01514g24409	837	862	1.90E-41	38.5
FXSY01001575.1:27805-28209	Minc3s01575g24833	494	628	6.90E-40	43.7
FXSY01001575.1:26928-27005	Minc3s01575g24833	837	862	6.90E-40	42.3
FXSY01001575.1:29733-29885	Minc3s01575g24834	97	147	6.90E-40	43.1
FXSY01001575.1:29568-29624	Minc3s01575g24834	173	191	6.90E-40	52.6
FXSY01001861.1:3614-4033	Minc3s01861g26817	495	634	4.50E-52	41.4
FXSY01001861.1:1726-1881	Minc3s01861g26817	96	147	4.50E-52	38.5
FXSY01001861.1:3184-3369	Minc3s01861g26817	415	476	4.50E-52	29
FXSY01001861.1:6047-6199	Minc3s01861g26817	1088	1138	4.50E-52	23.5
FXSY01001861.1:1948-2022	Minc3s01861g26817	173	197	4.50E-52	44
FXSY01001861.1:4796-4909	Minc3s01861g26817	825	862	4.50E-52	26.3
FXSY01001861.1:8436-8516	Minc3s01861g26818	1534	1560	4.50E-52	37
FXSY01001861.1:8604-8696	Minc3s01861g26818	1568	1598	4.50E-52	35.5
FXSY01001861.1:7575-7628	Minc3s01861g26818	1435	1452	4.50E-52	50
FXSY01001975.1:6826-6906	Minc3s01975g27537	1534	1560	2.50E-50	37
FXSY01001975.1:7698-7751	Minc3s01975g27537	1435	1452	2.50E-50	50
FXSY01001975.1:9126-9380	Minc3s01975g27538	1054	1138	2.50E-50	17.6
FXSY01001975.1:11227-11646	Minc3s01975g27539	495	634	2.50E-50	42.1
FXSY01001975.1:13421-13576	Minc3s01975g27539	96	147	2.50E-50	38.5
FXSY01001975.1:11926-12096	Minc3s01975g27539	420	476	2.50E-50	28.1

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01001975.1:13280-13354	Minc3s01975g27539	173	197	2.50E-50	44
FXSY01001975.1:10350-10463	Minc3s01975g27539	825	862	2.50E-50	28.9
FXSY01002166.1:6080-6484	Minc3s02166g28653	494	628	7.60E-56	43.7
FXSY01002166.1:4398-4550	Minc3s02166g28653	97	147	7.60E-56	45.1
FXSY01002166.1:8342-8524	Minc3s02166g28653	1090	1150	7.60E-56	27.9
FXSY01002166.1:7286-7363	Minc3s02166g28653	837	862	7.60E-56	42.3
FXSY01002166.1:7482-7568	Minc3s02166g28653	884	912	7.60E-56	37.9
FXSY01002166.1:9712-9765	Minc3s02166g28653	1435	1452	7.60E-56	55.6
FXSY01002166.1:4614-4715	Minc3s02166g28653, Minc3s02166g28654	158	191	7.60E-56	35.3
FXSY01002260.1:15771-15956	Minc3s02260g29127	1089	1150	6.00E-18	30.6
FXSY01002260.1:13867-13995	Minc3s02260g29127	586	628	6.00E-18	41.9
FXSY01002260.1:14797-14874	Minc3s02260g29127	837	862	6.00E-18	42.3
FXSY01002260.1:17150-17203	Minc3s02260g29127	1435	1452	6.00E-18	55.6
FXSY01002260.1:17899-17970	Minc3s02260g29127	1568	1591	6.00E-18	41.7
FXSY01002260.1:16808-16849	Minc3s02260g29127	1332	1345	6.00E-18	57.1
FXSY01002607.1:12894-13298	Minc3s02607g30865	494	628	1.40E-56	45.2
FXSY01002607.1:11224-11376	Minc3s02607g30865	97	147	1.40E-56	43.1
FXSY01002607.1:15073-15258	Minc3s02607g30865	1089	1150	1.40E-56	29
FXSY01002607.1:14098-14175	Minc3s02607g30865	837	862	1.40E-56	42.3
FXSY01002607.1:17208-17279	Minc3s02607g30865	1568	1591	1.40E-56	45.8
FXSY01002607.1:16454-16507	Minc3s02607g30865	1435	1452	1.40E-56	55.6
FXSY01002607.1:16112-16153	Minc3s02607g30865	1332	1345	1.40E-56	57.1
FXSY01002607.1:11440-11541	Minc3s02607g30865, Minc3s02607g30866	158	191	1.40E-56	35.3
FXSY01002607.1:11626-11712	Minc3s02607g30865, Minc3s02607g30866	200	228	1.40E-56	34.5
FXSY01003289.1:86-490	Minc3s03289g33421	494	628	4.60E-40	43.7
FXSY01003289.1:2018-2170	Minc3s03289g33421	97	147	4.60E-40	45.1
FXSY01003289.1:1853-1954	Minc3s03289g33421	158	191	4.60E-40	35.3
FXSY01006972.1:1222-1374	Minc3s06972g40563	97	147	1.20E-13	45.1
FXSY01006972.1:1057-1158	Minc3s06972g40563, Minc3s06972g40564	158	191	1.20E-13	35.3
FXSY01006972.1:886-972	Minc3s06972g40563, Minc3s06972g40564	200	228	1.20E-13	34.5
FXSY01007386.1:3392-3619	Minc3s07387g41062	1090	1165	2.10E-11	25
FXSY01007386.1:1376-1447	Minc3s07387g41062	1568	1591	2.10E-11	45.8
FXSY01007386.1:2214-2267	Minc3s07387g41062	1435	1452	2.10E-11	55.6
FXSY01007386.1:2586-2618	Minc3s07387g41062	1332	1342	2.10E-11	72.7
FXSY01008451.1:2583-2987	Minc3s08451g42267	494	628	8.00E-41	43.7
FXSY01008451.1:604-786	Minc3s08451g42267	1090	1150	8.00E-41	26.2
FXSY01008451.1:1707-1784	Minc3s08451g42267	837	862	8.00E-41	42.3
FXSY01008451.1:1512-1586	Minc3s08451g42267	888	912	8.00E-41	44
FXSY01009698.1:2274-2678	Minc3s09698g43458	494	628	5.00E-42	44.4
FXSY01009698.1:248-430	Minc3s09698g43458	1090	1150	5.00E-42	27.9
FXSY01009698.1:1383-1460	Minc3s09698g43458	837	862	5.00E-42	42.3
FXSY01009698.1:1179-1253	Minc3s09698g43458	888	912	5.00E-42	44
FXSY01010887.1:339-419	Minc3s10886g44441	1534	1560	0.0048	37

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01010887.1:168-260	Minc3s10886g44441	1568	1598	0.0048	35.5
FXSY01010887.1:1512-1565	Minc3s10886g44441	1435	1452	0.0048	50
FXSY01002166.1:10455-10526		1568	1591	7.60E-56	45.8
FXSY01010887.1:1910-1942		1332	1342	0.0048	63.6
Query Sequence: <i>C. elegans vit-2</i>					
FXSY01000180.1:1963-2115	Minc3s00180g06901	97	147	4.60E-49	45.1
FXSY01000180.1:5013-5417	Minc3s00180g06902	492	626	4.60E-49	43.7
FXSY01000180.1:7198-7350	Minc3s00180g06902	1087	1137	4.60E-49	39.2
FXSY01000180.1:6220-6300	Minc3s00180g06902	835	861	4.60E-49	40.7
FXSY01000180.1:8584-8637	Minc3s00180g06902	1433	1450	4.60E-49	55.6
FXSY01000286.1:9228-9632	Minc3s00286g09435	492	626	3.50E-50	42.2
FXSY01000286.1:7550-7702	Minc3s00286g09435	97	147	3.50E-50	47.1
FXSY01000286.1:11483-11674	Minc3s00286g09435	1088	1151	3.50E-50	28.1
FXSY01000286.1:10435-10515	Minc3s00286g09435	835	861	3.50E-50	44.4
FXSY01000286.1:10645-10719	Minc3s00286g09435	886	910	3.50E-50	44
FXSY01000286.1:7766-7867	Minc3s00286g09435	158	191	3.50E-50	35.3
FXSY01000286.1:12840-12893	Minc3s00286g09435	1433	1450	3.50E-50	55.6
FXSY01000369.1:11838-12257	Minc3s00369g11129	493	632	2.90E-49	43.6
FXSY01000369.1:9915-10070	Minc3s00369g11129	96	147	2.90E-49	40.4
FXSY01000369.1:11403-11558	Minc3s00369g11129	423	474	2.90E-49	32.7
FXSY01000369.1:14178-14432	Minc3s00369g11129	1052	1136	2.90E-49	18.8
FXSY01000369.1:13022-13138	Minc3s00369g11129	823	861	2.90E-49	30.8
FXSY01000369.1:10137-10211	Minc3s00369g11129	173	197	2.90E-49	44
FXSY01000369.1:16654-16728	Minc3s00369g11131	1532	1556	2.90E-49	40
FXSY01000805.1:30425-30829	Minc3s00805g17578	492	626	1.10E-50	43.7
FXSY01000805.1:28499-28651	Minc3s00805g17578	1087	1137	1.10E-50	39.2
FXSY01000805.1:29545-29625	Minc3s00805g17578	835	861	1.10E-50	44.4
FXSY01000805.1:27217-27270	Minc3s00805g17578	1433	1450	1.10E-50	55.6
FXSY01000805.1:32573-32725	Minc3s00805g17579	97	147	1.10E-50	45.1
FXSY01000805.1:32408-32509	Minc3s00805g17579	158	191	1.10E-50	35.3
FXSY01001515.1:30673-31077	Minc3s01514g24409	492	626	1.90E-39	42.2
FXSY01001515.1:28677-28826	Minc3s01514g24409	1088	1137	1.90E-39	36
FXSY01001515.1:29790-29870	Minc3s01514g24409	835	861	1.90E-39	40.7
FXSY01001515.1:29586-29660	Minc3s01514g24409	886	910	1.90E-39	44
FXSY01001515.1:27421-27474	Minc3s01514g24409	1433	1450	1.90E-39	55.6
FXSY01001575.1:27805-28209	Minc3s01575g24833	492	626	5.50E-40	42.2
FXSY01001575.1:26925-27005	Minc3s01575g24833	835	861	5.50E-40	44.4
FXSY01001575.1:29733-29885	Minc3s01575g24834	97	147	5.50E-40	45.1
FXSY01001575.1:29568-29624	Minc3s01575g24834	173	191	5.50E-40	52.6
FXSY01001861.1:3614-4033	Minc3s01861g26817	493	632	4.50E-52	41.4
FXSY01001861.1:1726-1881	Minc3s01861g26817	96	147	4.50E-52	40.4
FXSY01001861.1:3214-3369	Minc3s01861g26817	423	474	4.50E-52	32.7
FXSY01001861.1:5945-6199	Minc3s01861g26817	1052	1136	4.50E-52	20
FXSY01001861.1:1948-2022	Minc3s01861g26817	173	197	4.50E-52	44
FXSY01001861.1:4796-4912	Minc3s01861g26817	823	861	4.50E-52	28.2

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01001861.1:8436-8510	Minc3s01861g26818	1532	1556	4.50E-52	40
FXSY01001861.1:7575-7628	Minc3s01861g26818	1433	1450	4.50E-52	50
FXSY01001975.1:6832-6906	Minc3s01975g27537	1532	1556	9.00E-53	40
FXSY01001975.1:7698-7751	Minc3s01975g27537	1433	1450	9.00E-53	50
FXSY01001975.1:9126-9380	Minc3s01975g27538	1052	1136	9.00E-53	18.8
FXSY01001975.1:11227-11646	Minc3s01975g27539	493	632	9.00E-53	42.1
FXSY01001975.1:13421-13576	Minc3s01975g27539	96	147	9.00E-53	40.4
FXSY01001975.1:11926-12081	Minc3s01975g27539	423	474	9.00E-53	32.7
FXSY01001975.1:10347-10463	Minc3s01975g27539	823	861	9.00E-53	30.8
FXSY01001975.1:13280-13354	Minc3s01975g27539	173	197	9.00E-53	44
FXSY01002166.1:6080-6484	Minc3s02166g28653	492	626	1.00E-53	42.2
FXSY01002166.1:4398-4550	Minc3s02166g28653	97	147	1.00E-53	47.1
FXSY01002166.1:8342-8491	Minc3s02166g28653	1088	1137	1.00E-53	36
FXSY01002166.1:7286-7366	Minc3s02166g28653	835	861	1.00E-53	44.4
FXSY01002166.1:7482-7568	Minc3s02166g28653	882	910	1.00E-53	37.9
FXSY01002166.1:9712-9765	Minc3s02166g28653	1433	1450	1.00E-53	55.6
FXSY01002166.1:4614-4715	Minc3s02166g28653, Minc3s02166g28654	158	191	1.00E-53	35.3
FXSY01002260.1:15771-15923	Minc3s02260g29127	1087	1137	1.60E-16	39.2
FXSY01002260.1:13867-13995	Minc3s02260g29127	584	626	1.60E-16	39.5
FXSY01002260.1:14797-14877	Minc3s02260g29127	835	861	1.60E-16	44.4
FXSY01002260.1:17150-17203	Minc3s02260g29127	1433	1450	1.60E-16	55.6
FXSY01002260.1:16808-16849	Minc3s02260g29127	1330	1343	1.60E-16	57.1
FXSY01002607.1:12894-13298	Minc3s02607g30865	492	626	1.80E-54	43.7
FXSY01002607.1:11224-11376	Minc3s02607g30865	97	147	1.80E-54	45.1
FXSY01002607.1:15073-15225	Minc3s02607g30865	1087	1137	1.80E-54	37.3
FXSY01002607.1:14098-14178	Minc3s02607g30865	835	861	1.80E-54	44.4
FXSY01002607.1:16454-16507	Minc3s02607g30865	1433	1450	1.80E-54	55.6
FXSY01002607.1:16112-16153	Minc3s02607g30865	1330	1343	1.80E-54	57.1
FXSY01002607.1:11440-11541	Minc3s02607g30865, Minc3s02607g30866	158	191	1.80E-54	35.3
FXSY01002607.1:11626-11712	Minc3s02607g30865, Minc3s02607g30866	200	228	1.80E-54	34.5
FXSY01003289.1:86-490	Minc3s03289g33421	492	626	1.00E-39	42.2
FXSY01003289.1:2018-2170	Minc3s03289g33421	97	147	1.00E-39	47.1
FXSY01003289.1:1853-1954	Minc3s03289g33421	158	191	1.00E-39	35.3
FXSY01003289.1:2384-2479	Minc3s03289g33421	24	55	1.00E-39	34.4
FXSY01006972.1:1222-1374	Minc3s06972g40563	97	147	6.80E-14	47.1
FXSY01006972.1:1057-1158	Minc3s06972g40563, Minc3s06972g40564	158	191	6.80E-14	35.3
FXSY01006972.1:886-972	Minc3s06972g40563, Minc3s06972g40564	200	228	6.80E-14	34.5
FXSY01007386.1:3392-3619	Minc3s07387g41062	1088	1163	6.90E-09	26.3
FXSY01007386.1:2214-2267	Minc3s07387g41062	1433	1450	6.90E-09	55.6
FXSY01007386.1:2586-2618	Minc3s07387g41062	1330	1340	6.90E-09	72.7
FXSY01008451.1:2583-2987	Minc3s08451g42267	492	626	6.30E-41	42.2
FXSY01008451.1:637-786	Minc3s08451g42267	1088	1137	6.30E-41	34
FXSY01008451.1:1704-1784	Minc3s08451g42267	835	861	6.30E-41	44.4

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01008451.1:1512-1586	Minc3s08451g42267	886	910	6.30E-41	44
FXSY01009698.1:2274-2678	Minc3s09698g43458	492	626	3.90E-42	43
FXSY01009698.1:281-430	Minc3s09698g43458	1088	1137	3.90E-42	36
FXSY01009698.1:1380-1460	Minc3s09698g43458	835	861	3.90E-42	44.4
FXSY01009698.1:1179-1253	Minc3s09698g43458	886	910	3.90E-42	44
Query Sequence: <i>C. elegans vit-3</i>					
FXSY01000180.1:2023-2112	Minc3s00180g06901	117	146	7.80E-43	36.7
FXSY01000180.1:5013-5432	Minc3s00180g06902	491	630	7.80E-43	39.3
FXSY01000180.1:7198-7383	Minc3s00180g06902	1085	1146	7.80E-43	27.4
FXSY01000180.1:6202-6300	Minc3s00180g06902	826	858	7.80E-43	39.4
FXSY01000180.1:7935-8045	Minc3s00180g06902	1253	1289	7.80E-43	27
FXSY01000180.1:6418-6519	Minc3s00180g06902	883	916	7.80E-43	26.5
FXSY01000286.1:9228-9647	Minc3s00286g09435	491	630	5.40E-46	37.9
FXSY01000286.1:11483-11674	Minc3s00286g09435	1086	1149	5.40E-46	23.4
FXSY01000286.1:7610-7699	Minc3s00286g09435	117	146	5.40E-46	40
FXSY01000286.1:10417-10515	Minc3s00286g09435	826	858	5.40E-46	42.4
FXSY01000286.1:12180-12290	Minc3s00286g09435	1253	1289	5.40E-46	29.7
FXSY01000286.1:8698-8754	Minc3s00286g09435	354	372	5.40E-46	52.6
FXSY01000286.1:10645-10722	Minc3s00286g09435	883	908	5.40E-46	34.6
FXSY01000286.1:13582-13674	Minc3s00286g09435	1547	1577	5.40E-46	35.5
FXSY01000286.1:7766-7867	Minc3s00286g09435	155	188	5.40E-46	29.4
FXSY01000369.1:11832-12371	Minc3s00369g11129	490	669	4.70E-43	35
FXSY01000369.1:9978-10193	Minc3s00369g11129	117	188	4.70E-43	34.7
FXSY01000369.1:13034-13138	Minc3s00369g11129	824	858	4.70E-43	28.6
FXSY01000369.1:11388-11537	Minc3s00369g11129	417	466	4.70E-43	32
FXSY01000369.1:11138-11239	Minc3s00369g11129	354	387	4.70E-43	38.2
FXSY01000805.1:30410-30829	Minc3s00805g17578	491	630	2.20E-45	39.3
FXSY01000805.1:28466-28651	Minc3s00805g17578	1085	1146	2.20E-45	25.8
FXSY01000805.1:29545-29643	Minc3s00805g17578	826	858	2.20E-45	39.4
FXSY01000805.1:27809-27919	Minc3s00805g17578	1253	1289	2.20E-45	29.7
FXSY01000805.1:29327-29428	Minc3s00805g17578	883	916	2.20E-45	29.4
FXSY01000805.1:32576-32665	Minc3s00805g17579	117	146	2.20E-45	36.7
FXSY01000805.1:32408-32461	Minc3s00805g17579	171	188	2.20E-45	50
FXSY01001515.1:30658-31077	Minc3s01514g24409	491	630	2.60E-41	37.9
FXSY01001515.1:28644-28826	Minc3s01514g24409	1086	1146	2.60E-41	24.6
FXSY01001515.1:28015-28125	Minc3s01514g24409	1253	1289	2.60E-41	29.7
FXSY01001515.1:29790-29888	Minc3s01514g24409	826	858	2.60E-41	39.4
FXSY01001515.1:31550-31606	Minc3s01514g24409	354	372	2.60E-41	52.6
FXSY01001515.1:29583-29660	Minc3s01514g24409	883	908	2.60E-41	34.6
FXSY01001515.1:26644-26721	Minc3s01514g24409	1552	1577	2.60E-41	38.5
FXSY01001575.1:27790-28209	Minc3s01575g24833	491	630	1.30E-34	37.9
FXSY01001575.1:26925-27023	Minc3s01575g24833	826	858	1.30E-34	42.4
FXSY01001575.1:28682-28738	Minc3s01575g24833	354	372	1.30E-34	52.6
FXSY01001575.1:29736-29825	Minc3s01575g24834	117	146	1.30E-34	40
FXSY01001575.1:29568-29621	Minc3s01575g24834	171	188	1.30E-34	50

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01001861.1:3608-4147	Minc3s01861g26817	490	669	3.10E-44	35.6
FXSY01001861.1:1789-2004	Minc3s01861g26817	117	188	3.10E-44	34.7
FXSY01001861.1:3187-3348	Minc3s01861g26817	413	466	3.10E-44	31.5
FXSY01001861.1:4808-4912	Minc3s01861g26817	824	858	3.10E-44	28.6
FXSY01001861.1:2948-3049	Minc3s01861g26817	354	387	3.10E-44	38.2
FXSY01001975.1:11113-11652	Minc3s01975g27539	490	669	2.10E-43	34.4
FXSY01001975.1:13298-13513	Minc3s01975g27539	117	188	2.10E-43	34.7
FXSY01001975.1:10347-10451	Minc3s01975g27539	824	858	2.10E-43	28.6
FXSY01001975.1:12245-12346	Minc3s01975g27539	354	387	2.10E-43	38.2
FXSY01001975.1:11947-12096	Minc3s01975g27539	417	466	2.10E-43	30
FXSY01002166.1:6080-6499	Minc3s02166g28653	491	630	8.50E-50	37.9
FXSY01002166.1:8342-8524	Minc3s02166g28653	1086	1146	8.50E-50	24.6
FXSY01002166.1:4458-4547	Minc3s02166g28653	117	146	8.50E-50	40
FXSY01002166.1:7268-7366	Minc3s02166g28653	826	858	8.50E-50	42.4
FXSY01002166.1:9057-9167	Minc3s02166g28653	1253	1289	8.50E-50	29.7
FXSY01002166.1:5549-5605	Minc3s02166g28653	354	372	8.50E-50	52.6
FXSY01002166.1:7494-7571	Minc3s02166g28653	883	908	8.50E-50	34.6
FXSY01002166.1:4614-4715	Minc3s02166g28653, Minc3s02166g28654	155	188	8.50E-50	29.4
FXSY01002260.1:15771-15956	Minc3s02260g29127	1085	1146	1.30E-19	25.8
FXSY01002260.1:13867-14010	Minc3s02260g29127	583	630	1.30E-19	37.5
FXSY01002260.1:14779-14877	Minc3s02260g29127	826	858	1.30E-19	39.4
FXSY01002260.1:16506-16616	Minc3s02260g29127	1253	1289	1.30E-19	29.7
FXSY01002260.1:14994-15095	Minc3s02260g29127	883	916	1.30E-19	29.4
FXSY01002607.1:12894-13313	Minc3s02607g30865	491	630	2.80E-47	39.3
FXSY01002607.1:15073-15258	Minc3s02607g30865	1085	1146	2.80E-47	25.8
FXSY01002607.1:11284-11373	Minc3s02607g30865	117	146	2.80E-47	36.7
FXSY01002607.1:14080-14178	Minc3s02607g30865	826	858	2.80E-47	39.4
FXSY01002607.1:15805-15915	Minc3s02607g30865	1253	1289	2.80E-47	29.7
FXSY01002607.1:14297-14398	Minc3s02607g30865	883	916	2.80E-47	26.5
FXSY01002607.1:11488-11541	Minc3s02607g30865, Minc3s02607g30866	171	188	2.80E-47	50
FXSY01003289.1:71-490	Minc3s03289g33421	491	630	6.20E-30	37.9
FXSY01003289.1:2021-2110	Minc3s03289g33421	117	146	6.20E-30	40
FXSY01003289.1:1853-1954	Minc3s03289g33421	155	188	6.20E-30	29.4
FXSY01006543.1:166-276	Minc3s06543g39965	1253	1289	3.10E-03	29.7
FXSY01006543.1:1628-1720	Minc3s06543g39965	1547	1577	3.10E-03	35.5
FXSY01006543.1:472-504	Minc3s06543g39965	1328	1338	3.10E-03	72.7
FXSY01006972.1:1225-1314	Minc3s06972g40563	117	146	8.60E-05	40
FXSY01006972.1:1057-1158	Minc3s06972g40563, Minc3s06972g40564	155	188	8.60E-05	29.4
FXSY01007386.1:3392-3619	Minc3s07387g41062	1086	1161	5.00E-12	23.7
FXSY01007386.1:2811-2921	Minc3s07387g41062	1253	1289	5.00E-12	29.7
FXSY01007386.1:1376-1453	Minc3s07387g41062	1552	1577	5.00E-12	38.5
FXSY01007386.1:2586-2618	Minc3s07387g41062	1328	1338	5.00E-12	72.7
FXSY01008451.1:2568-2987	Minc3s08451g42267	491	630	4.10E-34	37.9
FXSY01008451.1:1704-1802	Minc3s08451g42267	826	858	4.10E-34	42.4

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01008451.1:1509-1586	Minc3s08451g42267	883	908	4.10E-34	34.6
FXSY01009698.1:2259-2678	Minc3s09698g43458	491	630	2.20E-38	37.1
FXSY01009698.1:248-430	Minc3s09698g43458	1086	1146	2.20E-38	24.6
FXSY01009698.1:1380-1478	Minc3s09698g43458	826	858	2.20E-38	42.4
FXSY01009698.1:1176-1253	Minc3s09698g43458	883	908	2.20E-38	34.6
FXSY01008451.1:5-85		1253	1279	4.10E-34	33.3
FXSY01002166.1:10449-10526		1552	1577	8.50E-50	38.5
Query Sequence: <i>C. elegans vit-4</i>					
FXSY01000180.1:2023-2112	Minc3s00180g06901	117	146	2.40E-43	36.7
FXSY01000180.1:5013-5432	Minc3s00180g06902	491	630	2.40E-43	39.3
FXSY01000180.1:7198-7383	Minc3s00180g06902	1085	1146	2.40E-43	29
FXSY01000180.1:6202-6300	Minc3s00180g06902	826	858	2.40E-43	39.4
FXSY01000180.1:7935-8045	Minc3s00180g06902	1253	1289	2.40E-43	27
FXSY01000180.1:6418-6519	Minc3s00180g06902	883	916	2.40E-43	26.5
FXSY01000286.1:9228-9647	Minc3s00286g09435	491	630	4.70E-47	37.9
FXSY01000286.1:11483-11674	Minc3s00286g09435	1086	1149	4.70E-47	25
FXSY01000286.1:7610-7699	Minc3s00286g09435	117	146	4.70E-47	40
FXSY01000286.1:10417-10515	Minc3s00286g09435	826	858	4.70E-47	42.4
FXSY01000286.1:12180-12290	Minc3s00286g09435	1253	1289	4.70E-47	29.7
FXSY01000286.1:13582-13674	Minc3s00286g09435	1547	1577	4.70E-47	38.7
FXSY01000286.1:8698-8754	Minc3s00286g09435	354	372	4.70E-47	52.6
FXSY01000286.1:10645-10722	Minc3s00286g09435	883	908	4.70E-47	34.6
FXSY01000286.1:7766-7867	Minc3s00286g09435	155	188	4.70E-47	29.4
FXSY01000369.1:11832-12371	Minc3s00369g11129	490	669	3.60E-43	35
FXSY01000369.1:9978-10193	Minc3s00369g11129	117	188	3.60E-43	34.7
FXSY01000369.1:13034-13138	Minc3s00369g11129	824	858	3.60E-43	28.6
FXSY01000369.1:11388-11537	Minc3s00369g11129	417	466	3.60E-43	32
FXSY01000369.1:11138-11239	Minc3s00369g11129	354	387	3.60E-43	38.2
FXSY01000805.1:30410-30829	Minc3s00805g17578	491	630	6.70E-46	39.3
FXSY01000805.1:28466-28651	Minc3s00805g17578	1085	1146	6.70E-46	27.4
FXSY01000805.1:29545-29643	Minc3s00805g17578	826	858	6.70E-46	39.4
FXSY01000805.1:27809-27919	Minc3s00805g17578	1253	1289	6.70E-46	29.7
FXSY01000805.1:29327-29428	Minc3s00805g17578	883	916	6.70E-46	29.4
FXSY01000805.1:32576-32665	Minc3s00805g17579	117	146	6.70E-46	36.7
FXSY01000805.1:32408-32461	Minc3s00805g17579	171	188	6.70E-46	50
FXSY01001515.1:30658-31077	Minc3s01514g24409	491	630	2.40E-42	37.9
FXSY01001515.1:28644-28826	Minc3s01514g24409	1086	1146	2.40E-42	26.2
FXSY01001515.1:28015-28125	Minc3s01514g24409	1253	1289	2.40E-42	29.7
FXSY01001515.1:29790-29888	Minc3s01514g24409	826	858	2.40E-42	39.4
FXSY01001515.1:31550-31606	Minc3s01514g24409	354	372	2.40E-42	52.6
FXSY01001515.1:29583-29660	Minc3s01514g24409	883	908	2.40E-42	34.6
FXSY01001515.1:26644-26721	Minc3s01514g24409	1552	1577	2.40E-42	42.3
FXSY01001575.1:27790-28209	Minc3s01575g24833	491	630	1.00E-34	37.9
FXSY01001575.1:26925-27023	Minc3s01575g24833	826	858	1.00E-34	42.4
FXSY01001575.1:28682-28738	Minc3s01575g24833	354	372	1.00E-34	52.6

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01001575.1:29736-29825	Minc3s01575g24834	117	146	1.00E-34	40
FXSY01001575.1:29568-29621	Minc3s01575g24834	171	188	1.00E-34	50
FXSY01001861.1:3608-4147	Minc3s01861g26817	490	669	2.40E-44	35.6
FXSY01001861.1:1789-2004	Minc3s01861g26817	117	188	2.40E-44	34.7
FXSY01001861.1:3187-3348	Minc3s01861g26817	413	466	2.40E-44	31.5
FXSY01001861.1:4808-4912	Minc3s01861g26817	824	858	2.40E-44	28.6
FXSY01001861.1:2948-3049	Minc3s01861g26817	354	387	2.40E-44	38.2
FXSY01001975.1:11113-11652	Minc3s01975g27539	490	669	1.60E-43	34.4
FXSY01001975.1:13298-13513	Minc3s01975g27539	117	188	1.60E-43	34.7
FXSY01001975.1:10347-10451	Minc3s01975g27539	824	858	1.60E-43	28.6
FXSY01001975.1:12245-12346	Minc3s01975g27539	354	387	1.60E-43	38.2
FXSY01001975.1:11947-12096	Minc3s01975g27539	417	466	1.60E-43	30
FXSY01002166.1:6080-6499	Minc3s02166g28653	491	630	7.40E-51	37.9
FXSY01002166.1:8342-8524	Minc3s02166g28653	1086	1146	7.40E-51	26.2
FXSY01002166.1:4458-4547	Minc3s02166g28653	117	146	7.40E-51	40
FXSY01002166.1:7268-7366	Minc3s02166g28653	826	858	7.40E-51	42.4
FXSY01002166.1:9057-9167	Minc3s02166g28653	1253	1289	7.40E-51	29.7
FXSY01002166.1:5549-5605	Minc3s02166g28653	354	372	7.40E-51	52.6
FXSY01002166.1:7494-7571	Minc3s02166g28653	883	908	7.40E-51	34.6
FXSY01002166.1:4614-4715	Minc3s02166g28653, Minc3s02166g28654	155	188	7.40E-51	29.4
FXSY01002260.1:15771-15956	Minc3s02260g29127	1085	1146	4.40E-20	27.4
FXSY01002260.1:13867-14010	Minc3s02260g29127	583	630	4.40E-20	37.5
FXSY01002260.1:14779-14877	Minc3s02260g29127	826	858	4.40E-20	39.4
FXSY01002260.1:16506-16616	Minc3s02260g29127	1253	1289	4.40E-20	29.7
FXSY01002260.1:14994-15095	Minc3s02260g29127	883	916	4.40E-20	29.4
FXSY01002607.1:12894-13313	Minc3s02607g30865	491	630	4.10E-49	39.3
FXSY01002607.1:15073-15258	Minc3s02607g30865	1085	1146	4.10E-49	27.4
FXSY01002607.1:11284-11373	Minc3s02607g30865	117	146	4.10E-49	36.7
FXSY01002607.1:14080-14178	Minc3s02607g30865	826	858	4.10E-49	39.4
FXSY01002607.1:15805-15915	Minc3s02607g30865	1253	1289	4.10E-49	29.7
FXSY01002607.1:14297-14398	Minc3s02607g30865	883	916	4.10E-49	26.5
FXSY01002607.1:17190-17279	Minc3s02607g30865	1548	1577	4.10E-49	33.3
FXSY01002607.1:11488-11541	Minc3s02607g30865, Minc3s02607g30866	171	188	4.10E-49	50
FXSY01003289.1:71-490	Minc3s03289g33421	491	630	5.20E-30	37.9
FXSY01003289.1:2021-2110	Minc3s03289g33421	117	146	5.20E-30	40
FXSY01003289.1:1853-1954	Minc3s03289g33421	155	188	5.20E-30	29.4
FXSY01006543.1:166-276	Minc3s06543g39965	1253	1289	9.00E-04	29.7
FXSY01006543.1:1628-1720	Minc3s06543g39965	1547	1577	9.00E-04	38.7
FXSY01006543.1:472-504	Minc3s06543g39965	1328	1338	9.00E-04	72.7
FXSY01006972.1:1225-1314	Minc3s06972g40563	117	146	8.10E-05	40
FXSY01006972.1:1057-1158	Minc3s06972g40563, Minc3s06972g40564	155	188	8.10E-05	29.4
FXSY01007386.1:3392-3619	Minc3s07387g41062	1086	1161	5.50E-13	25
FXSY01007386.1:2811-2921	Minc3s07387g41062	1253	1289	5.50E-13	29.7
FXSY01007386.1:1376-1453	Minc3s07387g41062	1552	1577	5.50E-13	42.3

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01007386.1:2586-2618	Minc3s07387g41062	1328	1338	5.50E-13	72.7
FXSY01008451.1:2568-2987	Minc3s08451g42267	491	630	3.40E-34	37.9
FXSY01008451.1:1704-1802	Minc3s08451g42267	826	858	3.40E-34	42.4
FXSY01008451.1:1509-1586	Minc3s08451g42267	883	908	3.40E-34	34.6
FXSY01009698.1:2259-2678	Minc3s09698g43458	491	630	7.00E-39	37.1
FXSY01009698.1:248-430	Minc3s09698g43458	1086	1146	7.00E-39	26.2
FXSY01009698.1:1380-1478	Minc3s09698g43458	826	858	7.00E-39	42.4
FXSY01009698.1:1176-1253	Minc3s09698g43458	883	908	7.00E-39	34.6
FXSY01008451.1:5-85		1253	1279	3.40E-34	33.3
FXSY01002166.1:10449-10526		1552	1577	7.40E-51	42.3
Query Sequence: <i>C. elegans vit-5</i>					
FXSY01000180.1:1963-2112	Minc3s00180g06901	97	146	1.70E-44	32
FXSY01000180.1:5013-5432	Minc3s00180g06902	491	630	1.70E-44	40
FXSY01000180.1:7198-7383	Minc3s00180g06902	1085	1146	1.70E-44	27.4
FXSY01000180.1:6202-6300	Minc3s00180g06902	826	858	1.70E-44	39.4
FXSY01000180.1:7935-8045	Minc3s00180g06902	1253	1289	1.70E-44	27
FXSY01000180.1:6418-6519	Minc3s00180g06902	883	916	1.70E-44	26.5
FXSY01000286.1:9228-9647	Minc3s00286g09435	491	630	3.60E-48	38.6
FXSY01000286.1:11483-11674	Minc3s00286g09435	1086	1149	3.60E-48	23.4
FXSY01000286.1:7550-7699	Minc3s00286g09435	97	146	3.60E-48	34
FXSY01000286.1:10417-10515	Minc3s00286g09435	826	858	3.60E-48	42.4
FXSY01000286.1:12180-12290	Minc3s00286g09435	1253	1289	3.60E-48	29.7
FXSY01000286.1:13582-13674	Minc3s00286g09435	1547	1577	3.60E-48	38.7
FXSY01000286.1:8698-8754	Minc3s00286g09435	354	372	3.60E-48	52.6
FXSY01000286.1:10645-10722	Minc3s00286g09435	883	908	3.60E-48	34.6
FXSY01000286.1:7766-7867	Minc3s00286g09435	155	188	3.60E-48	29.4
FXSY01000369.1:11832-12371	Minc3s00369g11129	490	669	2.90E-45	35.6
FXSY01000369.1:9915-10193	Minc3s00369g11129	96	188	2.90E-45	32.3
FXSY01000369.1:13034-13138	Minc3s00369g11129	824	858	2.90E-45	28.6
FXSY01000369.1:11388-11537	Minc3s00369g11129	417	466	2.90E-45	32
FXSY01000369.1:11138-11239	Minc3s00369g11129	354	387	2.90E-45	38.2
FXSY01000805.1:30410-30829	Minc3s00805g17578	491	630	4.90E-47	40
FXSY01000805.1:28466-28651	Minc3s00805g17578	1085	1146	4.90E-47	25.8
FXSY01000805.1:29545-29643	Minc3s00805g17578	826	858	4.90E-47	39.4
FXSY01000805.1:27809-27919	Minc3s00805g17578	1253	1289	4.90E-47	29.7
FXSY01000805.1:29327-29428	Minc3s00805g17578	883	916	4.90E-47	29.4
FXSY01000805.1:32576-32725	Minc3s00805g17579	97	146	4.90E-47	32
FXSY01000805.1:32408-32461	Minc3s00805g17579	171	188	4.90E-47	50
FXSY01001515.1:30658-31077	Minc3s01514g24409	491	630	8.90E-42	38.6
FXSY01001515.1:28644-28826	Minc3s01514g24409	1086	1146	8.90E-42	24.6
FXSY01001515.1:28015-28125	Minc3s01514g24409	1253	1289	8.90E-42	29.7
FXSY01001515.1:29790-29888	Minc3s01514g24409	826	858	8.90E-42	39.4
FXSY01001515.1:31550-31606	Minc3s01514g24409	354	372	8.90E-42	52.6
FXSY01001515.1:29583-29660	Minc3s01514g24409	883	908	8.90E-42	34.6
FXSY01001515.1:26644-26721	Minc3s01514g24409	1552	1577	8.90E-42	42.3

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01001575.1:27790-28209	Minc3s01575g24833	491	630	1.30E-35	38.6
FXSY01001575.1:26925-27023	Minc3s01575g24833	826	858	1.30E-35	42.4
FXSY01001575.1:28682-28738	Minc3s01575g24833	354	372	1.30E-35	52.6
FXSY01001575.1:29736-29885	Minc3s01575g24834	97	146	1.30E-35	34
FXSY01001575.1:29568-29621	Minc3s01575g24834	171	188	1.30E-35	50
FXSY01001861.1:3608-4147	Minc3s01861g26817	490	669	2.00E-46	36.1
FXSY01001861.1:1726-2004	Minc3s01861g26817	96	188	2.00E-46	32.3
FXSY01001861.1:3187-3348	Minc3s01861g26817	413	466	2.00E-46	31.5
FXSY01001861.1:4808-4912	Minc3s01861g26817	824	858	2.00E-46	28.6
FXSY01001861.1:2948-3049	Minc3s01861g26817	354	387	2.00E-46	38.2
FXSY01001975.1:11113-11652	Minc3s01975g27539	490	669	1.30E-45	35
FXSY01001975.1:13298-13576	Minc3s01975g27539	96	188	1.30E-45	32.3
FXSY01001975.1:10347-10451	Minc3s01975g27539	824	858	1.30E-45	28.6
FXSY01001975.1:12245-12346	Minc3s01975g27539	354	387	1.30E-45	38.2
FXSY01001975.1:11947-12096	Minc3s01975g27539	417	466	1.30E-45	30
FXSY01002166.1:6080-6499	Minc3s02166g28653	491	630	5.60E-52	38.6
FXSY01002166.1:8342-8524	Minc3s02166g28653	1086	1146	5.60E-52	24.6
FXSY01002166.1:4398-4547	Minc3s02166g28653	97	146	5.60E-52	34
FXSY01002166.1:7268-7366	Minc3s02166g28653	826	858	5.60E-52	42.4
FXSY01002166.1:9057-9167	Minc3s02166g28653	1253	1289	5.60E-52	29.7
FXSY01002166.1:5549-5605	Minc3s02166g28653	354	372	5.60E-52	52.6
FXSY01002166.1:7494-7571	Minc3s02166g28653	883	908	5.60E-52	34.6
FXSY01002166.1:4614-4715	Minc3s02166g28653, Minc3s02166g28654	155	188	5.60E-52	29.4
FXSY01002260.1:15771-15956	Minc3s02260g29127	1085	1146	2.80E-19	25.8
FXSY01002260.1:13867-14010	Minc3s02260g29127	583	630	2.80E-19	37.5
FXSY01002260.1:14779-14877	Minc3s02260g29127	826	858	2.80E-19	39.4
FXSY01002260.1:16506-16616	Minc3s02260g29127	1253	1289	2.80E-19	29.7
FXSY01002260.1:14994-15095	Minc3s02260g29127	883	916	2.80E-19	29.4
FXSY01002607.1:12894-13313	Minc3s02607g30865	491	630	3.00E-50	40
FXSY01002607.1:15073-15258	Minc3s02607g30865	1085	1146	3.00E-50	25.8
FXSY01002607.1:11224-11373	Minc3s02607g30865	97	146	3.00E-50	32
FXSY01002607.1:14080-14178	Minc3s02607g30865	826	858	3.00E-50	39.4
FXSY01002607.1:15805-15915	Minc3s02607g30865	1253	1289	3.00E-50	29.7
FXSY01002607.1:14297-14398	Minc3s02607g30865	883	916	3.00E-50	26.5
FXSY01002607.1:17190-17279	Minc3s02607g30865	1548	1577	3.00E-50	33.3
FXSY01002607.1:11488-11541	Minc3s02607g30865, Minc3s02607g30866	171	188	3.00E-50	50
FXSY01003289.1:71-490	Minc3s03289g33421	491	630	1.30E-31	38.6
FXSY01003289.1:2021-2170	Minc3s03289g33421	97	146	1.30E-31	34
FXSY01003289.1:1853-1954	Minc3s03289g33421	155	188	1.30E-31	29.4
FXSY01006543.1:166-276	Minc3s06543g39965	1253	1289	9.20E-04	29.7
FXSY01006543.1:1628-1720	Minc3s06543g39965	1547	1577	9.20E-04	38.7
FXSY01006543.1:472-504	Minc3s06543g39965	1328	1338	9.20E-04	72.7
FXSY01006972.1:1225-1374	Minc3s06972g40563	97	146	1.60E-06	34
FXSY01006972.1:1057-1158	Minc3s06972g40563, Minc3s06972g40564	155	188	1.60E-06	29.4

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01007386.1:3392-3619	Minc3s07387g41062	1086	1161	1.90E-12	23.7
FXSY01007386.1:2811-2921	Minc3s07387g41062	1253	1289	1.90E-12	29.7
FXSY01007386.1:1376-1453	Minc3s07387g41062	1552	1577	1.90E-12	42.3
FXSY01007386.1:2586-2618	Minc3s07387g41062	1328	1338	1.90E-12	72.7
FXSY01008451.1:2568-2987	Minc3s08451g42267	491	630	4.90E-34	38.6
FXSY01008451.1:1704-1802	Minc3s08451g42267	826	858	4.90E-34	42.4
FXSY01008451.1:1509-1586	Minc3s08451g42267	883	908	4.90E-34	34.6
FXSY01009698.1:2259-2678	Minc3s09698g43458	491	630	2.60E-38	37.9
FXSY01009698.1:248-430	Minc3s09698g43458	1086	1146	2.60E-38	24.6
FXSY01009698.1:1380-1478	Minc3s09698g43458	826	858	2.60E-38	42.4
FXSY01009698.1:1176-1253	Minc3s09698g43458	883	908	2.60E-38	34.6
FXSY01002166.1:10449-10526		1552	1577	5.60E-52	42.3
FXSY01008451.1:5-85		1253	1279	4.90E-34	33.3
Query Sequence: <i>C. elegans vit-6</i>					
FXSY01000180.1:1963-2199	Minc3s00180g06901	109	187	2.30E-56	26.6
FXSY01000180.1:2224-2280	Minc3s00180g06901	180	198	2.30E-56	57.9
FXSY01000180.1:5010-5426	Minc3s00180g06902	519	657	2.30E-56	40.3
FXSY01000180.1:4715-4894	Minc3s00180g06902	438	497	2.30E-56	35
FXSY01000180.1:8500-8637	Minc3s00180g06902	1444	1489	2.30E-56	43.5
FXSY01000180.1:6187-6318	Minc3s00180g06902	853	896	2.30E-56	31.8
FXSY01000180.1:8242-8283	Minc3s00180g06902	1362	1375	2.30E-56	78.6
FXSY01000180.1:7941-8039	Minc3s00180g06902	1277	1309	2.30E-56	30.3
FXSY01000286.1:9225-9641	Minc3s00286g09435	519	657	5.60E-58	38.1
FXSY01000286.1:7550-7753	Minc3s00286g09435	109	176	5.60E-58	30.9
FXSY01000286.1:8922-9107	Minc3s00286g09435	436	497	5.60E-58	35.5
FXSY01000286.1:12756-12893	Minc3s00286g09435	1444	1489	5.60E-58	47.8
FXSY01000286.1:10402-10533	Minc3s00286g09435	853	896	5.60E-58	29.5
FXSY01000286.1:12486-12524	Minc3s00286g09435	1362	1374	5.60E-58	84.6
FXSY01000286.1:7811-7867	Minc3s00286g09435	180	198	5.60E-58	57.9
FXSY01000286.1:12186-12284	Minc3s00286g09435	1277	1309	5.60E-58	30.3
FXSY01000369.1:11844-12320	Minc3s00369g11129	523	681	4.60E-60	35.2
FXSY01000369.1:9930-10100	Minc3s00369g11129	113	169	4.60E-60	42.1
FXSY01000369.1:11388-11582	Minc3s00369g11129	442	506	4.60E-60	32.3
FXSY01000369.1:15436-15501	Minc3s00369g11129	1364	1385	4.60E-60	68.2
FXSY01000369.1:10137-10220	Minc3s00369g11129	180	207	4.60E-60	42.9
FXSY01000369.1:15733-15861	Minc3s00369g11130	1447	1489	4.60E-60	48.8
FXSY01000805.1:30416-30832	Minc3s00805g17578	519	657	8.60E-59	39.6
FXSY01000805.1:30944-31123	Minc3s00805g17578	438	497	8.60E-59	36.7
FXSY01000805.1:27217-27351	Minc3s00805g17578	1445	1489	8.60E-59	46.7
FXSY01000805.1:27571-27612	Minc3s00805g17578	1362	1375	8.60E-59	78.6
FXSY01000805.1:29527-29658	Minc3s00805g17578	853	896	8.60E-59	31.8
FXSY01000805.1:27815-27913	Minc3s00805g17578	1277	1309	8.60E-59	30.3
FXSY01000805.1:32489-32725	Minc3s00805g17579	109	187	8.60E-59	26.6
FXSY01000805.1:32408-32464	Minc3s00805g17579	180	198	8.60E-59	57.9
FXSY01001515.1:30664-31080	Minc3s01514g24409	519	657	5.10E-47	38.1

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01001515.1:31196-31375	Minc3s01514g24409	438	497	5.10E-47	36.7
FXSY01001515.1:27421-27558	Minc3s01514g24409	1444	1489	5.10E-47	47.8
FXSY01001515.1:27781-27819	Minc3s01514g24409	1362	1374	5.10E-47	84.6
FXSY01001515.1:28021-28119	Minc3s01514g24409	1277	1309	5.10E-47	30.3
FXSY01001515.1:29772-29903	Minc3s01514g24409	853	896	5.10E-47	27.3
FXSY01001575.1:27796-28212	Minc3s01575g24833	519	657	1.60E-47	38.1
FXSY01001575.1:28328-28513	Minc3s01575g24833	436	497	1.60E-47	35.5
FXSY01001575.1:26907-27038	Minc3s01575g24833	853	896	1.60E-47	29.5
FXSY01001575.1:29682-29885	Minc3s01575g24834	109	176	1.60E-47	30.9
FXSY01001575.1:29568-29624	Minc3s01575g24834	180	198	1.60E-47	57.9
FXSY01001861.1:3620-4096	Minc3s01861g26817	523	681	3.50E-59	36.5
FXSY01001861.1:1741-1911	Minc3s01861g26817	113	169	3.50E-59	42.1
FXSY01001861.1:3199-3360	Minc3s01861g26817	442	495	3.50E-59	37
FXSY01001861.1:7203-7268	Minc3s01861g26817	1364	1385	3.50E-59	68.2
FXSY01001861.1:1948-2031	Minc3s01861g26817	180	207	3.50E-59	42.9
FXSY01001861.1:7554-7628	Minc3s01861g26818	1465	1489	3.50E-59	56
FXSY01001975.1:7698-7826	Minc3s01975g27537	1447	1489	5.00E-61	46.5
FXSY01001975.1:8057-8122	Minc3s01975g27538	1364	1385	5.00E-61	68.2
FXSY01001975.1:11164-11640	Minc3s01975g27539	523	681	5.00E-61	35.2
FXSY01001975.1:13391-13561	Minc3s01975g27539	113	169	5.00E-61	42.1
FXSY01001975.1:11902-12096	Minc3s01975g27539	442	506	5.00E-61	30.8
FXSY01001975.1:13271-13354	Minc3s01975g27539	180	207	5.00E-61	42.9
FXSY01002166.1:6077-6493	Minc3s02166g28653	519	657	4.90E-62	38.1
FXSY01002166.1:5774-5959	Minc3s02166g28653	436	497	4.90E-62	35.5
FXSY01002166.1:9628-9765	Minc3s02166g28653	1444	1489	4.90E-62	47.8
FXSY01002166.1:7253-7384	Minc3s02166g28653	853	896	4.90E-62	29.5
FXSY01002166.1:9367-9405	Minc3s02166g28653	1362	1374	4.90E-62	84.6
FXSY01002166.1:9063-9161	Minc3s02166g28653	1277	1309	4.90E-62	30.3
FXSY01002166.1:4398-4634	Minc3s02166g28653, Minc3s02166g28654	109	187	4.90E-62	29.1
FXSY01002166.1:4659-4715	Minc3s02166g28653, Minc3s02166g28654	180	198	4.90E-62	57.9
FXSY01002260.1:17066-17203	Minc3s02260g29127	1444	1489	2.10E-09	41.3
FXSY01002260.1:16808-16849	Minc3s02260g29127	1362	1375	2.10E-09	78.6
FXSY01002260.1:14764-14895	Minc3s02260g29127	853	896	2.10E-09	31.8
FXSY01002260.1:16512-16610	Minc3s02260g29127	1277	1309	2.10E-09	30.3
FXSY01002607.1:12891-13307	Minc3s02607g30865	519	657	8.50E-62	39.6
FXSY01002607.1:12600-12779	Minc3s02607g30865	438	497	8.50E-62	36.7
FXSY01002607.1:16373-16507	Minc3s02607g30865	1445	1489	8.50E-62	46.7
FXSY01002607.1:16112-16153	Minc3s02607g30865	1362	1375	8.50E-62	78.6
FXSY01002607.1:14065-14196	Minc3s02607g30865	853	896	8.50E-62	31.8
FXSY01002607.1:15811-15909	Minc3s02607g30865	1277	1309	8.50E-62	30.3
FXSY01002607.1:11224-11460	Minc3s02607g30865, Minc3s02607g30866	109	187	8.50E-62	26.6
FXSY01002607.1:11485-11541	Minc3s02607g30865, Minc3s02607g30866	180	198	8.50E-62	57.9
FXSY01003289.1:77-493	Minc3s03289g33421	519	657	2.80E-47	38.1
FXSY01003289.1:1934-2170	Minc3s03289g33421	109	187	2.80E-47	29.1

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
FXSY01003289.1:612-797	Minc3s03289g33421	436	497	2.80E-47	35.5
FXSY01003289.1:1853-1909	Minc3s03289g33421	180	198	2.80E-47	57.9
FXSY01006543.1:745-882	Minc3s06543g39965	1444	1489	8.00E-12	47.8
FXSY01006543.1:472-522	Minc3s06543g39965	1362	1378	8.00E-12	70.6
FXSY01006543.1:172-270	Minc3s06543g39965	1277	1309	8.00E-12	30.3
FXSY01006543.1:1103-1267	Minc3s06543g39965	1504	1558	8.00E-12	27.3
FXSY01006972.1:1138-1374	Minc3s06972g40563, Minc3s06972g40564	109	187	1.30E-11	29.1
FXSY01006972.1:1057-1113	Minc3s06972g40563, Minc3s06972g40564	180	198	1.30E-11	57.9
FXSY01007386.1:2214-2351	Minc3s07387g41062	1444	1489	1.10E-09	47.8
FXSY01007386.1:2580-2618	Minc3s07387g41062	1362	1374	1.10E-09	84.6
FXSY01007386.1:2817-2915	Minc3s07387g41062	1277	1309	1.10E-09	30.3
FXSY01008451.1:2574-2990	Minc3s08451g42267	519	657	3.70E-38	38.1
FXSY01008451.1:3106-3285	Minc3s08451g42267	438	497	3.70E-38	36.7
FXSY01008451.1:1686-1817	Minc3s08451g42267	853	896	3.70E-38	29.5
FXSY01009698.1:2265-2681	Minc3s09698g43458	519	657	2.00E-28	37.4
FXSY01009698.1:1362-1493	Minc3s09698g43458	853	896	2.00E-28	29.5
FXSY01010887.1:1512-1640	Minc3s10886g44441	1447	1489	3.50E-09	46.5
FXSY01010887.1:1871-1936		1364	1385	3.50E-09	68.2

Table S3.1.8: BLAST result. Query sequence: *Caenorhabditis elegans vit* 1-6 gene sequences. Subject: *Meloidogyne javanica* genome.

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Query Sequence: <i>C. elegans vit</i>-1					
Scaff10759:84-236	Scaff10759g059768	97	147	2.00E-07	45.1
Scaff1399:7357-7428	Scaff1399g015376	1568	1591	5.10E-57	45.8
Scaff1399:11379-11783	Scaff1399g015377	494	628	5.10E-57	43
Scaff1399:13307-13459	Scaff1399g015377	97	147	5.10E-57	45.1
Scaff1399:9361-9552	Scaff1399g015377	1090	1153	5.10E-57	26.6
Scaff1399:10499-10576	Scaff1399g015377	837	862	5.10E-57	42.3
Scaff1399:10289-10363	Scaff1399g015377	888	912	5.10E-57	44
Scaff1399:8145-8198	Scaff1399g015377	1435	1452	5.10E-57	55.6
Scaff1399:12971-13057	Scaff1399g015377	200	228	5.10E-57	37.9
Scaff1399:13142-13243	Scaff1399g015377	158	191	5.10E-57	35.3
Scaff1515:9630-9710	Scaff1515g016285	1534	1560	3.30E-55	37
Scaff1515:9459-9551	Scaff1515g016285	1568	1598	3.30E-55	35.5
Scaff1515:10498-10551	Scaff1515g016285	1435	1452	3.30E-55	50
Scaff1515:14520-14939	Scaff1515g016286	495	634	3.30E-55	43.6
Scaff1515:16706-16861	Scaff1515g016286	96	147	3.30E-55	38.5
Scaff1515:15219-15389	Scaff1515g016286	420	476	3.30E-55	29.8
Scaff1515:12321-12473	Scaff1515g016286	1088	1138	3.30E-55	23.5
Scaff1515:16565-16639	Scaff1515g016286	173	197	3.30E-55	44

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff1515:16439-16531	Scaff1515g016286	200	230	3.30E-55	32.3
Scaff1515:15453-15551	Scaff1515g016286	382	414	3.30E-55	36.4
Scaff15289:3806-4168	Scaff15289g071985	494	614	4.00E-38	43.8
Scaff15289:1885-2037	Scaff15289g071985	97	147	4.00E-38	41.2
Scaff15289:2143-2199	Scaff15289g071985	173	191	4.00E-38	52.6
Scaff15289:2284-2370	Scaff15289g071985	200	228	4.00E-38	34.5
Scaff15289:5147-5224	Scaff15289g071986	837	862	4.00E-38	38.5
Scaff18769:3307-3489	Scaff18769g079434	1090	1150	4.30E-10	27.9
Scaff18769:1295-1366	Scaff18769g079434	1568	1591	4.30E-10	45.8
Scaff18769:2078-2131	Scaff18769g079434	1435	1452	4.30E-10	55.6
Scaff18769:2450-2482	Scaff18769g079434	1332	1342	4.30E-10	72.7
Scaff21123:1092-1283	Scaff21123g083914	1090	1153	1.60E-11	26.6
Scaff21123:2438-2491	Scaff21123g083914	1435	1452	1.60E-11	55.6
Scaff21123:2087-2119	Scaff21123g083914	1332	1342	1.60E-11	72.7
Scaff21123:2354-2449	Scaff21123g083914	1406	1437	1.60E-11	31.2
Scaff21123:3312-3383	Scaff21123g083915	1568	1591	1.60E-11	45.8
Scaff21949:700-852	Scaff21949g085322	97	147	1.50E-13	45.1
Scaff21949:535-636	Scaff21949g085322	158	191	1.50E-13	35.3
Scaff21949:364-450	Scaff21949g085322	200	228	1.50E-13	34.5
Scaff2226:10416-10724	Scaff2226g021515	494	596	9.50E-50	46.6
Scaff2226:8738-8890	Scaff2226g021515	97	147	9.50E-50	43.1
Scaff2226:12593-12778	Scaff2226g021515	1089	1150	9.50E-50	29
Scaff2226:11618-11695	Scaff2226g021515	837	862	9.50E-50	42.3
Scaff2226:10730-10819	Scaff2226g021515	599	628	9.50E-50	40
Scaff2226:13970-14023	Scaff2226g021515	1435	1452	9.50E-50	61.1
Scaff2226:14728-14799	Scaff2226g021515	1568	1591	9.50E-50	45.8
Scaff2226:8954-9055	Scaff2226g021515	158	191	9.50E-50	35.3
Scaff2226:13634-13675	Scaff2226g021515	1332	1345	9.50E-50	57.1
Scaff23110:709-891	Scaff23110g087285	1090	1150	3.10E-09	27.9
Scaff23110:2456-2533	Scaff23110g087286	837	862	3.10E-09	42.3
Scaff23110:2249-2323	Scaff23110g087286	888	912	3.10E-09	44
Scaff242:10005-10424	Scaff242g003894	495	634	1.50E-49	41.4
Scaff242:8083-8238	Scaff242g003894	96	147	1.50E-49	38.5
Scaff242:9540-9725	Scaff242g003894	415	476	1.50E-49	29
Scaff242:12455-12607	Scaff242g003894	1088	1138	1.50E-49	23.5
Scaff242:8305-8379	Scaff242g003894	173	197	1.50E-49	44
Scaff242:14876-14956	Scaff242g003895	1534	1560	1.50E-49	37
Scaff242:15044-15136	Scaff242g003895	1568	1598	1.50E-49	35.5
Scaff242:14037-14090	Scaff242g003895	1435	1452	1.50E-49	50
Scaff25080:1306-1710	Scaff25080g090352	494	628	1.90E-33	43.7
Scaff25080:431-508	Scaff25080g090352	837	862	1.90E-33	42.3
Scaff25080:226-312	Scaff25080g090352	884	912	1.90E-33	37.9
Scaff30898:834-905	Scaff30898g098056	1568	1591	0.0026	45.8

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff30898:1616-1669	Scaff30898g098056	1435	1452	0.0026	55.6
Scaff30898:1658-1753	Scaff30898g098056	1406	1437	0.0026	31.2
Scaff349:14245-14649	Scaff349g005192	494	628	2.00E-52	44.4
Scaff349:12279-12464	Scaff349g005192	1089	1150	2.00E-52	30.6
Scaff349:11035-11088	Scaff349g005192	1435	1452	2.00E-52	61.1
Scaff349:13365-13442	Scaff349g005192	837	862	2.00E-52	38.5
Scaff349:10306-10377	Scaff349g005192	1568	1591	2.00E-52	45.8
Scaff349:16007-16063	Scaff349g005192	173	191	2.00E-52	52.6
Scaff349:16172-16312	Scaff349g005192, Scaff349g005193	101	147	2.00E-52	46.8
Scaff3500:3694-4098	Scaff3500g029487	494	628	1.20E-54	43.7
Scaff3500:5625-5777	Scaff3500g029487	97	147	1.20E-54	45.1
Scaff3500:1219-1410	Scaff3500g029487	1090	1153	1.20E-54	26.6
Scaff3500:2348-2425	Scaff3500g029487	837	862	1.20E-54	42.3
Scaff3500:2155-2229	Scaff3500g029487	888	912	1.20E-54	44
Scaff3500:5460-5561	Scaff3500g029487	158	191	1.20E-54	35.3
Scaff3500:4-48	Scaff3500g029487	1435	1449	1.20E-54	60
Scaff3500:370-402	Scaff3500g029487	1332	1342	1.20E-54	72.7
Scaff3804:9684-10088	Scaff3804g031270	494	628	9.30E-59	44.4
Scaff3804:8010-8162	Scaff3804g031270	97	147	9.30E-59	45.1
Scaff3804:11931-12113	Scaff3804g031270	1090	1150	9.30E-59	27.9
Scaff3804:10886-10963	Scaff3804g031270	837	862	9.30E-59	42.3
Scaff3804:11082-11168	Scaff3804g031270	884	912	9.30E-59	37.9
Scaff3804:8226-8327	Scaff3804g031270	158	191	9.30E-59	35.3
Scaff3804:13308-13361	Scaff3804g031270	1435	1452	9.30E-59	55.6
Scaff3804:12960-12992	Scaff3804g031270	1332	1342	9.30E-59	72.7
Scaff3804:14085-14156	Scaff3804g031271	1568	1591	9.30E-59	45.8
Scaff6405:5991-6395	Scaff6405g043787	494	628	3.80E-61	45.2
Scaff6405:4317-4469	Scaff6405g043787	97	147	3.80E-61	43.1
Scaff6405:8163-8345	Scaff6405g043787	1090	1150	3.80E-61	27.9
Scaff6405:7186-7263	Scaff6405g043787	837	862	3.80E-61	42.3
Scaff6405:9549-9602	Scaff6405g043787	1435	1452	3.80E-61	61.1
Scaff6405:10254-10325	Scaff6405g043787	1568	1591	3.80E-61	45.8
Scaff6405:4578-4634	Scaff6405g043787	173	191	3.80E-61	52.6
Scaff6405:4719-4805	Scaff6405g043787	200	228	3.80E-61	34.5
Scaff6405:9207-9248	Scaff6405g043787	1332	1345	3.80E-61	57.1
Scaff6405:4004-4066	Scaff6405g043787	24	44	3.80E-61	52.4
Scaff6405:7383-7484	Scaff6405g043787	888	921	3.80E-61	29.4
Scaff6533:2229-2309	Scaff6533g044375	1534	1560	1.20E-56	37
Scaff6533:2049-2141	Scaff6533g044375	1568	1598	1.20E-56	35.5
Scaff6533:3183-3236	Scaff6533g044375	1435	1452	1.20E-56	50
Scaff6533:4815-4967	Scaff6533g044377	1088	1138	1.20E-56	23.5
Scaff6533:7097-7516	Scaff6533g044379	495	634	1.20E-56	41.4
Scaff6533:7798-7986	Scaff6533g044379	414	476	1.20E-56	28.6

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff6533:6221-6334	Scaff6533g044379	825	862	1.20E-56	26.3
Scaff6533:9283-9438	Scaff6533g044380	96	147	1.20E-56	38.5
Scaff6533:9142-9216	Scaff6533g044380	173	197	1.20E-56	44
Scaff6533:9016-9108	Scaff6533g044380	200	230	1.20E-56	32.3
Scaff6982:2634-3038	Scaff6982g046249	494	628	4.10E-46	43.7
Scaff6982:4562-4714	Scaff6982g046249	97	147	4.10E-46	45.1
Scaff6982:594-671	Scaff6982g046249	837	862	4.10E-46	42.3
Scaff6982:390-476	Scaff6982g046249	884	912	4.10E-46	37.9
Scaff6982:1475-1879	Scaff6982g046249	494	628	1.10E-29	43.7
Scaff6982:4397-4498	Scaff6982g046249, Scaff6982g046250	158	191	4.10E-46	35.3
Scaff6982:4226-4312	Scaff6982g046249, Scaff6982g046250	200	228	4.10E-46	34.5
Scaff7843:1502-1684	Scaff7843g049685	1090	1150	1.80E-54	27.9
Scaff7843:257-310	Scaff7843g049685	1435	1452	1.80E-54	55.6
Scaff7843:629-661	Scaff7843g049685	1332	1342	1.80E-54	72.7
Scaff7843:3781-3984	Scaff7843g049686	561	628	1.80E-54	42.6
Scaff7843:3994-4182	Scaff7843g049686	494	556	1.80E-54	49.2
Scaff7843:5708-5860	Scaff7843g049686	97	147	1.80E-54	45.1
Scaff7843:2900-2977	Scaff7843g049686	837	862	1.80E-54	42.3
Scaff7843:2695-2781	Scaff7843g049686	884	912	1.80E-54	37.9
Scaff7843:5543-5644	Scaff7843g049686	158	191	1.80E-54	35.3
Scaff7843:5372-5458	Scaff7843g049686	200	228	1.80E-54	34.5
Scaff7843:1779-1925	Scaff7843g049686	1090	1138	0.0003	32.7
Scaff8571:5626-5679	Scaff8571g052471	1435	1452	3.50E-13	61.1
Scaff8571:4884-4955	Scaff8571g052471	1568	1591	3.50E-13	45.8
Scaff8571:7406-7591	Scaff8571g052472	1089	1150	3.50E-13	30.6
Scaff8571:8496-8573	Scaff8571g052472	837	862	3.50E-13	38.5
Scaff8571:6481-6522	Scaff8571g052472	1332	1345	3.50E-13	57.1
Scaff9762:999-1079	Scaff9762g056627	1534	1560	1.40E-45	37
Scaff9762:828-920	Scaff9762g056627	1568	1598	1.40E-45	35.5
Scaff9762:1871-1924	Scaff9762g056627	1435	1452	1.40E-45	50
Scaff9762:3760-3906	Scaff9762g056629	1090	1138	1.40E-45	22.4
Scaff9762:5860-6279	Scaff9762g056631	495	634	1.40E-45	42.9
Scaff9762:6559-6729	Scaff9762g056631	420	476	1.40E-45	29.8
Scaff9762:4959-5072	Scaff9762g056631	825	862	1.40E-45	28.9
Scaff9762:6793-6891	Scaff9762g056631	382	414	1.40E-45	36.4
Scaff30898:1991-2032		1329	1342	0.0026	57.1
Query Sequence: <i>C. elegans vit-2</i>					
Scaff10759:84-236	Scaff10759g059768	97	147	1.10E-07	47.1
Scaff1399:11379-11783	Scaff1399g015377	492	626	1.00E-54	41.5
Scaff1399:13307-13459	Scaff1399g015377	97	147	1.00E-54	47.1
Scaff1399:9361-9552	Scaff1399g015377	1088	1151	1.00E-54	28.1
Scaff1399:10496-10576	Scaff1399g015377	835	861	1.00E-54	44.4
Scaff1399:10289-10363	Scaff1399g015377	886	910	1.00E-54	44

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff1399:8145-8198	Scaff1399g015377	1433	1450	1.00E-54	55.6
Scaff1399:12971-13057	Scaff1399g015377	200	228	1.00E-54	37.9
Scaff1399:13142-13243	Scaff1399g015377	158	191	1.00E-54	35.3
Scaff1515:9636-9710	Scaff1515g016285	1532	1556	5.50E-55	40
Scaff1515:10498-10551	Scaff1515g016285	1433	1450	5.50E-55	50
Scaff1515:14520-14939	Scaff1515g016286	493	632	5.50E-55	43.6
Scaff1515:16706-16861	Scaff1515g016286	96	147	5.50E-55	40.4
Scaff1515:15219-15374	Scaff1515g016286	423	474	5.50E-55	32.7
Scaff1515:12321-12575	Scaff1515g016286	1052	1136	5.50E-55	18.8
Scaff1515:16565-16639	Scaff1515g016286	173	197	5.50E-55	44
Scaff1515:13615-13728	Scaff1515g016286	824	861	5.50E-55	28.9
Scaff1515:16439-16531	Scaff1515g016286	200	230	5.50E-55	32.3
Scaff15289:3806-4168	Scaff15289g071985	492	612	2.50E-41	43.8
Scaff15289:1885-2037	Scaff15289g071985	97	147	2.50E-41	43.1
Scaff15289:2143-2199	Scaff15289g071985	173	191	2.50E-41	52.6
Scaff15289:2284-2370	Scaff15289g071985	200	228	2.50E-41	34.5
Scaff15289:1575-1631	Scaff15289g071985	24	42	2.50E-41	52.6
Scaff15289:5147-5227	Scaff15289g071986	835	861	2.50E-41	40.7
Scaff18769:3340-3489	Scaff18769g079434	1088	1137	1.40E-07	36
Scaff18769:2078-2131	Scaff18769g079434	1433	1450	1.40E-07	55.6
Scaff18769:2450-2482	Scaff18769g079434	1330	1340	1.40E-07	72.7
Scaff21123:1092-1283	Scaff21123g083914	1088	1151	3.20E-08	28.1
Scaff21123:2481-2564	Scaff21123g083914	1423	1450	3.20E-08	39.3
Scaff21123:2087-2119	Scaff21123g083914	1330	1340	3.20E-08	72.7
Scaff21949:700-852	Scaff21949g085322	97	147	8.70E-14	47.1
Scaff21949:535-636	Scaff21949g085322	158	191	8.70E-14	35.3
Scaff21949:364-450	Scaff21949g085322	200	228	8.70E-14	34.5
Scaff2226:10416-10724	Scaff2226g021515	492	594	8.90E-46	45.6
Scaff2226:8738-8890	Scaff2226g021515	97	147	8.90E-46	45.1
Scaff2226:12593-12745	Scaff2226g021515	1087	1137	8.90E-46	37.3
Scaff2226:11618-11698	Scaff2226g021515	835	861	8.90E-46	44.4
Scaff2226:13970-14023	Scaff2226g021515	1433	1450	8.90E-46	61.1
Scaff2226:8954-9055	Scaff2226g021515	158	191	8.90E-46	35.3
Scaff2226:13634-13675	Scaff2226g021515	1330	1343	8.90E-46	57.1
Scaff23110:742-891	Scaff23110g087285	1088	1137	2.10E-10	36
Scaff23110:2453-2533	Scaff23110g087286	835	861	2.10E-10	44.4
Scaff23110:2249-2323	Scaff23110g087286	886	910	2.10E-10	44
Scaff242:10005-10424	Scaff242g003894	493	632	3.40E-51	41.4
Scaff242:8083-8238	Scaff242g003894	96	147	3.40E-51	40.4
Scaff242:9570-9725	Scaff242g003894	423	474	3.40E-51	32.7
Scaff242:12353-12607	Scaff242g003894	1052	1136	3.40E-51	20

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff242:11204-11320	Scaff242g003894	823	861	3.40E-51	28.2
Scaff242:8305-8379	Scaff242g003894	173	197	3.40E-51	44
Scaff242:14876-14950	Scaff242g003895	1532	1556	3.40E-51	40
Scaff242:14037-14090	Scaff242g003895	1433	1450	3.40E-51	50
Scaff25080:1306-1710	Scaff25080g090352	492	626	2.80E-33	42.2
Scaff25080:428-508	Scaff25080g090352	835	861	2.80E-33	44.4
Scaff25080:226-312	Scaff25080g090352	882	910	2.80E-33	37.9
Scaff349:14245-14649	Scaff349g005192	492	626	5.50E-52	43.7
Scaff349:12312-12464	Scaff349g005192	1087	1137	5.50E-52	39.2
Scaff349:13362-13442	Scaff349g005192	835	861	5.50E-52	40.7
Scaff349:11035-11088	Scaff349g005192	1433	1450	5.50E-52	61.1
Scaff349:16007-16063	Scaff349g005192	173	191	5.50E-52	52.6
Scaff349:16172-16312	Scaff349g005192, Scaff349g005193	101	147	5.50E-52	48.9
Scaff3500:3694-4098	Scaff3500g029487	492	626	2.40E-56	42.2
Scaff3500:5625-5777	Scaff3500g029487	97	147	2.40E-56	47.1
Scaff3500:1219-1410	Scaff3500g029487	1088	1151	2.40E-56	28.1
Scaff3500:2345-2425	Scaff3500g029487	835	861	2.40E-56	44.4
Scaff3500:2155-2229	Scaff3500g029487	886	910	2.40E-56	44
Scaff3500:5460-5561	Scaff3500g029487	158	191	2.40E-56	35.3
Scaff3500:4-48	Scaff3500g029487	1433	1447	2.40E-56	60
Scaff3500:5991-6086	Scaff3500g029487	24	55	2.40E-56	34.4
Scaff3500:370-402	Scaff3500g029487	1330	1340	2.40E-56	72.7
Scaff3804:9684-10088	Scaff3804g031270	492	626	8.60E-58	43
Scaff3804:8010-8162	Scaff3804g031270	97	147	8.60E-58	47.1
Scaff3804:11931-12080	Scaff3804g031270	1088	1137	8.60E-58	36
Scaff3804:10886-10966	Scaff3804g031270	835	861	8.60E-58	44.4
Scaff3804:11082-11168	Scaff3804g031270	882	910	8.60E-58	37.9
Scaff3804:8226-8327	Scaff3804g031270	158	191	8.60E-58	35.3
Scaff3804:13308-13361	Scaff3804g031270	1433	1450	8.60E-58	55.6
Scaff3804:7701-7796	Scaff3804g031270	24	55	8.60E-58	34.4
Scaff3804:12960-12992	Scaff3804g031270	1330	1340	8.60E-58	72.7
Scaff6405:5991-6395	Scaff6405g043787	492	626	2.00E-57	43.7
Scaff6405:4317-4469	Scaff6405g043787	97	147	2.00E-57	45.1
Scaff6405:8163-8312	Scaff6405g043787	1088	1137	2.00E-57	36
Scaff6405:7186-7266	Scaff6405g043787	835	861	2.00E-57	44.4
Scaff6405:9549-9602	Scaff6405g043787	1433	1450	2.00E-57	61.1
Scaff6405:4578-4634	Scaff6405g043787	173	191	2.00E-57	52.6
Scaff6405:4004-4060	Scaff6405g043787	24	42	2.00E-57	57.9
Scaff6405:4719-4805	Scaff6405g043787	200	228	2.00E-57	34.5
Scaff6405:9207-9248	Scaff6405g043787	1330	1343	2.00E-57	57.1
Scaff6533:2235-2309	Scaff6533g044375	1532	1556	3.80E-56	40
Scaff6533:3183-3236	Scaff6533g044375	1433	1450	3.80E-56	50
Scaff6533:4815-5069	Scaff6533g044377	1052	1136	3.80E-56	18.8

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff6533:7097-7516	Scaff6533g044379	493	632	3.80E-56	41.4
Scaff6533:7798-7953	Scaff6533g044379	423	474	3.80E-56	32.7
Scaff6533:6218-6334	Scaff6533g044379	823	861	3.80E-56	28.2
Scaff6533:9283-9438	Scaff6533g044380	96	147	3.80E-56	40.4
Scaff6533:9142-9216	Scaff6533g044380	173	197	3.80E-56	44
Scaff6533:9016-9108	Scaff6533g044380	200	230	3.80E-56	32.3
Scaff6982:2634-3038	Scaff6982g046249	492	626	1.80E-47	42.2
Scaff6982:4562-4714	Scaff6982g046249	97	147	1.80E-47	47.1
Scaff6982:591-671	Scaff6982g046249	835	861	1.80E-47	44.4
Scaff6982:390-476	Scaff6982g046249	882	910	1.80E-47	37.9
Scaff6982:4928-5023	Scaff6982g046249	24	55	1.80E-47	34.4
Scaff6982:1475-1879	Scaff6982g046249	492	626	1.40E-28	42.2
Scaff6982:4397-4498	Scaff6982g046249, Scaff6982g046250	158	191	1.80E-47	35.3
Scaff6982:4226-4312	Scaff6982g046249, Scaff6982g046250	200	228	1.80E-47	34.5
Scaff7843:1535-1684	Scaff7843g049685	1088	1137	1.40E-56	36
Scaff7843:257-310	Scaff7843g049685	1433	1450	1.40E-56	55.6
Scaff7843:629-661	Scaff7843g049685	1330	1340	1.40E-56	72.7
Scaff7843:3781-3984	Scaff7843g049686	559	626	1.40E-56	41.2
Scaff7843:3994-4182	Scaff7843g049686	492	554	1.40E-56	47.6
Scaff7843:5708-5860	Scaff7843g049686	97	147	1.40E-56	47.1
Scaff7843:2897-2977	Scaff7843g049686	835	861	1.40E-56	44.4
Scaff7843:2695-2781	Scaff7843g049686	882	910	1.40E-56	37.9
Scaff7843:5543-5644	Scaff7843g049686	158	191	1.40E-56	35.3
Scaff7843:6074-6169	Scaff7843g049686	24	55	1.40E-56	34.4
Scaff7843:5372-5458	Scaff7843g049686	200	228	1.40E-56	34.5
Scaff7843:1779-1925	Scaff7843g049686	1088	1136	1.20E-04	34.7
Scaff8571:5626-5679	Scaff8571g052471	1433	1450	6.70E-12	61.1
Scaff8571:7439-7591	Scaff8571g052472	1087	1137	6.70E-12	39.2
Scaff8571:8493-8573	Scaff8571g052472	835	861	6.70E-12	40.7
Scaff8571:6481-6522	Scaff8571g052472	1330	1343	6.70E-12	57.1
Scaff9762:1005-1079	Scaff9762g056627	1532	1556	5.20E-43	40
Scaff9762:1871-1924	Scaff9762g056627	1433	1450	5.20E-43	50
Scaff9762:3760-4014	Scaff9762g056629	1052	1136	5.20E-43	17.6
Scaff9762:5860-6279	Scaff9762g056631	493	632	5.20E-43	42.9
Scaff9762:6559-6714	Scaff9762g056631	423	474	5.20E-43	32.7
Scaff9762:4956-5072	Scaff9762g056631	823	861	5.20E-43	30.8
Query Sequence: <i>C. elegans vit-3</i>					
Scaff1399:7357-7449	Scaff1399g015376	1547	1577	3.40E-50	35.5
Scaff1399:11364-11783	Scaff1399g015377	491	630	3.40E-50	37.1
Scaff1399:9361-9552	Scaff1399g015377	1086	1149	3.40E-50	23.4
Scaff1399:13310-13399	Scaff1399g015377	117	146	3.40E-50	40
Scaff1399:10496-10594	Scaff1399g015377	826	858	3.40E-50	42.4
Scaff1399:8745-8855	Scaff1399g015377	1253	1289	3.40E-50	29.7

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff1399:12257-12313	Scaff1399g015377	354	372	3.40E-50	52.6
Scaff1399:10286-10363	Scaff1399g015377	883	908	3.40E-50	34.6
Scaff1399:13142-13270	Scaff1399g015377	146	188	3.40E-50	25.6
Scaff1515:14406-14945	Scaff1515g016286	490	669	5.90E-45	35
Scaff1515:16583-16798	Scaff1515g016286	117	188	5.90E-45	34.7
Scaff1515:13615-13719	Scaff1515g016286	824	858	5.90E-45	28.6
Scaff1515:15240-15389	Scaff1515g016286	417	466	5.90E-45	32
Scaff1515:15537-15638	Scaff1515g016286	354	387	5.90E-45	38.2
Scaff15289:3806-4159	Scaff15289g071985	491	608	1.80E-33	39.8
Scaff15289:1945-2034	Scaff15289g071985	117	146	1.80E-33	36.7
Scaff15289:2146-2199	Scaff15289g071985	171	188	1.80E-33	55.6
Scaff15289:5129-5227	Scaff15289g071986	826	858	1.80E-33	39.4
Scaff15289:4270-4362	Scaff15289g071986	600	630	1.80E-33	38.7
Scaff18769:3307-3489	Scaff18769g079434	1086	1146	6.00E-10	24.6
Scaff18769:2681-2791	Scaff18769g079434	1253	1289	6.00E-10	29.7
Scaff18769:1295-1372	Scaff18769g079434	1552	1577	6.00E-10	38.5
Scaff18769:2450-2482	Scaff18769g079434	1328	1338	6.00E-10	72.7
Scaff21123:1092-1283	Scaff21123g083914	1086	1149	2.90E-12	23.4
Scaff21123:1781-1891	Scaff21123g083914	1253	1289	2.90E-12	29.7
Scaff21123:2087-2119	Scaff21123g083914	1328	1338	2.90E-12	72.7
Scaff21123:2481-2564	Scaff21123g083914	1421	1448	2.90E-12	39.3
Scaff21123:3291-3383	Scaff21123g083915	1547	1577	2.90E-12	35.5
Scaff21949:703-792	Scaff21949g085322	117	146	1.10E-04	40
Scaff21949:535-636	Scaff21949g085322	155	188	1.10E-04	29.4
Scaff2226:10416-10724	Scaff2226g021515	491	593	6.90E-42	40.8
Scaff2226:12593-12778	Scaff2226g021515	1085	1146	6.90E-42	25.8
Scaff2226:8798-8887	Scaff2226g021515	117	146	6.90E-42	36.7
Scaff2226:11600-11698	Scaff2226g021515	826	858	6.90E-42	39.4
Scaff2226:13329-13439	Scaff2226g021515	1253	1289	6.90E-42	29.7
Scaff2226:10730-10834	Scaff2226g021515	596	630	6.90E-42	37.1
Scaff2226:11817-11918	Scaff2226g021515	883	916	6.90E-42	29.4
Scaff2226:9002-9055	Scaff2226g021515	171	188	6.90E-42	50
Scaff23110:709-891	Scaff23110g087285	1086	1146	1.60E-14	24.6
Scaff23110:95-205	Scaff23110g087285	1253	1289	1.60E-14	29.7
Scaff23110:2453-2551	Scaff23110g087286	826	858	1.60E-14	42.4
Scaff23110:2246-2323	Scaff23110g087286	883	908	1.60E-14	34.6
Scaff242:9999-10538	Scaff242g003894	490	669	1.30E-42	35.6
Scaff242:8146-8361	Scaff242g003894	117	188	1.30E-42	33.3
Scaff242:11216-11320	Scaff242g003894	824	858	1.30E-42	28.6
Scaff242:9543-9704	Scaff242g003894	413	466	1.30E-42	31.5
Scaff242:9304-9405	Scaff242g003894	354	387	1.30E-42	38.2

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff25080:1291-1710	Scaff25080g090352	491	630	2.10E-33	37.9
Scaff25080:428-526	Scaff25080g090352	826	858	2.10E-33	42.4
Scaff25080:2185-2241	Scaff25080g090352	354	372	2.10E-33	52.6
Scaff25080:223-300	Scaff25080g090352	883	908	2.10E-33	34.6
Scaff349:14230-14649	Scaff349g005192	491	630	4.90E-48	39.3
Scaff349:12279-12464	Scaff349g005192	1085	1146	4.90E-48	27.4
Scaff349:11627-11737	Scaff349g005192	1253	1289	4.90E-48	32.4
Scaff349:13362-13460	Scaff349g005192	826	858	4.90E-48	39.4
Scaff349:13143-13244	Scaff349g005192	883	916	4.90E-48	26.5
Scaff349:16007-16060	Scaff349g005192	171	188	4.90E-48	50
Scaff349:16175-16264	Scaff349g005192, Scaff349g005193	117	146	4.90E-48	36.7
Scaff3500:3679-4098	Scaff3500g029487	491	630	7.00E-50	37.9
Scaff3500:1219-1410	Scaff3500g029487	1086	1149	7.00E-50	23.4
Scaff3500:5628-5717	Scaff3500g029487	117	146	7.00E-50	40
Scaff3500:2345-2443	Scaff3500g029487	826	858	7.00E-50	42.4
Scaff3500:596-706	Scaff3500g029487	1253	1289	7.00E-50	29.7
Scaff3500:2125-2229	Scaff3500g029487	883	917	7.00E-50	28.6
Scaff3500:4573-4629	Scaff3500g029487	354	372	7.00E-50	52.6
Scaff3500:5460-5561	Scaff3500g029487	155	188	7.00E-50	29.4
Scaff3804:9684-10103	Scaff3804g031270	491	630	1.60E-51	37.9
Scaff3804:11931-12113	Scaff3804g031270	1086	1146	1.60E-51	24.6
Scaff3804:8070-8159	Scaff3804g031270	117	146	1.60E-51	40
Scaff3804:10868-10966	Scaff3804g031270	826	858	1.60E-51	42.4
Scaff3804:12651-12761	Scaff3804g031270	1253	1289	1.60E-51	29.7
Scaff3804:9161-9217	Scaff3804g031270	354	372	1.60E-51	52.6
Scaff3804:11094-11171	Scaff3804g031270	883	908	1.60E-51	34.6
Scaff3804:8226-8327	Scaff3804g031270	155	188	1.60E-51	29.4
Scaff3804:14082-14156	Scaff3804g031271	1553	1577	1.60E-51	40
Scaff6405:5991-6410	Scaff6405g043787	491	630	9.00E-50	39.3
Scaff6405:8163-8345	Scaff6405g043787	1086	1146	9.00E-50	24.6
Scaff6405:8894-9004	Scaff6405g043787	1253	1289	9.00E-50	32.4
Scaff6405:4377-4466	Scaff6405g043787	117	146	9.00E-50	36.7
Scaff6405:7168-7266	Scaff6405g043787	826	858	9.00E-50	39.4
Scaff6405:7383-7484	Scaff6405g043787	883	916	9.00E-50	29.4
Scaff6405:4581-4634	Scaff6405g043787	171	188	9.00E-50	50
Scaff6405:9207-9239	Scaff6405g043787	1328	1338	9.00E-50	72.7
Scaff6533:6983-7522	Scaff6533g044379	490	669	8.20E-46	35.6
Scaff6533:7819-7995	Scaff6533g044379	408	466	8.20E-46	30.5
Scaff6533:6218-6322	Scaff6533g044379	824	858	8.20E-46	28.6
Scaff6533:8117-8218	Scaff6533g044379	354	387	8.20E-46	38.2
Scaff6533:9160-9375	Scaff6533g044380	117	188	8.20E-46	33.3
Scaff6982:2619-3038	Scaff6982g046249	491	630	1.50E-38	37.9
Scaff6982:4565-4654	Scaff6982g046249	117	146	1.50E-38	40

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff6982:591-689	Scaff6982g046249	826	858	1.50E-38	42.4
Scaff6982:3510-3566	Scaff6982g046249	354	372	1.50E-38	52.6
Scaff6982:387-464	Scaff6982g046249	883	908	1.50E-38	34.6
Scaff6982:1460-1879	Scaff6982g046249	491	630	1.10E-25	37.9
Scaff6982:4397-4498	Scaff6982g046249, Scaff6982g046250	155	188	1.50E-38	29.4
Scaff7843:1502-1684	Scaff7843g049685	1086	1146	1.70E-48	24.6
Scaff7843:860-970	Scaff7843g049685	1253	1289	1.70E-48	29.7
Scaff7843:629-661	Scaff7843g049685	1328	1338	1.70E-48	72.7
Scaff7843:3766-3984	Scaff7843g049686	558	630	1.70E-48	38.4
Scaff7843:4003-4182	Scaff7843g049686	491	550	1.70E-48	41.7
Scaff7843:5711-5800	Scaff7843g049686	117	146	1.70E-48	40
Scaff7843:2897-2995	Scaff7843g049686	826	858	1.70E-48	42.4
Scaff7843:4653-4709	Scaff7843g049686	354	372	1.70E-48	52.6
Scaff7843:2692-2769	Scaff7843g049686	883	908	1.70E-48	34.6
Scaff7843:5543-5644	Scaff7843g049686	155	188	1.70E-48	29.4
Scaff7843:1779-1925	Scaff7843g049686	1086	1134	5.20E-03	28.6
Scaff8571:7406-7591	Scaff8571g052472	1085	1146	1.40E-15	27.4
Scaff8571:6719-6829	Scaff8571g052472	1253	1289	1.40E-15	32.4
Scaff8571:8493-8591	Scaff8571g052472	826	858	1.40E-15	39.4
Scaff8571:8274-8369	Scaff8571g052472	885	916	1.40E-15	25
Scaff8571:6490-6522	Scaff8571g052472	1328	1338	1.40E-15	72.7
Scaff9762:5746-6285	Scaff9762g056631	490	669	6.50E-35	35
Scaff9762:4956-5060	Scaff9762g056631	824	858	6.50E-35	28.6
Scaff9762:6580-6729	Scaff9762g056631	417	466	6.50E-35	32
Scaff9762:6877-6978	Scaff9762g056631	354	387	6.50E-35	38.2
Query Sequence: <i>C. elegans vit-4</i>					
Scaff1399:7357-7449	Scaff1399g015376	1547	1577	2.90E-51	38.7
Scaff1399:11364-11783	Scaff1399g015377	491	630	2.90E-51	37.1
Scaff1399:9361-9552	Scaff1399g015377	1086	1149	2.90E-51	25
Scaff1399:13310-13399	Scaff1399g015377	117	146	2.90E-51	40
Scaff1399:10496-10594	Scaff1399g015377	826	858	2.90E-51	42.4
Scaff1399:8745-8855	Scaff1399g015377	1253	1289	2.90E-51	29.7
Scaff1399:12257-12313	Scaff1399g015377	354	372	2.90E-51	52.6
Scaff1399:10286-10363	Scaff1399g015377	883	908	2.90E-51	34.6
Scaff1399:13142-13270	Scaff1399g015377	146	188	2.90E-51	25.6
Scaff1515:14406-14945	Scaff1515g016286	490	669	4.50E-45	35
Scaff1515:16583-16798	Scaff1515g016286	117	188	4.50E-45	34.7
Scaff1515:13615-13719	Scaff1515g016286	824	858	4.50E-45	28.6
Scaff1515:15240-15389	Scaff1515g016286	417	466	4.50E-45	32
Scaff1515:15537-15638	Scaff1515g016286	354	387	4.50E-45	38.2
Scaff15289:3806-4159	Scaff15289g071985	491	608	1.50E-33	39.8
Scaff15289:1945-2034	Scaff15289g071985	117	146	1.50E-33	36.7

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff15289:2146-2199	Scaff15289g071985	171	188	1.50E-33	55.6
Scaff15289:5129-5227	Scaff15289g071986	826	858	1.50E-33	39.4
Scaff15289:4270-4362	Scaff15289g071986	600	630	1.50E-33	38.7
Scaff18769:3307-3489	Scaff18769g079434	1086	1146	6.70E-11	26.2
Scaff18769:2681-2791	Scaff18769g079434	1253	1289	6.70E-11	29.7
Scaff18769:1295-1372	Scaff18769g079434	1552	1577	6.70E-11	42.3
Scaff18769:2450-2482	Scaff18769g079434	1328	1338	6.70E-11	72.7
Scaff21123:1092-1283	Scaff21123g083914	1086	1149	3.20E-13	25
Scaff21123:1781-1891	Scaff21123g083914	1253	1289	3.20E-13	29.7
Scaff21123:2087-2119	Scaff21123g083914	1328	1338	3.20E-13	72.7
Scaff21123:2481-2564	Scaff21123g083914	1421	1448	3.20E-13	39.3
Scaff21123:3291-3383	Scaff21123g083915	1547	1577	3.20E-13	38.7
Scaff21949:703-792	Scaff21949g085322	117	146	1.00E-04	40
Scaff21949:535-636	Scaff21949g085322	155	188	1.00E-04	29.4
Scaff2226:10416-10724	Scaff2226g021515	491	593	6.90E-44	40.8
Scaff2226:12593-12778	Scaff2226g021515	1085	1146	6.90E-44	27.4
Scaff2226:8798-8887	Scaff2226g021515	117	146	6.90E-44	36.7
Scaff2226:11600-11698	Scaff2226g021515	826	858	6.90E-44	39.4
Scaff2226:13329-13439	Scaff2226g021515	1253	1289	6.90E-44	29.7
Scaff2226:10730-10834	Scaff2226g021515	596	630	6.90E-44	37.1
Scaff2226:11817-11918	Scaff2226g021515	883	916	6.90E-44	29.4
Scaff2226:9002-9055	Scaff2226g021515	171	188	6.90E-44	50
Scaff2226:14710-14799	Scaff2226g021515	1548	1577	6.90E-44	33.3
Scaff23110:709-891	Scaff23110g087285	1086	1146	5.70E-15	26.2
Scaff23110:95-205	Scaff23110g087285	1253	1289	5.70E-15	29.7
Scaff23110:2453-2551	Scaff23110g087286	826	858	5.70E-15	42.4
Scaff23110:2246-2323	Scaff23110g087286	883	908	5.70E-15	34.6
Scaff242:9999-10538	Scaff242g003894	490	669	9.80E-43	35.6
Scaff242:8146-8361	Scaff242g003894	117	188	9.80E-43	33.3
Scaff242:11216-11320	Scaff242g003894	824	858	9.80E-43	28.6
Scaff242:9543-9704	Scaff242g003894	413	466	9.80E-43	31.5
Scaff242:9304-9405	Scaff242g003894	354	387	9.80E-43	38.2
Scaff25080:1291-1710	Scaff25080g090352	491	630	1.70E-33	37.9
Scaff25080:428-526	Scaff25080g090352	826	858	1.70E-33	42.4
Scaff25080:2185-2241	Scaff25080g090352	354	372	1.70E-33	52.6
Scaff25080:223-300	Scaff25080g090352	883	908	1.70E-33	34.6
Scaff349:14230-14649	Scaff349g005192	491	630	1.40E-48	39.3
Scaff349:12279-12464	Scaff349g005192	1085	1146	1.40E-48	29
Scaff349:11627-11737	Scaff349g005192	1253	1289	1.40E-48	32.4
Scaff349:13362-13460	Scaff349g005192	826	858	1.40E-48	39.4
Scaff349:13143-13244	Scaff349g005192	883	916	1.40E-48	26.5
Scaff349:16007-16060	Scaff349g005192	171	188	1.40E-48	50
Scaff349:16175-16264	Scaff349g005192, Scaff349g005193	117	146	1.40E-48	36.7

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff3500:3679-4098	Scaff3500g029487	491	630	2.00E-50	37.9
Scaff3500:1219-1410	Scaff3500g029487	1086	1149	2.00E-50	25
Scaff3500:5628-5717	Scaff3500g029487	117	146	2.00E-50	40
Scaff3500:2345-2443	Scaff3500g029487	826	858	2.00E-50	42.4
Scaff3500:596-706	Scaff3500g029487	1253	1289	2.00E-50	29.7
Scaff3500:2125-2229	Scaff3500g029487	883	917	2.00E-50	28.6
Scaff3500:4573-4629	Scaff3500g029487	354	372	2.00E-50	52.6
Scaff3500:5460-5561	Scaff3500g029487	155	188	2.00E-50	29.4
Scaff3804:9684-10103	Scaff3804g031270	491	630	1.40E-52	37.9
Scaff3804:11931-12113	Scaff3804g031270	1086	1146	1.40E-52	26.2
Scaff3804:8070-8159	Scaff3804g031270	117	146	1.40E-52	40
Scaff3804:10868-10966	Scaff3804g031270	826	858	1.40E-52	42.4
Scaff3804:12651-12761	Scaff3804g031270	1253	1289	1.40E-52	29.7
Scaff3804:9161-9217	Scaff3804g031270	354	372	1.40E-52	52.6
Scaff3804:11094-11171	Scaff3804g031270	883	908	1.40E-52	34.6
Scaff3804:8226-8327	Scaff3804g031270	155	188	1.40E-52	29.4
Scaff3804:14082-14156	Scaff3804g031271	1553	1577	1.40E-52	44
Scaff6405:5991-6410	Scaff6405g043787	491	630	1.10E-51	39.3
Scaff6405:8163-8345	Scaff6405g043787	1086	1146	1.10E-51	26.2
Scaff6405:8894-9004	Scaff6405g043787	1253	1289	1.10E-51	32.4
Scaff6405:4377-4466	Scaff6405g043787	117	146	1.10E-51	36.7
Scaff6405:7168-7266	Scaff6405g043787	826	858	1.10E-51	39.4
Scaff6405:7383-7484	Scaff6405g043787	883	916	1.10E-51	29.4
Scaff6405:4581-4634	Scaff6405g043787	171	188	1.10E-51	50
Scaff6405:9207-9239	Scaff6405g043787	1328	1338	1.10E-51	72.7
Scaff6405:10248-10325	Scaff6405g043787	1552	1577	1.10E-51	34.6
Scaff6533:6983-7522	Scaff6533g044379	490	669	6.30E-46	35.6
Scaff6533:7819-7995	Scaff6533g044379	408	466	6.30E-46	30.5
Scaff6533:6218-6322	Scaff6533g044379	824	858	6.30E-46	28.6
Scaff6533:8117-8218	Scaff6533g044379	354	387	6.30E-46	38.2
Scaff6533:9160-9375	Scaff6533g044380	117	188	6.30E-46	33.3
Scaff6982:2619-3038	Scaff6982g046249	491	630	1.10E-38	37.9
Scaff6982:4565-4654	Scaff6982g046249	117	146	1.10E-38	40
Scaff6982:591-689	Scaff6982g046249	826	858	1.10E-38	42.4
Scaff6982:3510-3566	Scaff6982g046249	354	372	1.10E-38	52.6
Scaff6982:387-464	Scaff6982g046249	883	908	1.10E-38	34.6
Scaff6982:1460-1879	Scaff6982g046249	491	630	9.70E-26	37.9
Scaff6982:4397-4498	Scaff6982g046249, Scaff6982g046250	155	188	1.10E-38	29.4
Scaff7843:1502-1684	Scaff7843g049685	1086	1146	5.00E-49	26.2
Scaff7843:860-970	Scaff7843g049685	1253	1289	5.00E-49	29.7
Scaff7843:629-661	Scaff7843g049685	1328	1338	5.00E-49	72.7
Scaff7843:3766-3984	Scaff7843g049686	558	630	5.00E-49	38.4
Scaff7843:4003-4182	Scaff7843g049686	491	550	5.00E-49	41.7

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff7843:5711-5800	Scaff7843g049686	117	146	5.00E-49	40
Scaff7843:2897-2995	Scaff7843g049686	826	858	5.00E-49	42.4
Scaff7843:4653-4709	Scaff7843g049686	354	372	5.00E-49	52.6
Scaff7843:2692-2769	Scaff7843g049686	883	908	5.00E-49	34.6
Scaff7843:5543-5644	Scaff7843g049686	155	188	5.00E-49	29.4
Scaff7843:1779-1925	Scaff7843g049686	1086	1134	1.90E-03	30.6
Scaff8571:4884-4973	Scaff8571g052471	1548	1577	1.50E-17	33.3
Scaff8571:7406-7591	Scaff8571g052472	1085	1146	1.50E-17	29
Scaff8571:6719-6829	Scaff8571g052472	1253	1289	1.50E-17	32.4
Scaff8571:8493-8591	Scaff8571g052472	826	858	1.50E-17	39.4
Scaff8571:8274-8369	Scaff8571g052472	885	916	1.50E-17	25
Scaff8571:6490-6522	Scaff8571g052472	1328	1338	1.50E-17	72.7
Scaff9762:5746-6285	Scaff9762g056631	490	669	5.20E-35	35
Scaff9762:4956-5060	Scaff9762g056631	824	858	5.20E-35	28.6
Scaff9762:6580-6729	Scaff9762g056631	417	466	5.20E-35	32
Scaff9762:6877-6978	Scaff9762g056631	354	387	5.20E-35	38.2
Query Sequence: <i>C. elegans vit-5</i>					
Scaff10759:87-236	Scaff10759g059768	97	146	3.70E-03	34
Scaff1399:7357-7449	Scaff1399g015376	1547	1577	2.20E-52	38.7
Scaff1399:11364-11783	Scaff1399g015377	491	630	2.20E-52	37.9
Scaff1399:9361-9552	Scaff1399g015377	1086	1149	2.20E-52	23.4
Scaff1399:13310-13459	Scaff1399g015377	97	146	2.20E-52	34
Scaff1399:10496-10594	Scaff1399g015377	826	858	2.20E-52	42.4
Scaff1399:8745-8855	Scaff1399g015377	1253	1289	2.20E-52	29.7
Scaff1399:12257-12313	Scaff1399g015377	354	372	2.20E-52	52.6
Scaff1399:10286-10363	Scaff1399g015377	883	908	2.20E-52	34.6
Scaff1399:13142-13270	Scaff1399g015377	146	188	2.20E-52	25.6
Scaff1515:14406-14945	Scaff1515g016286	490	669	3.70E-47	35.6
Scaff1515:16583-16861	Scaff1515g016286	96	188	3.70E-47	32.3
Scaff1515:13615-13719	Scaff1515g016286	824	858	3.70E-47	28.6
Scaff1515:15240-15389	Scaff1515g016286	417	466	3.70E-47	32
Scaff1515:15537-15638	Scaff1515g016286	354	387	3.70E-47	38.2
Scaff15289:3806-4159	Scaff15289g071985	491	608	1.90E-34	40.7
Scaff15289:1885-2034	Scaff15289g071985	97	146	1.90E-34	32
Scaff15289:2146-2199	Scaff15289g071985	171	188	1.90E-34	55.6
Scaff15289:5129-5227	Scaff15289g071986	826	858	1.90E-34	39.4
Scaff15289:4270-4362	Scaff15289g071986	600	630	1.90E-34	38.7
Scaff18769:3307-3489	Scaff18769g079434	1086	1146	1.70E-10	24.6
Scaff18769:2681-2791	Scaff18769g079434	1253	1289	1.70E-10	29.7
Scaff18769:1295-1372	Scaff18769g079434	1552	1577	1.70E-10	42.3
Scaff18769:2450-2482	Scaff18769g079434	1328	1338	1.70E-10	72.7
Scaff21123:1092-1283	Scaff21123g083914	1086	1149	8.20E-13	23.4

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff21123:1781-1891	Scaff21123g083914	1253	1289	8.20E-13	29.7
Scaff21123:2087-2119	Scaff21123g083914	1328	1338	8.20E-13	72.7
Scaff21123:2481-2564	Scaff21123g083914	1421	1448	8.20E-13	39.3
Scaff21123:3291-3383	Scaff21123g083915	1547	1577	8.20E-13	38.7
Scaff21949:703-852	Scaff21949g085322	97	146	2.00E-06	34
Scaff21949:535-636	Scaff21949g085322	155	188	2.00E-06	29.4
Scaff2226:10416-10724	Scaff2226g021515	491	593	5.30E-45	41.7
Scaff2226:12593-12778	Scaff2226g021515	1085	1146	5.30E-45	25.8
Scaff2226:8738-8887	Scaff2226g021515	97	146	5.30E-45	32
Scaff2226:11600-11698	Scaff2226g021515	826	858	5.30E-45	39.4
Scaff2226:13329-13439	Scaff2226g021515	1253	1289	5.30E-45	29.7
Scaff2226:10730-10834	Scaff2226g021515	596	630	5.30E-45	37.1
Scaff2226:11817-11918	Scaff2226g021515	883	916	5.30E-45	29.4
Scaff2226:9002-9055	Scaff2226g021515	171	188	5.30E-45	50
Scaff2226:14710-14799	Scaff2226g021515	1548	1577	5.30E-45	33.3
Scaff23110:709-891	Scaff23110g087285	1086	1146	1.50E-14	24.6
Scaff23110:95-205	Scaff23110g087285	1253	1289	1.50E-14	29.7
Scaff23110:2453-2551	Scaff23110g087286	826	858	1.50E-14	42.4
Scaff23110:2246-2323	Scaff23110g087286	883	908	1.50E-14	34.6
Scaff242:9999-10538	Scaff242g003894	490	669	8.10E-45	36.1
Scaff242:8083-8361	Scaff242g003894	96	188	8.10E-45	31.2
Scaff242:11216-11320	Scaff242g003894	824	858	8.10E-45	28.6
Scaff242:9543-9704	Scaff242g003894	413	466	8.10E-45	31.5
Scaff242:9304-9405	Scaff242g003894	354	387	8.10E-45	38.2
Scaff25080:1291-1710	Scaff25080g090352	491	630	2.50E-33	38.6
Scaff25080:428-526	Scaff25080g090352	826	858	2.50E-33	42.4
Scaff25080:2185-2241	Scaff25080g090352	354	372	2.50E-33	52.6
Scaff25080:223-300	Scaff25080g090352	883	908	2.50E-33	34.6
Scaff349:14230-14649	Scaff349g005192	491	630	3.50E-49	40
Scaff349:12279-12464	Scaff349g005192	1085	1146	3.50E-49	27.4
Scaff349:11627-11737	Scaff349g005192	1253	1289	3.50E-49	32.4
Scaff349:13362-13460	Scaff349g005192	826	858	3.50E-49	39.4
Scaff349:13143-13244	Scaff349g005192	883	916	3.50E-49	26.5
Scaff349:16007-16108	Scaff349g005192	155	188	3.50E-49	29.4
Scaff349:16175-16309	Scaff349g005192, Scaff349g005193	102	146	3.50E-49	35.6
Scaff3500:3679-4098	Scaff3500g029487	491	630	1.50E-51	38.6
Scaff3500:1219-1410	Scaff3500g029487	1086	1149	1.50E-51	23.4
Scaff3500:5628-5777	Scaff3500g029487	97	146	1.50E-51	34
Scaff3500:2345-2443	Scaff3500g029487	826	858	1.50E-51	42.4
Scaff3500:596-706	Scaff3500g029487	1253	1289	1.50E-51	29.7
Scaff3500:2125-2229	Scaff3500g029487	883	917	1.50E-51	28.6
Scaff3500:4573-4629	Scaff3500g029487	354	372	1.50E-51	52.6
Scaff3500:5460-5561	Scaff3500g029487	155	188	1.50E-51	29.4

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff3804:9684-10103	Scaff3804g031270	491	630	1.00E-53	38.6
Scaff3804:11931-12113	Scaff3804g031270	1086	1146	1.00E-53	24.6
Scaff3804:8010-8159	Scaff3804g031270	97	146	1.00E-53	34
Scaff3804:10868-10966	Scaff3804g031270	826	858	1.00E-53	42.4
Scaff3804:12651-12761	Scaff3804g031270	1253	1289	1.00E-53	29.7
Scaff3804:9161-9217	Scaff3804g031270	354	372	1.00E-53	52.6
Scaff3804:11094-11171	Scaff3804g031270	883	908	1.00E-53	34.6
Scaff3804:8226-8327	Scaff3804g031270	155	188	1.00E-53	29.4
Scaff3804:14082-14156	Scaff3804g031271	1553	1577	1.00E-53	44
Scaff6405:5991-6410	Scaff6405g043787	491	630	6.00E-53	40
Scaff6405:8163-8345	Scaff6405g043787	1086	1146	6.00E-53	24.6
Scaff6405:4317-4466	Scaff6405g043787	97	146	6.00E-53	32
Scaff6405:8894-9004	Scaff6405g043787	1253	1289	6.00E-53	32.4
Scaff6405:7168-7266	Scaff6405g043787	826	858	6.00E-53	39.4
Scaff6405:7383-7484	Scaff6405g043787	883	916	6.00E-53	29.4
Scaff6405:4533-4634	Scaff6405g043787	155	188	6.00E-53	29.4
Scaff6405:9207-9239	Scaff6405g043787	1328	1338	6.00E-53	72.7
Scaff6405:10248-10325	Scaff6405g043787	1552	1577	6.00E-53	34.6
Scaff6533:6983-7522	Scaff6533g044379	490	669	5.20E-48	36.1
Scaff6533:7819-7995	Scaff6533g044379	408	466	5.20E-48	30.5
Scaff6533:6218-6322	Scaff6533g044379	824	858	5.20E-48	28.6
Scaff6533:8117-8218	Scaff6533g044379	354	387	5.20E-48	38.2
Scaff6533:9160-9438	Scaff6533g044380	96	188	5.20E-48	31.2
Scaff6982:2619-3038	Scaff6982g046249	491	630	3.30E-40	38.6
Scaff6982:4565-4714	Scaff6982g046249	97	146	3.30E-40	34
Scaff6982:591-689	Scaff6982g046249	826	858	3.30E-40	42.4
Scaff6982:3510-3566	Scaff6982g046249	354	372	3.30E-40	52.6
Scaff6982:387-464	Scaff6982g046249	883	908	3.30E-40	34.6
Scaff6982:1460-1879	Scaff6982g046249	491	630	1.40E-25	38.6
Scaff6982:4397-4498	Scaff6982g046249, Scaff6982g046250	155	188	3.30E-40	29.4
Scaff7843:1502-1684	Scaff7843g049685	1086	1146	3.90E-50	24.6
Scaff7843:860-970	Scaff7843g049685	1253	1289	3.90E-50	29.7
Scaff7843:629-661	Scaff7843g049685	1328	1338	3.90E-50	72.7
Scaff7843:3766-3984	Scaff7843g049686	558	630	3.90E-50	38.4
Scaff7843:4003-4182	Scaff7843g049686	491	550	3.90E-50	43.3
Scaff7843:5711-5860	Scaff7843g049686	97	146	3.90E-50	34
Scaff7843:2897-2995	Scaff7843g049686	826	858	3.90E-50	42.4
Scaff7843:4653-4709	Scaff7843g049686	354	372	3.90E-50	52.6
Scaff7843:2692-2769	Scaff7843g049686	883	908	3.90E-50	34.6
Scaff7843:5543-5644	Scaff7843g049686	155	188	3.90E-50	29.4
Scaff7843:1779-1925	Scaff7843g049686	1086	1134	5.00E-03	28.6
Scaff8571:4884-4973	Scaff8571g052471	1548	1577	3.80E-17	33.3
Scaff8571:7406-7591	Scaff8571g052472	1085	1146	3.80E-17	27.4

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff8571:6719-6829	Scaff8571g052472	1253	1289	3.80E-17	32.4
Scaff8571:8493-8591	Scaff8571g052472	826	858	3.80E-17	39.4
Scaff8571:8274-8369	Scaff8571g052472	885	916	3.80E-17	25
Scaff8571:6490-6522	Scaff8571g052472	1328	1338	3.80E-17	72.7
Scaff9762:5746-6285	Scaff9762g056631	490	669	1.00E-34	35.6
Scaff9762:4956-5060	Scaff9762g056631	824	858	1.00E-34	28.6
Scaff9762:6580-6729	Scaff9762g056631	417	466	1.00E-34	32
Scaff9762:6877-6978	Scaff9762g056631	354	387	1.00E-34	38.2
Query Sequence: <i>C. elegans vit-6</i>					
Scaff10759:60-236	Scaff10759g059768	109	167	1.10E-07	33.9
Scaff1399:11370-11786	Scaff1399g015377	519	657	1.20E-61	37.4
Scaff1399:13223-13459	Scaff1399g015377	109	187	1.20E-61	29.1
Scaff1399:11904-12089	Scaff1399g015377	436	497	1.20E-61	35.5
Scaff1399:8145-8282	Scaff1399g015377	1444	1489	1.20E-61	47.8
Scaff1399:10478-10609	Scaff1399g015377	853	896	1.20E-61	29.5
Scaff1399:8511-8549	Scaff1399g015377	1362	1374	1.20E-61	84.6
Scaff1399:13055-13198	Scaff1399g015377	180	227	1.20E-61	31.2
Scaff1399:8751-8849	Scaff1399g015377	1277	1309	1.20E-61	30.3
Scaff1515:10498-10626	Scaff1515g016285	1447	1489	6.10E-63	48.8
Scaff1515:14457-14933	Scaff1515g016286	523	681	6.10E-63	35.8
Scaff1515:16676-16846	Scaff1515g016286	113	169	6.10E-63	42.1
Scaff1515:15195-15389	Scaff1515g016286	442	506	6.10E-63	32.3
Scaff1515:10857-10922	Scaff1515g016286	1364	1385	6.10E-63	68.2
Scaff1515:16556-16639	Scaff1515g016286	180	207	6.10E-63	42.9
Scaff15289:3803-4177	Scaff15289g071985	519	643	8.30E-48	42.4
Scaff15289:1885-2202	Scaff15289g071985	109	214	8.30E-48	25.5
Scaff15289:3508-3687	Scaff15289g071985	438	497	8.30E-48	36.7
Scaff15289:5114-5245	Scaff15289g071986	853	896	8.30E-48	31.8
Scaff18769:2078-2215	Scaff18769g079434	1444	1489	1.50E-09	47.8
Scaff18769:2444-2482	Scaff18769g079434	1362	1374	1.50E-09	84.6
Scaff18769:2687-2785	Scaff18769g079434	1277	1309	1.50E-09	30.3
Scaff21123:2354-2491	Scaff21123g083914	1444	1489	8.30E-12	47.8
Scaff21123:2087-2125	Scaff21123g083914	1362	1374	8.30E-12	84.6
Scaff21123:1787-1885	Scaff21123g083914	1277	1309	8.30E-12	30.3
Scaff21123:2772-2936	Scaff21123g083915	1504	1558	8.30E-12	27.3
Scaff21949:616-852	Scaff21949g085322	109	187	1.70E-11	29.1
Scaff21949:535-591	Scaff21949g085322	180	198	1.70E-11	57.9
Scaff2226:10425-10730	Scaff2226g021515	523	624	4.20E-54	43.1
Scaff2226:8738-8974	Scaff2226g021515	109	187	4.20E-54	26.6
Scaff2226:10114-10293	Scaff2226g021515	438	497	4.20E-54	36.7
Scaff2226:13886-14023	Scaff2226g021515	1444	1489	4.20E-54	43.5

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff2226:13634-13675	Scaff2226g021515	1362	1375	4.20E-54	78.6
Scaff2226:11585-11716	Scaff2226g021515	853	896	4.20E-54	31.8
Scaff2226:8999-9055	Scaff2226g021515	180	198	4.20E-54	57.9
Scaff2226:13335-13433	Scaff2226g021515	1277	1309	4.20E-54	30.3
Scaff242:10011-10487	Scaff242g003894	523	681	8.90E-63	36.5
Scaff242:8098-8268	Scaff242g003894	113	169	8.90E-63	42.1
Scaff242:9555-9749	Scaff242g003894	442	506	8.90E-63	32.3
Scaff242:13667-13732	Scaff242g003894	1364	1385	8.90E-63	68.2
Scaff242:8305-8388	Scaff242g003894	180	207	8.90E-63	42.9
Scaff242:13983-14090	Scaff242g003895	1454	1489	8.90E-63	58.3
Scaff25080:1297-1713	Scaff25080g090352	519	657	2.60E-38	38.1
Scaff25080:1831-2016	Scaff25080g090352	436	497	2.60E-38	35.5
Scaff25080:410-541	Scaff25080g090352	853	896	2.60E-38	29.5
Scaff30898:1616-1753	Scaff30898g098056	1444	1489	1.90E-08	47.8
Scaff349:14236-14652	Scaff349g005192	519	657	8.10E-60	40.3
Scaff349:14768-14947	Scaff349g005192	438	497	8.10E-60	36.7
Scaff349:11035-11172	Scaff349g005192	1444	1489	8.10E-60	41.3
Scaff349:13344-13475	Scaff349g005192	853	896	8.10E-60	31.8
Scaff349:11389-11430	Scaff349g005192	1362	1375	8.10E-60	78.6
Scaff349:16007-16063	Scaff349g005192	180	198	8.10E-60	57.9
Scaff349:11633-11725	Scaff349g005192	1279	1309	8.10E-60	29
Scaff349:16121-16327	Scaff349g005192, Scaff349g005193	108	176	8.10E-60	29
Scaff3500:3685-4101	Scaff3500g029487	519	657	1.10E-62	38.1
Scaff3500:5541-5777	Scaff3500g029487	109	187	1.10E-62	29.1
Scaff3500:4219-4398	Scaff3500g029487	438	497	1.10E-62	36.7
Scaff3500:10-132	Scaff3500g029487	1444	1484	1.10E-62	51.2
Scaff3500:2327-2458	Scaff3500g029487	853	896	1.10E-62	29.5
Scaff3500:364-402	Scaff3500g029487	1362	1374	1.10E-62	84.6
Scaff3500:5460-5516	Scaff3500g029487	180	198	1.10E-62	57.9
Scaff3500:602-700	Scaff3500g029487	1277	1309	1.10E-62	30.3
Scaff3804:9681-10097	Scaff3804g031270	519	657	1.90E-63	38.1
Scaff3804:8010-8246	Scaff3804g031270	109	187	1.90E-63	29.1
Scaff3804:9382-9567	Scaff3804g031270	436	497	1.90E-63	35.5
Scaff3804:13224-13361	Scaff3804g031270	1444	1489	1.90E-63	47.8
Scaff3804:10853-10984	Scaff3804g031270	853	896	1.90E-63	29.5
Scaff3804:12960-12998	Scaff3804g031270	1362	1374	1.90E-63	84.6
Scaff3804:8271-8327	Scaff3804g031270	180	198	1.90E-63	57.9
Scaff3804:12657-12755	Scaff3804g031270	1277	1309	1.90E-63	30.3
Scaff4688:11790-11927	Scaff4688g035870	1444	1489	1.40E-10	47.8
Scaff4688:12147-12197	Scaff4688g035870	1362	1378	1.40E-10	70.6
Scaff4688:12399-12497	Scaff4688g035870	1277	1309	1.40E-10	30.3
Scaff4688:11413-11577	Scaff4688g035870	1504	1558	1.40E-10	27.3

Genomic Location	Overlapping Gene(s)	Query start	Query end	E-value	%ID
Scaff6405:5988-6404	Scaff6405g043787	519	657	2.00E-62	39.6
Scaff6405:4317-4553	Scaff6405g043787	109	187	2.00E-62	26.6
Scaff6405:5689-5868	Scaff6405g043787	438	497	2.00E-62	36.7
Scaff6405:9465-9602	Scaff6405g043787	1444	1489	2.00E-62	41.3
Scaff6405:9207-9248	Scaff6405g043787	1362	1375	2.00E-62	78.6
Scaff6405:7153-7284	Scaff6405g043787	853	896	2.00E-62	31.8
Scaff6405:4578-4634	Scaff6405g043787	180	198	2.00E-62	57.9
Scaff6405:8906-8998	Scaff6405g043787	1279	1309	2.00E-62	29
Scaff6533:3183-3290	Scaff6533g044375	1454	1489	2.30E-65	58.3
Scaff6533:3543-3608	Scaff6533g044376	1364	1385	2.30E-65	68.2
Scaff6533:7034-7510	Scaff6533g044379	523	681	2.30E-65	36.5
Scaff6533:7807-7986	Scaff6533g044379	436	495	2.30E-65	35
Scaff6533:6218-6331	Scaff6533g044379	853	890	2.30E-65	21.1
Scaff6533:9265-9423	Scaff6533g044380	113	165	2.30E-65	43.4
Scaff6533:9133-9216	Scaff6533g044380	180	207	2.30E-65	42.9
Scaff6982:2625-3041	Scaff6982g046249	519	657	2.40E-50	38.1
Scaff6982:3157-3342	Scaff6982g046249	436	497	2.40E-50	35.5
Scaff6982:573-704	Scaff6982g046249	853	896	2.40E-50	29.5
Scaff6982:1466-1882	Scaff6982g046249	519	657	7.40E-37	38.1
Scaff6982:2000-2185	Scaff6982g046249	436	497	7.40E-37	33.9
Scaff6982:4478-4714	Scaff6982g046249, Scaff6982g046250	109	187	2.40E-50	29.1
Scaff6982:4310-4453	Scaff6982g046249, Scaff6982g046250	180	227	2.40E-50	31.2
Scaff7843:257-394	Scaff7843g049685	1444	1489	6.70E-60	47.8
Scaff7843:623-661	Scaff7843g049685	1362	1374	6.70E-60	84.6
Scaff7843:866-964	Scaff7843g049685	1277	1309	6.70E-60	30.3
Scaff7843:3772-3984	Scaff7843g049686	587	657	6.70E-60	35.2
Scaff7843:5624-5860	Scaff7843g049686	109	187	6.70E-60	29.1
Scaff7843:3997-4185	Scaff7843g049686	519	581	6.70E-60	44.4
Scaff7843:4303-4488	Scaff7843g049686	436	497	6.70E-60	35.5
Scaff7843:2879-3010	Scaff7843g049686	853	896	6.70E-60	29.5
Scaff7843:5543-5599	Scaff7843g049686	180	198	6.70E-60	57.9
Scaff8571:5626-5763	Scaff8571g052471	1444	1489	1.70E-10	41.3
Scaff8571:8475-8606	Scaff8571g052472	853	896	1.70E-10	31.8
Scaff8571:6481-6522	Scaff8571g052472	1362	1375	1.70E-10	78.6
Scaff8571:6725-6817	Scaff8571g052472	1279	1309	1.70E-10	29
Scaff9762:1871-1999	Scaff9762g056627	1447	1489	5.10E-51	46.5
Scaff9762:2227-2292	Scaff9762g056628	1364	1385	5.10E-51	68.2
Scaff9762:5797-6273	Scaff9762g056631	523	681	5.10E-51	35.8
Scaff9762:6535-6729	Scaff9762g056631	442	506	5.10E-51	32.3
Scaff9762:4956-5069	Scaff9762g056631	853	890	5.10E-51	21.1
Scaff4885:2106-2408		1494	1594	7.00E-03	26.7
Scaff30898:1973-2026		1361	1378	1.90E-08	66.7

Table S3.1.9: BLAST result. Query sequence: *Caenorhabditis elegans* vit 1-6 gene sequences. Subject: *Globodera ellingtonae* genome.

Query	Subject	Query Start	Query End	E-value	% ID
<i>C. elegans vit-1</i>					
	MEIZ01000007.1	501	862	2.09E-25	25.32
	MEIZ01000007.1	96	230	1.12E-07	24.46
	MEIZ01000007.1	421	476	0.004	33.93
	MEIZ01000016.1	425	628	1.81E-20	26.27
	MEIZ01000016.1	1535	1613	0.003	36.96
	MEIZ01000060.1	493	633	3.45E-16	31.92
	MEIZ01000060.1	176	230	0.027	33.33
	MEIZ01000060.1	1535	1613	0.045	31.37
	MEIZ01000069.1	384	463	8.77E-08	37.35
	MEIZ01000069.1	1533	1603	5.98E-05	34.34
	MEIZ01000069.1	180	224	0.004	39.13
<i>C. elegans vit-2</i>					
	MEIZ01000007.1	499	863	2.11E-26	25.18
	MEIZ01000007.1	96	230	3.17E-06	24.46
	MEIZ01000016.1	423	626	1.77E-19	25.85
	MEIZ01000060.1	491	631	1.71E-16	31.92
	MEIZ01000060.1	176	230	0.037	33.33
	MEIZ01000069.1	1531	1600	0.002	31.31
	MEIZ01000069.1	180	224	0.004	39.13
<i>C. elegans vit-3</i>					
	MEIZ01000007.1	502	860	2.53E-25	26.13
	MEIZ01000016.1	655	944	7.48E-10	22.95
	MEIZ01000016.1	1055	1239	8.37E-05	21.94
	MEIZ01000060.1	491	625	2.17E-16	30.37
	MEIZ01000060.1	1057	1239	8.23E-07	20.30
	MEIZ01000069.1	392	460	4.42E-05	38.89
	MEIZ01000069.1	1436	1590	6.36E-04	23.74
<i>C. elegans vit-4</i>					
	MEIZ01000007.1	502	860	1.57E-25	26.13
	MEIZ01000016.1	655	944	6.88E-10	22.95
	MEIZ01000016.1	1055	1239	2.84E-05	22.45
	MEIZ01000060.1	491	625	1.99E-16	30.37
	MEIZ01000060.1	1057	1239	3.16E-07	20.79
	MEIZ01000069.1	392	460	4.13E-05	38.89
	MEIZ01000069.1	177	224	0.01	42.86
<i>C. elegans vit-5</i>					
	MEIZ01000007.1	502	860	2.31E-25	26.58
	MEIZ01000016.1	655	944	7.48E-10	23.51
	MEIZ01000016.1	1055	1239	1.76E-05	22.45
	MEIZ01000060.1	491	625	1.96E-16	31.11
	MEIZ01000060.1	1057	1248	9.77E-08	21.18
	MEIZ01000069.1	392	460	4.61E-05	38.89
	MEIZ01000069.1	177	224	0.006	42.86
<i>C. elegans vit-6</i>					
	MEIZ01000007.1	534	924	2.81E-26	24.31
	MEIZ01000007.1	119	220	1.75E-04	24.03
	MEIZ01000007.1	1415	1489	6.20E-04	32.90
	MEIZ01000007.1	436	495	6.80E-04	33.33
	MEIZ01000016.1	446	656	2.57E-28	31.82
	MEIZ01000016.1	39	158	1.69E-05	23.33
	MEIZ01000016.1	1415	1489	3.03E-05	36.36
	MEIZ01000060.1	23	218	1.12E-04	24.02
	MEIZ01000060.1	446	495	1.61E-04	44.00
	MEIZ01000060.1	1415	1489	2.15E-04	34.21
	MEIZ01000060.1	119	159	0.001	43.90

Table S3.1.10: BLAST result. Query sequence: *Caenorhabditis elegans* vit 1-6 gene sequences. Subject: *Pratylenchus scribneri* genome.

Query	Subject	Query Start	Query End	E-value	% ID
<i>C. elegans vit-1</i>					
	JAQSEB010000001.1	219	633	1.22E-39	27.686
	JAQSEB010000001.1	1053	1245	3.56E-05	23.116
	JAQSEB010000001.1	1533	1614	0.001	28.713
	JAQSEB010000032.1	400	633	3.72E-29	30.566
	JAQSEB010000032.1	1051	1245	1.34E-04	24.257
	JAQSEB010000032.1	96	166	4.19E-04	32.895
	JAQSEB010000032.1	1533	1614	0.001	28.431
	JAQSEB010000147.1	493	576	2.02E-16	50
	JAQSEB010000147.1	91	200	5.67E-06	28.696
	JAQSEB010000147.1	1533	1613	1.44E-05	32.039
	JAQSEB010000017.1	493	576	3.82E-16	48.81
	JAQSEB010000017.1	96	173	2.82E-05	33.333
	JAQSEB010000017.1	1533	1613	2.62E-04	30.275
	JAQSEB010000019.1	497	576	4.05E-16	51.25
	JAQSEB010000019.1	493	576	1.47E-15	48.81
	JAQSEB010000019.1	493	576	6.04E-15	48.81
	JAQSEB010000019.1	1254	1362	2.88E-06	27.143
	JAQSEB010000019.1	91	170	4.14E-05	32.941
	JAQSEB010000019.1	96	172	7.77E-05	33.735
	JAQSEB010000019.1	1533	1613	1.27E-04	29.907
	JAQSEB010000111.1	493	576	6.70E-16	50
	JAQSEB010000111.1	1254	1458	3.24E-10	24.723
	JAQSEB010000111.1	82	147	2.54E-06	38.571
	JAQSEB010000064.1	493	576	8.83E-16	48.81
	JAQSEB010000064.1	1322	1458	1.47E-06	29.213
	JAQSEB010000064.1	93	147	3.48E-04	40
	JAQSEB010000053.1	497	576	1.32E-14	48.75
	JAQSEB010000053.1	1323	1458	4.44E-06	28.814
	JAQSEB010000053.1	91	170	3.74E-05	32.941
	JAQSEB010000053.1	1533	1613	5.18E-05	31.776
	JAQSEB010000053.1	1533	1613	6.50E-05	32.075
	JAQSEB010000130.1	96	147	0.01	36.538
<i>C. elegans vit-2</i>					
	JAQSEB010000001.1	219	631	1.62E-37	26.305
	JAQSEB010000001.1	1051	1243	3.82E-06	23.618
	JAQSEB010000001.1	96	147	2.10E-05	42.308
	JAQSEB010000032.1	421	631	9.58E-28	30.833
	JAQSEB010000032.1	96	147	1.40E-05	42.308
	JAQSEB010000032.1	1049	1243	1.60E-05	24.752
	JAQSEB010000032.1	96	166	9.72E-05	34.211
	JAQSEB010000147.1	491	574	3.41E-16	50
	JAQSEB010000147.1	91	200	1.68E-06	29.565
	JAQSEB010000147.1	1531	1610	0.003	28.155
	JAQSEB010000017.1	491	574	5.42E-16	48.81
	JAQSEB010000017.1	96	171	8.54E-06	35.526
	JAQSEB010000017.1	1531	1610	0.004	27.523
	JAQSEB010000019.1	495	574	6.14E-16	51.25
	JAQSEB010000019.1	491	574	2.26E-15	48.81
	JAQSEB010000019.1	491	574	7.88E-15	48.81
	JAQSEB010000019.1	1252	1360	5.03E-06	26.241
	JAQSEB010000019.1	91	170	1.25E-05	34.118
	JAQSEB010000019.1	96	172	2.30E-05	34.94
	JAQSEB010000019.1	1531	1610	0.002	27.778
	JAQSEB010000019.1	1531	1610	0.007	26.852
	JAQSEB010000111.1	491	574	7.33E-16	50
	JAQSEB010000111.1	1252	1456	6.48E-12	24.908
	JAQSEB010000111.1	82	147	6.19E-07	40
	JAQSEB010000111.1	1531	1610	0.003	28.037
	JAQSEB010000064.1	491	574	1.61E-15	48.81
	JAQSEB010000064.1	1320	1456	1.49E-06	28.652
	JAQSEB010000064.1	93	147	8.64E-05	41.818
	JAQSEB010000064.1	1531	1610	0.002	28.037
	JAQSEB010000053.1	495	574	1.96E-14	48.75
	JAQSEB010000053.1	1321	1456	3.45E-06	28.814
	JAQSEB010000053.1	91	170	1.00E-05	34.118
	JAQSEB010000053.1	1531	1610	0.005	29.245

Query	Subject	Query Start	Query End	E-value	% ID
	JAQSEB010000053.1	1531	1610	0.007	29.907
	JAQSEB010000130.1	96	147	0.006	38.462
C. elegans vit-3					
	JAQSEB010000001.1	240	632	9.61E-36	25.714
	JAQSEB010000001.1	826	993	6.76E-06	22.959
	JAQSEB010000001.1	1521	1601	0.014	29.703
	JAQSEB010000001.1	1319	1349	0.034	60.606
	JAQSEB010000032.1	417	630	5.82E-26	29.508
	JAQSEB010000032.1	1521	1601	0.003	29.412
	JAQSEB010000147.1	490	573	2.09E-14	41.667
	JAQSEB010000147.1	1038	1240	3.52E-08	22.43
	JAQSEB010000147.1	1521	1600	0.003	30.392
	JAQSEB010000017.1	490	573	2.57E-14	41.667
	JAQSEB010000017.1	1057	1240	9.95E-08	23.56
	JAQSEB010000017.1	1520	1600	0.017	28.44
	JAQSEB010000019.1	494	573	1.54E-14	43.75
	JAQSEB010000019.1	490	573	8.76E-14	40.476
	JAQSEB010000019.1	490	573	1.88E-13	40.476
	JAQSEB010000019.1	1057	1240	5.67E-08	23.56
	JAQSEB010000019.1	1034	1240	2.57E-07	22.374
	JAQSEB010000019.1	1034	1240	3.01E-07	21.918
	JAQSEB010000019.1	1250	1354	2.48E-05	27.206
	JAQSEB010000019.1	1521	1600	0.006	28.972
	JAQSEB010000019.1	1521	1600	0.006	28.302
	JAQSEB010000019.1	1520	1600	0.014	28.704
	JAQSEB010000111.1	490	573	1.84E-14	41.667
	JAQSEB010000111.1	1048	1240	1.76E-07	21.182
	JAQSEB010000111.1	1250	1354	2.40E-05	27.206
	JAQSEB010000111.1	1521	1600	0.01	29.245
	JAQSEB010000064.1	494	573	2.41E-14	43.75
	JAQSEB010000064.1	1037	1240	1.27E-08	22.581
	JAQSEB010000064.1	1520	1600	0.002	29.907
	JAQSEB010000053.1	494	573	1.21E-13	42.5
	JAQSEB010000053.1	1521	1600	0.006	30.189
	JAQSEB010000053.1	1520	1600	0.049	29.245
C. elegans vit-4					
	JAQSEB010000001.1	240	632	6.64E-36	25.714
	JAQSEB010000001.1	826	993	6.82E-06	22.959
	JAQSEB010000001.1	1319	1349	0.032	60.606
	JAQSEB010000032.1	417	630	4.80E-26	29.508
	JAQSEB010000147.1	490	573	2.29E-14	41.667
	JAQSEB010000147.1	1038	1240	1.42E-08	22.897
	JAQSEB010000017.1	490	573	2.62E-14	41.667
	JAQSEB010000017.1	1057	1240	3.49E-08	24.084
	JAQSEB010000019.1	494	573	1.48E-14	43.75
	JAQSEB010000019.1	490	573	8.61E-14	40.476
	JAQSEB010000019.1	490	573	1.85E-13	40.476
	JAQSEB010000019.1	1057	1240	2.29E-08	24.084
	JAQSEB010000019.1	1034	1240	9.46E-08	22.831
	JAQSEB010000019.1	1034	1240	1.36E-07	22.374
	JAQSEB010000019.1	1250	1354	2.83E-06	27.941
	JAQSEB010000111.1	490	573	1.97E-14	41.667
	JAQSEB010000111.1	1048	1240	6.71E-08	21.675
	JAQSEB010000111.1	1250	1448	1.30E-06	23.134
	JAQSEB010000064.1	494	573	2.66E-14	43.75
	JAQSEB010000064.1	1037	1240	5.70E-09	23.041
	JAQSEB010000053.1	494	573	1.18E-13	42.5
C. elegans vit-5					
	JAQSEB010000001.1	240	632	2.99E-36	26.593
	JAQSEB010000001.1	826	993	6.99E-06	22.959
	JAQSEB010000001.1	1319	1349	0.032	60.606
	JAQSEB010000032.1	417	630	5.22E-26	29.918
	JAQSEB010000147.1	490	573	1.07E-14	42.857
	JAQSEB010000147.1	1038	1240	2.83E-08	22.43
	JAQSEB010000017.1	490	573	1.21E-14	42.857
	JAQSEB010000017.1	1057	1240	7.42E-08	23.56
	JAQSEB010000019.1	494	573	7.31E-15	45
	JAQSEB010000019.1	490	573	4.26E-14	41.667
	JAQSEB010000019.1	490	573	8.61E-14	41.667
	JAQSEB010000019.1	1057	1240	4.37E-08	23.56

Query	Subject	Query Start	Query End	E-value	% ID
	JAQSEB010000019.1	1034	1240	1.93E-07	22.374
	JAQSEB010000019.1	1034	1240	2.44E-07	21.918
	JAQSEB010000019.1	1250	1354	2.25E-06	27.941
	JAQSEB010000111.1	490	573	9.80E-15	42.857
	JAQSEB010000111.1	1048	1240	1.20E-07	21.182
	JAQSEB010000111.1	1250	1448	1.43E-06	23.134
	JAQSEB010000064.1	494	573	1.24E-14	45
	JAQSEB010000064.1	1037	1240	9.74E-09	22.581
	JAQSEB010000053.1	494	573	5.42E-14	43.75
C. elegans vit-6					
	JAQSEB010000001.1	246	657	1.31E-44	28.151
	JAQSEB010000001.1	114	168	1.98E-07	49.091
	JAQSEB010000001.1	1454	1496	6.13E-05	44.186
	JAQSEB010000032.1	442	657	9.71E-34	33.195
	JAQSEB010000032.1	114	168	3.94E-07	49.091
	JAQSEB010000032.1	1454	1498	1.17E-04	42.222
	JAQSEB010000032.1	114	176	4.55E-04	33.846
	JAQSEB010000147.1	523	601	3.58E-12	43.75
	JAQSEB010000147.1	436	495	4.48E-06	39.344
	JAQSEB010000147.1	118	168	6.15E-04	39.216
	JAQSEB010000147.1	1454	1489	0.005	44.444
	JAQSEB010000017.1	519	597	2.12E-11	40.506
	JAQSEB010000017.1	436	495	9.86E-06	39.344
	JAQSEB010000017.1	118	168	1.66E-04	39.216
	JAQSEB010000017.1	1454	1489	0.005	44.444
	JAQSEB010000019.1	523	601	4.31E-12	41.25
	JAQSEB010000019.1	523	601	2.45E-11	41.25
	JAQSEB010000019.1	523	601	4.08E-11	41.25
	JAQSEB010000019.1	410	495	4.86E-07	34.783
	JAQSEB010000019.1	1273	1385	9.66E-07	29.927
	JAQSEB010000019.1	437	495	7.05E-06	40
	JAQSEB010000019.1	1362	1385	4.41E-05	75
	JAQSEB010000019.1	1249	1354	4.41E-05	22.523
	JAQSEB010000019.1	118	168	3.31E-04	39.216
	JAQSEB010000019.1	114	168	3.78E-04	36.364
	JAQSEB010000019.1	1454	1489	0.002	44.444
	JAQSEB010000019.1	1454	1489	0.003	44.444
	JAQSEB010000111.1	1269	1489	3.89E-12	26.182
	JAQSEB010000111.1	523	601	2.07E-11	42.5
	JAQSEB010000111.1	436	495	4.56E-06	39.344
	JAQSEB010000111.1	109	168	3.02E-04	33.333
	JAQSEB010000064.1	523	604	1.61E-12	43.373
	JAQSEB010000064.1	1362	1489	2.48E-07	31.25
	JAQSEB010000064.1	436	495	4.56E-06	39.344
	JAQSEB010000064.1	114	168	2.37E-04	36.364
	JAQSEB010000053.1	523	619	1.70E-11	33.333
	JAQSEB010000053.1	435	495	1.34E-06	40.323
	JAQSEB010000053.1	1364	1489	1.30E-04	27.215
	JAQSEB010000053.1	118	168	3.33E-04	39.216
	JAQSEB010000130.1	118	168	0.003	37.255

Supplementary File 3.2: Putative vitellogenin gene sequences included in protein domain alignment and phylogenetic tree or in genome comparison.

>*C. briggsae* 002

MRSIIIASLVALLASSPAFERSFEPETDYHYKLDGLVLSGLPTESSEQSQTRISARARIQSDNDRIYIH
QLVNIRLAASHLPESQMPALNSMEPRELSEYTHMLERPIRAQLRNGLISEIQFEKEDAEWSKNMKR
AVLNMISFNPVAPRDEMEEMESYRKESSSFEENTSFFTNEKTLEGDCQVAYTVVREQKTIITKSVNF
DKCTERSEIAYGLRFSSECPECEKDEMIQPQTVYTYVMENNELKESEVRSRYTVNVNGQEVMTKET
RSKLVLEESHSHKSHIKKANGEKEDIYSSRWEQLVEDFYKNGDKAEFSPFEKFPLDKKIQLIKTITEQVQ
EVENNMPETAHFLARLVIRFTTSISQLKEIHEKIYAKADRKIQSMIESALAISGKTNTIQHLLNHIENEDIT
PVEAAKLIKSIQETPFPSQSIAESLLKFAQSRVAKTNQIIRQSVWLATGSVVRGVDIRSLRPLIREDKRE
MKDKFFRTFMKQFQDAETTYEKVLALKTIGNAGLDISVNELNDIIVDKRQPLPVRKEAIDALRLLKDSM
PRKIQKVLLPIYKNRQYEPEIRILALWRMMHTRPEEPLLQVVSQMEKESNQQVAALTYQMIRQFSQS
TNPCYQRVATECAKVLSTTRYQPQEQMIASAYALPLFMQNAFSGAQFDFAAIFEKNSVLPKDLHASL
DAVFGGNWNKYFAQVGFSQENMDQYIYKTLQKLESLEKESTTVVRGRRIQTGIRMLKELAQKMNIRA
RPSTRNEKNAYAMIYLRKDMDFAVLPIDSQLVEDVVERFVSSGKIEFSAIRRVLNKDSEFEAHHAFF
YEA VRKFP TTLGLPLTISGKIPTVMSAEGQISLEMDGTELRLTVEARPSVAATHVYEMRMFTPLFEQGI
KTLQSIRAYTPLKIQAVAGLKKNFELVYKIVPESQKSIVSLSTRPVVFLRHPGFSKFYEIEAEERTVAVP
QWEQKTQEVERTENYLGLEFSTRGNILRQHSVENWLLAEQDFEVTVENKNRPAEFVFRMTASPLEKT
PLSEIKISNEIFEKEFELEQERSENRRRESFSKMIKNIQKEQGYKSLITLKFEAPRDYQMRSELTTVCDKQ
VRMCQWTVENVRSPIFAEEKDWTLSQLLAIRPEMPSSLRQLREQPHREVQLSLTSTWGSQKNE
LNINAQLQQSKEQKKYERNMEREVRGMPEYELLIKAARLNQVMVADYKFTKENEETMSRLFDLVKA
YNYWTVSSRPEDNEENRVVLQLTVEPISRQYVNVTVQTPEGRVEMKNTQIPRVHLPSIAQRATKYQY
TEASGAVCKVQKNQIRTFDDVLYKTPLTTCYSLIAKDCSEPTFAVLSSKKVEKNSEEMMIKVLARGEQEI
VAQLQNEEIRVKVDGKKIQSEDFVDYQIERLGESAIVIELPEGEVRFDGTYTIKTQLPSYSRQQQLCGLC
GNNDDESTNEFMTADNTETEDMEEFHRSYLLKNEECETEEERLTEKKNYRKYERDEEESDESDES
DNDEEPQYTKSNKKSEKKNQIVEKTQIKFESHRIKCSVEPVAECRRGYEAEQSQQRKIRFTCLPRHN
RDASRLLKESRQQPLELNDYPVSFVEAVKVPTACVAF

>*C. briggsae* 006

MKFFIALALLGAALASSHFDRSLERNIQESSFRAGREYRYLFNGQLSAGLPIPSTPQGSRQLQTQVNLQ
WTDGNTVRMQLEKTRFATSQQETNSQKMLPFERFEEVDREHMQQLSYPVEFDYEHGLVHNIRF
AEDDQPWSENIRKRAVINMLQVNILKKEKHDAEKSDNQEQLFAFTNVERTLEGECEVLYTVEENQKEE
NQRWAKSINFDKCTRRPYIHHVQVPVCKDCEQTLDDQKMSSTVLNYNISGTTSSFLINSVELRSQHLLF
APISEKHQLVSAFTLNTLELIYAGEKKSEIKVRSEKTSSELYNQEWAEQQWAQTGEEKYLRQMPQ
WSENKVMVQKMFSLMAKQIEQGEAELEAAHTVARIVKVLQRCNEEQLEQIYRHVAEHKDEKIAEQL
RSIYFNTLALAGTRVTIQGFVDKVQSRKNITPLKASVAIKTLVDMRYPSLAIAEDVARLCESDVSSSFPAL
RQSCWLTYGAVNGVCGQTPRVFVQKNGVKMCPRDSKQRIVDKLVQQFESASSRYEKVLALKTLAN
AGLDLCVYPLEKILNEQHETTVAQAIESLRRLRHQMPVKIQRVLLPIYLNRRQQPQHIRMSSALHQLIYT
QPEWSVLSQIGNQLRQERNQQVRAFTLSLLRSYANNESPCEQSFSSRVQSLLNNMPFSSQEIDRFES
VYGKYSTYSRRHQSGFEANFASLFTSESVLPTEMMALEGLVSGEWNQYFAQVGFTQNNMEKIIKK
LLTSVQEKGLEQIVVRGKRATGSFKPTEFLNSLLEKLRTTRRQPSQDPHAVVYIRHRDMDYAFLPIDA
ESIPEVVRSMIQGGRLEIGDIERVLAQGVHFSASNAAFLYETVRRIPSPMGLPVQFTSKMPTISSIRGKI
TFELEPKNGKSFDDLRLRLQAEPRVASTHVLSTVICPIAEVGTFLHQAVALNTPIDTEVKMNWEDKIVI
RSLYTAPSEEKRIALIQSRPVTFVRTVVPPEARQYPEPEMPTYMLPENQQLSQSLDHEYKQFRVQGTNLN
RPTSGRVPQWIVDNTVELFYKPHIEKYEAVLELDMYNNYKMEKNYEKEYKKHTGKRYLESEPEYDEE
EHQEIQTKKFEWLQNEKVYQHAKFEVKPSVLKVDVEAVCNKDFHFKCTQLRGEDILATVQYVYPQT
PRTVEELKEQKYRQLVVLGEMTYGENTIQUININGQQSQEQKKFVKQIEQAPHEHTLLEASRLDQYQAV
VEYELEPKSAQYVARYWNLVQAYFRTQYPWTSRIKSREESSRKNVIRATFNVEPRQRQTVNMTIETPA
EKTILERAELPFRLPTAQIHYQPRNSRFEQKPMLEKIVRHTTKQANCVVKSTKIQTDFEVAYKNQFTPC
YSVLAKDCGSEKNEPRFVVLMMKMNKKEWKNVKKVYGENEIEMYKTEEGLVCRVNGEEIEYQPES
EIEKKQYSLIWLKNTVKFESDDVTVQFDGVNARIHLSALYRNQQCGLCGHYDNEKETEFYDAENQE
NTIPKFAKSYLYKDSKCNERNVFEKEENFRRIERDEEEQEQEMNYEESRREEQDDEPTEKIVAIERQ
HEICFTQKPVLRQCNGKSKKQKTDVYCLPSSNSWARRQVREIRREPLTKWSEERLENLRDQPK
MEERTVRVPVDQKCEKFDY

>*C. briggsae* 14203

MKSIIIASLVALAIAASPALDRPFSPKSEYIYQFDGLLLSGLPTTSSDASQTRISCRALIQAIIDRYIHLQLT
DVTYSASHIPQTEKMPKMNSLEQRELSDEHKEMLELPFRAQLRNLISEIQFSTEDA EW SMNSKRVII
NLLSLRTSAPVDEMIQEEKYMETEKDSRFFTVNEKTMEGDCEVAYTIVEEGEKTIIYTKSVNFDKCITRP
EMAYGLRFGSECQQCEKEGQFVKPQTVYTYTFKNQQLEQSEVHSVYTLNVNGQEVVKSETRSKIVF
VEENKINKEIKKVTGPKESLVYSAGNEWLIEQFYQQGDKVEVNPFKVISPEQKFEELREIFRQVQELEE
NTPETVHMLARAVRIVRMLSIEELKKVHTKIYTESENKVQRLIESACAIAGTKNTIQHLIHHFHKMTVSP
RASELLKSIQETLYPSEQIADLLIELAKSPLAEKNEPLRQSAWLAAGSIVRGFSAKTQDPLTRPASRQL
KEYYVRVFMQLFRSAESTYEKVLALKTLGNAGIDLSVNELVQLVQDPRQPLAIRTEAVDALRLLKDIMP
RKIQKILLPVYKNRQNKPELRMAALFRMMHTLPEEPILAHIVAQMEKESNQHVAAFTYNVIRQFARSTN
PCYQPLAVRCSKVLLFTRYQPQEQLLSTYSQLPIFHSEMLSGVQFDFATIFEKNAFLPKEVQASFESVF
GGNWNKYLQVVGFSQQNFEQVIVKALERLSRYGKQSNELRSRRVQSGIQMLQEIVKKMNIRPRVHRT
DAQSAHAVFYLRKEMDYIVPLDMETLDNLEKYVRNGEFDIKSVLSFLNADSKFELHRAHYIHESVR
RIPTSIGMPLTISGKMPTVLSFTGKVTVMQKLGARLILDVPSVATTHVTEMRFWSPRFEQGVKSLQS
ARVHTPLRLESTVELKKNTLEITHKFAVPENKKTIVSVHSRPVAFLRIPKNDETQYTESEEKTISHSEYQ
WTTEEIDRQYEVGLKVNAGGNILSQWSLPKVLMTEQDFEYTLNKNRPAEFTARMTIGNLEKTDLSE
IKFDKIFEKEFDSENTESESRRQYFHKMIREIKAEQGYMNLITLLEAPQQTYFNSELRTTCDQWVRM
CRFEMDSRRSPISEETKEWTLRTELVAARPQMPSSLRQLREQPHREVQLALNAKWGSSKKSEMTVT
AQLEQSKEQKKYIRNIERVFNGLPEYELLIKAARLNQINVVSEYKLTPEAEHTFSRLFDLVKAYNFWTVS
EKRVENEERRVTLQLTVEPMSRQYVNMTIQTPQQEVEIKNIRVPRVVLPTIARRAMIQQAWEKTETTC
KVDQSEVSTFDDVIYRAPLTTCFSLIAKDCSEPTFAVLAKRMNKNSEDELLVKVIGREQEVIVKKTSEEF
LVKVDGKNIPQSELEQYIEILGDNLVVIRLPQGEVRFDGYTIKTNLPSVASKSQLCGLCGNNDGERDN
EFMTADNYETEDVEEFHRSYLLKNEECEVENDRLSEKKNYRNKWDREEKKRDQSSSESYYYEEES
EKSDKKSVEKTQIKEFSHRVCFSLPVAECRRGYKSGKTSTKKIRFTCMPRHSNARRLLKEAREQNI
LDLSDFEVSYSVESVKVPTTCVEAF

>*C. briggsae* 14234

MKSIIIASLVALAIAASPALDRPFSPKSEYVYQFDGLLLSGLPTTFADASQTRISCRVRLQAVDDRYIHLQ
LTNIEYSASHVPQNEQWPKMNSLEQRELSDELKELLELPFRAQLRNLISEIQFSTEDA EW SMNSKKV
IINLLSLRTSAPIDEMIQGEKYMETEKDSRFFTVNEKTMEGDCEVAYTIVEEGEKTIIYTKSVNFDKCITRP
EMAYGLHFGSECQQCEKEGQFVKPQTVYTYTFKNQQLEQSEVHSVYTLNVNGQEVVKSETRSKIVF
VEENKINKDIKKATGPKTDIVFSAQDEILMERFCQQGDKAENVNPFVMPSSQQKIEQLEKIFRQMQELEE
NTPETVHLIARAVRLFRMCTIEELKKFHTKVYTESETKVQLLVQSSLAAGTRNTIQHLIHHFHKMTVSP
RAAELLKSIQETISPSEEIADLVIELAKSPLAEKNEPLRQSAWLAAGSIVRGFSTKTQDPLTRPASRQL
KEYYVRVFMQLFRSAESTYEKVLALKTLGNAGIDLSVNELVQLIQDPRQPLAIRTEAVDALRLLKDIMP
KIQKILLPVYKNRQNKPELRMTALFRMMHTLPEEPILAHIVAQMEKESNQHVAAFTYNVIRQFARSTNP
CYQPLAVRCSKVLLFTRYQPQEQLLSTYSQLPIFHSEMLSGVQFDFATIFEKNAFLPKEVQASFESVFG
GNWNKYLQVVGFSQQNFEQVIVKALERLSRYGKQSNELRSRRVQSGIQMLQEIVKKMNIRPRVQRTD
AQSAHAVFYLRKEMDYIVLPLDMETLDNLEKYVRNGEFDIKSVLSFLNADSKFELHRAIFYESIRRI
PTTSGVSLTISGKMPTVLSFTGKVAVEMQKLGTRLILDVPSVATSHVTEMRFWNPIDQGVKSLQSAR
VHTTFRLESTVELKKNTLEITHKFAVPENKKTIVSVHSRPVASLRVPRNDQYVEAEKKTISHSEYQWTT
EIDRKYEVGLLRVNAQGNILSQWSLPKVLMTEQDFEYTLNKNRPAEFTARMTIGNLEKTDLSEIKFD
KIFEKEFDSENTESESRRQYFHKMIREIKAEQGYMNLITLLEAPQQTYFNSELRTTCDQWVRMCRFE
MDSRRSPISEETKEWTLRTELVAARPQMPSSLRQLREQPHREVQLALNAKWGSSKKSEMTVTAQLE
QSKEQKKYIRNIERVFNGLPEYELLIKAARLNQINVVSEYKLTPEAEHTFSRLFDLVKAYNFWTVSEKRV
ENEERRVTLQLTVEPMSRQYVNMTIQTPQQEVEIKNIRVPRVVLPTIARRAMIQQAWEKTETTCKVDQ
SEVSTFDDVIYRAPLTTCFSLIAKDCSEPTFAVLAKRMNKNSEDELLVKVIRREQEVIVKKTSEEFV
DGKNIPQSELEQYIEILGDNLVVIRLPQGEVRFDGYTIKTNLPSVASKSQLCGLCGNSDGEKDNEFMT
ADNYETEDVEEFHRSYLLKNEECEVENDRLSEKKNYRNKWDREEKKRDQSSSESYYYEEESEKSE
KKSVEQTQIKEFSHRVCFSLPVAECRRGYKTDDETLTKIRFTCMPRHSNNARRFLKEAREQTVLDLS
DFEVSYSVEAVKIPTACVAF

>*C. briggsae* 16767

MRSIIFASLLALAFATTPAFERLFEPKTDYHYKFDGLVLSGLPISSTEQSQTRISARARIQSNDRDILLQLV
DIQLAASHLSDELMPVMNSLEQRELSNVYKKILELPVRAQLKNGLIFELQFEEEDTEWSENMKRAILN

MISFNPIAPRDEKLESDKEDNTAFFTNEKTLEGNCQVAYSVVREQKTIIVTKSVNFDKCTERS
 ETAYGLRYTSECPECDNQTEVIRPQTVYTYVLENNELKQSEVRSLYTINVNGQEVMTETRSKLVYEE
 SHSIKNIKKVNGEKEDIIYSSRWEQLVEDFYKNGDKSNAQPFVFPLEKKIQLVKSMIDRIQEEQNPV
 ETAHFFARIVRIFRTCSVSQLEKIHDSLYIKADKKIQSLLDNLAVSGTKNTIDHLLTHAEIDEVSPINAKL
 LKTIQETPFPPSEIAETLIKFAQSRPTKNNQVIRQSAWLASGSVVRGIVDIRSVRPFLREDKREIKQKFL
 RTFMKQFENSETTYEKILALKTIGNAGLDISVNLNDIIVDKRQPLPVRKEGIDALRLLKDSMPRKIQKVL
 LPIYKNRQYEPEIRMLALWRMMQTRPEEPLLVQVFSQMEKESNQQAALTYQMIRQFSQSTNPCYTR
 VAKDCSKILSFTRFQPQSAFTSTYSLPLFSQEAQFSGAQFDFAAIFEKNSFLPKDLHATLGTVLGGNW
 NEYFAQVGFSGENMDQYIFKTLQKLESLEKKSTTVVRGRRIQAGIRMLKELAQKMNIRARPSTLNEKD
 AFAMVYLRYKSMDYAVLPVDSQLIEDLVERFVNGKLFHSEIRRLFNQDHFKAHSAFFYEVIRKFTT
 TFGLPLVISGKIPTFMSVEGQFSLDLEGTELRTAEARPSVAATHVYEMRMFTPLFEQGVKTLQSLRA
 YTPIKFQVDADLKKNFDLVYKVVIPENEKSIVTISTRPIIFLQHPGFAKFNNDSEERTVSIPQWQQKTQ
 EIERTENFLGLEVSTSGNLRQHSVENWLVAEQDFEVTFENKNRPAEFVFRMTVSPLEKTPLSEIKINN
 DIFEKEFELEQETSENRRSEFSKMKVNIQKEQGYKSFITLKFEAPRDYQMRSELTTVC DKQVRMCQW
 TVEVNRSPIFAEKKDWTLRSQLLAIRPVMPSLRQLREQPHREVQLSLTSTWGSQKKNELNINTQLQ
 QSKEQKKYERQMEREVRGIPEYELLIKAARLNQVNMVAEYKLTKENDETMSRLFNLVKAYNYWTVSS
 RPEDNEKNRVVLQLTVEPISRQYVNVTVQTPEGRVEMKNTQIPRVHLPSIAQRATKYQYTESGAVCK
 VQKNQIRTFDDVLYKTPLTTCYSLIAKDCSEPTFAILSKKVEKNSEEMMLKVLRGEQEIVAQIQNEEIR
 VKVDGKKIQSEDFADYQIETLGESAIVIELPEGEVRFDDGYTIKTQLPSYSRQQQLCGLCGNNDDESTN
 EFMTADNIETEDMEEFHRSYLLKNEECETEEERLTEKKNYKKYERDEDESDSYENNEEPEYTKKYN
 KKSEMKNQLVEKTQIKEQSHRICFSMEPVAECCRGRGFEPRNLTCKIRFTCMPRHSINARRLMKEVRQE
 PLQLDDYTTSFVESVKVPTACVAYKTSSTGVFV

>*C. elegans vit-1*

MRSIIIASIVALAIAFSPAFAFERTFEPKIDYHYKFDGLVLSGLPTASSELSSQSRFSARVRIQAVDDRIHLQL
 VNIHMAASHLPESEQIPSLNSMEQRELSEEYKQMLKLPLRAQLRNLGLIAELQFDKEDAEWSKNMKRA
 VVNMISFNPIAPRNEIEKIESSYDKEEQSEENTSFFTNEKTLEGDCQVAYTVIREQKTIITKSINFDKCT
 ERSEIAYGLRYSSECPECEKDTVLIRPQTVYTYILENEELKESEVRSLYTVNVNGQEVMTETRSKLV
 EENHSIKSHIEKVNGEKESIIYSSRWEQLVEDFFKNGDKAEFAPFEKFPLDKKMHLIKTITEQIQEVENNI
 PETSHFLARLVIRFTTSTSQLKEIHETLYVKADKKIQLSMEHALAIAGTKNTIQHILVHIENEDIVPLEAA
 QLLKSIQETPFPSQTIAEALIKFAESRVSKNNQVVRQSAWLAAGSVVRGIVDYKNIRPLVREDKRELKE
 KFLRVFMQQYKDAETTYEKILALKSIGNAGLDISVNLNEIIVDKRQLLPVRKEAIDALRLLKDTMPRKIQ
 KVLLPIYKNRQYEPEIRMLALWRMMHTRPEESLLVQVVSQMEKETNQQAALTHQMIRHFAKSTNPC
 YQRAIVCCKSVLSFTRYQPQEQAASSYAQLPLFLQNSFSGAQFDFAAIFEKNSFLKDLHASLDAVFG
 GNWNKYFAQIGFSQQHMDKYVQMALEKLESIEKESTTVVRGRRIQTGITLLKELALKMNIRARPANYN
 EKDAFAMVYLRYKDMDYAILPVDTLQIEKIEKYISNGKVQFSEIRRLNQEHEFETHHAAYFYEAIRKF
 PTTLGLPLIVSGKIPTVFAEGQFSLGLEETELRLTVEARPSVAATHVYEMRMFTPLFEQGVKSVQSVR
 AYTPIKIQAVVGMKRNFEIVYKVVVPENQKSIISLTTRPVVFLRFPFGSKFEYIEAEERTVVVPWQQKT
 QEIEKVFNFLGLEVSTRGNILNQHTLENWLLAEQDFEVSVENKYRPAFTARLTVGQLEKTELSQIKYN
 KIFEKEFELEQENTESRREYFTKMVKSQKEQGYKSVVSLRLEAPRDYTMNTEVTTVC DKQVRMCQW
 EVEIRRSPILEETKEWTLRSQLLVVRPEMPSSLRQLHDQPHREVQLSLTSTWGSQKKSEVTVNAQLQ
 QSKEQKKYERNMDRHFNGMPEYELLIKAARLNQINAVAEYKLTRETEQVLARYFDLVKAYNYWTVSS
 RPENNENDRVVVQLTVEPMSRQYVNITMQSPIERVELKNVQVPRVYLPISIAQRSVKHLLNEASGSVC
 KVQKNQIRTFDDVLYNTPLTTCYSLIAKDCSEPTFAVLSSKTEKNSEEMIIKVIRGEQEIVAQLQNEEIR
 VKVDGKKILSEDYSAHQIERLGESDIVIELPEGEVRFDDGYTIKTQLPSYSRKNQLCGLCGNNDDESTNE
 FYTSDNTETKDIEEFHRSYLLKNEECEAEERLSEKKNYRKYDERKYESEEYSFEETYDYEQENTNKK
 QKNQRSQKKS DLVEKTQIKEFSHRICFSVEPVAECCRGRGYEAVEQQQRKVRFTCLPRHSSEARRLVK
 EARQGTQVQLDDHKISFVHSVQVPVACVAY

>*C. elegans vit-2*

MRSIIIASLVALALASSPAFAFERTFEPKTDYHYKFDGLVLSGLPSASSELSSQSRISARARIQAVDDRYIHLQ
 LVNIRMAASHLPESEQMPSLNSMEQRELSEEYKQMLELPLRAQLRNLGLISELQFDKEDAEWSKNMKR
 AVVNMISFNPIAPRNEIEKIESSYDKEEQSEENTSFFTNEKTLEGDCQVAYTVIREQKTIITKSINFDKC
 TERSEIAYGLRYSSECPECEKDTTELIRPQTVYTYVLENEELKESEVRSLYTVNVNGQEVMTETRSKL
 VLEENHSIKSHIKKVNGEKESIIYSSRWEQLVEDFFKNGDKAEFAPFEKFPLDKKMHLIKTITEQIQEVE

NNMPETSHFLARLVIRFTTSTSQLKEIHETLYVKADKKIQSLMEHALAIAGTKNTIQHILVHMENEDILP
 LGQILKTIQETPFPSQSIAEALIKFAESRVAKNNLVVRQAAWLAAGSVVRGIVDYKNIRPLVREDKRELK
 EKFLRVFMQQYKDAETTYEKILALKTIGNAGLDISVNQLNEIIVDKRQPLPVRKEAIDALRLKDTMPRKI
 QKVLLPIYKNRQYEPEIRMLALWRMMHTRPEESLLVQVVSQMEKETNQQVAALTHQMIRHFAMSTNP
 CYQ RVAIVCSKVLSTRYQPQEQMIASSYAQLPLFLQNSFSGAQDFAAIFEKNSFLPKDLHASLDAVF
 GGNWNKYFAQIGFSQQHMDKYVQMALEKLESLEKESTTVVRGRRIQTGIKLLKELAQKMNIARPAT
 YTEKDAFAMVYLRYKMDYAFLPIDRQLVENLIEKFTSNGKVQFSEIRLLNQELEFETHHAAYFYEAIR
 KFPTTLGLPLTISGKIPTVISAEGQFSLEEGTELRLTVEARPSVAATHVYEMRMFTPLFEQGVKSVQSV
 RAYTPIKIQAVAGMKRNFEIVYKVVPENQKSIVSLTTRPVVFLRFPGFSKFEYIEAEERTVVVPQWQQ
 KTQEIEKVFNLGLEVSTRGNILNQHTLENWLLAEQDFEVSVENKNRPAEFTARLTVGQLEKTELSQIK
 YNKIFEKEFELEQENTESRREYFNKMVKNIQKEQGYKSVISLKLEAPRDYTMNTELTTCVCDKQVRMCQ
 WEVEIRRSPILEETKEWTLRSQLLVVRPEMPSSLRQLRDQPHREVQLSLTSTWGSQKKSEVTVNAQL
 QQSKEQKKYERNMDRQFNGMPEYELLKAARLNQINAVAKEYLTRETEQVLARYFDLVKTYNYWTVS
 SRPENNENDRVVVQLTVEPMSRQYVNITMQSPMERIELKNVQVPRVYLPISIAQRSVKHQLTEASGSV
 CKVQKNQIRTFDDVLYNTPLTTCYSLIAKDCSEPTFAVLSSKTEKNSEEMIIK VIRGEQEIVAQLQNEEI
 RVKVDGKKIQSESDYSAYQIERLGESAIVIELPEGEVRFDDGYTIKTQLPSYSRKNQLCGLCGNNDDEST
 NEFYTSDNTETEDIEEFHRSYLLKNEECEAEERLSEKKNYRKYERDEEQSDEYSSEETYDYEQENT
 KKSQKNQRSQKKSDDLVEKTQIKEFSHRICFSVEPVAECRRGYEVEQQQQRKIRFTCLQRHNRDASRL
 LKESRQQPLQLDDYPVSFVESVKVPTACVAY

>*C. elegans vit-3*

MKSIIIASLVALAIAASPALDRTFSPKSEYVYKFDGLLLSGLPTTFSDASQTLISCRTRLQAVDDRYIHLQL
 IDIQSASHIPQSEQWPKISLEQRELSDELKELLELPFRAQIRNGLVSEIQFSSDAEWSKNAKRSILN
 LFSLRKSAPVDEMSQDQKDMESDKDSLFFNVHEKTMEGDCEVAYTIVQEGEKTIYTKSVNFDKCITRP
 ETAYGLRFGSECKEKEGQFVKPQTVYTYTFKNEKLQSEVHSIYTLNVNGQEVVKSETRSKVTFFV
 EESKINREIKKVSGPKKEIVYSMENEKLEQFYQQGDQAEVNPFAIEMEQQVEQLDEIFRQIQEHEQN
 TPETVHLIARAVRMFRMCTIEELKKVHTTIYTKAEKKVQLVIETTLAVAGTKNTIQHLIHHFEKKSITPLRA
 AELLKSVQETLYPSEHIADLLIQLAQSPLEKYEPLRQSAWLAAGSVVRGFASKTQDLPLIRPASRQTK
 EKYVRVFMQHFRNADSTYEKVLALKTLGNAGIDLSVYELVQLIQDPRQPLSIRTEAVDALRLLKDVMPR
 KIQKVLLPVYKNRQNKPELRMAALWRMMHTIPEEPVLAHIVSQMENESNQHVAAFTYNVLRQFSKST
 NPCYQQLAVRCSKVLLFTRYQPQEQMLSTYSQPLFNSEWLSGVQDFATIFEKNAFLPKEVQASFET
 VFGGNWNKYFAQVGFSSQNFQVILKTEKLSLYGKQSDRLSRRVQSGIQMLQEIVKKMNIRPRVQ
 QTDSQNAHAVFYLRKEMDYIVLPIDMETIDNVVEKYVRNGEFDIKSLTFLTNDSKFELHRALFFYEAE
 RRIPTTIGMPLTISGKMPTILSINGKVSIELEKLGARLVLDIVPTVATTHVTEMRFWYPVIEQGVKSLQSA
 RLHTPLRFESTVELKKNTLEITHKFVVPENKKTTSVHTRPVAFIRVPKNQDSEYVETEETKISHSQYQ
 MSTEEIDRQYETFLRINAQGNVLSQWTLPMVLMTEQDFEFTLENKNRPVEFTARVTIGNLEKTDLSE
 IKFDKIFEKEFDLENNESENRRQYFHKMIREIQSEGGFKNLITLKEAPQQMYWNTELRTVCDKWIRM
 CKVEMDARRSPMEHENKEWTLRTELLAARPQMPSSLRQLREQPHREVQLALNAKWGSSKKSEITFN
 AQLEQSTEQKKFLRNIEREYKGIPEYELLKAARLNQVNVVSEYKLTPESEYTFSRIFDLIKAYNFWTVS
 EKRQVQNEEDRRVVLQLSVEPLSRQYMNMTIQTPEQEVELKNVRIPRVLPPTIARRAMFQQTWEKTGAT
 CKVGQSEVSTFDNVIYRAPLTTCSYSLVAKDCSEQPRFAVLAKKINKNSEELLVKVVRREEEIVVKSDD
 KFLVKVDEKKVNPTLEQYNIIEILGDNLIVIRLPHGEVRFDDGYTVKTNMPSVASQNQLCGLCGNNDGE
 RDNEFMTADNYETEDVEEFHRSYLLKNEECEVENDRISEKKNYRNKWNREEKKS DYVSSSDYENNY
 DEKETENQLFKKTLIKEFSNRVCFSEIPVSECRRGLESEKTSNEKIRFTCMPRHSKNARRFLKEAREQ
 TVADLVDFPVSFVESVKIPTACVAY

>*C. elegans vit-4*

MKSIIIASLVALAIAASPALDRTFSPKSEYVYKFDGLLLSGLPTTFSDASQTLISCRTRLQAVDDRYIHLQL
 IDIQSASHIPQSEQWPKIESLEQRELSDELKELLELPFRAQIRNGLVSEIQFSSDAEWSKNAKRSILN
 LFSLRKSAPVDEMSQDQKDMESDKDSLFFNVHEKTMEGDCEVAYTIVQEGGKTIYTKSVNFDKCITR
 PETAYGLRFGSECKEKEGQFVQPTVYTYTFKNEKLQSEVNSIYTLNVNGQEVVKSETRAKVTF
 VEESKINREIKKVSGPKKEIVYSMENEKLEQFYKQGDQAEVNPFAIIEEQVEQLEEIFRQIQEHEQN
 TPETVHLIARAVRMFRMCTIEELKKVHTTIYTKAEKKVQLVIETTLAVAGTKNTIQHLIHHFEKKSITPLRA
 AELLKSVQETLYPSEHIADLLIQLAQSPLEKYEPLRQSAWLAAGSVVRGFASKTQDLPLIRPASRQTK
 EKYVRVFMQHFRNADSTYEKVLALKTLGNAGIDLSVYELVQLIQDPRQPLSIRTEAVDALRLLKDVMPR

KIQKVLLPVYKNRQNKPELRMAALWRMMHTIPEEPVLAHIVSQMENESNQHVAAFTYNVLRQFSKST
 NPCYQQLAVRCSKVLLFTRYQPQEQMLSTYSQPLFNSEWLSGVQDFATIFEKNAFLPKEVQASFET
 VFGGNWNKYFAQVGFSQQNFEQVILKTLEKLSLYGKQSDRLSRRVQSGIQMLQEIVKKMNIRPRVQ
 QTDSQNAHAVFYLRKEMDYIVLPIDMETIDNVVEKYVRNGEFDIKSLLTFLTNDSEKLFHRAFFYEAE
 RRIPTTIGMPLTISGKMPTILSINGKVSIELEKLGARLVLDIVPTVATTHVTEMRFWYPVIEQGVKSLQSA
 RLHTPLRFESTVELKKNLEITHKFVVPENKTTVSVHTRPVAFIRVPKNQDSEYVEAEKTIHSQYQ
 MSTEEIDRQYETFGRLINAQGNVLSQWTLPMVLMTEQDFEFTLENKNRPVEFTARVTIGNLEKTDLSE
 IKFDKIFEKEFDLENNESENRRQYFHKMIREIQSEQGFKNLITLKEAPQQMYWNTTELRTVCDKWIRM
 CKVEMDARRSPIEHENKEWTLRTELLAARPQMPSSLRQLREQPHREVQLALNAKWGSSKKSEITFNA
 QLEQSTEQKKFLRNIEREYKGIPEYELLIKAARLNQVNVVSEYKLTPESEYTFSRIFDLIKAYNFWTVSE
 KRVQNEEDRRVVLQLSVEPLSRQYMNMTIQTPEQEVELKNVRIPRVVLPTIARRAMFQQTWEKTGATC
 KVDQSEVSTFDNVIYRAPLTTCYSLVAKDCSEQPRFAVLAKKINKNSEELLVKVVRREEEIVVKKSDDK
 FLVKVDGKKVNPTELEQYNIEILGDNLIVIRLPHGEVRFDGYTVKTNMPSVASQNQLCGLCGNNDGER
 DNEFMTADNYETEDVEEFHRSYLLKNEECEVEKDRISEKKNYKNKWNREEKKS DYESSDYESNYD
 EKETEKELVKKTLIKEFSNRVCFSEIPVSECRRGLESEKTSNKKIRFTCMPRHRSKNARRFLKEAREQT
 ADLVDFPVSVFVESVKIPTACVAY

>*C. elegans vit-5*

MKSIIIASLVALAIAASPALDRTFSPKSEYVYKFDGLLSGLPTASSDASQTLISCRTRLQAVDDRYIHLQL
 TDIQYSASHIPQSEQWPKIESLEQRELSDEFKELLELPFRAQIRNGLISEIQFSSSEDAEWSKNAKRSILN
 LFSLRKSAPVDEMNDQDKDMESDKDSLFFNVHEKTMEGDCEVAYTIVQEGETIYTKSVNFDKCITRP
 ETAYGLRFGSECKEKEGQFVKPQTVYTYTFKNEKLQSEVHSVYTLNVNGQEVVKSETRAKVTFV
 EESKINREIKKVSGPKKEIVYSMENEKLEQFYQQGDKAENVPFKAIMEQKVEQLQEIFRQIQEHEQN
 TPETVHLIARAVRMFRMCTIEELKKVHTTIYTKAEKKVQLVIETTLAVAGTKNTIQHLIHHFEKKSITPLRA
 AELLKSVQETLYPSEHIADLLIQLAQSPLEKYEPLRQSAWLAAGSVVRGFASKTQDLPLIRPASRQTK
 EKYVRVFMQHFNRADSTYEKVLALKTLGNAGIDLSVYELVQIIQDPRQPLSIRTEAVDALRLLKDVMPR
 KIQKVLLPVYKNRQNKPELRMAALWRMMHTIPEEPVLAHIVSQMENESNQHVAAFTYNVLRQFYKST
 NPCYQQLAVRCSKILLFTRYQPQEQMLSTYSQPLFNSEWLSGVQDFATIFEKNAFLPKEVQASFET
 VFGGNWNKYFAQVGFSQQNFEQVILKTLEKLSLYGKQSDRLSRRVQSGIQMLQEIVKKMNIRPRVQ
 QTDSQNAHAVFYLRKEMDYIVLPIDMETIDTLVEKYVRNGEFDIKSLLTFLTNDSEKLFHRAFFYEAE
 RRIPTTIGMPLTISGKMPTILSINGKVSIELEKLGARLVLDIVPTVATTHVTEMRFWYPVIEQGVKSLQSA
 RLHTPLRFESTVELKKNLEITHKFVVPENKTTVSVHTRPVAFIRVPKNQDSEYVEAEKTIHSQYQ
 MSTEEIDRQYETFGRLINAQGNVLSQWTLPMVLMTEQDFEYTLLENKNRPVEFTARVTIGNLEKTDLSE
 IKFDKIFEKEFDLENNESENRRQYFHKMIREIQSEQGFKNLITLKEAPQQMYWNTTELRTVCDKWIRM
 CKVEMDARRSPMEHENKEWTLRTELLAARPQMPSSLRQLREQPHREVQLAFNAKWGSSKKSEITVN
 AQLEQSTEQKKFIRNIEREYKGIPEYELLIKAARLNQVNVVSEYKLTPESEYTFSRIFDLIKAYNFWTVS
 EKRQVQNEEDRRVVLQLSVEPLSRQYMNMTIQTPEQEVELKNVRIPRVVLPTIARSAMFQQTWEKTGAT
 CKVDQSEVSTFDNVIYRAPLTTCYSLVAKDCSEQPRFAVLAKKINKNSEELLVKVVRREEEIVVKKSDDK
 KFLVKVDGKKVNPTELEQYNIEILGDNLIVIRLPQGEVRFDGYTVKTNMPSVASQNQLCGLCGNNDGE
 RDNEFMTADNYETEDVEEFHRSYLLKNEECEVENDRISEKKNYRNKWNREEKKS DYESSDYESNY
 DEKETEKELVKKTLIKEFSNRVCFSEIPVSECRRGLESEKTSNKKIRFTCMPRHRSKNARRFLKEAREQT
 VAEVLDFPVSVFVESVKIPTACVAY

>*C. elegans vit-6*

MKFFIALALLGAALASTHLDYNSIERNIQESSFRAGREYRYLFNGQLSAGLPVPSTPQGSRQLQSQVT
 LQWTDGNTVRMQLQKTRFATSQQESNSMKMLPFERFEVERMNREHQELLSMPVEFDYEHGLVREI
 RFAENDQPWSENKRAVINMLQVNILKKEYEGAESDNQEPTFSFTNVERTLEGECEVLYTVEEIKKE
 DEQRWAKSINFDKCTRRPYIHHVQTPVCKDCQQTLEQDKMSSTVLNYNITGTSSSFLINSVELRSQHL
 FAPISEKHQLVSAFTLNTMELIYAGEKKTEIKQVRNEKTSSELVYNQESSEWAEQQWAQTGEEKYLRQLP
 QWTENKVEMIKKMFSLMAKQIEQGEAELEAAHTVARIVKVLRECNEEQLEQIYRHVAEHKDEKIAEQL
 RSIYFNTLALAGTRVTIQFVDKVQSRKNIAPLKASVAIKTLVDMRYPSLAIAEDIARLCESDVSSSPAL
 RQSCWLTYGAVNGVCGQTPRVFVQKNGVKMCPRDAKQRIVDKLVQQFESASTRYEKLALKTLANA
 GLDLSVYPLEKIILNEQHETTIRTQAIESFRRLRTQMPTKIQRVLMPVYLNRRQPQHIRMALHQQIYTQ
 PEWSVLSQIGNQLRQERNQQVRAFTLSLLRSYANNESPCEQTFSSRVQSLNIPFSSQEIDRFESVY
 GKWSTYSRRHQSGFEANFASLFTTESVLPTEMMASIEGVLSGEWNQYFAQIGFTQKNMEKIIKKLLSN

VQEKGLEQIVVRGKRASGSFQPTFLSNLLEKLRI TRRQSSEQDPHAFVYIRHRDMDYAFLPIDADSIP
EVVRSMIQGGRLEIGDIERVLAQGIHFSASNAAFLYETVRRVPTPMGLPVQFTSKMPTISSIRGQVTFE
LEPKNGKSF DGLRLRVQAGPRVASTHVL SLRVICPIAEVGTKFLHQAVLNTVPDTEIRMNWEDKVVIRA
IYNTPSEEEKRIAMIQSRPVTFTRTVAPDARQYPEPIEMTYMLPAHKQLSQSLDREYPQIRVQGT LNRP
SVRIPQWIVDSNVEVYKPNVEQYEAIFELNLYNNYKMEKNYEKVYKKHNGRRYLEAEPEYDEEEHR
EQITKKFEWLQNEKVYQHVAKFEIKPEVVKMEVEAVCNNDHFCKTQIRGEELKATIQYVYPQTPRTV
EELKEQKYRQLVVMGEMNYGENTI HININGQQSQEQKKFVKQIEQAPEHETLLEASRLDQYQTVVEY
EFEPKPAQYFARYWNMVQAYLRTQYPWTSRIETREEPSRKNMIRATINVEPRQRLTVNMTIETPMETT
VLERVELPFR LPTAQIHYQPRNSRYEQKPVMEKIAHHASKQANC VVKSTKINTFDQVAYRNQFTPCYS
VLAKDCGSEKSEPRFVVLMMKKINEKKEWKNVKV VYGENEIMYKTEEGLICRVNGEEIEYQPESEIEK
KQYNIWL NKNLTKFDSDDVT VQFDGVNARIHLSALYRNQQCGLCGHYDNEKETEFYDAENQENTIPK
FAKSYLYKDSKCN YEREMFEKEENFQRIEKNQEEEDQEMNYEESRREQDDEPTEQVAIVERQHEIC
FTQKPVLR CQNGKSQESKKQKTSVYCLPSSNSWARRQMREIRREPLAQWPEHKLRLNLRDQPMEE
RTVRVAVDQKCDKFDY

>*G. ellingtonae* 016

MRTLVL LAFVAAFLCLSAGRQYNSNPTRAGFLGFNPNSALKPFQ QKNIEYSFLYNAQIASGLQSEEN
AANPQQKAVTRIQCLAKIQFASERHGQLQLQCCQAGQQNAQMSEAEVQPMGA FEQTTPSQTKQQ
LQKPCQFSYSYGVIERVQFQND DHVWSKNFKKAVLNIIQVLADLNLKRNNAAQLRRSDVRDEQEQT
EQDAWQAKQGKTFEIPEITLEGACQTTYTINKAWSGQQYNVTKTINFKKCQKIADVANGFQTDQPQA
QCAQCQQYWAQQQNGQNLWPGQSDVNNAKFRDEQQNPCAKCDPKEVKENELDRSTVIRCQIQGE
PTEQYALHGCELRSQYLYRNSKSAVGNNGVAMHTIVAANLQALDAKPMQQQQQIP AISASKDETLMY
SNEKEVDEKRFYRNGD VDFSTDGKDSPFAQVPKVQQATQALLKLMQNTQDKHKGIDQQVPVQQQR
LVEMLRMCTQQELKRVEQQCPGT TAQQKEQCKQIFVNSLAQCGTLNCVAELAKRIKNKQVSQATSVK
ALLALNQLPAPSDSILDELEQLCKSEAVNSQPTSKQSCWLTFGALVNEVCQH KTEKNAQECVGETAN
CVQSGFHKKGQCPT EKKQKYKEAILDVYQTNECIYDKVVAMSALGNAGLHNALT DLEKIIKNPRESRV
VRIMAIEATRRLRVNAPWEVQNILLPVFLN TREQPQVRTAAAFANLLNTQPGPQVVDQLMYTMAKEPNK
EVQSYVYRTMKALAKSQNPSENKIAKHIKSALKMANVDEQQLRSSGKWQYSAFNAEERE GVFSFA
QAVSSRSALPAYAHAQLD SHFNDQYQLNNIKVFM LQQEAEENWYGNLADHFIKQKHTRAQR RNQQN
SQRLHNIYEALGIKSRAEFGT ELDDSKFEDKNQQNIKPFAMLC LRSNNVDKAFLALDEHTMEEQTRK
ILNGEEKANWALLFGQQQLNTAIVQNLNEKEAKVPTSVGVPLRILNSMPVLANIEGQIKIKVESARLTS
PLGIQFDVNAKASGR LAQIQKMELWTPFAASGVECVRSMQVNLPLQGELNANS DQGISLKAKLPQQK
TDLASAHS LPITYIAKVDPKTNLVREPRQILTIHDPKLEQQQH EVNVVQGQKNLAMPLHFRGHFWPA
EPTS YQQFAQLLLATENAVHVHYQPSRDSPREVIFRVTGEHFQQVAQHEKPDVQLKFFAE EFHNAY
SCNAEQCQQSGDCEQCRKFEQQQQEDAQEI EGQEQRQQMLNNFIGAYKPRQQYKHS LKLI AQTN
GAKDAKAQC DIQAQCDEQMNYCKVQLKAERPPIANENMKWTLNAQSHLVMPETAASVEQLQKPKQ
QLKQFIAQAE AQWGHGQPKQNVNIRMHGTQAHTQQVRAVLQQQNKRLISSQQKRN FQRRSAFL
NKFQIEAAYS LKSNAQNAINRVLEV LKSTYFWNTKSQMLLSGQNAQKNGKMAATILIDPMTQKNANVTI
RTPAQQVRMQQIELPYKVRPFS LRQQSARQTNSFSQLFHNFNINNRAQCTVDGRNVNTFDDLIYTAP
LSSNCHSVLAKDCGSDQPQFVVLMMKVQQQNNKKS AQNAEKKVLKVITPEHQIECQPKTNANSPR
QLICRV DGKQITGDEQCQDN YCVQFNNPQQNDLTVQTEGIAVRFN GRKAWLKISQAYKNTQCGLCG
HYDDSDDQEEELRMANGQTTSNLAEFHRSYVLEDSEQCSQQVLN NFYNEQSSEQDTYYQQYNQQF
DNNSEQEKNEAEDEDQ QDQVDWQKEQFDGQFNSLEEDDD EDNVNRPLEIIPDISAQKCCCCC DDED
GQCQNCQDNDECQEKCDQCEQSPQCCCCCQEGDSQCQEKCTSEQCDANCKDGRCAEQ LAV
WPTWTKIHEESHAVCFSTRPVKQCPQGAQAYSSLTNDYASSVEQKSKKVN FVCMERSEPETRLL
SLLRRGQRVVDVNARGASFTQNVSEPQQCWRNRY

>*G. ellingtonae* 060

MPFFLLKAMLAVYQNHG CIFDKAVALGALGNAGLHNALNDLERIMKNSQEPRVVRVMAIDATRRLGVQ
MPRKVRNVLLPIFLN TREQPQLRTAAAFANLLNAQTEPQVVDQLLFAIAREPNKEVQSYAYRTMKTLAKS
QNPNERRIAKHIKSALKMANVDEQQLRYS GKWQFSLYNSKERE GAFISLAQAVSSRSALPAYAHAQLD
SHFNDQYQLSNVKVFM LQQEAEKWYGNLADRFIKQQQQQTRAQRQNLQYSTKLRSIHENLG IKNRR
SLSYDNEMDGSMKIGTYKQQNIQPFVMLCLRSNNVDKAFLALDEHIMEEQTRKMLNGQEKANWALLF
GQQQQLNTAILQNLNEKEAK IPTSAGVPLRILNSMPILANIEGQLKIKTESARLTSPLGLQLDINAKASAR
MAQIQKMELWTPFAASGVECVRSVQVNLPLQGELNANS DQGISLKAKLPQQKTQLASAHS LPITYIAK

VDPKTNLVREPRQILTIHDPKLEKQQHEVNAVQGQKNWAMPLHIQGHYHWPTPTSQQFAQLLLAT
 ENAVQVHYQPGQNSPREAILRITGEHFQKAAQNEQPGTQLRNFFAEFTSPYSCNVEQCQQDGDCE
 QCSEFEQQQKEDAEVEGQQVRQQQLHSFINAYKPRQQYKNSLKLVAQTNGGAKDAEAQCDIQTQ
 CDEQMNYCKVQVQAERPPIGSENEKWTFFNAQAHVLPETAASVEQLQRKPRQFKQFIAQTKAQWG
 QQQQSQQHVNVRLGTQAHTQHVRALQQRQNKRLISAQQQRNFQRRSAFLNKLQIEAAYSLKAYA
 QNAINRVLEVLTSTYFWNTKSQLLSGQNAQKNGKVAIIVIDSMTLKNANVTIRTPAQKVLMMQIELPS
 KVRPFSLRHQSAQQIDSLVQLFHHFSIKNRPQCTVDGRSVDTFDDFIYTAPLSSNCHSVLAKDCGSE
 EPQFVVLMMKKVPQQQNNNNNNKKSALTAEKMLVVIT-
 PKQQIECQPKTNAKGPQQRRICKVNGQLVAGDEQCQGRSCVQFNDPQQSDVTIRTEDIAVRFNGR
 KAWLKVSHAFKNTQCGLCGHYDDSDQEEELRMANGKFASNLAEFHRSYALKDSELCEQALDNFY
 SEQNGEQDSFLEQYNQLDNDSEQNNEAEGEALESQWQKEQFDGKLDSEKHDEEDIGNNLSNISE
 PNGPSQPPAPQCCCCCQEGGDEQCQKCHTSKQCDVSDRINSFPPIPSPICFLCSATWPTWTK
 IQEQNHAVCFSTQPVKQCPGAYSSVSSDYGSDEEEESNYDEQQNEQKSKRVSFVCMERADPET
 RRLRLRRQGRVVNVSGHGASFTEIVSEPQQCWHSRF

>G. pallida 0128

MRTLVLAFVAAFLCLSAGRQYNSNSPARAGFLGFNPNSALKPFQQKNIEYTFLYNAQIASGLQSEEN
 AANPQQKAVTRIQCLAKIQFASERHGQLQLQCCQAGQQNAQMSEAEVQPMGAFEQTTPSQTKQQ
 LQRPCQFSYSNGVIERVQFQNDHVVWSKNFKKAVLNIIQLNLKRNNAAQGLRKSVDREDEQEGTEQD
 AWQAKQGKTFEIPETLEGACQTTYTINKAWSEQQYNVTKTINFKKCQKIADVANGFQTDQPPAQCA
 QCQQYWAQQQTQDNLWPGQSDVNSAQFRDEQQNPCAKCDPKEVKENELDRSTVIRCQIEGEPTEQ
 YALHGCELRSQYLYRNSKSAVGNNGAAMHTIVAANLQALDAKPMQQQIPASASKDETLMYSNEKE
 VDEKRFYRNGDVFSDGKDSQFAQVPKVQATQALLKLMQNTQDKHKGIDQQVPVQQQRLVEML
 RMCTQQELRQVEQQFPGTTAQKQEQCKQIFVNSLAQCGTLNCVAELAKRIKNKQVSQATSVKALLAL
 NQLPAPSDSILDELEQLCKSEAINSQPTSKQSCWLTFGALVNEVCQHKTEKNAQECVEETANCVQSG
 FHKKGQCPTTEKKQKYKEAMLDVYQTNECIYDKVAMSALGNAGLHNALTDLEKIIKNPRESRVVIMAI
 EATRRLQVNAPWEVQNILLPVFLNTREQPQVRTAAAFANLLNTQPGPQVVDQLMYTMAKEPNKEVQSY
 VYRTMKALAKSQNPSEKKIAKHIKSALKMANVDEQQLRSSGKWQYSVFNAEEREQGVFSFAQQVSS
 RSALPAYAHAQLDSHFNDQYQLNNIKVFMQQEAENWYGNLADRFKQKQHTRAQRRNQNSQRLH
 NIYEALGIKNRRSDFGTELDSDKEFDKQQNIKPFAMLSLRSNNVDKAFLALDEHTMEEQTRKILNGE
 EKANWALLFGQQQLNTAIVQNLNEKEAKVPTSVGVPLRILNSMPVLNIEGQIKIKVESARLTSPLGIQ
 FVDNAKASGRLAQIKMELWTPFAASGVECVSRMQVNLPLQGELNANSDDQGISLKAKLPQQKTDLAS
 AHSPLITFIAKVDPKTNLVREPRQILTIHDPKLEKQQHEVNAVQGGQKNLAMPLHFRGHFWPAEPTS
 QQFAQLLLATENAVHVHYQPSRSDPREVIFRVTGEHFQQAQHEKPDMQLRKFFAEFEHNAYSCNAE
 QCQQGGDCEQCSKFEQQQEDAEIEGQEQRQMLNFIKAYKPRQQYKHSCLKIAQTNGGAKDA
 KAQCQDIAQCDEQMNYCKVQLKAERPPIANENMKWTGADQGGQPKQHVNIRMHGTQAHTQQVRAVL
 QQQNKRLISAQQQRNFQRRSAFLNKFQIEAAYSLKSKAQNAINRVLEVLTSTYFWNTKSQMLLSGQ
 NAQKNGKMAATILIDPMTQKNANVTIRTPAQQVRMQQIELPYKVRPFSLRQQSARQTNSFSQLFHN
 NNNRPQCTIDGRNVNTFDDLIYTAPLILQLSRVSGQLRFGSAPIRRADEEGSATTEQQRNTAQNAEK
 VLKVITPEHQIECQPKTNANSPRQLICRVDGKQITGDEQCQDNYCVQFNNPQQNDLTVQTEGIAVRFN
 GRKAWLKISQAYKNTQCGLCGHYDDSDQEEELRMANGRTTSNLAEFHRSYVLEDSEQCSQQVLN
 NFYNEQSSEQDTNEAEDEDQQDQVWQKEQFDGQFNSLEEDDEDTVNRPLEIISDISAQKCCCCCD
 NEDGQCQQNCQDNDECQEKCDQCEQSPQCCCCCEEGDSQCQCEKQTSKQCDANCKDGRCAE
 QLAVWPTWTKIHEESHAVCFSTRPVKQCPQGAQAYSSLTNDYATGVEQNKSKKVNFCMERSEPE
 TRRLSLLRRGQRVVDVNARGASFTQNMAATILIDPMTQKNANVTIRTPAQQVRMQQIELPYKVRPFS
 LIRQQSARQTNSFSQLFHNFNINNRQCTIDGRNVNTFDDLIYTAPLSSNCHAVLAKDCGSDQPFV
 LMKKVQQQNNKKSQAQNAEKVLKVITPEHQIECQPKTNANSPRQLICRVDGKQITGDEQCQDNYCV
 QFNNPQQNDLTVQTEGIAVRFNGRKAWLKISQAYKNTQCGLCGHYDDSDQEEELRMANGRTTSNL
 AEFHRSYVLEDSEQCSQQVLNMFYNEQSSEQDTYYQQYNQQFDNNSEQEKDEAEDEDQQDQVWQ
 KEQFDGQFNSLEEDDEDTVNRPLEIISDISAQKCCCCCDNEDGQCQQNCQDNDECQEKCDQCEQ
 SPPQCCCCCEEGDSQCQCEKQTNKQCDANCKDGRCAEQLAVWPTWTKIHEESHAVCFSTRPVKQ
 CPQGAQAYSSLTNDYATGVEQNKSKKVNFCMERSEPETRRLSLLRRGQRVVDVNARGASFTQNV
 SEPQQCWNRNY

>G. pallida 0564

MLSTGSAFLFDLNDLQFLLHPTPVFAHPLLFGNRTKNAIPSASKQQQLSDKAETIEAQCLKNEMEANG
GGNDDEKQQHFVDPLNASAIDEKTTQLEKRHSVHQHNGFASGGDTPTAILFTERLNNTLAECESEQW
PLDGVDPSLVGHRPWRVLRQRILDTFSSTTDRLSLESSFLRVNKISGGEGDDSAQRARSLRKRNGGT
GRKGQKNFVTEQQQPVSKEFHELTQDEWTRKLKELRRILIVKWEENRVECIKLTAEVAALISTISSAQF
YPAQFVHVTDLFDLFGHLVYERLFRMAGEERRAAGLGVLP AHFDNDVLETTRMKTRNWF GKINETV
ALLAPRIYLLAALLPCQRFMDPLGAAQENVMELCQSVKQVQHPLISSYLRCYICRVAMRLNAADRAPH
WKCLNDWLQTYSSQPTFLLAPAMEWIVQCVAFNATSRYSMRRCPPLFRLFALASLLVVVTCRSNYA
SKYDDKFRHYHHEQNGQQDDQDHDYGFQVAREREGHLSNDYLKGKFGMSSDKRTPTFKSFQPG
YKFQFLYDSQTTSGYAPWWSIQRAMGRFQALATFEAIGPGEASMLSECLIGQLNGQSPTAGVWNPG
QLLPVRVFEVSMHFSHVQLEQPVRFITYSRDGIHHIYFDPMEPAWSRNYKKAILSTGQFRLGAKDDGD
EADQNQTNQTERSEEWKTKWQQTFSVHETSIEGNCETDYTIVPFSRELISVTKSINFDRCTRLADIAY
GMADEEAKQRGQHNNKPYQEQRERTDNEQQNLGGANPMISRQFLDRSTTIYLFVNQLKNNSM
LDHSSLLPHSVELLSQYTLKGSEAEQSQAMQTVAVAKLTYIETKQNRQKKNKRQQQDDKSGLEREQ
LLYTGESEQKEKCFMFGEDETSETVRCQRRAKAFNNDLHGVPLVISGSRMPSPGSPFPLNFPNSTT
KMRAEFVVRNVAQAQEDNDGTLQAMSQADEFVQLLRFSTEQELKQLKERLDNDRIAAQLFA
DALAAAGTRNTLKLTLTRIEQGEFTELKAAQMLKTF SRGSGLPAPSDRIADALLALCKTEHVQNSELLR
QTTAGTRNTLKLTLTRIEQGEFTELKAAQMLKTF SRGSGLPAPSDRIADALLALCKTEHVQNSELLRQT
SWLSFGAIVGELMQLKRIQRQSPVGTPLGEKIEQCPDHKMELYTKALLEQFSTGAGDDPMEKIVALKA
IGNAALDEVPEVLATLISPWNSEQQNDEQPTSFAKQLQEMQRQNAIEALRRIPSARVQQLLPVYTQ
QELPQTRTVALNVLMQSQLRPSDAVIEQIAFTMAKERCLQVLAYSYTTLKALSQSPAKRHQHSARIK
EALRVSGLDELELEKRIRHKDIVGLSRFVQTNALYSAEEGQILLQASNVFTDKTSALPNHAFSLSSSL
FNGQMSEEQLKLTVSQWNIESLLFTDAGLKELLKNGQENHQQQQRNGTEQQQQPSEESSSSSSSWT
SSAAEQLRSLWRRLSIVGRSNEEQSTDQMQQQKPVMLLALRVNNIDTVVLPFGIGGQQRSNSLISAL
KKLLAVGIGSISKLTGTSQHHRQPAVDTSGLFWHERAARVPMSTGQQLHVQHKSFLYGSIRANAERL
GSHIGLRLHPSATFGRIQKMKVWASPVHSSGVKNIGAIEANMPLKAQLNTLSQQKQVKLNKVPQTP
TRLFSMHTMPMTFVGTEDTQNNLSQHPSVFSHFGEQCKAVQNQQLLSMAKTFDKHSNTEHSLQLP
FHARGFLHMPRRVDKLKEWVKVLLSAENHYQVMLEPQAGRSPPHEIAFYLGWNLFQPHPTIGGLSRA
ELKQFLNQFTLVTPKSGDGEEADREEQWWTAAASSASSSAEESQKEEQQQRRIGAYLDKLGKAY
KHTLKASMETVGSQDQNRRAEFKLDNVCDERMRSCTLHAQVQHLGQQQQPQWSADLRVQSLYPEQ
APDLEHVMLHSREDKPPQQQQMFLCQAELEWNANGQLKEMIKLNIKGEPSSTEMLR LAHQFVQQQS
SGSNVANSVLPNTAGHPFVDKYTVVNVWQLRAAAQNAFVRYLETMKCSHFFNVDSKIVMSSSNND
DDSEEENSMISTGRQQQQQQQSVKQGTITITLSIDPITHQHINMTVRTADQLTVFRMVKMGSPVRPF
PLVRSSPAGTFSNAAPARVVSISRVVHQAHPFRSRAVCRLDTSKRQIRTFDGAERMDTPIPKCWTVL
AKDCSAEGQSQYAVLIKSSSSDNNQMKLVIAPIGILLVEPHAQQPDTKKQPQLQLQLHINGKQQQQ
KRTFTSDTPLSDLPDGI FAHSSSTEKEAVIETTT SINGAEPIAGITGALSVHFDGQKVRLTNGQQNR
QKPCGLCGNWNDAEDEL SMPDGSQTESAKAFNRAYS LIDDEGEEGGECEHNGILEKFHAQQEEPE
LDGQQLNTSGTVLEQIEPVLKTALIEPHQLCFSLKPVKQCPSGSLTARNIAGELMDQKQKQQQQQEA
EEEEEEKAPKQKVPFTCIERHSVEAKRFKAQIRRG EAVQLGDHRPISFQEVII EADRCIKFDG

>*G. pallida* 7024

MRTLALLALVAVFLSVSTSRHYTPGDYLEYAPSDGFGHARSNPNSASKPFQKQNT EYTFIYNAQIASG
LQSEENVANPQQKAITRVQCLAKIQFASERH GQLQLQCCRAGQQNEQMSEPQEMQPIEAFEQTKLQ
LQQPCQFSYSDGVVERVQCQQDEDWSKNFKKAVLNMIQLNLKRNSAQKEPKQQSTEQYTWETKMG
KTFTTPEITLEGECQTTYTINKARAEQQYSVTKTINFKKCQKIADVAKFQTEQPQAQCAHCQQYWAQ
QQDEQNLHANAAQQQQQNEEQHPCAKCDPKEVKENELDRSTVIRCRIQGEPT EQYALNGCELRSQ
YVYRNSKSADGNYGAAMQTVVTATLQALDVQLKQQQIP AISANDEETLLYSNGKDVDEKRFYRNGDE
NFGTNGKDSPFARVPKVQAVQALNKL VQNMQDKQKGIDQQAPIQQQRLVEMLRMCTQQELKQVE
QQCPGTTAQQREQCKQIVVIALAQCGTMNCVAELAKRIKNKQVPKAVVVQALLSLNQLPAPSDGILDE
MEQLCKSEAVNSQPM PKQSCWLTFGALVNEVCQH KTEKNAQECVGSSANCVQSGFYEKDQCPT EK
KQKYKEAMLAVYQNHGCI FDKAVALGALGNAGLHNALNDLEKILKNPQEPRVVRVMAIDATRRLGVEM
PRKVRNVLLPVFLNTREQPQLRIAAFANLLNAQTEPQVVDQLMFAIAREPNKEVQSYAYRTMKTLAKS
QNPNEKRIAKHIKSALKMANVDEQQLRYS GWQFSLYNAKEREGAFISLARAVSSRSALPAYAHAQLD
SHFNDQYQLSNV KQQQQQTRAQRQNLQYSTKLRSIYEDLGIKNRRSLFYDNEMDGSTKIGTYQQQNI
KPFVMLCLRSNNVDKAFLALDENIMEEQTRKMLNGQEKANWALLFGQQQLNTAILQNLNEKEAKIPT
SAGVPLRILNSMPILANIEGQLKIKTESARLTSPLGFQLDINAKASARMAQIQKMELWTPFAASGVECV
RSVQVNLPLQGELNANWDQGISLKAKLPQKQTQLANAHS LPITYIAKVSNTNLVREPRQILTIHDPKL

ERQQHEVNAVQGQKNWAMPLHIQGHYHWPTPTSYYQQAQLLLATEKCGASALPTRPEQSARGHL
AHYRRAPFKRRHRTSSRRARSCGISSRNSSARTVATSNSASRTAIVSSAANLSSCDEQMNYCKVQVQ
AEPHVLMPETAASVEQLQRKPQQFKQFIAQAKAQWGQQGQSQQHVNVRMLGTQAHTQHVRALQ
QRQDKRLISVQQQRNFQRRSAFLNKLQIEAAYSLKAYAQNAINRVLEVFKSTYFWNTKSQLLSGQNA
QKNGKVAIIVIDSMTQKNANVTIQTPAQKVRMQQIELPYKVRPFSLRQQSAQQIDSIAQLFHHFNIKN
RPQCTVDGRSVDTFDDFIYTAPLSSNCHSVLAKDCGSEEPQFVVLMMKKVPQQQNNNNNNKKSAPNAE
KMLVVITPKQIECQPKTNAKGPQQRRLICKVNGQQIAGDEQCQCGKSCVQFNDPQQNDVTIRTEDIAV
RFNGRKAWLKVSRAFKNTQCGLCGHYDDSDDQEEELRMANGKFASNLAEFHRSYALKDSELCSQQ
ALDNFYSEQNGEQDSFLEQYNQLDNDSEQNNEAEGEGLESDWQKEQFDGKLDSEKHEEEDIDNN
LSDNISEPNGPSQPPAPKCCCCQEEGDEQCQKCHTSKQCDVVCKSGRCAAEQPIWKATWPTW
KTKVQEQNHAVCFSTQPVKQCPSGAYSYSMSDDYGSDEEEDSNNDDEQQNEQKSKKVSFVCMERS
DQETRLLRLLRQGQRVVNVSGHGASFTETVNEPQQCWHSRF

>*G. pallida* 9459

MRTLALLVLVTALLSVGTGKRHINWNDESKDYSFGRRTSADAELKPFQKQDTEYSFIYNAQIASGLHS
EEDSAVNPPQKAMTRIQC VANIQFASERHAQLQLQQCRVGQQNEQISEPQEVQPMGAFEQTNIPEQ
TMQQQLQKPCQFSYADGVIERVQCQSDGEDWSKNFKKAVLNMIQMNLKRNNAAQGLHRSQVRDEQQQ
VQQQQQDAEQGITFNIPEITLEGACQTTYTINKARRSGQQQYNVTKTINFKKCQKIADVANGFQTDQP
QAQCEQCHQYWAQQQGDQQNIQTGAATEQQSGEEHPCAKCDPKEVKENELDRSTVIRCQIQGEP
EQYTLKGCELRSQYVYRNSKSADGSYGAAMQTVVAATLKALDVKPKKQIPALSSTEDETLMHSNEK
EVDEKRFFMNGDKDFNSKGSFPARVPKVQQAQKALLKLEQSMQDKHKGIDQQAPVQQQRLVEMLR
MCTQQELKQVEQQQCSSKATAQQAQEQCKQIVANALAQCGTLNCVTELAKKIKTKQVSKANAVQALLA
LNQLPAPSDSILDEMEQLCQSEAVQNNQMIKQSCWLTFGALVNELCQHKAEKNAQECVGTANCVC
SGFFFKGQCPMDKKQKYKEAMLAVYQNNQNCIYDKAVALGALGNAGLDIALTDLEQIIKDPRESRVVR
VIAVDATRRLRTKLPRVQNVLLPVFLNTREQPQVRIAFAFANVMDAQPAQQVIDQLMFAITKEPNKEVQ
SYAFRTMQTVAKSQKPSESQIAKHIKSALKMANVDEQQLRASGLWQFGAICAEQREGVFSLAQAVS
PRSALPAYAHAQLDSDFNDQYQHSNIKVFMLQQEAQEKWYGNMANRFFTKQQQQQPTRAERGKPPQ
QQNNQKLRLNIYEALGIKSRRSDFGSTDVDSKEMETYQQNTRAFALLGLRSNNVDKFLALDEHIMEE
QTRRMLNGQEKANWAYLFGQQGQPLLNTVLVQNMNEKEAKIPTSAGVPLRILNSMPMLANIEGQLKI
KADSANPASPLGVQFDINAKASARIAQIQKMEIAPFATSGVECVRSMHVNLP LHGELNANTNHGISVK
AKLPQKQTQLASASQLPVTFIASVDPKSNLVREPRQILTVDHLKLERQQHDMNTVLGQKDWALPVHIR
GHYWPVAVPSSYKQFAQLLMATENAVHVHYQPSQDSPRELTLRATGEQFQKAAQIEQPDQLRNFYA
EEFHNAYSCNAEQCQQTGECEQCNEFEQQQRQDAQEVESQQVRQQLLHNFINAYKPRQQYKHSIK
LIAQTNGGAKDATAQCEIQACDERMSYCKLQTTVNANQKWSLNAQAHVVMPEPAASVQQLQKQPE
QLKQDSPNSTSTCAFTARKHTLNRFANSSSSSNKQQQQQQQKSLISAQQQRNFQRRSAFLNKFEL
EATYSLKAYAQNTINRALEMLKSTYFWNTKSQLLNGGQSVQKNGKIAAIIAIDPLTLKTANVTIRTPTQH
MRMQQIELPAKVRPFSLRVQQSAQQLHSLAQLFQQYSFTKRPQCTVDGRSVNTFDNLVYTAPLSSSC
LSVLAKDCGSEEPQFVVLMMKIMRPEQQQQNAKKSQAQNAKVLKVITADQQIECQPKTANQQQQQRLI
CKVNGQQVTEDEQCQDGQCVQFNDPQQTDTVIRAEGIAVRFNGRKAWLVSEAFKNTQCGLCGHY
DDSDDQEQELRMANGQIASGLAEFHRSYALQDSEQCTQQVLNKFYSEKNNNDGDDDDDEQDTFRQQ
YNQLDDDNNEQYNNEAEENEQNRDWQKEQFDGKFGSLEQDDDEDINSDFSWSIDTSDQQCCCC
CNNDGEDNDDDKQCKQNCQDNDQCQRKCEKQCEQQQQQQQPAPPCRAGDDKCMKKCQNGEE
AACTKEAGWPQWKTIVQEQNHAVCFSTEPVKQCPRGFSPYDSYSSDDDEVEQQQQQQQQQNVK
KSKKVSFVCM DRAGSETRLLRRLRRGQRVVNVSARGASFTETVDEPQQCWRRARN

>*G. rostochiensis* 2537

MRTLVLAFVAAFLCLSAGRQYNSNPPARAGFLGFNPNSALKPFPQKNIEYSFLYNAQIASGLQSEEN
AANPQQKAVTRIQCLAKIQFASERHGQLKLQCCQAGQQNAQMSEAEVQPMGAFEQTTPSQTQKQ
LQRPCQFSYSNGVIERVQFQNDHDVWSKNFKKAVLNIIQLNLKRNNAAQGLRGSQVRDEQEQTQD
AWQAKQGKTFEIPETLEGACQTTYTINKAWSGQQYNVTKTINFKKCQKIADVAYGFQTDQPQTQCAQ
CQQYWAQQQNGQNLWPGQSDVNNAQFRDEQQNPCAKCDPKEVKENELDRSTVIRCQIQGEPTQ
YALHGCELRSQYLYRNSKSAVGNNGAAMHTIVAANLQALDAKPMQQQQQQQIPASASKDETLMYSNE
KEVDEKRFYRNGDVDFSTDGKDSFPAQVPKVQQAQALLKLMQNTQDKHKGIDQQVPVQQQRLVE
MLRMCTQQELKRVEQQCPGTTAQKQEQCKQIFVNSLAQCGTLNCVAELAKRIKNKQVSQATSVKALL
ALNQLPAPSDSILDELEQLCKSEAVNSQPTSQKQSCWLTFGALVNEVCQHKTEKNAQECVGETANCVC

SGFHKKGQCPTTEKKQKYKEAILDVYQTNECIYDKVVAMSALGNAGLHNALTDLEKIIKNPRESRVVRIL
AIEATRRLRVNAPWEVQNILLPVFLNTREQPQVRTAAAFANLLNTQPGPQVVDQLMYTMAKEPNKEVQ
SYVYRTMKALAKSQNPSENKIAKHIKSALKMANMDEQQLRSSGKWQYSVFNAEEREGVVFVSAQAV
SSRSALPAYAHAQLDSHFNDQYQLNNIKVFMQLQEAENWYGNLADRFIKQKQHTRAQRNRNQQNSQR
LHNIYEALGIKNRRADFGTELDDSKFEFDKNQQNIKPFAMLCRLSNNVDKAFLALDEHTMEEQTRKILN
GEEKANWALLFGQGQQLNNTAIVQNLNEKEAKVPTSIGVPLRILNSMPVLANIEGQIKIKVESARLTSPLG
IQFDVNAKASGRLAQIQKMEIWPFAASGVECVSRSMQVNLPLQGELNANSDDQGISLAKLPQQKTDL
ASAHSLPITYIAKVDPKTNLTNLVREPRQILTIHDPKLEQQQHEVNVVQGQKNLAMPLHFRGHFWPA
EPTSYYQFAQLLLATENAVHVHYQPSRDSPREVIFRVTGEQFQQVAQHEKPDMLRKFFAEEFHNAY
SCNAEQCQQGGDCEQCSKFEQQQQEDAQIEGQEQRQQMLNFIGAYKPREQYKHSCLKIAQTNG
GAKDAKAQCDIQAQCDCEQMNYCKVQLKAERPPIANENMKWTLNAQSHLVMPETAASIEQLQQKPPQ
LKQFIAQAEAQWGHGQPKQHVNIRMHGTQAHTQQVRAVLQQQNKRLISAQQKRNFRRRSAFLN
KFQIEAAYSLKSNAQNAINRVLEVLIKSTYFWNTKSQMLLSGQNAQKNGKMAATILIDPMTQKNANVTIR
TPAQQVRMQQIELPSKVRPFSLIRQQSARQTNSFSQLFHNFNINNRPQCTVDGRNVNTFDDLIYTAPL
SSNCHSVLAKDCGSDQPQFVVLMMKVPQQQNNKKSQAQNSEKVLKVITPEHQIECQPKTNANSRQLI
CRVDGKQITGDEQCQDNYCVQFNNPQQNDLTVTQTEGIAVRFNNGRKAWLKISQAYKNTQCGLCGHYD
DSDDDQEEELRMANGQTTSNLAEFHRSYVLEDESEQCSQQVLNNFYNEQSSEQDYYQQYNQQFDNN
SEQEKNEAEDEDQQDQVWQKEQFDGQFNSLEEDNDEDNVRPLEIIPNISAQKCCCCCDEDEDGQC
QQNCQDNDECQEKCDQCEQSPQCCCCCQEGDNQCQEKQTSKQCDANCKDDRCAEQLAVWPT
WTKTIHEESHAVCFSTRPVKQCPQGAQAYSSLTNDYATGVEQKSKKVNFCMERSEPETRLLSLLR
RGQRVVDVNARGASFTQNVIEKMLQQLASFNIIPRIKLDEFVDRLCYRYTPMLLALFMFITGTVDYLG
KPISCMPLNEYSEAWVYVYHQYCYVSGTYVKDYMHIEDSSQPYRISVHYQWVPFVIAQALLIRLPF
FFWRLGQRIHDVDFNHVNELCWDARLSSGEKREKNLNEVSDFIFQTAKARGGGLGCWACFMYLIFKL
MNIVLIGAQLFFLTIFYGYKDWRWGIKLAMSTWQNNGWQEFTYFPRISYCEYGRDFLGHADGGLL
SEVRCVLSVNMINEKMFLLFYFWYFALIFLAFNTFFYIVRLMRPFQLYSAKSFLKLVPTKVDGWSVRQ
FEDELLGLDGLLLMRVFRQNSGPLVAGEVGARLYENLHRTDSKNEDELDVE

>*G. rostochiensis* 2714

MRTLALLVLVTALLSVGTGKRHIKWNDENGGYSFGGRTSADSELKPFQQKDTEYSFIYNAQIASGLQ
SEEDSAVNPQQKAMTRIQCQVANIQFASERHAQLQLQKCRVGQQNEQISEPQEVQPMGAFEQTNIPD
QTMQQQLQKPCQFSYADGVIERVQCQSDGEDWSKNFKKAVLNMIQMNLKRNNAAQGLHRSQVRDEQQ
QVQQQQQDAEQGITFNIPEITLEGACQTTYTINKARRSGQQQYNVTKTINFKKCQKIADVANGFQTDQ
PQAQCEQCHQYWAQQQGGEQNIQTGATTEQQSGEEHPCAKCDPKEVKENELDRSTVIRCQIQGEP
SEQYTLKGCELRSQYVYRNSKSADGNYGAAMQTVVAATLKALDVKPKKQQVPALSSTEDETLMHSN
EKEVDEKRFFMNGDKDFNSKGSFARVPKVQQAQALLKLEQSTQDKHKGIDQQAPVQQQRLVEML
RMCTQQELKQVEQQQCSSTTAQQTEQCKQIVANALAQCGTLNVCVELAKKIKTQVSKANAVQALL
ALNQLPAPSDSILNELEQLCQSEAVQNNQMIKQSCWLTFGALVNELCQHKAEKNAQECVGNANTANCVQ
SGFFKKGQCPMDKKQKYKEAMLALYQNNQNCIYDKAVALGALGNAGLDIALTDLEQIIKDPRESRVVR
VIAVDATRRLRTKLPRVQNVLLPVFLNTREQSQVRIAFAFANVMDAQAQAVQVIDQLIFAITKEPNKEVQ
SYAFRTMKTVAKSQKPSESQIAKHIKSALKMANVDEQQLRSSGLWQFGAISAVQREGVFFSLAQAA
PRSALPAYAHAQLDSHFNDQYQHSNIKVFMQLQEAQWYGNMANRFFNKQQQPTRAERGKPPQQQ
NNQKLNIYEALGIKSRRSDFGSTDADSKEMETYQQNNRAFAMLGLRSNNVDKLFWALDEHIMEEQ
RRMLNGQEKANWAYLFGQQGQPLLNTAIVQNMNEKEAKIPTSAGVPLRILNSMPMLANIEGQLKIKAD
SANPASPLGVQLDINAKASARIAQIQKMEIAPFATSGVECVSRSMHVNLPLHGELNANTNHGISVKAKL
PQQKTQLASASQLPVTFIASVDPKSNLVREPRQILTVHDLKLERQQHEMNTVLGQKDWALPVHIRGH
HYWPAMPTSYYQFAQLLMATENAVHVHYQPSQSDSPRELTLRATGEQFQKAAQIEQPDTPLRNFYAE
FHNAYSCNAEQCQQAGECEQCNEFEQQQRQDAQEVESQQVRQQLLHNFINAYKPRQYKHSIKLIA
QTNGGAKDATAQCEIQAQCDERTSYCKLQTTVNANQKWTNLNAQAHVVMPEPAASVQQLQQKPEQL
KQFVAQTKAQWGAQGGQPKQHVNLRHGTQAHTAQIRKFLQQQQQGGKQLISAQQQRNFQRRSAFLN
KFELEATYSLKAYAQNAINRVLEVLIKSTYFWNTKSQLLNGGQSVQKNGKIAAILDPLTLKTANVTIRTP
TQHMRMQQIELPAKVRPFSLVRQQSAQQLHSLAQLFQQYSLTKRPQCTVDGRSVNTFDNLVYTAPLS
SSCLSVLAKDCGSEQPQFVVLMMKIMRPQQQQQQNARKSAQNANKVLKVITADKQIECQPKTANQQQ
QQQQRICKVNGQQVTEDEQCQDGQCVQFNDPQQTDVTIRAEGIAVRFNNGRKAWLKVSQAFKNTQ
CGLCGHYDDSDDDQEQELRMANGQIASNLAEFHRSYALQDSEQCTQQVLNKFYSEKNTKDDDDNDE
QDTRQQYNQLDEDDDDNDDNNNEQYNNEAEENEQNRDWQKEQFDGQFGSLEQADDEDIDSDSFS
WISDTSQQCCCCCNNDGEDNDDDDDDDKQCKQNCQDNDQCQRKCEKQQCEQQQQQPPPC

RAGDDKCMKKCQNGEAACPKETGWPMWKTIVQEQNHAVCFSTEPVKQCPRGSSPYDSDYNSDDD
EVQQQQKQQLQQLQNVKSKKVSFVCM DRAGSETRLLRQLRRGQRVVNV SARGASFTETVDEPQ
QCLRARN

>*G. rostochiensis* 6238

MRTLALLALVAVFLSVSAGRHYTPGDYLEYAPSDGFGHSRSPNSVSEPFQQKNTEYTFIYNAQIASG
LQSEENVANPQQKAITRVQCLAKIQFASNRLAQLQLQQCRAGQQNEQMSEPQEMQPIEAFEQANIPS
QAKLQLQQPCQFSYSDGVVERVQCQQDEDWSKNFKKAVLNMIQLNLKRNSPQKDPKQQNTEQYAW
ETKMGKTFTTTEITLEGECQTTYTINKARAGHQYSVTKTINFKKCQKIADVAKGFQTDQPQAQCAHCQ
QYWAQQQDEQNLQQVANAPQQQQQNGEQHPCANCDPKEVKENELDRSTVIRCRIQGEQNEQYALN
GCELQSQYMYRNSKNADGNYGAAMQTVVATLQALDVQLKQQQIPASASDEETLLYSNGKDVDEKR
FYKNGDKDFGTNGKDSPFARVPKVQAVQALNKLQNMQDRQKQKIDQQAPIQQQRLVEMLRMCTQ
QELKQVEQQCPGTTAQQRQCKQVVVIALAQCGTMNCVAELAKRIKNQVPKAVVVQALLSLNQLPA
PSDGILDQMEQLCKSEAVNSQPMPKQSCWLTFGALVNEVCQHKTENAQECVGSTANCVQSGFYEK
DQCPTKQKQYKEAMLAVYQNHGCFDKAVALGALGNAGLHNALNDLEKIMKNSQEPRVVRVMAIDA
TRRLGVEMPRKVRNVLLPVFLNTRREQPQLRTAAAFANLLNAQTEPVVDQLMFAIAKEPNKEVQSYAY
RTMKTAKSQPNERNRIAKHISVLKMANVDEQQLRYSQKQWQFSLYNAKEREAFISLAQAVSSRSA
LPAYAHAQLDSHFNDQYQLNNVKVFMQLQQAQEKWYGNLADRFTTQQQQQTRAQRQNLQYSTKLNI
HENLGIKNRRSLSYDNEMDGS MKIGTYKQNIQPFVMLCLRSNNVDKAFLALDEHIMEEQTRKMLNG
QEKANWALLFGQQQQLNTAILQNLNEKEAKIPTSAGVPLRILNSMPVLANIEGQLKIKTESARLTSPGL
QLDINAKASARMAQIQKMEWTPFAASGVECVRSVQVNLPLQGELNANSQQGSLKAKLPQQKTQLA
NAHSLPITYIAKVDPKTNLVREPRQILTIHDPKLERQQHEVNAVQGGQKNWAMPLHIQGHYHWPTPTS
YQQFAQLLLATENAVQVHYQPGQNSPREAILRITGEHFQKAAQNEQPGTQLRNFFAEFTSPYSCNV
EQCQQDGDCEQCSKFEQQQKEDAEVEGQQVRQQQLHSFINAYKPRQQYKNSLKLVAQTNGGAK
DAKAQCDIQTQCDEQMNYCKVQVQAERPPIGSENEKWTFFNAQAHVLPETAASVEQLQRKPQQFK
QFIAQAKAQWGGQGSQQHVNVRLGTQAHTQHVRALQQRQDKRLISAQQQRNFQRRSAFLNK
LQIEAAYS LKAYAQNAINRVLEV LKSTYFWNTKSQLLSGQNAQKNGKMAAIIVIDSMTQKNANVTIRTPA
QKVLMMQIELPSKVRPFSLIRHQSAQQIDSLVQLFHHFSIKNRPQCTVDGRSVDTFDDFIYTAPLSSNC
HSVLAKDCGSEEPQFVVLMMKKVPQQQNNNNKSASTAEKMLVVITPKQQIECQPKTNTKGPQQRRLIC
KVNGQQVAGDEQCQGTSCVQFNPPQSDVTIRTEDIAVRFNNGRKAWLKVSRAFKNTQCGLCGHYD
DSDDQEEELRMANGKFASNLAEFHRSYALKDSELCSQQALDNFYSEQNGKQDSFLEQYNQLDNDSE
QNNEAEGEALES DWQKEQFDGKLDSS EEHDEEDIGNNLSDNISEPDGPSQPPPPAPKCCCCQEGE
DEQCQQKCHTSKQCDVVC KGGRC AAEQPIWKATWPTWTKIQEQNHAVCFSTQPVKQCPSGAYS
SSVSSDYGSDEEEDSNDDEQQNEQKSKRVSFVCMERADPETRLLRLLRQGQRVVNVSGHGASFT
ETVNEPQQCWHSRF

>*H. glycines* 13367

MRLLPFLT VFLVLCGHSSEYDDKFDDENILWSDQRVGAFDKRNTKPIKSFQPGFVFQFRYDSQITSGYA
PWVSVQRAMGRFQALVSFEMVQPG EAKMQLSECLVGHLNGHPPSDVAWNPAHLLPVGEFESVTQF
SQVHLERPIRFTFSREGVHNVFFEPMEPPWSRNLK KAILSTAQFRTEKEEEGKEKEEDKEKMRLQKQ
PSEWEKEEEEKRRMEWQWKRSYTVHETSIEGTCDADYTVVPLSHEQISVTKSINYERCTRLADIVYG
MQADEEAKQRGEQRQKWNKAQKREEEETDASGASPMIARRFLDRSTTVRFLFSANAQNSSNSLG
LDLNWPFAHSVEMVSQYALKGSEAEQSQAMQTVAVAKIKHIGTKLKQNNKL RQHGEEGEGEKLGM
KEQLMYNGEMEQQEKCFMFGEQSETLRCYRRAKAFTNDEL RGGPKPSAGASRMSSSESPFPLN
FPNSTAKIRLAEY CIRQVVQLQND ETVQAMSRTLMDEFVQILRFSTEKELQQLKQQLDDHPIASQLFA
DSLAAAGTRNALKMLLTRIEQGEFTL KAAQMLKTF SRGNGLPAPSDRIADALLALCKSGRVQDSELL
RPMWSL SFGAIVGELMQLKRLLRISPVGTPLGEKIEQCPDHKMELYTKALLEQFTSIADDNLVEKTIALK
AIGNAALDEVPKLLVDLINPNGTEEPKAEDQKRAMGEILRQNAIEALRRIPSQSVHQLLLPVFTNRQEW
PQTRISALDVLMQSQLSPSNSVIDQIAFAMAKENC SLLAYAHSM LKALSQSPAKKHQKIAQRIKEALL
MAGVDTLELEQRIRRN DIVGISRFSQTNALFSEEEGQ GILLQASRVFTDKTSALPNHAFVLSLSSLFNGE
MSEENLKM TISQRNIERLFT EAGIIELLRNEQETEEQSNANEQKKDRQGQKEGQQQNGTQQKPFSTD
EESSSSSLLSSAADQLRSLWRRLSIVGRSDEEQKDELKHQPM LLLALRVNNIDTVVLPFGTDGKKRPN
SLISTMKMLTVGIGSIGRLTGIGQQKRSIDASGILFWHERAARVPMSTGQQLHIQQKSFAFGTIKAKA
QRLGLSQIGLQLHPSATFGRIQKMKAWASPVHSLGVKHIGAIEANLPLYAQLNTLSQQKQVTLNVQVP
KTPTRLFSLHTLPMFTVGTENVQNARPQQRSAFLHFGEQCKAVQNNQQLLSMAMAFDKHSDTEQSLL

QLPFHVRGILHRPRRMDKLKEWVKVLLSTENHFQLSLVSTSSLSPEIAFSIGWDLSRPQPTISEALSH
 IGFKWPKAFTSTASDIERKDNENEETEEDDGQWGTEATAASSSEEEKTTQRTKNRQIGEYLKKLGA
 EKTHKMLKFAMGTIGNDENRRAELTLNLTCDGRLRFCALRAKTEQSVVGSQRPQWKADLRVQFLF
 PEQAPDLEQVGQHNRTSESEQQKMFMRTEFEWKDGLPKQTINVSIIIGEPSEEMLQLAHQFAEQS
 FSRRNGTDHKKHLPITAGHPFVNKYKMAADWTFRMDTQNAFLRGLAVKSAHFFNAASKIILSKNG
 EDSEKSDRKMGRKERTDRGRTLATLSIDPLTHRLDASVRTSAQLTFFRMVVPASPLRPFPLVRPSHR
 PFSRASSASLPQFFHNIAHPLRSRAVCRVDSPARQLRTFDGAQLDVHFSPCWTVLAKDCSAEGRAQF
 AVLVKAKANSEQMMLKMISHEIGILVVEPHAQQPDSKLQVQISGKDQKRTVLSDTPLSDIPNGIQFDR
 SLPTKKEALIVEWTTSNANSEGSVGATGAISVHFDGQKVQILLDGQQNRHHSCGLCGDWNGNAGDE
 FRMPNGGQTENAKEFHQAYSLVDGEGEEDGECEDSAILDKFHSEKTEENETNEPKQQKDSEEKPKM
 EQIAAVRKTIVIEHPHQLCFSVHPTKQCPSGTFAIDTDKKEKRREGKGKKEGKEEKEKEEKPQRKMP
 FACVERHSTEAKRLQSQTRRGEVVQMGDGLPISFHEMVIAEQCVKVQKGNNA

>*H. glycines* 23424

MRTFALLALVAAILCLSAGRRTFPNTSPRVGFLNFGTRNSPNSAIKPFQQKNTEYTFIYNAQIASGLQS
 EENAANPQQKAVTRIQCLAKIQFASERHGQLQLEECQAGQQNEQMSEAQEVQPMELFEQTPIPSQT
 KLQLHKPCQFTSADGVIERVQCQEDETETWTKNFKKSVLNIIQLNLKRNNAAQLSRNDERQQNTAEKE
 SWQTKQGSFTIQEITLEGACQTTYINRAWSGQQFNVTKTINFKKCQKIADVANGFQTDPPQAQCA
 QCQQYWAQQQSGQNLQQSDSETDNAQNGEQNPCAKCDPKEVKEHELDRTVIRCQIQGEPTKQYT
 LHSCMRSQYLYRNSKSAVGNYGAAQMOTIVAATLHAMDSKHTYQQIPAMSASTEETLMYSNSKEVDE
 KRFRNGDDDFGTDGKGSPPFAQVPKVQQAQALIKLVQNVQDKLKGIDQQAPVQQQRLVEMLRMC
 TQQELNQVERQCPGNTGQQKEQCKQMFVNGLAECGTKNCAELAKRIKSKQVSQANAVKALLALTQ
 LPAPSDNILEEMEQLCSSEPVESQPTTKQSCWLTFGALVNEVCQHKTEKNAEECAGETANCAQSGFS
 KKGQCPMDKKQKYKEAMLA VYQNSGCIYDKIVALSGNAGMDNSLKELENIIDKPREPRVVRVMAIH
 ATRRLQVTEPWEVQSVLLPVFLNTREQPHVRIAFAFANLMNENGQPEPHVVDQLMYTIANEPNKEVQA
 YAYRMMKAMAKSQKPNEKQIAKHRSALKMANVDEQQLRSSGKWEFSLFNAEEREQVGVSLAQAVS
 ARSALPAYAHAQLDSHFNDQYQLGNVKVFMQQEVDKQWYGNMADRFIKQQQQQQSQHTRGQRRS
 QPNSEKLNDIYKALGIKSRSSFFGTDEDDSNIGKYQQNVQPFAMLALRSNNVDKAFIALDEHLMGEQ
 TRRILNGQEKADWALLLGGQKQLNMAILQNLNEKEAKIPTSVGVQLRILNSMPILANIEGQMKIKAESA
 SLMSPLGIQFDVNTLASARLSQIQKMELWTFPASSGVESVRSQVNLPLQAELENDSDQGISLKANLP
 KQKTQLASVHSLPITFTAKVDQNTKLVREPRQIVTIHDPKLERQQYELNSVQGGQKNLAMPLHFRGHYH
 WPAEPTSYYQFAQLLLATENAVNIHYQPSRESPREAIKRVGTGQHFEQLAQSEQPDALRNFFSEKFA
 NAYSCDDEQCQQRGDCEQCEQIEQQQREDAQELEDQQMRQQSLHSFISAYKPRQQYKHSKLMVAQ
 TNGGAKDAKAQGDIAICDEQTAYCKVLVVERPPIGNENQKWMLNAQAHVLPETAASVEQLQTQP
 KQLKQFIAQAKAQWGQQGQQTQQVNIRVHGAQAHTQQVRAVLQQQGSQQLISAQHLRNMQRRSF
 LNKQFIEAIYSLKANANINRVLELFKSAYFWNTKSQLMSGQSAQKDGKIAATILIDPVTQKSANITIRT
 PAQQLRMEQIEVPYKVRPFALIRQSAQQTHSLAQLFHQFSIKYRPQCTVDGRKVNTFDDLIYTAPLS
 SNCYSVVAKDCGSDQPQFVVLMMKVAQQQQQQTNKRAQNAAKVLKVITPEQQIECHPKTNAAGIPT
 QCLICKVNGQQVTGDEQCQDSSSCVQFNNKQQDDVTIQTEAIAVRFNNGRKAWLKVSQAYKNTQCGLC
 GHYDDSDDQGEELRMANGQRASDLAEFHRSYALEDAEECPQQVMNNFYSEQQQQNSEQDTFYGQ
 FNQFEDEDEADDEQQQQSEKDEWQNEQFNGQFEEQNDGQVTDGGNPFEGIFNDQQQQQKCCCC
 CDDDEDEEGTNERANADKTAKTAGNAQRSATKTTKSNAESSDDEQCKGICNDGDEACAEQQQWQK
 AVWPILRTKTHEQNHAVCFSTRPVKHCPPDAYPYNPMNGYEQQQQPKTKKVSFVCMERAEPETRRL
 LRSLRRGQRVVDVNARGASFTENVSEPHQCWRRRY

>*H. glycines* 6510

MRTAALLVLVTAFLSVGTSGIRHSNRMSFGGFLSSANANFGTARVGPTSGDNDAAELLQPFQQKDTE
 YSFIYNAQIASGLQSAEQSAANPQQKAVTRIQCLANIHFASEHAQLRLEQCRVGQQNGKMSQPQEV
 QPMEAFEETNIAEKEKQQLLTPCQFSFADGVIERVQCPQDAEWSKNVKKAVLNIIQLNLKQNNAAQL
 HKSSASHAEQQQQEATGQGSFKIPEITVEGACQTTYINKAQRIHQQQQQQFNVTKTINFKKCQKI
 ADVANGFQADQPQAQCAQCQQYWAQQQSDQQNSPTDQKTNWQMGESDQHPCAKCDPKEVK
 ENEMDRSTVIRCRIAGDPADQYALRGCELRSQYIYRSWKSAAEAGTNGAVMQTVVVANLEAVAVPKPS
 QQIPAINGASEHDTLMYSNDKAVDEKRFFANGDNEFGTDGKGSPPFARVPKVQQAQALLKLEQYTEN
 NHKGIDQQTPVQQQRLVEMLRMCTEKELKQVEEHCAANTAAQQQKQKRGEKCHQMMANALAQCGTR
 NCVAELARKIKNGQVVQSIQVQSLALNQLPSPSDTILDEVQQLCQSQTVAGSQALKQSCWLAFGALV

NEVCQHTTAQNAKECAGEGGAGATANCVQSGFVDKDKCPADKKQKYKETLFAMYKQQQNEQQQTQ
 CIYDKAVALGALGNAGLDIALADLERIIKDTREARVVRMMADATRRMRTQMPHSVRAMLLPVFLNTGE
 QPQVRITAFANILGAQPTAEVVDQLVYAIGKEPNREVQAYCYRTMKTMANSSQKPDQIAKHVKSSLK
 QLANAVDEQQLRTSGKWQFEQLYNNAQREGVFMFVHAVSSRSALPAYAHVQLDSHFNDHYTYNKL
 NIYMVQQEAEQWYERAMNSFGNSKRKQKQQQHSGSKQLSSIYEQLGIKSHRSDFGTVPFTAADNDNEA
 YQQHQQQQQQNNMRPFAMLGRLSNNVDKAFIALDEHTIGEQTTRMLNGEEMKPNWAAILFGHGAQ
 QQHLDMAIVQNMGEKEAKIATSAGVPLRILSTTPMMANVEGRVGAKAAPKMGPSAAGHGLSAQFEF
 SAMSARMAQIQKMEITPFAASGVESIRSLHINLPLQGELSLVNSEQQQGISIKVKLPQRGTQLASAHS
 LPFTFIKAEPSAQFVREPRQIIAVHDPKLDQQQREVNAVLGQKALALPLQIRGHYHVPAPESNLLQFA
 QLLMATENAVHVQYQPNESPRELTIRASGEHFRKAAQMQLPDTPLHNFYAQQFHSSSSCDSDDS
 VNAANCEENDATEQQQGENGGEQYISNDNQQQQLNSFISAYKPRQQYAHALRLSAQTSGGARDATA
 HANIQAQCDERMRYCTVQADANKNQWTLQAQAQVMPDTPVPSVEQLQQPTMLKQFIALAKAQWG
 DQQDKQHMMNVIRGTQAHTEQTRAVMAQIQQQHEGKRHGFSAQQQRKMLQRS AFLNQFHVEATYA
 LKAYAQNMMNRGIELLSVYFWNTKSQLLGGPNAPKNGKIAATISIDPLQPSNANVTIRTPAQQQLHMQ
 QIQMPAAVGRPFPLIRQQSAQQTSLAQLFYKQFSLKSRAQCIVDGRNVDTFDGFAPLSSNCFVS
 LAKHCSSPMEEQHQQQPQFVVMLKKVPGKQNANKSAQKADKVLKVITAEQQIECQPKTAQGGKKGQ
 MICKVNGQPISGDEQCQDELQHQQCQVQFNDPRQTDVTIRTGPIAVRFNGRKAWLKVSQAYKNAQCG
 LCGHYDDADEQEQLRMANGQIASNLVEFHRSYALKDAEQCTQQALNSFYNEEQQQQNVAYGDE
 QQTDNFYGGYNQWKITNENGEDEEMEYSDREENNNDSSGEAEEEEEGQKEQFDGQIEDISDEN
 WAGEEEDIEMSGEEQPEKCCCCCAKEKDEARKQKCRKCCQSDCRRKCCQKQCQQQQQQRKM
 GKQKQKQSSALEGQWEAVMPTMNTKVYEQNHALCFSTEPVKQCPRGTTTNGGEEADQQLSGEK
 PKRVHFCMDRAESEARRLMRQLRRGQRSVKVSSAQGESFSEMMDEPQQCWRRRN

>*H. schachtii* 007

MRTFALLALVAAILCLSAGRRTFPNTSPRVGFLNFGTRNSPNSAIKPFQKQNTYTFIYNAQIASGLQS
 EENAANPQQKAVTRIQCLAKIQFASERHGQLQLEECQAGQQNEQMSEAQEVQPMELFEQTPIPSQT
 KLQLHKPCQFTSADGVIERVQCQEDETEWTKNFKKSVLNIIQLNLKRNNAAQLSRNDRQQNTAE
 AWQTKQKGSFTIQEITLEGACQTTYINRAWSGQQFNVTKTINFKKCQKIADVANGFQTDSPQAQCA
 QCQQYWAQQQSGQNLQQSDSETDNAQNGEQNPCAKCDPKEVKEHELDSTVIRCQIQGEPTKQYT
 LHSCEMRSQYLYRNSKSAVGNYGAAMQTIVAATLHAMDSKHTYQQIPAMSASTEETLMYSNSKEVDE
 KRFRYRNGDDDFGTDGKGSFPAQVPKVQQAQALIKLVQNLQDKLKGIDQQAPVQQQRLVEMLRMCT
 QQELNQVERQCPGNTQQKEQCKQMFVNGLAECGTKNCAELAKRIKQVVSQANAVKAMLALTQL
 PAPSDNILEEMEQLCSSEPVENQPTTKQSCWLTFGALVNEVCQHKTENAECAGETANCVQSGFS
 KKGQCPMDKKQKYKEAMLAVYQNSGCIYDKMVALSALGNAGMDNSLKELEKIIKDPREPRVVRVMAI
 HATRRLQVNAPWEVQSVLMPVFLNTREQPHVRIAFAFANLMNENGQPEPHVVDQLMYTIANEPNKEV
 QAYAYRTMKAMAMSKPNEKQIAKHRSALKMANVDEQQLRSSGKWEFSLFNAEEREGVVFVSLAQA
 VSARSALPAYAHAQLDSHFNDQYQLGNVKVFMQQEVEKQWYGNMANRFIKQQQQSQHTRGQRRSQ
 PNSEKLNDIYKALGIKSRSSFFGTDEDDSNIGKYQQNVQPFAMLALRSNNVDKAFIALDEHLMEEQT
 RRILNGQEKADWALLLGGKQLNMAILQNLNEKEAKIPTSVGVQLRILNSMPILANIEGQMKAESAS
 LMSPLGIQFDVNALASARLSQIQKMEICTPFASNGVETVRSQVNLPLQAE LNANSDDQGISLKANLPK
 QKTQLASVHSLPITFTAKVDQNTKLVREPRQIVTIHDPKLERQQYELNSVQGGQKNLAMPLHFRGHYH
 WPAEPTSYQQFAQLLLATENAVYIHYQPSRESPREAFRVTGQHFEQLAQSEQPDALRNFFSEKFAN
 AYSCDDEQCQQRGNCEQCEIEQQQREDAQELEDQQMRQQSLHSFISAYKPRQQYKHS LKMVAQT
 NGGAKDAKAQGGDIQAMCDEQTAYCKVLVKVERPPIGNENQKWLNAQAHVLMPEAASVEQLQTQP
 KQLKQFIAQAKAQWGQQGQQTTQQVNIRVHGAQAHTQQVRAVLQQQGNQLISAQHLRNMQRSSA
 FLNKFQIEAIYSLKANAAQNAINRVLELFKSAYFWNTKSQ LMSGQSAHKDGKIAATILIDPVTQRSANITIR
 TPAQQLRMEQIEVPYKVRPFALIRQRSAQQTSLAQLFHQFSINYRPQCTVDGRKVNTFDDLIYTAPL
 SSNCYSVVAKDCGSDQPQFVLMKKVAQQQQQTNKRAQNAAKVLKVITPEQQIECHPKTNAAGIPTQ
 LICKVNGQEVTGDEQCQDSSSCVQFNKQDDVTIQTEIAVRFNGRKAWLKVSQAYKNTQCGLCG
 HYDDSDDDQGEELRMANGQRASDLAEFHRSYALEDAEECPQQVMNNFYSEQQQQNSEQDTFYGGF
 NQFEDEDEADDEQQQQSEKDEWQNEQFNGQFEEQNDGQVTDGGNPFEGIFNDQQQQQKCCCCC
 DDEDEEGTNEQSQRQNCQNGGECPEKCNKNDQKQCRKSSPPQPPQQQNCKGSDDEQCKGIC
 NDGDEACAEQQQQQQWKA VWPILRTKTHEQNHAVCFSTRPVKHCPPDAYPYNPMNGYEQQQPKT
 KKVSVFCMERAEPETRLLRLLRRGQRVVDVNARGASFTENVSEPHQCWRRRY

>*H. schachtii* 19700

MRLLSLFLFTFSAFLTFFVLCGHLSEYDDKFDEENIWWSDQRFQAFDKRKTPIKSFQPGFVFQFRYD
 SQTTSQYAPWVSVQRAMGRFQALVSFEMVQPGQAKMQLSECLVGHNLNGHPPSDVAWNPAHLLPVG
 EFESVTQFSQVHLERPIRFTFSREGVHNHVFEPMESAWSRNKAILSTAQFRTEKEEEGKEKEEDEKE
 KMRQQTTPREGEKEEEEKRRREWQRSYTVHETSIEGTCADYTVVPLSHEQISVTKSINYERCTRLA
 DIVYGMQADEEAKQRGEQQKKWNKAQKRDEEEPTDASGASPMIARRFLDRSTTVRFLFSANAQNSS
 NSLGLDLNWPFAHSVEMVSQYVLKGSEAEQSQAMQTVAVAKIKHIGTKLKQHNKRRQHGEEGEGEK
 LGMDKEQLMYNGEMEQQKCFMFQDEQSEILRCYRRAKFTNDELRGGPKPSAGASRMSSSESP
 FPLNFPNSTAKIRLAEYCIRQIVQLQNDQTVQAMSRTLMDEFVQILRFSTEKELQQLKQQLDDHPIAS
 QLFADSLAVAGTRNTLKMILLTRIEQGEFTLKAQMLKTFSRGNGLPAPSDRIADALLALCKSGRVQD
 SELLRPMWSLWFGAIVGELMQLKRLLRISPVGTPLGEKIEQCPDHKMELYTKALLEQFTSIADDNLVEK
 TIALKAIGNAALDEVPKLLADLINPNGTEEPKAEDQKRAMDEVLRQNAIEALRRIPSQSVDHQLLLPVFTN
 RQEWPTTRISALDVLMSQSLSPSNSVIDQIAFAMAKENCSSQLLAYAHSMKALSQSPAKRHQKIAQRI
 KEALLMAGVDLTLELEQRIRRNNDIVGISRFSQTNALFSEEEGQGILLQASRVFTDKTSALPNHAFVSLSS
 LFNGEISEENLKMNTISQRNIERLFTAGIIELLRNEQETEEQSNANEQKKDRQGQEQKQNGTQQKP
 LSAAEESSSSSSLFSSAADQLRSLWRRLSIVGRSDEEQKDELKHQPMILALRVNNIDTVVLPFGTDGK
 KRPNSLITTMKKMLTVGIGSIGRLTGIGQQKRSIDASGILFWHERAARVPMSTGQQLHIQQKSFAFGTIK
 AKAQRLGHSQIGLQLHPSATFGRIQKMKAWASPVHSLGVKHIGAIANLPLYAQLNTLSQQKQVTLNV
 QVPKTPTRLFFLHTLPMTFVGTENVQNARPPQQRSAFLHFGEQCKAVQNQQLLSMAMAFDKHSDTEQ
 SLLQLPFHARGILHRPRRMDKLKEWVKVLLSTENHFQLSLMPTSSISPHEIAFSIGWDLSPQPTISEA
 LSHIGFKEWPKAFTSTASDIERKNNENEETEEDDGQWGTEATAASSSEEEKTTQRTKNRQIGEYLLKL
 GAEKAYKHMMLKFEMGTIGNDENRRAELTLNTLDCSRLRFCALRAKTEQSVVGDQRPQWTADLRVQF
 LFPEQAPDLEQVQGHNRTKPEQQQKMFMRTEFEWKGDLGPKQTINVSIGEPSEEMLQLAHQFAE
 QSFSSRRNGTDHGHQHLPIAGHPFVNKYKMAADWTLRTDTQNAFLRGLEAVKSAHFFNAASKVILSK
 NGEDSEKSDRKMGRKERTDRGRTLATLSIDPLTHRLDASVRTSAQLTFFRTVPVASPLRPFLVRPS
 HRPFSRASSASLPQFFHNIAHPLRSRAVCRVDSAPARQLRTFDGAQLDVHFSPCWTVLAKDCSAEGR
 QFAVLLKTKANSEQMMLKVISPEIGILVVEPHAKQPNSKLQVQISGKDQKQKRVSSDTPLSDIPNGIQFD
 RSLPTKKEALIVEWTTSNANSEGSVGATGAISVHFDGQKVQILLDGGQNRHHSCGLCGDWNGNAGD
 EFRMPNGGQTESAKEFHRAVSLVDGEGEEDGECEDSAILDKFHSEKAEENETNEPKQKQSSEKPK
 TERIAAVRKTIVIEHPHQLCFVHPTKQCPSTGFAIGTDKKEKRRREGKEKKEGKEEEEKEKEEPEKQR
 KMPFACVERHSTEAKRLQSQTRRGEVVQMGDGLPISFHEIVIEAEQCVKVQKGN

>*M.hapla* 3631

MRTIVFIILLGLAGRSSGLLDQSSDDYYSTSKQSPKFNFFYREYNTEAEQIKTPSFQEDNEYVYNYD
 AQVENGILLSSVDLTETKEESQQQAIRIQSIKMYFSSERRGQLCMTQLKFGQFNDLKQMESQQCC
 DPMQMFEPKIPQETRQKLELCCQFDYVDGMVVRIQFDKQDVSWSENIKRGVLNMLQMNLRQNNNGQ
 MDNEGGNKNVDRLAHGFTVPEITVEGECQTSYINKGSMQCQENQQPCSFNITKSINFKKCSKIADLT
 NGFQTEQPQLKCSQCPVNEEVKEGSEEQDGTGENQCACDCTRELNNMLDRQTIMRAMVKCGTE
 QKQQNNNCPNCCQLDNSEMSNYIYKNTKIESEGYPYGSALRTKTCARLSLATVQTVQKPAPLAATSEN
 DETLLYSNEWGMACRRFMYGDEEFPDANSSPYKMPNVEQATQALQAIQVQSTSDHHHGKTKHTINL
 NRLVDKLRMCRYEDLKQIEQNAKNQNGEKSQKTATDVFDVSLATCGTRNCVKLLAEIENEEVDENKA
 AFAINQLKLNLPISNSILKQVKQLCQNVNVQNKPTVRQAALWTFGALVNELCQHKSRKMMETSQDISK
 SIQEQCTDDKKEEYKRTLMEQYNQATTIYDKVLALSALGNAGIDTTTPYIDQIINDKSQHFIRVKAIDAL
 RRLYTKQPRTIQILLPIFLNNREDTSVRSFGLSTLLRTQPEPSVVDQIFYAMRQEPNKRKAYTYQTM
 KLIAESKNPADQMQSKHVKHAMQTVNINKEDLKFNMWQIPAYCSKHQEGFLFFKLGAQQAQRGWMPN
 FSYLNADSYLKEQSNIDLISIVMQGSEPYSYEETRNYSSTTQNNQHNDNSKTLKQINKMLRIKSRNP
 TSMTDDEQSSHRQYSLRNQQDPSSILCVRMGDVHACYGLNQNWDMSSNTKNYPTAPPMLEEIL
 QSIRRGEMPKFNSMTSKFGLQQLQSFQNHFTSAIYNEREAIPTSIGLPVRVNSVPMLMSVEGSTKM
 QNREKNEETKVQIDARIMAGISHVQRIGSWTPFLITGASSIRTVEVNLVPQAELSGNTETSNSMQSFN
 VKVKMPTETTRLLGVHSHQYTYKTGINSTTNMPMGRDYRTIRNPKLDRLQYQMDKVIGSKNVGVPVH
 VYAHYHWPSSADNIVQIIKMLTSTENAMHVTYQPGKGTAREVMFSMNTENFQPVKSFQPKMDNFFER
 SAMSEEEEMPENQQQYYANQQQKMQKFLSKYQKAQKHQKHYSLEMSVKTFFGAKPFCQSAQLQ
 AQCNQKMQVCQVKMEAQRSAMNEERDKWTKMGHTEMVMPGGVDSLEEDNSGSDQFLCFANCQW
 GANTQKQHVNMRIVGEQASDRTTGWYMFQRRQEEGEQQRGELQGDEQRMPEQQDGENQNSQLN
 MRLQRRAAFINLYNCRMQYTNLSPSTRNVFERAVELVSKYFWNSKTELMSESNMNPANGKVYCTLS
 IDSLTQNHANITLMTPEQIVHLQQIELPIKLRPFLMRRNARRIQSFDQFYQQQEVKNRAECSVDGRTK
 LKILTHGQTVQCQHETSNTHKVRCMLNRRYMNEDEENNVMYNNNAKKSDDLTVHMNEVSVRFNGYKA

WIKLSGAYRNTQCGLCGHYDDATNDENEMIMANNEFTSNLAQFHRSYSLSTSHDSDDQMCNEQELN
SFYEENKNKFGSSDNEMEEEMDEENEKENKNNNRMSKYNEWEDMFDTGDEENEDEAYFSDNY
EGQTNMRNMNGNKRIMAPHGMTRMIEMENEICFSKQPIKQCPSGSYPTGWDNTGNNENQENQES
TKNENFTTKNVPFICLQRSLETRQLLRQYKQNDYQIHMDAYKQHSFVEKVHEPKECRR

>*M. incognita* 11129

MKALVFILLLLSLSGSSSIPIDDRSSNEYYSMTKSAPFKNFFYREYNTAAEQIKTPNFQENNEYVYSYN
AQVENGMLSSVDSSGTKEESQQQAITRIQSQIRMYFSSERRGQLCMTQLKFGQFNDMRQMKEKQC
YDPMGMFEPLQIPQETRQKLELCCQFDYVDGMVTRIQFEKQDVSWSKNIKRGVLNMLQMNLRHNEE
GQNAMGENKNVDRLAQSFVTPETTVEGECQTIYTINKGSKQCQESQQPCSFNVTKSINFKKCSKIAD
STNGFQTEQPQLKCSQCPMSEEVGRNVEQEGTAENQCADCDCPREMTELMLDRTVIRSMKCGTE
QKQNNACPTCCQLDNSEMTSNYIYKNTKIESEGLYGSALRTKCAKLSLETVQNIQTPAPLAATAESDE
ALLYSNEWGMNLRRFYMFGDEAFPDVNSSPYKMPKVEEATQALRSIVQSTMDHHHGIKTKHTIDLSR
LVNKMRCMSYEELKQIEQNAKNQNEEKQKQVAMDVFDCLATSGTRNTVKLLAEIENEEIDENKAAY
AINQLKNLPIPSNSILKQVKQLAQNVNVQNKPMVRQAAWLTFGALVNELCQHKGRNMMESSQDVSKS
VQQQCTDDKKEEYKRTLLEQYNQATTFYDKFLALSALGNAGIDTTTPYIDQIINDKSQDYLRVKAIDAL
RRLFTKQPRTIQQILLPIFLNNQEEAYVRSSALTTLLRTQPEPSVIDQVVYSMRQEPNKRVKASTYQTM
KMAAESRNPVDQQVSKHIKNALKSVNIDNEEMKNYKLWQIPAFCSKQQEGVFFTIGARKAQRWMPN
FSYLKTDTYLNEKSNLDSLTVYVLQDSESYDSYGETRNYTASTNNNDQHSSENSKTLKQINQMLQIKPR
NAFKYGYGDEQQSSSNPLKYGYPYGQQSSHYSLRNQNPSSVLCVRIGDVTDFHALNQNWDMIM
SPQNWDMSSQTSIPTTSPMLEEILQSMQKGEMLFNRLISKFGLQQLQSVQNHTYSIAFNEREALIPT
SVGLPIKIVNSVPMLMNVEASTQLQNRQSQAQKMQIDARIMAGISHVQRIGSWTPFLIIGVSNTRSVEV
NLPIQAELSANTEASRQSLTVKVKMPRETTLLGAHSHQYTYKAGIDQNTYLPMEYRTISNPKLDR
LQYQFNKVGSKNVGVPVHVYAHYHWPSSNAANIVQILKMLTSTENAMHVTYQPGQGTASEVFNMT
ENFQPVKAFQSKMDSFFERSATSNEEGNPQEGIMPEDQQQYANQQQKLQTLFSNYQKSQKHQKH
YKHSMELSAKTIGASKQFKASVQLQAQCNDKMQVCQLRMGAERSKMNEERGMMWAMKGQVEMVM
PGGAEEENSNSYASDQFLSFANLHWGDNTRKQQVNMRIVGEQAPERTSGWYMFQNNQEGENQEG
EQQNEENQNSDQNSPPFNTKLQRRATAFINLYNCRMEYTNLTGTRNVFERAVELAKSKYFWNSKTEI
KSESSLNPENGKVYCTLSIDSLTQNHANITLMTPEQIVHLQQIEMPFKLRPFSLRNRNVQSIQSFDQFY
SQQEIKNRAECSVDGRTVRTFDGKKYNAPLSSNCYSVLAKDCGRMTDVPDFVMMMDTGKGSQKS
QKAN

>*M. incognita* 17578

MPCSFKVTKSINFKQCSRVDISAGFQTEQPQARCAQCKMDWYKQKERDQQQTGTAEHPCAKCDP
KRVKEETLDRQTVMRAMLKCGSEQQMIEGKQQEIQNSCQLDSSEMRSIQAKTNRYPHRHWTTSLSLW
YTQSRLTQTLHASICLEMKNSQIQIHHSPQYQKWNKLSAQSSQDQVPTQTLDNDESLVYTVQTYT
DFARFYMFGDEEFDPDTPNSPFSSVPKVEQAGQALRALVQSTLDHHHGIGTKHTIQLSRLVEMLRMCS
YEDLKQIEQNSQQYHQNNQQQTAKQILVDAMASGTRNCVVRMAEIINNQEVSPPHKAAYTISQLTGLQS
PSNLIVKEVQKLCQSEQVKNRPIVQQSACLTGALVNELCSQRRSEKTPDLTHSIPEQCTEDKKEQYK
QVLVEQYQKATTLQEKLNALASLGNAGIDTSTPQLEQIHKDRSEHLSRAKAIDALRLTTKEPMTLQH
MLLPIYLNQEDPNVRIVALQILMRTQPEPSVIDQIYTMSQEPNKRVKAYTYQTMKMAKSKNPADRQ
MSKHVKNAMQSANVDEQQLWSYAKWQIPVYSPEQEEGIFLTATEYSQRSWMPSLSYMKIDSYFKD
QSNVNDLVFYLMKNQPLYSQDYTKSRNRRTVPIRIDRRKFEQENNENTQQLKNMDSQLGIKSRHGGY
SFNDRYSGSGQQQQQHIHNNNNINVFCTRIGDVDHSCFEIQNSYLNQTTQRQQQQQTSSMPTLPPL
FTELIQSIQNRKPSLAKMFSSELKQLQQQKVHQYTIASILNDKEAVVPTCVGLPLGIANFVPTILNVEGN
VNLQKNTMGVKVQINGRIMGGMAHIQKMRWCWTPFLVSGIKSIRSIEINIPTQLDINVQTANNGKQSLT
MSVKLPKEKTRLVGAHSHPYVYADVDPKTGSVIEPRKVRTITNPKLDRQYQMNTVIGGKNLGIPVN
VHAHYHWPSSSANIEQICKMLMATENAIHVTYQPRGEESAKEVVFQMNLESFQPTNNYQLKMNDLYL
NKFKNIEEEEEGGNNYEEENEQNSKRQKMSKFLTNYSQQQKYYKHNLMSAKTVGGPEEFKF
QTKLEAKCEEQMTCQLEIQAERSPINEEQKWTLKAQAEMVMPENVENNEEFNEEEEEENKIKNNLF
SCKANCYWGSDGGNKQTINVRINGEQSTTRRKEWNSNNQQRMNKEFEIKLRRTAFINKFNCEAKY
TNLKPTRNAFERALELVKSKYFWNTKSQLIDQSTEGNPANGEAYLNIVIDPITQKHANMTVKTPPEQLIR
FQQIELPIQMRPFLLKQQNVYQNIQSFGQFLTRQEVSRAECSVDGQTVETFDEREYRSPSTNICY
SVLAKDCGADYMGVPRFVVLMAIERGNKYSQKKLKIISPQMVVEQCRESLNNHNSPIKCKINGKEEN
NYQQQQQLIQYNNQMNDLIINWDKLAIRFNGFKAWIKISSDYRNTQCGLCGHYDDASNDENELLMAN

NEIAPSLTQFHRSYSLSNSYDNDDKLCNAQELDKFYEENS NKFGYLNEMEEEEERE EGKENKVYQNW
DEEGFDGNFDRNEENKRFYNEEWY GQNEREEINYQQKNQLAMTKSGNRNLKGRFVPILIRTKIMEM
ENDICFSKQPIKQCPLGSYPTGWFEERENNSKHFTTKNIPFICLPRSDSETRNLLNQYRQLMKNGNN
QQLELNNYKQHSIVEKVLEPKECRRL

>*M. incognita* 24409

MLRMCSYEDLKQIEQNAQQYHQNQQETAKQILVDAMATSGTRNCVVRVAEMIKNQEVSPHKAAYTIS
QLTGLPSPSNSIVKEVQKLCQSEQVQNQPIVKQSACLTFGALVNELCSQRRSEKTPDLTHPIPEQCTD
DKKEQYKQVLVEQYKATTLQEKLNALASLGNAGIDTSTPQLEQIIKDRSEHSLCRAKAIDALRRLTTS
QQMIVQHIILPIYLNQEDPNVRIVALQILMRTQPEPSVIDQIIYTMSQEPNKRVKAYTYQTMKMVAKSK
NPADRQMAKHVKNAMQSANVDEQQLWSYGKWQIPVYSPEQEEGIFLTATEYSQRSWMPSLSYMKI
DSYFKDQSNVNDLVFYL MNQRLYSQDYTTKSRNRRSAAIRGFERREFEQNNENTQQLKDMHSQL
GIKSRHGGYSFNDRYSRSGQQQQQQQSYNNINVFCTRIGDVDHSCYEIQNSYLNQTQQQQQSQS
SSMPTLPPMFTELIQSIQNRKPSLAKMFSELEQLQQQKVHQYTIASILNDKEAVIPTCVGLPLGIVNSV
PTILNVEENVNLQKNTEMGVKVQINGRIMGGMAHIQKMRCWTPFLVSGVKSIRSIEINVPIKLDINAQTT
TTNNGKQSLIMSVKLPNEKTRLVGAHSHPFVYTADVDPKTNLVIEPRQVVRTIRNPKLDRLQYQMNTLIG
GKNLGIPVNVHAHYHWPSSSVNFEQICKMLMATENAIHVITYQPRGEESASQVFRMNLESFQPTNDY
QLKMNDFYLNKFKNIENEEEEEEEEGGNNNYEEEEENKQNSKRQQKMSKFLTNYSKQQKYKHNL
MSAQTVGGLKEFKFQTKLEAKCEEMMQTCQLEIQAE RSPINNEQQKWTLKAQAEMVMPENVENNEE
FNEEEENLYSCKANCYWGNDDDRSKQTINVRINGEQSMFRGGMEQRMNKKLRRTAFINKFNCEA
KYTNLKPTRNVFERVLELVKSKYFWNTKSQ LIDSSRGGNPANGEAYLNIVIDPITQKHANMTVKTP
EQ LIRFQQIELPIQMRPFLLRQQNVHQNIQSFGQFLTRQEVESRAECSDGQTVETFDEREYRSPLSTN
CYSVLAKDCGASYMGGVPRFVLMKAIERGNKY SQKKLKLSPQMSVEQCQRETPNNHNSPIKCKING
KEENNQQQQQLIQYNNNQKNDLIINWDKLAVR FNGYKVKWIKISSDYRNTQCGLCGHYDDASNDENEL
LMANNEIASSLTQFHRSYSLSNSYDNDDKLCNAQKLDK FYEENS NKFGYLNEMEEKEEEEEEEENN
KLNQNWDEEGFNGNFDEENKYFYNEEWYGENEEEEWKEEGKENNNNQKMATTKSGNRNLKRRG
GLLPILRTKIMEMENEICFSKQPIKQCPMGTYPTGWFE EEEENKNANKHFTTKNIPFICMPRSDSEARN
LLNQYRQLIGNNQQQQQLELNNYKQHSIVEKVLEPKECRRL

>*M. incognita* 28653

MRINSILLLLAIGLLGPSNFVVGDEYPATSTLKRFFYRHFFMDNAASEQTRAPNFQEGTVYHYSYDAQ
VESGLSTVDES GTTTQTS DNGQQA VTRIQSQVKVYFASQRRALLCMTQNKFGQFNEQRKQTD SNQC
YEPLSIFEPKQIPQEQRQKLELCCQFDYVDGVIVRVQFDKQDEPWSKNMKRGVLNMIQMNLRQNN
QELATEQSRNQEGNIEENESTQMRDQLAHSFTIPEVTVEGECQTTYTINTANPSKHCTQKSSDESN
KQQQMPCSFNVTKSINHKQCSRIGDISAGFQTEQPQARCAQCKMNWYKQKEGDQQQSGTAEHPCA
KCDPKRIKEETLDRQTVMRAMLKCGSEQQMIEGKQQETPNSCQLDSSEMRSMYVYKNTKTESGPY
GSAMRTEVCAKLQLRSAQSIQDQQVPTQTLDNDESLVYTQTFSDYARFYMFGDEEFPDSTSSPFYS
VPKVEQAGQALRALVQSTLDHHHGIDTKHATQLSRLVRMLRMCSYEDLKQIEQNAQQYHQNQQETA
KQILVDAMATSGTRNCVVRVAEMIKNQEVSPHKAAYTISQLTGLPSPSNSIVKEVQKLCQSEQVQNQPI
VKQSACLTFGALVNELCSQKRSEKTPDLTHPIPEQCTDDKKEQYKQVLVEQYKATTLQEKLNALASL
GNAGIDTSTPQLEQIIKDRSEHSLCRAKAIDALRRLTTNQQMIVQHIILPIYLNQEDPNVRIVALQILMRT
QPEPSVIDQIIYTMSQEPNKRVKAYTYQTMKMVAKSKNPADRQMAKHVKNAMQSANVDEQQLWSYG
KWQIPVYSPEQEEGIFLTATEYSQRSWMPSLSYMKIDSYFKDQSNVNDLVFYL MNQRLYSQDYTT
KSRNRRRTAAIRGFERREFEQENNENTQQLKDMHSQLGIKSRHGGYSFNDRYSTSGQQQQQQQSY
NNINVFCTRIGDVDHSCYEIQNSYLNQTQQQQQSQSSMPTLPPMFTELIQSIQNRKPSLAKMFSEL
EQLQQQKVHQYTIASILNDKEAVIPTCVGLPLGIVNSVPTILNVEGNVNLQKNTEMGVKVQINGRIMGG
MAHIQKMRCWTPFLVSGVKSIRSIEINVPIKLDINAQTTTTNNGKQSLIMSVKLPNEKTRLVGAHSHPFV
YTADVDPKTNLVIEPRQVVRTIRNPKLDRLQYQMNTLIGGKNLGIPVNVHAHYHWPSSSVNFEQICKML
MATENAIHVITYQPRGEESASQVFRMNLESFQPTNDYQLKMNDFYLNKFKNIENEEEEEEEEENKQ
NSKRQQKMSKFLTNYSKQQQKYKHNLMSAQTVGGSKEFKFQTKLEAKCEEMMQTCQLEIQAE
RSPINNEQQKWTLKAQAEMVMPENVENNEEFNEEEENLYSCKANCYWGNDDDRDKQTINVRINGE
QQSTMMRGGMEQRMNKKLRRTAFINKFNCEAKYTNLKPTRNVFERVLELVKSKYFWNTKSQ LID
SGRGGNPANGETYLNIVIDPITQKHANMTVKTP EQLIRFQQIELPIQMRPFLLRQQNVHQNIQSFGQF
LARQEVESRAECSDGQTVETFDEREYRSPLSTNCYSVLAKDCGASYMGGVPRFVLMKAIERGNK
YSQKKLKLSPQMSVEQCQRETPNNHNSPIKCKINGKEENNQQQQQLIQYNNNQKNDLIINWDKLAVR

FNGYKVIKISSDYRNTQCGLCGHYDDASNDENELLMANNEIASSLTQFHRSYSLSNSYDNDKLCN
AQKLDKFEENSNNKFGYLNEMEEKEEEREEENNNKLNQNWDEEGFNGNFDEENKYFYNEEWYGE
NEEEWKEEGKENNNNQKMATTKSGNRNLKRRGGLLPILRTKIMEMENEICFSKQPIKQCPMGTYP
TGWEFGRKQKCK

>*M. incognita* 30865

MRINSIFLLLIIGLFGPSNLVVGDEYPATSTLKRFFYRHHFADNSASEQTRAPNFQEGTVYHYSYDAQV
ESGLSTLDES GTTTQTS DNGQQA VTRIQS QVKVYFASQ TRAQLCMTQNRFGQLNEQTKQTDSSNQC
YEPLSTFEPKQIPQEQRQKLELCCQFDYVDGVVVRVQFDKQDEAWSKNMKKGVNLMIQMNLRQNN
QELASEQSRNQDGNIEENEATQMRDQLAHSFTIPEVTVEGECQTTYTINTANPSKNCTQKGYDGSN
KQQQMPCSFVKTSINFKQCSRVDISAGFQTEQPQARCAQCKMDWYKQKERDQQQTGTAEHPCA
KCDPKRVKEETLDRQTVMRAMLKCGSEQQMIEGKQQEIQNSCQLDSSEMRSYVYKNTKTESGPY
GSAMRTEVCAKLQLRSAQSSQDQVPTQTLDNDESLVYTVQTYTDFARFYMFGDEEFDPDTNSPFS
SVPKVEQAGQALRALVQSTLDHHHGIGTKHTIQLSRLVEMLRMCSYEDLKQIEQNSQQYHQNQQT
KQILVDAMASGTRNCVVRMAEIINNQEVS PHKAAYTISQLTGLQSPSNLIVKEVQKLCQSEQVKNRP
VQQSACLTFGALVNELCSQRRSEKTPDLTHSIP EQCTEDKKEQYKQVLVEQYEKATTLQEKLNALASLG
NAGIDTSTPQLEQIIKDRSEHSLSRKAIDALRRLTTKEPMTLQHMLLPYILNSQEDPNVRIVALQILMRT
QPEPSVIDQIIY TMSQEPNKRVKAYTYQTMK MVAKSKNPADRQMSKHVKNAMQSANVDEQQLWSNT
KWQIPVYSPEQEEGIFLTATEYSQRSWMPSLSYM KIDS YFKDQSNVNDLVFYL MKNQPLYSQDYTK
SRNRRTVPIRIDRRKFEQENNENTQQLKNMDSQLGIKSRHGGYSFNDRYSGSGQQQQQYIHNNNI
NVFCTRIGD VDHSCFEIQNSYLNQTTQRQQQQQTSSMPTLPPLFTELIQSIQNRKPSLAKMFSELKQ
LQQQKVHQYTIASILNDKEAVVPTCVGLPLGIANFVPTILNVEGNVNLQKNTEMGVKVQINGRIMGGM
AHIQKMRCWTPFLVSGIKSIRSIEINIPTQLDINIQT TNGKQSLTSLVKLPKEKTRLVGAHSHPFVYTAD
VDPKTGSVIEPRKVRTITNPKLDR LQYQMNTVIGGKNLGIPVNVHAHYHWPSSSANVEQICKMLMATE
NAIHVTYQTRGEESAKEVVFQMNLESFQPTNNYQLKMND FYLNKFKNIEEEEEGGNNYEEENEQN
SKRQQKMSKFLTNYSSQQQKYKHNLMSAKTVGGPKFQFQ TILEAKCEEGMQTCQLEIQAERSPI
NNEQQKWTLKAQAEMVMPENVENNEEFNEEEENKMKNNLFSC KANCYWGSDGNGKQTINVRING
EQSTTRRKEWNSNNQQRMNKEFEIKLRRTAFINKFNCEAKYTNLKP TTRNAFERALELVKSKYFWN
TKSQLIDQSTEGNPANGEAYLNIVIDPITQKHANMTVKTP EQLIRFQQIELPIQMRPFLLKQQNVYQNI
QSFGQFLTRQEVE SRAECSVDGQTVETFDEREYRSPLSTNCYSVLAKDCGADYMGVPRFVVLMAI
ERGNKYSQKKLKIISPQMVVECQRESLNNHNSPIKCKINGKEENNYQQQQQLIQYNNQMNDLIINWDK
LAIRFNGFKAWIKISSDYRNTQCGLCGHYDDASNDENELLMANNEIAPSLTQFHRSYSLSNSYDNDK
LCNAQELDKFEENSNNKFGYLNEMEEEEEEGKENKVYQNWDEEGFDGNFDRNEENKRFYNEEWY
GQNEREEINYQQKNQLAMTKSGNRNLKGRFVPILIRTKIMEMENDICFSKQPIKQCP LGSYPTGWEFE
ERENNSKHFTTKNIPFICLPRSDSETRNLLNQYRQLMKNGNNQQQLELNNYKQHSIVEKVLEPKECR
RL

>*M. incognita* 6902

MLRMCSYEDLKQIEQNSQQYHQNQQTAKQILVDAMASGTRNCVVRMAEIINNQEVS PHKAAYTISQ
LTGLQSPSNLIVKEVQKLCQSEQVKNRSIVKQSACLTFGALVNELCSQRRSEKTPDLTHSIP EQCTED
KKEQYKQVLVEQYEKATTLQEKLNALASLG NAGIDTSAPQLEQIIKDRSEPSLSRAKAIDALRRLTTKEP
MTLQHMLLPYILNNQEDPNVRIVALQILRTQPEPSVIDQIIY TMSQEPNKRVKAYTYQTMK MVAKSKNP
ADRQMSKHVKNAMQSANVDEQQLWSYAKWQIPVYSPEQEEGIFLTATEYSQRSWMPSLSYM KIDS
YFKDQSNVNDLVFYL MKNQPLYGQDYTKSRNRRTVPIRIDRRKFEQENNENTQQLKDMHSQLGIKSR
HGGYSFNDRYSRSGQQQQQKYNNNNNNNVFCTRIGD VDHSCFEIQNSYLNQTTQQQQQESSMPTL
PPMFAELIQSIQNRKPSLAKMFSELKQLQQQKVHQYTIASILNDKEAVVPTCVGLPLGIANFVPTILNV
QGNVNLQKNTEMGGKVQINGRIMGGM AHIQKMRCWTPFLVSGIKSIRSIEINIPTQLDINIQT TNGKQ
SLTSLVKLPKEKTRLVGAHSHPFVYTADVDPKTESVIEPRKVRTITNPKLDR LQYQMNTVFGGKNLGIP
VNVHAHYHWPSSSANVEQICKMLMATENAIHVTYQPRGEESAKEVVFQMNLESFQPTNNYQLKMND
FYLNKFTNIENEEGEEGNNYEEENEQNSKRRQQKMSKFLTNYSSQQQKYKHNLMSAKTVGGPKF
FKFQTKLEAKCEE TMQTCQLEIQAERSPINNEQQEWTLKAQAEMVMPENVENNEEFNEEEENKIKN
NLFSC KANCYWGNDGNGKQTINVRINGEQSTTRRKEWNSNNQQRMNKEFEIKLRRTAFINKFNCE
AKYTNLKTSTRNVFERALELVKSKYFWNTKSQLIDQNREGNPANGEAYLNIVIDPITQKHANMTVKTPL
QLIRFQQIELPIQMRPFLLRQQNVYQNIQSFGQFLTRQEVE SRAECSVDGQTVETFDEREYRSPLST
NCYSVLAKDCGADYMGVPRFVVLMAIERGNKYSQKKLKIISPQMVVECQRESLNNHNSPIKCKINGK

EENNYQQQQLIQYNNNNQKSDLIINWDKLAIRFNGFKVWIKISSEYRNTQCGLCGHYDDASNDENELLM
ANNEIAPSLTQFHRSYSLSNSYDNDKLCNAQELDKFYEENS NKFGYLNEMEEEEEEGKENKVYQ
NWDEEGFDGNFDKNEENKRFYNEEWYQNEREDNNYQQKNQLATIKSGNRNLKRRFVPILIRTKIM
EMENDICFSKQPIKQCPLGSYPGWEFEEENNKNKHFTTKNIPFICLPRSDSETRNLLNQYRQLMKN
GNNQQQLELNNYKQHSIVEQVLEPKECRRL

>*M. incognita* 9435

MDNAASEQTRAPNFQEGTVYHYSYDAQVESGLSTVDES GTTTQTSDNGQQA VTRIQSQVKVYFASQ
RRALLCMTQNKFGQFNEQRKQTD SNQCYEPLSIFEPKQIPQEQRQKLELCCQFDYVDGVIVRVQFDK
QDEPWSKNMKRGVLNMIQMNLRQNNAQELATEQSRNQYGNIEENESTQMRDQLAHSFTIPEVTVE
GECQTTYTINTANPHKHCMQKSSDES NKQQMPCTFNVTKSINHKKQCSRIGDISAGLQTEQPQARCA
QCKMDWYKQKEGDQQQSGTAEHPCAKCDPKRIKEETLDRQTVMRAMLKCGSEQQMIEGKQQETP
NSCQLDSSEMRSMYVYKNTKTESGLYGSAMRTEVCAKLQLRSAQSTQDQQVPTQTLDNDES LVYTE
QTFSDYARFYMFGDEEFPDPTSSPFYSVPKVEQAGQALRALVQSTLDH HHGIDTKHATQLSRLVRML
RMCSYEDLKQIEQNAQQYHQNNQKETAKQILVDAMATSGTRNCVVRVAEMIKNQEVS PHKAAYTISQL
TGLPSPSNSIVKEVQKLCQSEQVQNQPIVKQSACLTFGALVNELCSQKRSEKTPDLTHPIPEQCTDDK
KEQYKQVLEQYQEKATTLQEKLNALASLGNA GIDTSTPQLEQIIKDRSEHSLCRAKAIDALRRLTTNQQ
MIVQHIILPIYLNQEDPNVRIVALQILMRTQPEPSVIDQIIYTMSQEPNKRVKAYTYQTMKMKVAKSKNPA
DRQMAKHVKNAMQSANVDEQQLWSYGKWQIPVYSPEQEEGIFLTATEYSQRSWMP SLSYMKIDSY
FKDQSNVNDLVFYLMKNQRLYSQDYTTKSRNRRTAAIRGFERREFEQENNENTQQLKDMHSQLGIKS
RHGGYSFNDRYSTSGQQQQYSYNNNINVFCTRIGD VDHSCYEIQNSYLNQTQQQQQQQSSSMP
TLPPMFTELIQSIQNRKPSLAKMFSELEQLQQQKVHQYTIASILNDKEAVIPTCVGLPLGIVNSVPTILN
VEGNVNLQKNTEMGVKVQINGRIMGGM AHIQKMRCWTPFLVSGVKSIRSIEINVPIKLDINAQTTTTNN
GKQSLIMSVKLPNEKTRLVGAHSHPFVYTA DVPKTNLVIEPRQVVRTIRNPKLDRLQYQMNTLIGGKNL
GIPVNVHAHYHWPSSSVNFEQICKMLMATENAIHVTYQPRGEESASQVVF RMNLESYQPTNDYQLK
MNDFYLNKFKNIENEEGEEEEENKQNSKRQQKMSKFLTNYSKQQKYKHNL EMSAQTVGGLKEFK
FQTKLEAKCEEMMQTCQLEIQAERSPINNEQQKWTLKAQAEMVMPENVENNEEFNEEEENKNNLYSC
KANCYWGNDDDRAKQTINVRINGEQSMRGGMEQRMNKMIRRTAFINKFNCEAKYTNLKP TTRNVF
ERVLELVKSKYFWNTKSQLIDSSRGGNPANG EANLNIVDPITQKHANMTVKTP EQLIRFQQIELPIQMR
PFPLLRQQNVHQNIQSFGQFLTRQEVE SRAECSVDGQTVETFDEREYRSPLSTNCYSVLAKDCGAS
YMGGVPRFVVLMAIERGNKYSQKCLKLLSPQMSVE CQRETPNNHNSPIKCKINGKEEENQQQQQ
QQLIQYNNNNQKNDLIINWDKLAVRFNGYK VWIKISSDYRNTQCGLCGHYDDASNDENELLMANNEIAS
SLTQFHRSYSLSNSYDNDKLCNAQELDKFYEENS NKFGYLNEMEGEDEERWEEERKEENKNLNQ
NWDEEGFNGNFDENKYFYNEEWYGENEEEEWKEEGKEENNNNQKMATTTKSEN RNLKRRGGL
LPILKTKIMEMENEICFSKQPIKQCPMGTYPTGWEFEEEEKGENKHFTTKNIPFICM PRSDSEARNLLNQ
YRQLIGNNQQQQLELNNYKQHSIVEK VLEPKECRRL

>*M. javanica* 15377

MRINSLVLLLAIGLLSSSNLVVGDEYPATSTLKRFFYRHFFMDNAASEQTRAPNFQEGTVYHYSYDAQ
VESGLSTVDES GTTTQTSENGQQA VTRIQSQVKVYFASQRRALLCMTQNKFGQFNEQRKQTD SNQ
YEPLSIFEPKQIPQEQRQKLELCCQFDYVDGVIVRVQFDKQDEPWSKNMKRGVLNMIQMNLRQNN
AQELATELSRNQQEGNIEENEATQMRDQLAHSFTIPEVTVEGECQTTYTINTANPSKHCTQKSSDES NK
QQQMPCVFNVTKSINFKQCSRIGDISAGLQTEQPQSRCAQCKMDWYKQKEGDQQQSGTAEHPCAK
CDPKRIKEETLDRQTVMRAMLKCGSEQQMIEGKQQETPN SCQLDSSEMRSMYVYKNTKTESGPYG
SAMRTEVCAKLQLRSAQSAQDQQVPTQNLDNDES LVYTEQTFSDYARFYMFGDEEFPDSTSSPFYS
VPKVEQAGQALRALVQSTLDH HHGIDTKHTTQLSRLVRMLRMCSYEDLKRIEQNAQQYHQNNQETA
KQILVDAMATSGTRNCVVRVAEMIKNQEVS PHKAAYTISQLTGLPSPSNSIVKEVQKLCQSEQVQNQPI
VKQSACLTFGALVNELCSQKRSEKTPDLTHPIPEQCTDDKKEQYKQVLEQYQEKTTTLQEKLNALASL
GNAGIDTSTPQLEQIIKDRSEHSLCRAKAIDALRRLTTNQQMIVQHIILPIYLNQEDPNVRIVALQILMRT
QPEPSVIDQIIYTMSQEPNKRVKAYTYQTMKMKVAKSKNPA DRQMAKHVKNAMQSANVDEQQLWSYG
KWQIPVYSPEQEEGIFLTATEYSQRSWMP SLSYMKIDSYFKDQSNVNDLVFYLMKNQRLYSQDYTT
KSRNRRTAAIRGFERREFEQENNENTQQLKDMHSQLGIKS RHGGYSFNDRYSRSGQQQQQYSYNN
NINVFCTRIGD VDHSCYEIQNSYLNQTQQQQQQQSSSMP TLPPMFTELIQSIQNRKPSLAKMFSEL
EQLQQQKIHQYTIASILNDKEAVIPTCVGLPLGIVNSVPTILNVEGNVNLQKNTEMGVKVQINGRIMG
MAHIQKMRCWTPFLVSGVKSIRSIEINVPIKLDINAQTTTTNN GKQSLIMSVKLPNEKTRLVGAHSHPFV

YTADVDPKTNLVIEPRQVRTIRNPKLDRLQYQMNTLIGGKNLGPVNVHAHYHWPSSSVNFEQICKML
MATENAIHVITYQPRGEESASQVFRMNLSEFQPTNDYQLKMNDFYLNKFKNIENEEGEEEEENKQ
NSKRQQKMSKFLTNYSKQQKYKHNLMSAQTVGGLEKFKFQTKLEAKCEEMMQTCQLEIQAERSP
INNEQQKWTLKAQAEMVMPENVENNEEFNEEEKNNLYSCKANCYWGNDDDRQKTINVRINGEQS
MRGGMEQRMNKMIRRTAFINKFNCEAKYTNLKPTRNVFERVLELVKSKYFWNTKSQIDSSRGGN
PANGEANLNIVIDPITQKHANMTVKTPQLIRFQQIELPIQMRPFLLRQQNVHQNIQSFGQFLTRQEV
ESRAECSVDGQTVETFDEREYRSPLSTNCYSVLAKDCGASYMGEIPRFVLMKAIERGNKYSQKKLK
LLSPQMSVECQRETPNNHNSPIKCKINGKEEENNNQQQQQLIQYNNNQKNDLIINWDKLAVRFNGYK
VWIKISSDYRNTQCGLCGHYDDASNDENELLMANNEIAPSLTQFHRSYSLSNSYDNDKLCNAQELD
KFYEENS NKFGYLNEMEEKEEEERMVWRK

>*M. javanica* 21515

MRINSILLLLAIGLLGPSNLVVGDEYPATSTLKRFFYRHFFADNSASEQTRAPNFQEGTVYHYSYDAQV
ESGLSTLDES GTTQTSDNGQQAIRIQSQVKVYFASQTRAQLCMTQNRFGQLNEQTKQTDSSNQ
YEPLSTFEPKQIPQEQRQKLELCCQFDYVDGVIVRVQFDKQDEAWSKNMKGVLNMIQMNLRQNN
QELASEQSRNQDGNIEENATQMRDQLAHSFTIPEVTVEGECQTTYTINTANPSKNCTQKGYDGSN
KQQQMPCSFNLTKSINFKQCSRIGDISAGSQTEQPQARCAQCKMDWYKQKERDQQQTGTAEHPCA
KCDPKRVKEETLDRQTVMRAMLKCGSEQQMIEGKQKETQNSCQLDSSEMRS MYVYKNTKTESGPY
GSAMRTEVCAKLQLRSAQSNQDQVPTQTLDSDESLVYTVQTYTDFARFYMFGDEEFPDSTNSPFS
SVPKVEQAGQALRALVQSTLDHHHGIGTKHTIQLSRLVEMLRMC SYEDLKQIEQNSQQYHQNNQQT
KQILVDAMAI SGTRNCVVRMAEIINNQEVS PHKAAYTISQLTGLQSPSNLIVKEVQKLCQSEQVKNRPIV
KQSACLTFGALVNELCSQRRSEKTPDLSHSIPEQCTEDKKEQYKQVLVEQYKATTLQEKLNALASLG
NAGIDTSTPQLEQIIKDRSEHSLSRAKAIDALRRLTTKEPMTLQHMLLPYILNSQEDPNVRIALQILMRT
QPEPSPNKRKAYTYQTMKMKVAKSKNPADRQMSKHVKNAMQSANVDEQQLWSYAKWQIPVYSPEQ
EEGIFLTATEYSQRSWMPSLSYMKIDSYFKDQSNVNDLVFYL MNQPLYGQDYTKSRNRRTVPIRID
RRKFEQENNENTQQLKDMHSQLGIKSRHGGYSFNDRYSGSGQQQEQYSHNNNINVFCTRIGD VDH
SCFEIQNSYLNQTTQQQQQESSMPTLPPMFTELIQSIQNRKPSLAKMFSELKQLQQQKVHQYTIASI
LNDKEAVVPTCVGLPLGIANFVPTILNVEGNVNLQKNTMG MKVQINGRIMGGMAHIQKMRCWTPFL
VSGIKSIRSIEINIPTQLDINVQTANNGKQSLTMSVKLPKEKTRLVGAHSHPYVYTADVDPKTGSVIEPR
KVKTITNPKLDRLQYQMNTVIGGKNLGPVNVHAHYHWPSSSANVEQICKMLMATENAIHVITYQPRGE
ESAKEVV FQMNLSEFQPTNNYQLKMNDFYLNKFKNIENEEEGGNNYEEENEQNSKRQQKMSKFLT
NYSQQQQKYKHNLMSAKTVGGPKFKFQTKLEAKCEEGMQTCQLEIQAERSPINNEQQKWTLKA
QAEMVMPENVENNEEFNEEEDENKMKNLFSCKANCYWGSDGGNKQTINVRINGEQSTTRRRNEW
NNNQQRMNKEFEIKLRRTAFINKFNCEANYTNLKPATRDVFERALELVKSKYFWNTKSQ LIGQSREG
NPANGEAYLNIVIDPITQKHANMTVKTPQLIRFQQIELPIQMRPFLLKQQNVYQNIQSFGQFFTRQE
VESRAECSIDGQTVETFDEREYRSPLSTNCYSVLAKDCGADYMGVPRFVLMKAIERGNKYSQKKLK
IISPQMVVECQRESLNNHNSPIKCKINGKEGNNYQQQLIQYNNNQKNDLIINLEKLAVRFNGFKVWIKISS
DYKNTQCGLCGHYDDASNDENELLMANNEIAPSLTQFHRSYSLSNSYDNDKLCNAQELDKFYEEN
SNKFGYLNEMEEEEEEEEKENKLYQNWDEEGFDGNF D KNEENKRFYN EEWYQNEREEENNYQQ
KKQLAAIKSGNRNLKRRFVPILKRTKIMEMENDICFSKQPIKQCPLGSYPTGWFEENNSKHFTTKNIP
FICLPRSDSETRNLNLYRQLMKNGNNQQQLELNNYKQHSIVEKVLEPKECRRL

>*M. javanica* 29487

MRINSLVLLLAIGLLGPSNLVVGDEYPATSTLKRFFYRHFFMDNSASEQTRAPNFQEGTVYHYSYDAQ
VESGLSTVDES GTTQTSENGQQA VTRIQSQVKVYFASQRRAVLCMTQNKFGQFNEQKKQTDSNQ
YEPLSIFEPKQIPQEQRQKLELCCQFDYVDGVIVRVQFDKQDEPWSKNMKRGVLNMIQMNLRQNN
QELATEQSRNQQEGNIEENESTQMRDQLAHSFTIPEVTVEGECQTTYTINTANPNKHCTQKSSDES
KQQQMPCAFNVTKSINH KQCSRIGDISAGLQTEQPQARCAQCKMDWYKQKEGDQQQSGTAEHPCA
KCDPKRIKEETLDRQTVMRAMLKCGSEQQMIEGKQKETSN SCQLDSSEMRS MYVYKNTKTESGPY
GSAMRTEVCAKLQLRSAQSAQDQVPTQTLDNDES LVYTVQTFSDYARFYMFGDEEFPDSTSSPFY
SVPKVEQAGQALRALVQSTLDHHHGIDTKHATQLSRLVRMLRMCSYEDLKQIEQNAQQYHQNNQKET
AKQILVDAMATSGTRNCVVRVAEMIKNQEVS PHKAAYTISQLTGLPSPSNSIVKEVQKLCQSEQVQNNQ
PIVKQSACLTFGALVNELCSQKRSEKTPDLTHPIPEQCTDDKKEQYKQVLVEQYKATTLQEKLNALAS
LGNAGIDTSTPQLEQIIKDRSEHSLCRAKAIDALRRLTTNQQMIVQHILPIYLNNQEDPNVRIVALQILMR
TQPEPSVIDQIIYTMSQEPNKRKAYTYQTMKMKVAKSKNPADRQMAKHVKNAMQSANVDEQQLWSY

GKWQIPVYSPEQEEGIFLTATEYSQRSWMPSLSYMKIDSYFKDQSNVNDLVFYLMKNQRLYSQDYT
TKSRNRRSAAIREPNKRVKAYTYQTMKMKVAKSKNPADRQMAKHVKNAMQSANVDEQQLWSYGKW
QIPVYSPEQEEGIFLTSTEYSQRSWMPSLSYMKIDSYFKDQSNVNDLVFYLMKNQQLYSQDYTTKSR
NRRTAIRGFERREFEQENNENTQQLKDMHSQLGIKSRHGGYSFNDRYSTSGQQQQQSYNNNNINVF
CTRIGDVDHSCYEIQNSYLNQTQQQQQSSSSMPTLPPMFTELIQSIQNRKPSLAKMFSEIEQLQQ
QKIHQYTIASILNDKEAVIPTCVGLPLGIVNSVPTILNVEGNVNLQKNTEMGVKVQINGRIMGMAHIQK
MRCWTPFLVSGVKSIRSIEINVPIKLDINAQTTTTTNNKGQSLIMSVKLPNEKTRLVGAHSHPFVYADV
DPKTNLVIEPRQVRTIRNPKLDRLQYQMNTLIGGRNLGIPVNVHAHYHWPSSSVNFEQICKMLMATEN
AIHVTYQPRGEESASQVVFRMNLESFQPTNDYQLKMNDFYLNKFKNIENEEGEGEEENKQNSKRQ
QKMSKFLTNYSKQQQQQKYYKHNLMSAQTVGGKLEFKFQTKLEAKCEEMMQTCQLEIQAERSPI
NNEQQKWTLKAQAEMVMPENVENNEEFNEEEENLNLYSCKANCYWGNDDDRAKQTINVRINGEQS
KMRGGMEQRMNKKLKRRTAFINKFNCEAKYTNLKPTTRNVFERVLELVKSKYFWNTKSQIDSSRGG
NPANGEAYLNIVIDPITQKHANMTVKTPQLIRFQQIELPIQMRPFLLRQQNVHQNIQSFQGLTRQEV
ESRAECSDGQTVETFDEREYRSPLSTNCYSVLAKDCGANYMGGVPRFVVLMAKIERGNKYKENNQ
QQQQQQLIQYNNNNQKNDLIINWDKLAVRFNGYKVVWIKISSDYRNTQCGLCGHYDDASK

>*M. javanica* 3127

MRINSLLLLLLAIGLLSSSKLVVGDEYPATSTLKRFFYRHFFMDNAASEQTRAPDFQEGTVYHYSYDAQ
VESGLSTVDESGLTTTQSDNGQQAVTRIQSQVKVYFASQRRALLCMTQNKFGQFNEQRKQTDNSNQ
YEPLSIFEPKQIPQEQRQKLELCCQFDYVDGVIVRVQFDKQDEPWSKNMKRGVLNMIQMNLRQNN
QELATEQSRNQEGNIEENESTQMRDQLAHSFTIPEVTVEGECQTTYTINTANPSKHCTQKSSDESN
KQQQMPCSFNVTKSINHKQCSRIGDISAGLQTEQPQARCAQCKMNWYKQKEGDQQQSGTAEHPCA
KCDPKRIKEETLDRQTVMRAMLKCGSEQQMIEGKQQESPNSCQLDSSEMRSMYVYKNTKTESGPY
GSAMRTEVCAKLQLRSAQSTQDQVPTQNLNDSELVYTVQTFSDYARFYMFGDEEFPDSTSSPFY
SVPKVEQAGQALRALVQSTLDHHHGIDTKHATQLSRLVRMLRMCSYEDLKQIEQNAQQYHQNQQUET
AKQILVDAMATSGTRNCVVRVAEMIKNQEVSPHKAAYTISQLTGLPSPSNSIVKEVQKLCQSEQVQNNQ
PIVKQSACLTFGALVNELCSQKRSEKTPDLTHPIPEQCTDDKKEQYKQVLVEQYKATTLQEKLNALAS
LGNAGIDTSTPQLEQIIKDKSENSLCRAKAIDALRRLTTNQQMIVQHIILPIYLNQEDPNVRIVALQILMR
TQPEPSVIDQIIYTMSQEPNKRKAYTYQTMKMKVAKSKNPADRQMAKHVKNAMQSANVDEQQLWSY
GKWQIPVYSPEQEEGIFLTATEYSQRSWMPSWSYMKIDSYFKDQSNVNDLVFYLMKNQRLYSQDYT
TKSRNRRTAIRGSRREFEQENNENTQQLKDMHSQLGIKSRHGGYSFNDRYSTSEQQQQQQSYNN
NNINVFCTRIGDVDHSCYEIQNSYLNQTQQQQQSSSSMPTLPPMFTELIQSIQNRKPSLAKMFSEL
EQLQQQKVHQYTIASILNDKEAVIPTCVGLPLGIVNSVPTILNVEGNVNLQKNTEMGVKVQINGRIMG
MAHIQKMRCWTPFLVSGVKSIRSIEINVPIKLDINAQTTTTTNNKGQSLIMSVKLPNEKTRLVGAHSHPFV
YTADVDPKTNLVIEPRQVRTIRNPKLDRLQYQMNTLIGGKNLIPVNVHAHYHWPSSSVNFEQICKML
MATENAIHVTYQPRGEESASQVVFRMNLESFQPTNDYQLKMNDFYLNKFKNIENEEGEEEEENKQ
NSKRQQKMSKFLTNYSKQQQQKYYKHNLMSAQTVGGSKLEFKFQTKLEAKCEEMMQTCQLEIAE
RSPINNEQQKWTLKAQAEMVMPENVENNEEFNEEEENLNLYSCKANCYWGNDDDRKLQKTINVRINGE
QQSTMMRGMEQRMNKKLKRRTAFINKFNCEAKYTNLKPTTRNIFERVLELVKSKYFWNTKSQIDSS
GRGGNPANGETYLNIVIDPITQKHANMTVKTPQLIRFQQIELPIQMRPFLLRQQNVHQNIQSFQGL
ARQEVESRAECSDGQTVETFDEREYRSPLSTNCYSVLAKDCGASYMGGVPRFVVLMAKIERGNKY
SQKKLKLSPQMSVEQCQRETPNNHNSPIKCKINGKEENNQQQQQQLIQYNNNNQKNDLIINWDKLAVR
FNGYKVVWIKISSDYRNTQCGLCGHYDDASNDENELLMANNEIASSLTQFHRSYSLSNSYDNDKLCN
AQELDKFYEENS NKFGYLNEMEEKEEEEEENNNKLNQNWDEEGFNGNFDEENKYFYNEEWYGE
NEEEWKEEGKGNNNNQKMATTKSGNRNLKRRGGGLCKYLIKFK

>*M. javanica* 3894

MKTLVFILFLLSLSGSSSIPIDDRSSNEYYSMTKSAPFKNFFYREYNTAAEQIKTPNFQENNEYVYSYN
AQVENGMLSSVDSSGTKEESQQQAITRIQSQIRMYFSSERRGQLCMTQLKFGQFNDMRQMKEKQC
YDPMGMFEPLQIPQETRQKLELCCQFDYVDGMVTRIQFEKQDVSWSKNIKRGVLNMLQMNLRHNEE
AQNAMGENKNVDRLAQSFTEPETTVEGECQTIYTINKGSKQCQESQQPCSYNVTKSINFKKCSKIAD
STNGFQTEQPQLKCSQCPMNEDVKVNEEHEGTAENQCADCDPREMTEQMLDRQTVIRAMLKCGDQ
QKQNNACPTCCQLDNSEMTSNYIYKNTKIESEGLYGSALRTKTKAKLVLETQVQNVQTPAPLAATAESD
EALLYSNEWGMNLRFFYMFGEAFPDVNSSPYKMPKVEEATQALRSIVQSTIDHHHGKTKHTIDLSR
LVNKMRCMSYEELKQIEQNAKDQDEEKQKQVAMDVFDCLATSGTRNTVKLLAEIENEEIDETKAAF

AINQLKNLPIPSNSILKQVKQLAQNVNVQNKPMVRQAAWLTFGALVNELCQHKGRNMMESSQDVSKS
VQQQCTDDKKEEYKRTLLEQYNQATTFYDKFLALSALGNAGIDTTTPYIDQIISDKSQDNIRVKIDAL
RRLFTKQPRTIQQILLPIFLNNQEDAYVRSSVLTLLRTQPEPSVIDQIVYSMRQEPNKRVKASTYQTLK
NVAESRNPVDQQVSKHVKNALKSVNIDNEEMKNYKLWQIPAFCSKQQEGVFFTIGARKAQRWMPNF
SYLKTDTYLNKESNLDSTLVYVLQASEGYNSYEEAHNYTASTDQHSSENSKTLKQINQMLQIKTRNPFK
YGYGDEQQSSSNPLNYGYPYGGQSSHYSLRNQNPSSVLCVRMGDGDHSHALDQNWDMDSKEI
GIQNWGIMSSQTSIPTTSPMLEEILQSIQRGEMPKFNRLISKFGLQQLQSVQNHTYSAIFNEREALPTS
VGIPVKIVNSVPMLMNVEASTQLQDRQNQEAKMQIDARIMAGVSHVQRIGSWTPFLITGVSNIERSVEV
NLPIQAELSANTEESRQSLTVKVKMPRETTLLGAHSHQYTYRAGIDQNTYLPMEYRTIRNPKLDR
LQYQKNKVGSKNVGVPVHVYAHYHWPSSNAANIVQILKMLTSTENAMHVTYQPGQGTAREVVFNMNT
ENFQPVKAFQSKMDSFFERSATSNEEGNPQEGIMPEDQQQYANQQQKLQSFLSNYQKAQKHQKH
YKHSMELSVKTIGASKQFKASVQLQAQCNDKMQVCQLKMGAEERSRMNEERGMWAMKGQVEMVM
PGGAEEENSNSYASDQFLSFANLHWGDNTRKQQVNMRIVGEQAPERTSGWYMFQNNQEGENQEG
EQNEENQNSEQNSSPFNTKLQRRTAFINLYNCRMEYTNLTPGTRNVFERAVELAKSKYFWNSKTEIK
SESSLNPENGKVYCTLSIDSLTQNHANITLMTPEQIVHLQQIEMPFKLRPFPLRRNRNVQSIQSFDQFYS
QQEIKNRAECSVDGRTVRTFDGKKYNAPLSSNCYSVLAKDCGRMTDAPDFVMMMATGKGSKHSQ
KAN

>*M. javanica* 43787

MRINSILLLLAIGLLGPSTLVVGDEYPATSTLKRFFYRHFFADNSASEQTRAPNFEEGTVYHYSYDAQV
ESGLSTLDESGTTTQASDNGQQA/TRIQSQVKVYFASRTRAHLCMTQNRFGQLNEQTKQTDSSNQC
YEPLSTFEPKQIPQEQRQKLELCCQFDYVDGVVIRVQFDKQDEAWSKNMKKGVNLMIQMNLQNNNA
QELASEQSRNQDGNIDENEATQMRDQLAHSFTIPEVTVEGECQTTYTINTANPSKNCTQKGYDGSN
KQQQMPCSFNLTKSINFKQCSRIGDISAGFQTEQPQARCAQCKMDWYKQKERDQQQTGTAEHPCA
KCDPKRVKEETLDRQTVMRAMLKCGSEQQMIEGKQKETQNSCQLDSSEMRSMYVYKNTKTESGPY
GSAMRTEVCAKLQLRSAQSNQDQVPTQTLDSDESLVYTVQTYTDFARFYMFGDEEFPDPTNSPFF
SVPKVEQAGQALRALVQSTLDHHHGIGTKHTIQLSRLVEMLRMCSYEDLKQIEQNSQQYHQNNQQTA
KQILVDAMASGTRNCVVRMAEIIINNQEVSHPKAAYTISQLTGLQSPSNLIVKEVQKLCQSEQVKNRPIV
KQSACLTFGALVNELCSQRRSEKTPDLSHSIPEQCTEDKKEQYKQVLVEQYEKATTLQEKLNALASLG
NAGIDTSTPQLEQIIKDRSEHSLSRAKAIDALRRLTTKEPMTLQHMLLPIYLNQEDPNVRIVALQILMRT
QPEPSVIDQIYTMSQEPNKRVKAYTYQTMKMKVAKSKNPADRQMSKHVKNAMQSANVDEQQLWSYT
KWQIPVVSPEQEEGIFLTATEYSQRSWMPSLSYMKIDSYFKDQSNVNDLVFYLMKNQPLYGQDYTK
SRNRRTVPIRIDRRKFEQENNENTQQLKDMHSQLGIKSRHGGYSFNDRYSGSGQQQQQYSHNNNIN
VFCTRIGDGDHSCFEIQNSYLNQTTQQQQQESSMPTLPPLFTELIQSIQNRKPSLAKMFSELKQLQQ
QKVHQYTIASILNDKEAVVPTCVGLPLGIANFVPTILNVEGNVNLQKNTEMGMKVQINGRIMGMAHIQ
KMRCWTPFLVSGIKSIRSIEINIPTQLDINVQTTNNGKQSLTMSVKLPKEKTRLVGAHSHPHYVYTADVDP
KTGSVIEPRKVKTITNPKLDRQYQMNTVIGGKNLGIQVNVHAHYHWPSSSANVEQICKMLMATENAI
HVTYQPRGEESAKEVVFQMNLESFQPTNNYQLKMNDFYLNKFKNIEEEEEGGNNYEEENEQNSKR
QQKMSKFLTNSQQQQKYKHNLMSAKTVGGSKEFKFQTKLEAKCEEQMTCQLEIQAERSPINN
EQQKWLKAQAEMVMPENVENNEEFNEEEENKMKNNLFSCKANCYWGSDDGKQNTINVRINGEQ
LTRRGNEWNNNNQQRMNKEFEIKLRRTAFINKFNCEANYTNLRPTTRNAFERALELVKTKYFWNTK
SQLIEQSREGNPANGAYLNLVIDPITQKHANMTVKTPEQLIRFQQIELPIQMRPFPLKQQNVYQNIQS
FGQFLTRQEVESRAECSVDGQTVETFDEREYRSPLSTNCYSVLAKDCGADYMGVPRFVVLKMAIER
GNKYSQKKLKIISPQMVVEQCRESLNHNHNSPIKCKINGKEENNYQQQQLIQYNNNNQKSDLIINWDKLA
RFNGFKVWIKISSEYKNTQCGLCGHYDDASNDENELLMANNEIAPSLTQFHRSYSLSNSYDNDKLC
NAQELDKFYEENS NKFGYLNEMEEEEEREKGKENLHQNWDEEGFDGNFDKNEENKRFYNEEWYG
QNEREDNNYQQKNQLATIKSGNRNLKRRFVPIIRTKIMEMENDICFSKQPIKQCPLGSPSGWEFEE
ENNKKNKHFTTKNIPFICLPRSDSETRNLLNQYRQLMKNGNNQQQLELNNYKQHSIVEQVLEPKECRR
L

>*M. javanica* 5192

MKKGVNLMIQMNLQNNNAQELASEQSRNQGGNIDENEATQMRDQLAHSFTIPEVTVEGECQTTYTI
NTANPSKNCTQKGYDGSNKQQQMPCSFNLTKSINFKQCSRIGDISAGFQTEQPQARCAQCKMDWY
KQKERDQQQTGTAEHPCA KCDPKRVKEETLDRQTVMRAMLKCGSEQQMIEGKQKETQNSCQLDS
EMRSMYVYKNTKTESGPYGSAMRTEVCAKLQLRSAQSSQDQVPTQTLDNDESLVYTVQTYTDFAR

FYMFGEDEFPDPTNSPFSSVPKVEQAGQALRALVQSTLDHHHGIGTKHTIQLSRLVEMLRMC SYEDL
KQIEQNYQQYHQNQQQTAKQILVDAMAI SGRNCVVRMAEIINNQEVS PHKAAYTISQLTGLQSPSNLI
VKEVQKLCQSEQVKNRPIVKQSACLTFGALVNELCSQRRSEKTPDL SHSIPEQCTEDKKEQYKQVLVE
QYEKATTLQEKLNALASLGNAGIDTSAPQLEQIIKDRSEPSLSRAKAIDALRRLTTKEPMTLQHMLLPY
LNNQEDPNVRIVALQILIRTQPEPSVIDQIIYSMSQEPNKRVKAYTYQTMKMVAKSKNPADRQMSKHVK
NAMQSANVDEQQLWSYAKWQIPVYSPEQEEGIFLTATEYSQRSWMPSLSYMKIDSYFKDQSNVND
LVFYLMKNQPLYGQDYTKSRNRRTVPIRIDRRKFEQENNENTQQLKDMHSQLGIKSRHGGYSFNDRY
SRSGQQQQQQYNHNNNNINVFCTRIGDVDHSCFEIQNSYLNQTTQQQQQESSMPTLPPMFAELIQSIQ
NRRKPSLAKMFSELKQLQQQKVHQYTIASILNDKEAVVPTCVGLPLGIANFVPTILNVQGNVNLQKNTE
MGGKVQINGRIMGMAHIQKMRCWTPFLVSGIKSIRSIEINIPTQLDINIQTNNNGKQSLTSLVKLPKEK
TRLVGAHSHPFVYTADVDPKTESVIEPRKIKTITNPKLDRLQYQMNTVFGGKNLGIPVNVHAHYHWP
SSANVEQICKMLMATENAIHVTYQPRGEESAKEVVFQMNLESFQPTNNYQLKMND FYLNKFKNIENE
EGEEGNYYEEENEQNSKRRQQKMSKFLTNYSKQQQKYKHNLMSAKTVGGPKFQKFKQTKLEAKC
EETMQTCQLEIQAERSPINNEQEWTLKAQAEMVMPENVENNEEFNEEEENKIKNNLFSCKANCY
WGNNDGNKQTINVRINGEQSTTRRKEWNSNNQQRMNKEFEIKLRRTAFINKFNCEAKYTNL KSTTR
NVFERALELVKSKYFWNTKSQLIDQNREGNPANGEAYLNLVIDPITQKHANMTVKTP EQLIRFQQIELPI
QMRPFPLLRQQNVYQNIQSFGQLLTRQEVVSRAECSDVGQTVETFDEREYRSP LSTNCYSVLAKDC
GADYMGVPRFVVLMAIERGNKYSQKKLKIISPQMVVEQCRESLNHNHNSPIKCKINGKEENNYQQQQ
LIQYNNNQKSDLIINWDKLAI RFNGFKVWIKISSEYKNTQCGLCGHYDDASNDENELLMANNEIAPSLT
QFHRSYLSNSYDNDKLCNAQELDKFYEENS NKFGYLNEMEEEEEREEGKENKLHQNWD EEGFDG
NFDKNEENKRFYN EEWYQNEREDNNYQQKNQLATIKSGNRNLKRRFVPILIRTKIMEMENDICFSK
QPIKQCPLGSPSGWEFEED EENNSKHFTTKNIPFICLPRSDSETRNLLNQYRQLMKNGNNQQQLEL
NNYKQHSIVEQVLEPKECRRL

>*P. scribneri* 001

MRILASLLLVALCLGTGLSDQQSGQQQQPSPFRKDNVYIFQYNSQISSGLSAPSDENSADPQQNSAIR
IQAQAKITFASDRHGQLQLEQIRVGQLNEQLNEPKKVQPM AIFEPKQIPDEKQRQLELPVQFH YDEGV
VERVQFNEEDDVWSKNVKRAVLNMLQLNLKRNN AQLRTGNEQEWQNEEENSSENGQQMSELGQ
TFNIPEITLEGVCQTTYAINAVNNGQWQQDEQEN EKATALNVTKTINFKQCAKIADVGYGFQTEQPQP
QCAQCKQYWSKNQQQQNGQNGQQSGQRHPCAKCDPKKVKEQLLDRSTVLR YQLVCPSGECGP
SNSQKYALKGTELLSQYVFKQTKSQTGKFSSAMQTIVAATLKFRS QEQTESVPSMSGGKDESLMFS
SEIDLDVKRFYMYGDEEFSQP ESSPFQQVPKVQQAQLALRLKSQTM TGNKQTGIEVQAGYEVQRLV
ELLRMSRREELEEIEQNAINNNGNGGEQERQKAKQIYTD SLAICGTRNCIQKLAEKIQNKEMSQT KAV
QALRTVNQVPAPSDQIVKIVQRLCKQENVHNQRSVKQSCWLTFGALVNELCQH KTPKKGTQKFAFEA
KSSLDKQECPKEKKEEYKQTIMDLYNSAATIIDKVL AISALGNAGIENAVPEIEQIIKDQTESCLVRCK
AIDALRRLRTQMPRKIQQILLPIFQNVQQRSEVRASALYILLTTQPDSAMIDQIAFTISKEGNRQIQSFAY
QQMKT LAKSKNPLDQQMAKHVKNALRLTNVDEQQLRMSGKWQFPVYSEQLQEGAFLELAHLISPRS
NAPAYSHAQIDSFFNDQFQTGQMKVYIAQQEAEDWYQRTSSNNAQNAHSPAQR RSFNGQSLRNIF
TVLGIKKRRNDEENNRPQPFAMFALRTRDQAFALD GSSKEQQDVQLQEMLGQFISGQSKPTLERF
LSQFHQSQQQNQNNQQQKEFRGILVTNLNEQQAK IPTSAGLPLRILTSNPMLSTIEGQTQLSKESDHL D
SPLGVQTQISVHSMASLVHIQKMESALPFMVAGVESVRSLELNVPLQAE LNANPSQNQMSVKVLP
QKHTLLGFHSLPFTYTAQVDVNTKLVHEPRHFRAIHHRELERMQHEPNEETPREMIFRISGNQFEKT N
GQRP ELDDFYREEMGS AFESHEDNFEMETTSRRNQKLQNFIGNYQSNQQQYKHSVKLTAQTVGG
PKEAKAQAELAASCDAKMQSCNTKLEAQFGKL RAGSEKWTMKIQTNVAMPETVQSVDELVELEQRN
NKFQAKIQAEWGTGTQREN LNIRLQGEQAKRRENQNN EQNVQTKRREQWRQAF LNKFDVEAQYQ
MKA KTMNYFVKVSEWLKSTYYWNSQQQSNNNQNGQMMQMRGQEKT LISLVIDPITQKHANVSIRTP
NEQIRLQQIELPMPIRRPFPLVRQQRQSIHSFGQFIDRIQVPGRAECTVDGRDVNTFDNVEFRAPLS D
ACYTVLAKDCSSEEPKFVVLMAQDNIDSGSQNSPKKLKVITPETEIVCEPKSSNGNQQKIVCKENG S
PVENDNDSDAVEFNNAKSDVTVHVGRGISVRFN GRKAWLKVSSVYKNGQCGLCGHYDDSEQDEWR
MSNNQNTNDLVEFHRSYTLNGAQQGEQCNEEELN NFYRQNGNKIRKNGRNQKNGKNSEEV DDEDD
DGQNDWMDEQFDGHW NENGRNNRRNGNQQKRIETPVKKT KLVELSNRICVSQKPIKQCPKGYIPIY
SSNEDSQEDPFDQQNGNGEGKNVQFSCFNRTAEARDILRQMRNGNDVVHF EKREHSMAEKIHEP
KKCVRGRY

>*P. scribneri* 017

MRPLLLLMAVGLCLFGNVALAKQHQNRFSPFNSNEQD TDGQQQKMRAPAFQPGRTYQFQYNAQIA
SGLIAQDEQSQADAQQKAITRIQAQAKITFGSERHAQLCLRQIRTGQLNEPVEDEQQVQPM SIFEPKQ
IPSDKKRLLERCCQFDYADGVIERVQFDQEDETWSKNMKRGVLQMIQLNLKRNNAAQGLRNEMNDQQ
QGEEDQDEDNEQENGQQQDKDTLAMS FNIP EIAVEGQCQTQYTINTANPSQRCTPDDQTPCSFNV
KSINFKQCSKIADVAFGYQTQQQQPQCAQCQQQFQEIQQEGCPRGNCFFNAAGQSQSNAQHPCA
KCDPKEVKEQKVDRTTVLRYVLKCGSESSRKSEEKCNCCQLERTELLSQYVYKNSKVENGQYGS
MHTVVS AQLQLRQVEQQQEVP TQSASQDESLMYSNEWETEEKRFYMYGDDQYPDPKSQPFQV
PKVNQAGQALRKL VQATQDQQQGIEPQQAIELQKLVQMLRMSSGQDLQQIQQHTVNQHEQQDRQK
AKQIFMDALATCGTRNCDMSTLKCAQSLNQLTNLPAPSDQIVQVQHLCTQSERVKQCPALRQSCWLT
FGALVNELCQHKTQKSAQQSVYGMQSGFAQQEQCSADKKQYKQTMLNEFKNSETTTQKVLALSA
LGNAGIDTSTPQLAEI IKDKQEQRVVRVKAIDALRRLRTQQPRQIQQILLPIFQNTHEQPEVRTAAFAILM
YTLPEQAVLDQLAYTISQEPNKQVQAFAYKTMKMLANSQNPVHEQFSKHAKNALKLANVDEQQLRAS
GKWEIPTCSKQQQEGLFISLATTISQRSSLPAYAHAQIDSVLNDQFHINDLKVYMMQQDAEHWYEQAG
YQFGAFSTKKT RGQRSSDDESSDTIEYNENSRQSLRNIYTALGIKSRRSYPTSTAQNQQQQGQQP
FAMACLRISDVTACIALDEQAGPEQMGAPIQAVLDSLKSGQKPSLARIFQFVQDGGQIRSTMAISLLE
KEAVIPMALGIPLRMTNTVPVLATIQTAKIERAESGHGPNAFKIKIEARPMVNGVHIQKMEAWTPFLVS
GVESIRSVEANIPLNVELSNNQHGLQAKVKVPEGNTKLFGAHSPLVSYCGEVQQSSGLVNEPRHVR
AIQNPKLNRQQHEKNTVVGGQSF GIPLHVQAHYHWPSQPSSFPEFTQMLMATENAIHVTCKSTQSSP
QEVIFRLNGEHFQPDNQNQHLEMSQSLNDFYASKFEQADSDENDYEDYQQEDNNQRGQKLQNFLT
TYQPRKQYKHGLKATVLTIGGVKEFRASAEVQAHCDAKLQVCQVQMDAQRNKIGSEQQRWTMQAK
AQLVLPETVSSTSEMAQLIAEKNAKFACKAECQWGADQKQKIQIRIQGEQVQRS AWRQQLQQQGN
RQQQQEQQLKRRSAFLNKFDLEAQYQGLKPSTQNTFARALELVKSAYYWNTNSELIPFNQRQQN
QGENGKARPFPLVRQNAQSIRSIGQFIGRQELMSRAECTVDGNRVNTFDNQEYSSPLSTKCYSVLAK
DCGSDQSD FVVL MKALQQGKNGQKKLKVISPMATIECQTKSCANGQCKSQSEQRAGEQIQCEVNG
QTVKTDDESQYVDFNNAQKSDVTVHVPGAISVRFN GRKAWLKVNGAYKNAQCGLCGHYDDQQDGE
NEWRMSNNKIASDLATFHRSYALQCQGE CNQQDLDDFYQENGQKIRQRYQRKQQQNDDSD EYNN
GNDDEEEWELENFDGNFQSDEWQNRNNNDYGCESDDN CN SFNNDNNQKNGKNQNGRKGK
KISPVKKT KVEMNH DICSKEPVKQCPRGTVPSDRTTSDFGSSSSSSSQQD DDDDEDNAEGETTK
NVSYICMNRSSSEARRLLRQYRQGQRLNVQSHGHSMVLPTKQPRVCRQA

>*P. scribneri* 019

MRPLLLLMAVGLCLFGNVALAKQHQNRFSPFNSNEQD TDGQQQKMRAPAFQPGRTYQFQYNAQIA
SGLIAQDEQSQADAQQKAITRIQAQAKITFGSERHAQLCLRQIRTGQLNEPVEDEQQVQPM SIFEPKQ
IPSDKKRLLERCCQFDYADGVIERVQFDQEDETWSKNMKRGVLQMIQLNLKRNNAAQGLRNEMNDQQ
GEEDQDEDNEQENG RQQQDKDTLAMS FNIP EIAVEGQCQTQYTINTANPSQRCTPDDQTPCSFNV
TKSINFKQCSKIADVAFGYQTQQQQPQCAQCQQQFQEIQMQQEGCPRGNCFYNAAGQSQSNAQHP
CAKCDPKEVKEQKVDRTTVLRYVLKCGSQSSRKSEEKCNCCQLERTELLSQYVYKNSKVENGQY
SAMHTVVS AQLQLIQVEQQQEVP TQSASQDESLMYSNEWETEEKRFYMQVPKVNQASQALRKL
VQATQDQQQGIEPQQAIELQKLVQMLRMSSGQDLQQIQQHAVNQQHEQQQDRQKAKQIFMDALATC
GTRNCDMSTLKCAQSLNQLTNLPAPSDQIVQTVQHLCTQSERVKKCPALRQSCWLTFGALVNELCQH
KTQKSAQQSVYGMQSGFAQQEQCSADKKQYKQTMLNEYKNAESISQKVLALSALGNAGIDTATPQ
LAEI IKDKQEQRVVRVKAIDALRRLRTQQPRQIQQILLPIFQNTREQPEVRTAAFAILMYTLPEQAVLDQL
AYTISQEPNKQVQAFAYKTMKMLANSQNPVHEQISKHAKNALKLANVDEQQLRASGKWEIPTCSKQQ
QEGLFISLATTISQRSSLPAYAHAQIDSVLNDQFHINDLKLYMMQQDAEQWYEQAGYQFGAFSTKKT
RGQRSSDETSSDATEYNENSRQSLRNIYSALGIKSRRSYPTSSAQNQQQQGQQQP FAMACLRISD
VD TACIALDEQAGPEQMGAPIQAVLDSLKSGQKPSLARIFQLVQDGGQIRSTMAISLLEKEAVIPMALGIPL
RMMNTVPVLATIQTAKIERAESGSGSNAFKVIEARPMVNGVHIQKMEAWTPFLVSGVESIRSVEANI
PLNAELSNSNQHGLQAKVKVPEGNTKLFGAHSPLVSYCGEVQQSSGLVNEPRHVRAIQNPKLNRQQ
HEKNTVVGEQSFGIPLHVQAHYHWPSQPSSFPEFTQMLMATENAIHVTCKSTQNSPQEVVFRNGE
HFQPGNQNHREMSQNLNDFYANKFEQADSDENDYEDYQQEDNNQRGQKLQNFLT TYQPRKQYK
HGLKATVLTIGGAKEFRASAEVQAHCDAKLQVCQVQMDAQRNKIGSEQQRWTMQAKAQLVMPETVS
STAEMAQLIAEKNAKFACKAECQWGADQKQKIQIRIQGEQVQRS AWRQQLQQQGN SRQQQQEQQR
LKRRSAFLNKFDLEAQYSGLPSTQNTFARALELVKSAYYWNTNSELIPFNQRQQQNGQENGKILAT
MVIDPITQKHANISVKTPSQLVRLQQIQLPIQARPFPLVRQNAQSIRSIGQFIGRQELMSRAECTVDGN
RVNTFDNQEYSSPLSTKCYSVLAKDCGSDQSD FVVL MKALQQGKNGQKKLKVISPMATIECQTKSCA
NGQCKSQSQQRAGEQIQCEVNGQTVKTDDESQYVDFNNAQKSDVTVHVPGAISVRFN GRKAWLKV

VNGAYKNAQCGLCGHYDDQQDGENEWRMSNNKIASDLATFHRSYALQSQGEQCNQQDLDDFYQEN
GQKIRQRYQRKQQQNDSDSEYNNNGNDDEEEWELENFNNGNFQSDERQNRNNNNNDYGCDDDDDDNC
NSFNNNNQKNGKNQKNGRKQGGKISPVKTKVVMNHDICFSKEPVKQCPKGTVPDRTTSDFGSN
SSSSSSQQDDDDDEDNAEGETKNVSYICMNRSSSEARRLLRQYRQGQVRVNVQSHGHSMVLPTKQP
RACRRA

>*P. scribneri* 064

MRPLLLLMAVGLCLFGNVTAKQHQNFRSPFNSNEQDTDGQQQKMRAPAFQPGRTYQFQYNAQIA
SGLIAQDEQSQADAQQKAITRIQAQAKITFGSDRHAQLCLRQIRTGQLNEPVEDEQQVQPMSEFEPKQ
IPSDKKRLLERCCQFDYADGVIERVQFDQDEETWSKNMKRGVLQMIQLNLKRNNAAQGLRNEMNEQQ
QGEEDQEEDNEQQENGQQDKDTLAMSFINPEIAVEGQCQTQYTINTANPSQRCTPDDQTPCSFNVT
KSINFKQCSKIADVAFGYQTQQQQPQCAQCQQQFQEQMQQEGCPRGNCFFNAAGQSQSNAQHPC
AKCDPKEVKEQKVDRTTVLRYVLKCGSESSRKSEEKCCQNCCLERTELLSSQYVYKNSKVENGGYGS
AMHTVVSQAQLQLSQVEQQQEVPQTQSAGQDESLMYSNEWETEEKRFYMQVPKVNQAGQALRKL
QATQDQQQGIQEPQQAIELQKLQMLRMSSGQDLQQIQQHAVNQHEQQDQKAKQIFMDALATCGTR
NCVKKLAEQIKKEASGFLGWNWEENDMSTLKCAQSLNQLTNLPAPSDQIVQTVQHLCTQSERVKQC
PALRQSCWLTFGALVNELCQHKTQKSAQQSVYGMQSGFAQQEQCSADKKQYKQTMLNEFKNAET
TTQKVLALSALGNAGIDTATPQLAEIHKDKQEQRVVRVKAIDALRLRTQQPRQIQILLPIFQNTREQP
EVRTAAFAILMYTLPEQAVLDQLAYTISQEPNKQVQAFAYKTMKMLANSQNPVHEQISKHAKNALKLAN
VDEQQLRASGKWEIPTCSKQQQEGLFISLATTISQRSSLPAYAHAQIDSVLNDQFHINDLKVYMMQQN
AEQWYEQAVYQFGAFSTKKTRGQRRSSDDESSDATEYNENSRLNIYTALGIKSRRSYPTSSAQ
NQQQQGQQQPFAMACLRISDVTACIALDEQAGPEQMGAPIQAVLDSLKSGQKPSLARIFQFVQDG
QQIRSTMAISLLEKEAVIPMALGIPLRMTNTVPVLATIQGMAKIERAESGNGPNFAKVKIEARPMVNGV
HIQKMEAWTPFLVSGIESIRSVEANIPLNAELSDNKQHGLQAKVKVPEGNTKLFGAHSPLVSYCGEVQ
QSSGLLNEPRHVRAIQNPKLNRQQHEKNTVVGGQSFQIPLHVEAHYHWPSQPSSFPFTQMLMATE
NAIHVTCKSTQNSPQEVVFRNLGEHFQPGNQNHREMSQSLNDFYASKFEQADSNDYEDYQQE
DNNQRGQKLQNLFTTYQPRKQYKHALKATVLTIGGAKEFRASAELQAHCDALQVCQVQMDAQRNKI
GNEQQRWMTMQAKAQLVMPETVSSTSEMAQLIAEKNAKFAKCAECQWGADQKQKIQIRIQGEQVQRS
AWRQQLQQGNSLQQQEQEQRLKRRSAFLNKFDELAQYSGLKPSTQNTFARALELVKSAYYWNTN
SELIPFNQRQQNGQENGKILATMVIDPITQKHANISVKTPSQLVRLQQIQLPIQARPFPLVRQNAQSIR
SIGQFIERQELMSRAECTVDGNRVNTFDNQEYSSPLSTKCYSVLAKDCGSDQSDQFVLMKALQQGK
NGQKKLVISPRANIECQTKSCANGQCKSQSEQQQQRAGEQIQCEVNGQTVKTDDDESQFVDFNNA
QKSDVTVHVPGAISVRFNGRKAWLVNGAYKNAQCGLCGHYDDQQDGENEWRMSNNKIASDLATF
HRSYALQSQGEQCNQQDLDDFYQENGQKIRQRYQRKQQQNDSDSEYNNNGNDDEEEWELENFDGNF
QWDERRIGNNNNDYGCDDDDDDNCNSFNQKNGKNQKNGRKQGGKISPVKTKVVMNHDICFSKE
PVKQCPRGTVPSDRTTSDFGSSSSSSSSQQDNDDEDNAEGETKNVSYICMNRSSSEARRLLRQYRQ
GQVRVNVQSHGHSMVLPTKQPRACRRA

>*P. scribneri* 111

MGIPSSSKLNSINSIFHSYSQKQISSDKRLLERCCQFDYADGVIERVQFDQDEETWSKNMKRGVLQ
MIQLNLKRNNAAQGLRNEMNDQQQGEDQQDEDNEQENGRQQQQDKDTLAMSFTIPEIAVEGQCQTQ
YTISTANPSQRCTPDDQTPCSFNVTKSINFKQCSKIADVAFGYQTQQQQPQCAQCQQQFQEQMQQE
GCPRGNCFYNAAGQSQSNAQHPCAKCDPKEVKEQKVDRTTVLRYVLKCGSQSSQKSEEKCCQNC
CLERTELLSSQYVYKNSKVENGGYGSAMHTVVSQAQLQLSQVEQQQEVPQTQSASQDESLMYSNEW
TEEKRFYMQVPKVNQASQALRKLQATQDQQQGIQEPQQAIELQKLQMLRMSSGQDLQQIQQHAVN
QQHEQQQDRQKAKQIFMDALATCGTRNCDMSTLKCAQSLNQLTNLPAPSDQIVQTVQHLCTQSERV
KQCPALRQSCWLTFGALVNELCQHKTQKSAQQSVYGMQSGFAQQEQCSADKKQYKQTMLNEFKN
AETTTQKVLALSALGNAGIDTATPQLAEIHKDKQEQRVVRVKAIDALRLRTQQPRQIQILLPIFQNTRE
QPEVRTAAFAILMYTLPEQAVLDQLAYTISQEPNKQVQAFAYKTMKMLANSQNPVHEQISKHAKNALK
LANVDEQQLRASGKWEIPTCSKQQQEGLFINLATTISQRSSLPAYAHAQIDSVLNDQFHINDLKLYMMQ
QDAEQWYEQAGYQFGAFSTKKTRGQRRSSDETSSDATEYNENSRLNIYSALGIKSRRSYPTSS
AQNQQQQGQQQPFAMACLRISDVTACIALDEQAGPEQMGAPIQAVLDSLKSSQKPSLARIFQFVQD
GQQIRSTMAISLLEKEAVIPMALGIPLRMMNTVPVLATIQGTAKIERAESGSGSHAFKVKIEARPMVNGV
HIQKMEAWTPFLVSGVESIRSVEANIPLNVELSNNQHGLQAKVKVPEGNTKLFGAHSPLVSYCGEV
QQSSGLVNEPRHVRAIQNPKLNRQQHEKNTVVGGQSFQIPLHVEAHYHWPSQPSSFPFTQMLMAT

ENAIHVTCKSTQNSPQEVIFRLDGEHFQPGNQNHREMSQSLNDFYANKFEQADSDENDYEDYQQE
DNNQRGQKLQNFLLTTYQPRKQYKHALKATVLTIGGAKEFRASAEVQAHCDALQVCQVQLDAQRNKI
GSEQQRWTMQAKAQLVMPETVSSTAEMAQLIAEKNAKFACKAECQWGADQKQKIQIRIQGEQVQRS
AWRQQLQQQGNSRQQQQEQQLRKRRSAFLNKFDELAQYSGLKPSTQNTFARALELVKSAYYWNTN
SELIPFNQRQQQNGQENGKILATMVIDPITQKHANISVKTPSQLVRLQQIQLPIQARPFPLVRQNAQSIR
SIGQFIGRQELMSRAECTVDGNRVNTFDNQEYSSPLSTKCYSVLAKDCGSDQSDFFVLMKALQQGK
NGQNKLVIFPMATIECQTKSCANGQCKSQSQQQQRAGEQIQCEVNGQTVKTDDESQYVDFNNAQ
KSDVTVHVPGAISVRFNGRKAWLKVNGAYKNAQCGLCGHYDDQQDGENEWRMSNNKIASDLATFH
RSYALQSQGEQCNQQDLDDFYQENSQKIRQRYQRKQQQNDSDSEYNNGNNDDEEWELENFDFGNFQ
SDERRSGNNNNNDYGCDDDDDDNCNSFNNNNNQKNGKNEKNRQKQKKISPVKTKVEMNHDICFSK
EPVKQCPQGTVPSPDRTTSDFGSSSSSSSQDDDDDDNAEGETTKNVSYICMNRSSSEARRLLRQYR
QGQPTKALIHPPVEAIIQPPIAANDNDEVQKNSTKAMLHPPVFEANDNDEA

>*P. scribneri* 147

MSTLKCAQSLNQLTNLPAPSDQIVQTVQHLCTQSERVKQCPALRQSCWLTFGALVNELCQHKTQKSA
QQSVYGMQSGFAQQEQCSADKKQYKQTMLNEFKNAETTTQKVLALSALGNAGIDTATPQLAEIHKD
KQEQRVVRVKAIDALRRLRTQQRPIQQILLPIFQNTREQPEVRTAAFAILMYTLPEQAVLDQLAYTISQ
EPNKQVQAFAYKTMKMLANSQNPVHEQISKHAKNALKLANVDEQQLRASGKWEIPTCSKQQQEGFLI
SLATTISQRSSLPAYAHAQIDSVLNDQFHINDLKLYMMQQDAEQWYEQAGYQFGAFSTKKTGRQRRS
SDDESSSDATEYNENSQRSLRNIYTALGIKSRRSYPTSSAQNNQQQQGQQQPFAMACLRISDVTACI
ALDEQAGPEQMGAPIQAVLDSLKSGQKPSLARIFQFVQDGGQIRSTMAISLLEKEAVIPMALGIPLRMT
NTVPVLATIQGTAKIERAESGNGPSAFKVKIEARPIVNGVHIQKMEAWTPFLVSGVESIRSVEANIPLNA
ELSNDKQHGLQAKVKVPEGNTKLFGAHSPLVSYCGEVQQSSGLVQEPRRVRAIQNPKLNRQQHEKN
TVVGQQSFGIPLHIEAHYHWPSQPSSFPEFTQMLMATENAIHVTCKSTQNSPQEVVFRNLNGEHFQPG
NQNHREMSQSLNDFYANKFEQADSDENDYEDYQQEDNNQRGQKLQNFLLTTYQPRKQYKHALKAT
VLTIGGAKEFRASAEVQAHCDALQVCQVQMDAQRNKGSEQQQRWTMQAKAQLVMPETVSSTAEM
AQLIAEKNAKFACKAECQWGADQKQKIQIRIQGEQVQRSARWQQLQQQGNSRQQQQEQQLRKRRS
AFLNKFDELAQYSGLKPSTQNTFARALELVKSAYYWNTNSELIPFNQRQQQNGQENGKILATMVIDPIT
QKHANISVKTPSQLVRIQQIQLPIQARPFPLVRQNAQSIRSIGQFIGRQELMSRAECTVDGNRVNTFDN
QEYSSPLSTKCYSVLAKDCGSDQSDFFVLMKALQQGRNFQKKLVISPMATIECQTKSCANGQCKS
QSEQQQQRAGEQIQCEVNGQTVKTDDESQYVDFNNAQKSDVTVHVPGAISVRFNGRKAWLKVNGA
YKNAQCGLCGHYDDQQDGENEWRMSNNKIASDLATFHRSYALQSQGEQCNQQDLDDFYQENGQKIR
QRYQRKQQQNDSDSEYNNGNNDDEEWELENFDFGNFQSDQWQNRNNNDYGCDDDDDDNCNAFNN
QQKNGRKQKQKISPVKTKVEMNHDICFSKEPVKQCPQGTVPSPDRTTSDFGSSSSSQDDDDDDN
EGETKNVSYICMNRSSSEARRLLRQYRQGQVRVNVQSHGHSMVLPTKQPLACRRA

>*P. scribneri* 032

MRILASVLLVALCLGAGLSQQSGQQQQPSPFRKDNVYIFQYNSQISSGLSAPSDENSADSQQNSAI
RIQAQAKKQIPDEKQRQLELPVQFHYDEGVVERVQFNEEDDVWSKNVKRAVLNMLQLNLKRNNAAQG
LRTGNEQEWQNEEENSSENGQQMSELGQTFNIPEITLEGVCQTTYAINAVNNGQWQQDEQENEKAT
ALNVTKTINFKQCAKIADVGYGFQTEQPQPQCAQCKQYWSKNQQQQNSQNGQQGQSAQQQHPCA
KCDPKKVKEQLLDRSTVLRVQLVCPSGGECGPSNSQKYALKGTELLSQYVFKQTKSQTGKFSSAMQ
TIVAATLKFRSQEQKTESVPSMSGGKDESLMFSSEIDLVDKRFYMYGDEEFSQPESSPFQQVPKVQQ
AQQALRKLSQTMGTGNKQGTGIEVQAGYEVQRLVELLRMSRREELEEIEQNAITNNGNGGEQERQKAK
QIYTDLSAICGTRNCIQKLAEKIQNKEMSQTKAVQALRTVNQVPAPSDQIVKIVQRLCKQENVHNQRSV
KQSCWLTFGALTIMDLYNSAATIIDKMLAISALGNAAIENAVPEIEQIHKDQTESCLVRCKAIDALRRLRT
QMPRKIQQILLPIFQNAQQRSEVRASALYILLTTPDQSAVIDQIAFTISKERNRQIQSFAYQQMKTAKSK
NPLDQQMAKHVKNALRLTNVDEQQLRMSGKWQFPVYSEQLQEGAFLELAQLISPRSNAPAILHAKID
SFFNDQFQTDQMQUYFVQQAEDWYQRTSSNNAQNGHSPAQRSSFNGQSLRNIFTVLGIKKRRN
DEENNRQPFAMFALRTRDQVDAFLAIDGSSKEQQDVQLQEMLGQFISGQSKPSLERFLSQFHQNNQ
QNKQKQKEFRGILVTNLNEQQAKIPTSAGLPLRLTNSPMLSTIEGQTQLSKESDHLDSPLVGTQINVH
SMASLVHIQKMESALPFMVAGVESVRSLELNVPLQAEELNANPSQNMVSVKVLPRQKHTLLGFHSLP
FTYTAQVDVNTKLVHEPRHFHAIHHRELERMQHEVSRCKLTEQFGIPLQIQGHYHWPTQSSTGIQQF
YQMLMATENVIHIVYQPNEETPREMIFRISGNHFEKTNGQRPELDDFYREEMGSASFESHLDHVKLT
AQTVGGPKEAKAQAELVASCDAMQSCNTKLEAQFGKLRAGNEKWTMKIQTNVAMPETVQSVDELV

ELEQRNNKFQAKIQAEWGTGTQRENLNIRLQGEQAKKRENQNNEQNMQTKRREQWRQAFLNKFDV
EAQYQMKARTMNYFVKVSEWLKSTYYWNSQQQSNNNQNGQMMQMRGQEKTLISLVIDPITQKHAN
VSIRTPNEQIRLQQIELPIHIRRPFPLVRQQRQSIHSFGQFIDRIQVPGRAECTVDGRDVNTFDNVEFRA
PLSDACYTVLAKDCSSDEPKFVFLMKAQDNIDSGSQNSPKKLKVLILTPETEIVCEPKSSNGNQQQIV
CKENGSPVENDGDSDAVEFNNAKSDVTVHVRGISVRFNGRKAWLKVSSVYKNGQCGLCGHYDDS
EQDEWRMSNNQNTNDLVEFHRSYTLNGAQQGEQCNEEELNNFYRQNGNKIRYNSRNQKNGKNSD
EVDDDEEDGQNDWMDEQYDGHWNENGRNNRRNGNQKRIETPVKTKLVELSNRICVSQKPIKQCP
KGYIPIYSSNEDSQEDPFDQQNGNNGEGKNVQFSCFNRTAEARDILRQMRNGNDVVHFEKREHSM
AEKIHEPKKCVRGRY

> *G. pallida* 4789

MRRCLPLFRLFALASLLVVVTCRSNYASKYDDKFRHYHHEQNGQQDDQDHDYGFQRVAREREGHL
SNDYLKKGKFGMSSDKRTPTFKSFQPGYKFQFLYDSQTTSGYAPWVSVQRAMGRFQALATFEAIGPG
EASMLSECLIGQLNGQSPTAGVWNPGQLLPVRVFESVMHFSHVQLEQPVRFYTSRDGIHHIYFDPM
EPAWSRNYKKAILSTGQFRLGAKDDGDEADQNQTNQTERAEWKTQWQQTFVHETSIEGNCETDY
TIVPFSRELISVTKSINFDRCTRLADIAYGMADEEAKQRGQHNNKPYQEQRERTDNEQQLNGGANP
MISRQFLDRSTTIRYLFVNQLKNNSMDLHSSLLPHSVELLSQYTLKGSEAEQSQAMQTAVAKLTYIE
TKQNRHKKNRQQQQDDKSGLEREQLLYTGESEQKECFFMFGDETSETVRCQRRAKAFNNDELH
GVPLVISGSRMPSPGSPFPLNFPNSTTKMRAEFCVRNVAQAQAEDNDGTLQAMSQADEFMDEFVQL
LRFSTEQELKQLKERLDNDRIAQLFADALAAAGTRNTLKLTLTRIEQGEFTELKAAQMLKTFSRGSL
PAPSDRIADALLALCKTEHVQNSELLRQTSWLSFGAIVGELMQLKRIQRQSPVGTPLGEKIEQCPDHK
MELYTKALLEQFSTGAGDDPMEKIVALKAIGNAALDEVPEVLATLISPWSEQQNDEQPKSFAKQLQE
MQRQNAIEALRRIPSARVQQLLPVYTNQQELPQTRTVALNVLMQSQLRPSDAVIEQIAFTMAKERCL
QVLAYSYTTLKALSQSPAKRHQHSQAQRIKEALRVSGLDELELEKRIRHKDIVGLSRFVQTNALYSAEEG
QGILLQASNVFTDKTSALPNHAFLSLSSLFNGQMSEELKLTVSQWNIESLLFTDAGLKELLKNGQEN
HQQQQRNGTEQQQQPSEESSSSSSWTSSAAEQLRSLWRRLSIVGRSDDEQSTDQMQQQKPVMLL
ALRVNNIDTIVLPFGIGGQQRSNSLISALKKLLAVGIGSISKLTGTSQHHRQPAVDTSILFWHERAARV
PMSTGQQLHVQHKSFLYGSIRANAERLGHSHIGLRLHPSATFGRIQKMKVWASPVHSSGVKNIGAIEA
NMPLKAQLNTLSQQKQVKNLIKVPQTPTRLFSMHTMPMTFVGTEDTQNNLSQHPSVFSHFGEQCKA
VQNQQLLSMAKTFDKHSNTEHSLQLRFHARGFLHMPRRVDKLKEWVKVLLSAENHYQVTLEPQAG
RSPHEIAFYLGWNLFQPHPTIGGLSRAELKQFLNQFTLVTPKSGDGEEADREEQWWTAAASSASSA
EESQKEEQQRRIGAYLDKLGERKAYKHTLKASMETVGSQDNRRAEFKLDNVCDERMRSCTLHAQV
QHLGQQQQPQWSADLRVQSLYPEQAPDLEHVMLHSREDKPPQQQQMFLCQAELEWNANGQLKE
MIKLNKIGEPSTEMLRLAHQFVQQQSSGSNVANSALPTNAGHPFVDKYTVVVDWQLRAAAQNAFV
RYLETMKCSHFFNVDSKIVMSSSNDDDDSEENSMISTGRQQQQQPGVKQGTITTITLSIDPITHQHI
NMTVRTADQLTVFRMVKMGSPVRPFPLVRSSPAGTFSNAAPARVVSISRVVHQMAHPFRSRAICRLD
TSKRQIRTFDGAERMDTPIPKCWTVLAKDCSAEGQSQYAVLIKSSSSDNNQMMLKVIAPDIGILLVEPH
AQQPDTKKQPQLQLHINGKQQQQKRTFTSDTPLSDLPDIEFAHSSSTEKEAVIETTTISINGAEPAGT
TGALSVHFDGQKVRITLNGQQNRQKPCGLCGNWNDAEDELSPDGSQTESAKAFNRAYSILIDDEG
EEGGECEHNGILDKFHAQQEETELDGQQLNISGTVLEQIEPVLTVLIEYPHQLCFSNPNVKQCPSGS
LTARNIAGELMDQQKQQQQQEAEEEEEEKAPKQKVPFTCIERHSVEAKRFKAQIRRGAEAVQLGDHR
PISFQVREIDVSLHYLSPTGYS

> *G. pallida* 5968

MRTLALLVLVTALLSVGTAKRHINWNDESKDYSFGRRTSADAELKPFQKQDTEYSFIYNAQIASGLHSE
EDSAVNPQQKAMTRIQCIVANIQFASERHAQLQLQQCRVGQQNEQISEPQEVQPMGAFEQTNIPQET
MQQLQKPCQFSYADGVIERVQCQSDGEDWSKNFKKAVLNMIQVASIGNNAQGLHRSQVRDEQQQV
QQQQQQDAEQGITFNIPEITLEGACQTTYTINKARRSGQQQYNVTKTINFKKCQKIADVANGFQTDQP
QAQCEQCHQYWAQQQGDQQNIQTGAATEQQSGEEHPCAKCDPKEVKENELDRSTVIRCQIQGEPS
EQYTLKGCELRSQYVYRNSKSADGSYGAAMQTVVAATLKALDVKPKKQIPALSSTEDETLMHSNEK
EVDEKRFFMVQTATKTSTAKGSPFARVPKVQQAQALLKLEQSMQDKHKIDQQAPVQQQRLVEML
RMCTQQELKQVEQQQCSSKATAQQTEQCKQIVANALAQCGTLNCVTELAKKIKTKQVSKANAVQALL
ALNQLPAPSDSILDEMEQLCQSEAVQNNQMIKQSCWLTFGALVNELCQHKAEKNAQECVGNTANCV
QSGFFKKGQCPMDKKQKYKEARSATPASTSRLTDLEQIIKDPRESRVVRVIAVDATRRRLTKLPQRVQ

NVLLPVFLNTREQPQVRIAFAFANVMDAQAQPAQQVIDQLMFAITKEPNKEVQSYAFRTMQTVAKSQKPSE
 SQIAKHIKSALKMANVDEQQLRASGTAVVWGLRREARGRVRLARPSCLAALCTACLRARAVGLGTS
 TTSTSTATSKCSCCNRRQRNGTATWPTASSPSSNSSNPRELNAANRSSSRTRRSCATSTRRWASKV
 AGSDFGSTDVDSKEMETYQQNTRAFALLGLRSNNVDKLFALDEHIMEEQTRRMLNGQEKANWAYL
 FGQQGQPLLNTVIVQNMNEKEAKIPTSAGVPLRILNSMPMLANIEGQLKIKADSANPASPLGVQFDINA
 KASARIAQIQKMEIAPFATSGVECVSRSMHVNPLHGLNANTNHGISVKAKLPQQKTQLASAQSLPVT
 FIASVDPKSNLVRERQILTVHDLKLERQQHDMNTVLGQKDWALPVHIRGHHYWPAPSSYKQFAQLL
 MATENAVHVHYQPSQDSPRELTLRATGEQFQKAAQIEQPDTQLRNFYAEFHNAYSCNAEQCQQTG
 ECEQCNEFEQQQRQDAQEVESSQVRQQLHNFAYKPRQQYKHSIKLIAQTNGGAKDATAQCEIQA
 QCDERMSYCKLQTTVNANQKWSLNAQAHVVMPEPAASVQQLQKPEQLKQFVAQTKAQWGAQQGQ
 PKQHVNLRIHGTQAHTEQIRKFLQQQQQQQQQQQQKSLISAQQQRNFQRRSAFLNKFEEATYSL
 KAYAQNTINRALEMLKSTYFWNTKSQLLNGGQSIAAIIAIDPLTLKTANVTIRTPTQHMRMQQIELPAKV
 RPFSLVRQSSAQQLHSLAQLFQQYSFTKR PQCTVDGRSVNTFDNLVYTAPLSSSCLSVLAKDCGSEQ
 PQFVVLKIMRPEQQQNAKKSQAQNAKVKLVITADQQIECQPKTANQQQQQRICKVNGQQVTE
 EQCQDGCQVQFNDPQQTDVTIRAEGIAVRFNGRKAWLKVSEAFKNTQCGLCGHYDDSDDEQELR
 MANGQIGDDELAFRFGRVPSLLRPAGLGTVHAAGVEQVLQRKEQQRRTTMTSKTLSASSTTSWTT
 TTNEQYNNEAEENEQNRDWQKEQFDGKFGSLEQDDDEDINSDFSWSIDTSDDQCCCCCNNDGE
 DNDDDKQCKQNCQDNDQCQRKCEKQQCEQQQQQQQPAPPCRAGDDKCMKKCQNGEEAACTKE
 AGWPQWKTIVQEQNHAVCFSTEPVKQCPRGFSPYSDSYSSDDDEVEQQQQQLQNVKSKKVSFVC
 MDRAGSETRLLRLRRGQRVVNVSARGASFTETVDEPQQCWRARN

> *G. pallida* 7045

MSEAEVQPMGAFEQTTIPSQTKQQLQRPCQFSYSNGVIERVQFQNDHVVWSKNFKKAVLNIIQLNL
 KRNNAAQLRKSDVRDEQEQQTEQDAWQAKQKGKTFEIPETLEGACQTTYTINKAWSGQQYNVTKTIN
 FKCCQKIADVANGFQTDQPPQAQCAQCQYWAQQQTDQNLWPGQSDVNSAQFRDEQQNPCAKCD
 PKEVKENELDRSTVIRCQIEGEPTQYALHGCELRSQYLYRNSKSAVGNNGAAMHTIVAANLQALDAK
 PMQQQQIPASASKDETLMSNEKEVDEKRFYRNGDVFDSKDGKDSFPAQVPKVQQATQALLKLMQ
 NTQDKHKGIDQVPVQQQRLVEMLRMCTQQELRRVEQQFPGTTAQQKEQCKQIFVNSLAQCGLTN
 CVAELAKRIKNQVSQATSVKALLALNQLPAPSDSILDELEQLCKSEAINSQPTSKQSCWLTFGALVNE
 VCQHKTEKNAQECVEETANCVQSGFHKGQCPTKKQYKEMLDVYQTNECIYDKVVAMSALGNA
 GLHNALTDLEKIIKNPRESRVVRIMAIEATRRLQVNAPWEVQNILLPVFLNTREQPQVRTAAAFANLLNTA
 AGAASGRSVDVHHGQRAKQGSAAVRLQNHEGVGQITEPQ

> Predicted GP_90

MPFFLLKAMLAVYQNHGCIKFDKAVAGALGNAGLHNALNDLEKILKNPQEPRVVRVMAIDATRRLGVK
 MPRKVRNVLLPVFLNTREQPQLRIAFAFANLLNAQTEPQVVDQLMFAIAREPNKEVQSYAYRTMKTAK
 SKNPNEKRIAKHIKSALKMANVDEQQLRYSGKWQFSLYNAKEREGAFISLARAVSSRSALPAYAHAQL
 DSHFNDQYQLSNVKVFMQQAQEKWYGNLADRFIKQQQQQQQTRAQRQNLQYSTKLRSIYEDLGKN
 RRSIFYDNEMDGSTKIGTYQQQNIKPFVMLCLRSNNVDKAFLALDENIMEEQTRKMLNGQEKANWAL
 LFGQQGQQLNTAILQNLNEKEAKIPTSAGVPLRILNSMPILANIEGQLKIKTESARLTSPPLGFQLDINAKAS
 ARMAQIQKMELWTPFAASGVECVRSVQVNLPLQGELNANWDQGISLKAKLPQQKTQLANAHSLPITY
 IAKVDSNTNLVREPRQILTIHDPKLERQQHEVNAVQGGQKNWAMPLHIQGHYHWPTPTSYYQQAQLL
 LATENAVQVHYQPGQNSPREAILRITGEHFQKAAQNEQPGTQLRNFFAEFISPYSCNVEQCQDGD
 CEQCSEFEQQQKEDAEVEGQVVRQQLHNFAYKPRQQYKNSLKLVAQTNGGAKDAKAQCDIQT
 QCDEQMNYCKVQVQAERPPIGSENEKWTLNQAHVLMPTAASVEQLQRKPQQFKQFIAQAKAQW
 GQQGQSQQHVNIIRMLGTQAHTQHVRALQQRQDKRLISVQQQRNFQRRSAFLNKLQIEAAYSLKAY
 AQNAINRVLEVLTSTYFWNTKSQLLSGQNAQKNGKVAIIVDSMTQKNANVTIQTAPQKVRMQQIELP
 YKVRPFSLRQSSAQQIDSLAQLFHHFNIKNRPQCTVDGRSVDTFDDFIYTAPLSSNCHSVLAKDCGSE
 EPQFVVLKMKVPQQQNNNNNNKKSAPNAEKMLVVITPKQQIECQPKTNAKGPQQRRLICKVNGQQIAG
 DEQCQKSCVQFNDPQQNDVTIRTEDIAVRFNGRKAWLKVSRAFKNTQCGLCGHYDDSDDEEEL
 RMANGKFASNLAEFHRSYALKDSELCSQQALDNFYSEQNGEQDSFLEQYNQLDNDSEQNNEAEGE
 GLESDWQKEQFDGKLDSSEKHEEEDIDNNLSDNISEPNGPKSTATGTVLLLLPGRRGRAVPAEMPY
 EQAEQNHAVCFSTQPVKQCPSGAYSYSMSDDYGSDEEEDSNNDEQQNEQKSKKVSFVCMERSD
 PETRLLRLRLRQGRVVNVSGHGASFTETVNEPQQCWHSRF

> *G. rostochiensis* 16066

MRTLALLALVAVFLSVSAGRHYTPGDYLEYAPSDGFGHSRSPNSVSEPFQQKNTYEYTFIYNAQIASG
LQSEENVANPQQKAITRVQCLAKIQFASNRLAQLQLQQCRAGQQNEQMSEPQEMQPIEAFEQANIPS
QAKLQLQQPCQFSYSDGVVERVQCQDQEDWSKNFKKAVLNMIQLNLKRNSPQKDPKQQNTEQYAW
ETKMGTFTTTTEITLEGECQTTYTINKARAGHQYSVTKTINFKKCQKIADVAKGFQTDQPQAQCAHCQ
QYWAQQQDEQNLQQVANAPQQQQQNGEQHPCANCDPKEVKENELDRSTVIRCRIQGEPEQYALN
GCELQSQYMYRNSKNADGNYGAAMQTVVATLQALDVQLKQQQIPASASDEETLLYSNGKDVDEKR
FYKNGDKDFGTNGKDSPFARVPKVQQAVQALNKLQNMQDRQKQIDQQAPIQQQRLVEMLRMCTQ
QELKQVEQQCPGTAAQQRQQCKQVVVIALAQCGTMNCVAELAKRIKNKQVPKAVVVQALLSLNQLPA
PSDGILDQMEQLCKSEAVNSQPMQKSCWLTFGALVNEVCQHKTENAEQCVGSTANCVQSGFYEK
DQCPTKQKQYKEAMLAVYQNHGCFDKAVALGALGNAGLHNLNDLEKIMKNSQEPRVVRVMAIDA
TRRLGVEMPRKVRNVLLPVFLNTREQPQLRTAAAFANLLNAQTEPQVVDQLMFAIAKEPNKEVQSYAY
RTMKTAKSQPNERRIAKHISVLKMANVDEQQLRYSQGWQFSLYNAKEREGAFISLAQAVSSRSA
LPAYAHAQLDSHFNDQYQLNNVKVFMQLQEAQWYGNLADRFITQQQQQTRAQRQNLQYSTKLRLNI
HENLGIKNRRSLSYDNEMDGSIMKIGTYKQNIQPFVMLCLRSNNVDKAFLALDEHIMEEQTRKMLNG
QEKANWALLFGQQQLNTAILQNLNEKEAKIPTSAGVPLRILNSMPVLNIEGQLKIKTESARLTSPGLL
QLDINAKASARMAQIQKMEWLTPFAASGVECVRSVQVNLPLQGELNANSDDQGISLAKLPQQKTQLA
NAHSLPITYIAKVDPKTNLVREPRQILTIHDPKLERQQHEVNAVQGGQKNWAMPLHIQGHYHWPTPTS
YQQFAQLLLATENAVQVHYQPGQNSPREAILRITGEHFQKAAQNEQPGTQLRNNFAEEFTSPYSCNV
EQCQDGDCEQCSKFEQQKEDAEVEGQQVRQQQLHSFINAYKPRQQYKNSLKLVAQTNGGAK
DAKAQCDIQTQCDEQMNYCKVQVQAERPPIGSENEKWTFFNAQAHVLMPEAASVEQLQRKPQQFK
QFIAQAKAQWGGQGSQQHVNVRMLGTQAHTQHVRALQQRQDKRLISAQQQRNFQRRSAFLNK
LQIEAAYSLSKAYAQNAINRVLEVLKSTYFWNTKSQLLSGQNAQKNGKMAAIIVIDSMTQKNANVTIRTPA
QKVLMMQIELPSKVRPFSLRHQSAQQIDSLVQLFHHFSIKNRPQCTVDGRSVDTDFDIYTAPLSSNC
HSVLAKDCGSEEPQFVVLMMKVPPQQNNNNKSASTAEKMLVVITPKQQIECQPKTNTKGPQQRRLIC
KVNGQQVAGDEQCQGTSCVQFNPPQSDVTIRTEDIAVRFNKRKAWLKVSRFAKNTQCGLCGHYD
DSDDEEELRMANGKFASNLAEFHRSYALKDSELCSQQALDNFYSEQNGKQDSFLEQYNQLDNDSE
QNEAEGEALESQWQKEQFDGKLDSSSEHDEEDIGNNLSNISEPDGPSQPPPPAPKCCCCQEGE
DEQCQKCHTSKQCDVVCKGGRCAAEQPIWKATWPTWTKIQEQNHAVCFSTQPVKQCPSGAYS
SSVSSDYGSDEEEDSNDDEQQNEQKSKRVSVFCMERADPETRRLRLRQGGQRVNVVSGHGASFT
ETVNEPQQCWHSRF

> *G. rostochiensis* 6962

MLSTGSAFSFDLNDLQFLHPTPVFAHPLLFGNRTKNAIPSASKQQQLSDKAETIEAQCLNNEREANG
GGNDDQHFVDPLNASAIQLEKRSVHQHNGFASGGGTPTAILFTERLNNTLAECESEQWPLD
GVDPSLVGHRPWVHLRQRILDTFSSTTDRLSLESSFLRANKVSGGEGDDLAQRGRSLRKRNGGTGR
KGQKNFVIEQQQPVSKFHELSQDEWTRKLKELRRILIVKWEENRLECIKLSAEVAALISTISSAQFYPA
QFVHVTDFDLDFGHLVYERLLRMAGEERRAAGLVLPVHFDNDVLETTMRKTRNWFVKINETVALL
APRIYLLAALLPCQRFMDPLGAAQENVMELCQSVKQVQHPLISSYLRCYICRVAMRLNAADRAPHWK
CLNDWLQTYSQPTFLAPAMEWIVQCFAFNAATYADLHPLWEYCRLPERRPVVLTPLMLNALPRYL
GEHALEACRIVTGSEMVSQDELCLVLANLLETDPVQARKPILKATWKCVMLASLKHVMCCCEW
VEFVARYFGTNELNVMLDNMLSKLNQQQQPDQYGLYQQLLAVLRRVDFHSRDIAFLAMDVFSTFV
DAFNAGPVREEAFLILRAVVSKEHAEASFTDIDLAYQILSMCQSLHDCVEPSLTDECVESVISLISAALDR
VDLSADPDQCLNFLVACRADLANIDGLQKYVLCRVQAFTHIARSVRYSNSRAAFLQGCLANLFIIPS
LSDPLDQIRWDDNPLNIMDDQFLDQFSVDSFVSLDFTDGLPRPTAALLPDIQQLGTNLFATLVHIPL
EQQQNEEDDDDEAAHETNGRLGEDEGKPTTKWRGPKLHFRFLHFVLCSPPLTSAQLDFVRLVR
CLHYLCIVWPSESRGLPCGVGISCKDQNDHPLLLGDNGHVTHQLQQQLQLHDDDDIDQQQPNVADL
RSICAESAQIADALEPHLTDGNASDEMCSVGVLELIEAQFGPTNAAVDADTDEEDASSALPDALLAL
AKANDLFIQSSILFDPLTSRYSMRRCPLPLFRFALASLLVVTCRSNYASKYDDRFRRHYNREQNGQ
QEDQDHDYGFQVRAREGHLSDNYLKGKFGTTSKRTPTFKSFQPGYKFQFLYDSQITSGYAPWV
SVQRAMARFQALATFEAIGPGEASMLSECLIGQLNGQSPTAGVWNPGQLLPVRVFEVMHFSHVQ
LEQPVRFTYSRDGIHHIYFDPMEPAWSRNYKKAILSTGQFRLGAKDDGDEADQNQTNTQTERAEW
PKWQQTFSVHETSIEGNCETDYTIVPFSRELISVTKSINFDRICRLADIAYGMQADEEAKQRGQHNNK
PYQEQRERTDNEQQLNGGANPMISRQFLDRSTTIRYLFVNQLKNNSMDLDYSSLLPHSVELLSQYTL
KGSEAEQSQAMQTVAVAKLTYIETKQNRHKKNKRRQQQQDDKSGLEREQLLYTGEMEQQEKCFMF

DETSETVRCQRRAKAFNDELRGVPLVISGSRMPSPGSPFPLNFPNSTTKMRRAEICVRNVAQAQAE
DNDGTLQAMSQAEFMDEFVQLLRFSTEQELKQLKERLDNDRIAALFADALAAAGTRNTLKLTLTRIE
QGEFTELKAAQMLKTF SRGSGLPAPSDRIADALLALCKTEHVQNSELLRQTSWLSFGAIVGELMQLKR
IQRQSPVGTPLGEKIEQCPDHKMELYTKALLEQFSTGAGDDKMEKIVALKAIGNAALDEVPEVLATLIN
PNWSEQQNDEQPKSFAKQLQEMQRQNAIEALRRIPSARVQQLLLPVYTNQQELPQTRIAALGVLMQS
QLLPSDAVIEQIAFTMAKERCPQVLAYSYTTLKALSQSPAKRHQHSARKEALRVSGLDELELEKRIR
HKDIVGLSRFVQTNLSYSAEEGQGILLQASNVFTGKTSALPNHAFSLSSLFNGQMSEEQLKLTVSQW
NIERLLFTDAGLKELLKNGQEDHQQQQRNDTQQQQQPSEESSSSSSWTSSAAEQLRSLWRRLSIVG
RSEDEQSTDQMQQKPVMLLALRVNNIDTVVLPFGIGGQQRNSNLISALKKLLAVGIGSISKLTGTSQH
RQPVVDTSGILFWHERAARVPMSTGQQHLVQLKSFLYGSIRANAERLGHSHIGLRLHPSATFGRIQKM
KVWASPVHSSGVKNIGAIEVNMPLKAQLNLSQQKQVKLNIQVPQTPTRLFSMHTMPMTFVGMEDT
QNNLSQQPSVFSHFGEQCKAVQNQQLLSMAKTFDKHSNTEQSFLQLPFHVRGFMHMPRRMDKLKE
WVKVLLSAENHYQVTLEPQAGRSPHEIAFYLGWNLFQPHPTIGGLSRAELKQFLNLTITPKSGDGE
EADREEQWWTAAASSASSSAEESQKEEQQRRIGAYLDKLGGERKAYKHTLKASMEIVGSDQNRQAE
FKLDSVCDERMRSCTLHAQVQHSAGVQQQQPQWSADLRVQSLYPEQAPDLEHVMLHSREDKRQQ
QQQMFLCRAELEWNANGQLKEMIKNIKGEPSTEMLR LAHQFVQQQSSGSNVVANS AVLPTNAGHP
FVDKYTAVVDWQLRPAAQNCFIYLETIKSACFFNADSKIVMSSNDDDDSEENSMISTGRQQQQQQ
QGVKQGTITITLSIDPITHQHINMTVRAADQLTVFRMVKMGTPVRPFPLVRSSPAGAFSNAAPARVMSI
SRLVHQMAHPFRSRAICRLDTSKRQIRTFDGAERIDTPIPKCWTVLAKDCSAEGQSQYAVLIKSSSSSD
NNQMKLVIAPEIGILLVEPRAQQPDTNKQPQLQLQHINEKQQQQKRTFTSDTPLSDLPDGIEFAHSS
STGKEAVIETTT SINGAEPIAGTTGALSVHFDGQKVRITLNGQQNRQKPCGLCGNWNDAEDELSP
DGSQTESAKAFNRAYS LIDDEGEEGGECEHNGILDKFHAQQEETELYGQQQQQLNTSGTVLEQIEPV
LKTVLIEYPHQLCFSLKPVKQCPSSSLTARNIAGELMDQQKQQQQQEAEEEEEEEPKQKQVPFTCIE
RHNVEAKRFKAQARRGEAVQLGDHRPISFQEVIIHADRCIKFDG

> *G. rostochiensis* 8204

MRTLVLAFVAAFLCLSAGRQYNSNPPARAGFLGFNPNSALKPFPQKNIEYSFLYNAQIASGLQSEEN
AANPQQKAVTRIQCLAKIQFVSERHGQLQLQQCQAGQQNAQMSEAEVQPMGAFEQTTPSQTQKQ
QLQRPCQFSYSNGNFKKAVLNIIQLNLKRNNAGQLRGSDVRDEQEQQTEQDAWQAKQGKTFEIPEIT
LEGACQTTYTINKAWSGQQYNVTKTINFKKCQKIADVAYGFQTDQPQTQCAQCQQYWAQQQNGQN
LWPGQSDVNNAQFRDEQQNPCAKCDPKEVKENELDRSTVIRCQIQGEPTQYALHGCELRSQYLYR
NSKSAVGNGAAMHTIVAANLQALDAKPMQQQQQQQIPASASKDETLMYSNEKEVDEKRFYRNGDV
DFSTDGKDSPFAQVPKVQATQALLKLMQNTQDKHKGIDQQVPVQQQRLVEMLRMCTQQELKRVE
QQCPGTTAQQKEQCKQIFVNSLAQCGTLNCVAELAKRIKNKQVSQATSVKALLALNQLPAPSDSILDE
LEQLCKSEAVNSQPTSKQSCWLTFGALVNEVCQHKTENAQECVGETANCVQSGFHKKGQCPTTEKK
QKYKEAILDVYQTNECIYDKVVAMSALGNAGLHNALTDLEKIIKNPRESRVVRILAIEATRRLRVNAPWE
VQNILLPVFLNLTREQPQVRTAAAFANLLNTQPGPQVVDQLMYTMAKEPNKEVQSYVYRTMKALAKSQN
PSENKIAKHIKSALKMANMDEQQLRSSGKWQYSVFNAEEREGVVFVSFAQAVSSRSALPAYAHAQLDS
HFNDQYQLNNIKVFMQLQQAENWYGNLADRFIKQKQHTRAQRNQQNSQRLHNIYEALGIKNRRAD
FGTELDDSKFQKQNIKPFAMLCRLSNNDVKAFLALDEHTMEEQTRKILNGEEKANWALLFGGQQQL
NTAIVQNLNEKEAKVPTSIGVPLRILNSMPVLANIEGQIKIKVESARLTSPLGIQFDVNAKASGRLAQIQK
MELWTPFAASGVECVRSMQVNLPLQGELNANSDDQGISLKAKLPQQKTDLASAHSPLITYIAKVDPKAN
LVREPRQILTIHDPKLEQQQHEVNVVQGQKNLAMPLHFRGHFWPAEPTSYYQFAQLLLATENAVHV
HYQPSRSDSPREVIFRVTGEQFQQVAQHEKPDMQLRKFFAEFHNAYSCNAEQCQGGDCEQCSKF
EQQQQEDAQEIEGQEQRQMLNNFIGAYKPREQYKHSCLKLIAQTNGGAKDAKAQCDIQAQCDEQMN
YCKVQLKAERPPIANENMKWTLNAQSHLVMPETAASIEQLQQKPQQLKQFIAQAEAWGHGQPKQ
HVNIRMHGHTQAHTQQVRAVLQQQNKRLISAQQKRNFRRS AFLNKFQIEAAYS LKSNAQNAINRVL
EVLKSTYFWNTKSQMLLSGQNAQKNGKMAATILIDPMTQKNANVTIRT PAQQVRMQQIELPSKVRPF
SLIRQQSARQTNSFSQLFHNFNINNR PQCTVDGRNVNTFDDLIYTAPLSSNCHSVLAKDCGSDQPQF
VVLMMKKVPQQQNNKKS AQNSEKVLKVITPEHQIECQPKTNANS PRQLICRV DGKQITGDEQCQDNYC
VQFNNPQQNDLT VQTEGIAVRFNGRKAWLKISQAYKNTQCGLCGHYDDSDDQEEELRMANGQTTSN
LAEFHRSYVLEDSEQCSQQVLNNFYNEQSSEQDTYYQQYNQQFDNNSEQEKEAEDEDDQDQVDW
QKEQFDGQFNSLEEDNEDNVNRPLEIIPNISAQKCCCCCDDDEDGQCQQNCQDNDECQEKCDQCE
QSPPQCCCCCQEGDNQCQEKQCTSKQCDANCKDDRCAEQLAVWPTWTKIHEESHAVCFSTRPV
KQCPQGAQAYSSLTNDYATGVEQKSKKVNFCMERSEPETRRLSLLRRGQRVVDVNARGASFTQN
VSEPQQCWNRNY

> *G. rostochiensis* 8560

MRTLALLVLVTALLSVGTGKRHIKWNDENGGYSFGGRTSADSELKPFQQKDTEYSFIYNAQIASGLQ
SEEDSAVNPQQKAMTRIQC VANIQFASERHAQLQLQKCRVGQQNEQISEPQEVQPMGAFEQTNIPD
QTMQQLQKPCQFSYADGVIERVQCQSDGEDWSKNFKKAVLNMIQMNLKRNNAAQGLHRSQVRDEQQ
QVQQQQDAEQGITFNIPEITLEGACQTTYTINKARRSGQQQYNVTKTINFKKCQKIADVANGFQTDQP
QAQCEQCHQYWAQQQGGDEQNIQTGATTEQQSGEEHPCAKCDPKEVKENELDRSTVIRCQIQGEPS
EQYTLKGCELRSQYVYRNSKSADGNYGAAMQTVVAATLKALDVKPKKQQVPALSSSTEDETLMHSNE
KEVDEKRFFMNGDKDFNSKGS PFARVPKVQQAQALLKLEQSTQDKHKGIDQQAPVQQQRLVEMLR
MCTQQELKQVEQQQCSSKTTAQQTEQCKQIVANALAQCGTLNCVTELAKKIKTKQVSKANAVQALLA
LNQLPAPSDSILDELEQLCQSEAVQNNQMIKQSCWLTFGALVNELCQHKA EKNAQECVGNTANCVQS
GFFKKGGQCPMDKKQKYKEAMLALYQNNQENCIYDKAVALGALGNAGLDIALTDLEQIIKDPRESRVVRI
AVDATRRLRRTKLPQRVQNVLLPVFLNTREQSQVRIAAFANVMDAQPAQQVIDQLIFAITKEPNKEVQSY
AFRTMKTVAKSQKPSESQIAKHIKSALKMANVDEQQLRSSGLWQFGAISAVQREGVFFSLAQAA SPR
SALPAYAHAQLDSHFNDQYQHSKIKVFM LQQEAEKWYGNMANRFFNKQQQPTRAERGKPPQQQNN
QKLRNIYEALGIKSRRSDFGSTDADSKEMETYQNNRAFAM LGLRSNNVDKLFWALDEHIMEEQTRR
MLNGQEKANWAYLFGQQGQPLLNTAIVQNMNEKEAKIPTSAGVPLRILNSMPMLANIEGQLKIKADSA
NPASPLGVQLDINAKASARIAQIQKMEIAPFATSGVECVRSMHVN LPLHGELNANTNHGISVKAKLPQ
QKTQLASASLPVTFIASVDPKSNLVREPRQILTVHDLKLERQQHEMNTVLGQKDWALPVHIRGHHY
WPAMPTS YQQFAQLLMATENAVHVHYQPSQDSPRELTLRATGEQFQKAAQIEQPDTPLRNFYAE EFH
NAYSCNAEQCQAGECEQCNEFEQQQRQDAQEVESSQVRQQLLHNFINAYKPRQQYKHSIKLIAQT
NGGAKDATAQCEIQAQCDERTSYCKLQTTVNANQKWTLNAQAHVVMPEPAASVQQLQKQPEQLKQ
FVAQTKAQWGAQGGQPKQHVNLRHGTQAHTEQIRKFLQQQQQ GKQLISAQQQRNFQRRSAFLNKFE
LEATYSLKAYAQNAINRVLEV LKSTYFWNTKSQLLNGGQSVQKNGKIAAILAIDPLTLKTANVTIRTPTQ
HMRMQQIELPAKVRPFS LVRQQSAQQLHSLAQLFQQYSLTKRPQCTVDGRSVNTFDNLVYTAPLSSS
CLSVLAKDCGSEQPQFVLMKIMRPQQQQQQQQNARKSVQNANKVLKVITADQQIECQPKTANQQQ
QQRICKVNGQQVTEDEQCQDGGCVQFNDPQQTDVTIRAEGIAVRFN GRKAWLKVVSQAFKNTQCGL
CGHYDDSDDDQEQELRMANGQIASNLA EFHRSYALQDSEQCTQQVLNKFYSEKNTKDDDDNDEQDT
FRQQYNQLDEDEDDNNNNEQYNNEAENEQNRDWQKEQFDGQFGSLEQDDDEDIDSDSFSWIS
DTSDDQCCCCCNNDGEDDDDKQCKQNCQDNDQCQRKCEKQQCEQQQQQQPAPPCRAGDDKCM
KKCQNGEAACPKETGWPMWKTIVQE QNHAVCFSTEPVKQCPRGSSPYDSYSSNDDEVEQQQK
QQLQLQNVKSKKVSFVCM DRAGSETRLLRQLRRGQRVVNVSARGASFTETVDEPQQCLRARN

> *G. rostochiensis* 10898

MRTLALLVLVTALLSVGTGKRHIKWNDENGGYSFGGRTSADSELKPFQQKDTEYSFIYNAQIASGLQ
SEEDSAVNPQQKAMTRIQC VANIQFASERHAQLQLQKCRVGQQNEQISEPQEVQPMGAFEQTNIPD
QTMQQLQKPCQFSYADGVIERVQCQSDGEDWSKNFKKAVLNMIQMNLKRNNAAQGLHRSQVRDEQQ
QVQQQQDAEQGITFNIPEITLEGACQTTYTINKARRSGQQQYNVTKTINFKKCQKIADVANGFQTDQP
QAQCEQCHQYWAQQQGGDEQNIQTGATTEQQSGEEHPCAKCDPKEVKENELDRSTVIRCQIQGEPS
EQYTLKGCELRSQYVYRNSKSADGNYGAAMQTVVAATLKALDVKPKKQQVPALSSSTEDETLMHSNE
KEVDEKRFFMNGDKDFNSKGS PFARVPKVQQAQALLKLEQSTQDKHKGIDQQAPVQQQRLVEMLR
MCTQQELKQVEQQQCSSKTTAQQTEQCKQIVANALAQCGTLNCVTELAKKIKTKQVSKANAVQALLA
LNQLPAPSDSILDELEQLCQSEAVQNNQMIKQSCWLTFGALVNELCQHKA EKNAQECVGNTANCVQS
GFFKKGGQCPMDKKQKYKEAMLALYQNNQENCIYDKAVALGALGNAGLDIALTDLEQIIKDPRESRVVRI
AVDATRRLRRTKLPQRVQNVLLPVFLNTREQSQVRIAAFANVMDAQPAQQVIDQLIFAITKEPNKEVQSY
AFRTMKTVAKSQKPSESQIAKHIKSALKMANVDEQQLRSSGLWQFGAISAVQREGVFFSLAQAA SPR
SALPAYAHAQLDSHFNDQYQHSKIKVFM LQQEAEKWYGNMANRFFNKQQQPTRAERGKPPQQQNN
QKLRNIYEALGIKSRRSDFGSTDADSKEMETYQNNRAFAM LGLRSNNVDKLFWALDEHIMEEQTRR
MLNGQEKANWAYLFGQQGQPLLNTAIVQNMNEKEAKIPTSAGVPLRILNSMPMLANIEGQLKIKADSA
NPASPLGVQLDINAKASARIAQIQKMEIAPFATSGVECVRSMHVN LPLHGELNANTNHGISVKAKLPQ
QKTQLASASLPVTFIASVDPKSNLVREPRQILTVHDLKLERQQHEMNTVLGQKDWALPVHIRGHHY
WPAMPTS YQQFAQLLMATENAVHVHYQPSQDSPRELTLRATGEQFQKAAQIEQPDTPLRNFYAE EFH
NAYSCNAEQCQAGECEQCNEFEQQQRQDAQEVESSQVRQQLLHNFINAYKPRQQYKHSIKLIAQT
NGGAKDATAQCEIQAQCDERTSYCKLQTTVNANQKWTLNAQAHVVMPEPAASVQQLQKQPEQLKQ
FVAQTKAQWGAQGGQPKQHVNLRHGTQAHTEQIRKFLQQQQQ GKQLISAQQQRNFQRRSAFLNKFE
LEATYSLKAYAQNAINRVLEV LKSTYFWNTKSQLLNGGQSVQKNGKIAAILAIDPLTLKTANVTIRTPTQ

HMRMQQIELPAKVRPFSVLRQQSAQQLHSLAQLFQQYSLTKRPQCTVDGRSVNTFDNLVYTAPLSSS
CLSVLAKDCGSEQPQFVVLKIMRPQQQQQQQQNARKSVQNANKVLKVITADQQIECQPKTANQQQ
QQRICKVNGQQVTEDEQCQDGGCVQFNDPQQTDTVIRAEGIAVRFNGRKAWLKVSQAFKNTQCGL
CGHYDDSDDEQEQLRMANGQIASNLAEFHRSYALQDSEQCTQQVLNKFYSEKNTKDDDDNDEQDT
FRQQYNQLDEDEDDNNNNEQYNNEAEENEQNRDWQKEQFDGQFGSLEQDDDEDIDSDSFSWIS
DTSQQCCCCCNNDGEDDDDKQCKQNCQDNDQCQRKCEKQQCEQQQQQPAPPCRAGDDKCM
KKCQNGEAACPKETGWPMWKTIVQEQNHAVCFSTEPVKQCPRGSSPYDSYSSNDDEVEQQQK
QQLQLQNVKSKKVSFVCMRAGSETRLLRQLRRGQRVVNSARGASFTETVDEPQQCLRARN

> *G. rostochiensis* 18657

MLSTGSAFSFDLNDLQFLLHPTPVFAHPLLFGNRTKNAIPSASKQQQLSDKAETIEAQCLNNEREANG
GGNDDQHFDPLNASAIERTAQLEKRHSVHQHNGFASGGGTPTAILFTERLNNTLAECESEQWPLD
GVDPSLVGHRPWVHLRQRILDTFSSTTDRLSLESSFLRANKVSGGEGDDLAQRGRSLRKRNGGTGR
KGQKNFVIEQQPVSKFHLSQDEWTRKLKELRRLIVKWEENRLECIKLSAEVAALISTISSAQFYPA
QFVHVTDFDLFGHLVYERLLRMAGEERRAAGLGVLPHFDNDVLETTMRMKTNRNWFVKINETVALL
APRIYLLAALLPCQRFMDPLGAAQENVMELCQSVKQVQHPLISSYLRCYICRVAMRLNAADRAPHWK
CLNDWLQTYSSQPTFLLAPAMEWIVQCFAFNAATYADLHPLWEYCRLPERRPVVLTPLMLNALPRYL
GEHALEACRIVTGSEMVSGDELCVLGANLLETDPVQARKPILKATWKCVMRLASLKHVVMCCECW
VEFVARYFGTNELNVMLDNMLSKLNQQQQPDKYGLYQPQLLAVLRRVFDHSRDIAFLFAMDVFSTFV
DAFNAGPVREEAFLILRAVVSXKHAESFTDIDLAYQILSMCQSLHDCVEPSLTDECVESVISLISAALDR
VDLSADPDQCLNFLVACRADLANIDGLQKYVLCRVQAFTHIARSVRYSNSRAAFLQGCLANLFIIPS
LSDPLDQIRWDDNPLNIMDDQFLDQFSVDSFVSFLDTFDGLPRPTAALLPDIQQLGTNFLATLVHIPL
EQQQNEEDDDDEAAHETNGRLGEDEGTKPTTKWRGPKLHFRFLHFVLCSPPLTSAQLDFVRLVR
CLHYLCIVWPSERSGLPCGVGISCKDQNDHPLLLGDNGHVTHQLQQQLLHDDDDIDQQQPNVADL
RSICAESAQIADALEPHLTDGNASDEMCSVGVELLELIAQFGPTNAAVDADTDEEDASSALPDALLAL
AKAVYKQCKRRVASDPDITSRYSMRRCPLPLFRLFALASLLVVVTCRSNYASKYDDRFRHYNREQNG
QQEDQDHDYGFQVRAREREGHLSNDYLGKFGTTSKRTPTFKSFQPGYKFQFLYDSQITSGYAPW
VSVQRAMARFQALATFEAIGPGEASMLSECLIGQLNGQSPTAGVWNPQQLLPVRVFESVMHFVSHV
QLEQPVRFTYSRDGIHHIYFDPMEPAWSRNYKKAILSTGQFRLGAKDDGDEADQNQTNQTERAEW
KPKWQQTFVSVHETSIEGNCETDYTIVPFSRELISVTKSINFDRICRLADIAYGMQADEEAKQRGQHNN
KPYQEQRERTDNEQQNLGGANPMISRQFLDRSTTIRYLFVNQLKNNSMDLDYSSLLPHSVELLSQYT
LKGSEAEQSQAMQTVAVAKLTYIETKQNRHKKNKRQQQQQDDKSGLEREQLLYTGEMEQQEKCFMF
GDETSETVRCQRRAKAFNDELRGVPLVISGSRMPSPGSPFPLNFPNSTTKMRRAEICVRNVAQAQA
EDNDGTLQAMSQADEFMDEFVQLLRFSTEQELQKLERLDNDRIAQLFADALAAAGTRNTLKLTLTRI
EQGEFTELKAAQMLKTFSRGSGLPAPSDRIADALLALCKTEHVQNSELLRQTSWLSFGAIVGELMQLK
RIQRQSPVGTPLGEKIEQCPDHKMELYTKALLEQFSTGAGDDKMEKIVALKAIGNAALDEVPEVLATLI
NPNWSEQQNDEQPKSFAKQLQEMQRQNAIEALRRIPSARVQQLLPVYTNQQELPQTRIAALGVLM
QSQLLPDAVIEQIAFTMAKERCPQVLAYSYTTLKALSQSPAKRHQHSARKEALRVSGLDELELEKR
IRHKDIVGLSRFVQTNLSLYSAEEGQGILLQASNVFTGKTSALPNHAFSLSSSLFNGQMSEEQLKLTVSQ
WNIERLLFTDAGLKELLKNGQEDHQQQQQRNDTQQQQQPSEESSSSSSWTSSAAEQLRSLWRRLSIV
GRSEDEQSTDQMQQQKPVMLLALRVNNIDTVLPFGIGGQQRSNSLISALKKLLAVGIGSISKLTGTSQ
HRQPVVDTSGILFWHERAARVPMSTGQQLHVQLKSFLYGSIRANAERLGHSHIGLRLHPSATFGRIQK
MKVWASPVHSSGVKNIGAIEVNMPLKAQLNTLSQQKQVKLNIQVPQTPTRLFSMHTMPMTFVGMED
TQNNLSQQPSVFSHFGEQCKAVQNQQLSMAKTFDKHSNTEQSFLQLPFHVRGFMHMPRRMDKLK
EWWKVLLSAENHYQVTLEPQAGRSPHEIAFYLGWNLFQPHPTIGGLSRAELKQFLNFTLITPKSGDG
EADREEQWWTAAASSASSSAEESQKEEQQRIGAYLDKLGERKAYKHTLKASMEIVGSDQNRQA
EFKLDVCDERMRSCTLHAQVQHSAGVQQQQPQWSADLRVQSLYPEQAPDLEHVMLHSREDKRQ
QQQQMFLCRAELEWNANGQLKEMIKLNIKGEPESTEMLRLAHQFVQQQSSGSNVVANSVAVLPTNAGH
PFVDKYTAVVDWQLRPAAQNCFIRYLETIKSACFFNADSKIVMSSNNDDDDSEEENSMISTGRQQQQQ
QQGVKQGTITITLSIDPITHQHINMTVRAADQLTVFRMVKMGTPVRPFPLVRSSPAGAFSNAAPARVM
SISRLVHQMAHPFRSRAICRLDTSKRQIRTFDGAERIDTPIPKCWTVLAKDCSAEGQSQYAVLIKSSSS
SDNNQMKLVIAPEIGILLVEPRAQQPDNTKQPQLQLQLHINEKQQQQKRTFTSDTPLSDLPDGIEFAH
SSSTGKEAVIETTTISINGAEPAGTTGALSVMHFDGQKVRLTLNGQQNRQKPCGLCGNWNDAEDEL
LPDGSQTESAKAFNRAYSLIDDEGEEGGECEHNGILDKFHAQQEETELYGQQQQQLNTSGTVLEQIE
PVLKTVLIEYPHQLCFSLKPVKQCPSSSLTARNIAGELMDQQKQQQQQEAEEEEEEEEAPKQKVPFT
CIEHNVAKRFKAQARRGEAVQLGDHRPISFQEVIIIEADRCIKFDG

> *G. rostochiensis* 2260

MRTLALLALVAVFLSVSAGRHYTPGDYLEYAPSDGFGHSRSPNSVSEPFQQKNTYEYTFIYNAQIASG
LQSEENVANPQQKAITRVQCLAKIQFASNRLAQLQLQQCRAGQQNEQMSEPQEMQPIEAFEQANIPS
QAKLQLQQPCQFSYSDGVVERVQCQQDEDWSKNFKKAVLNMIQLNLKRNSPQKDPKQQNTEQYAW
ETKMGTFTTTTEITLEGECQTTYTINKARAGHQYSVTKTINFKKCQKIADVAKGFQTDQPQAQCAHCQ
QYWAQQQDEQNLQQVANAPQQQQQNGEQHPCANCDPKEVKENELDRSTVIRCRIQGEPEQYALN
GCELQSQYMYRNSKNADGNYGAAMQTVVATLQALDVQLKQQQIPASASDEETLLYSNGKDVDEKR
FYKNGDKDFGTNGKDSPFARVPKVQQAVQALNKLQNMQDRQKQKIDQQAPIQQQRLVEMLRMCTQ
QELKQVEQQCPGTAAQQRQQCKQVVIALAQCGTMNCVAELAKRIKNKQVPKAVVQALLSLNQLPA
PSDGILDQMEQLCKSEAVNSQPMQKQSCWLTFGALVNEVCQHKTENAEQCVGSTANCVQSGFYEK
DQCPTKQKQYKEAMLAVYQNHGCFDKAVALGALGNAGLHNALNDLEKIMKNSQEPRVVRVMAIDA
TRRLGVEMPRKVRNVLLPVFLNTREQPQLRTAAAFANLLNAQTEPQVVDQLMFAIAKEPNKEVQSYAY
RTMKTAKSQNPNERIAKHISVLKMANVDEQQLRYSQGWQFSLYNAKEREGAFISLAQAVSSRSA
LPAYAHAQLDSHFNDQYQLNNVKVFMQLQQAQEKWYGNLADRFITQQQQQTRAQRQNLQYSTKLNI
HENLGIKNRRSLSYDNEMDGSMKIGTYKQNNIQPFVMLCLRSNNVDKAFLALDEHIMEEQTRKMLNG
QEKANWALLFGGQQQLNTAILQNLNEKEAKIPTAGVPLRILNSMPVLANIEGQLKIKTESARLTSPGL
QLDINAKASARMAQIQKMEWLTPFAASGVECVRSVQVNLPLQGELNANSDDQGISLAKLPQQKTQLA
NAHSLPITYIAKVDPKTNLVREPRQILTIHDPKLERQQHEVNAVQGGQKNWAMPLHIQGHYHWPTPTS
YQQFAQLLLATENAVQVHYQPGQNSPREAILRITGEHFQKAAQNEQPGTQLRNFFAEFTSPYSCNV
EQCQDGDCEQCSKFEQQQKEDAEVEGQQVRQQQLHSFINAYKPRQQYKNSLKLVAQTNGGAK
DAKAQCDIQTQCDEQMNYCKVQVQAERPPIGSENEKWTFFNAQAHVLMPEAASVEQLQRKPQQFK
QFIAQAKAQWGQQGQSQQHVNVRLGTQAHTQHVRALQQRQDKRLISAQQQRNFQRRSAFLNK
LQIEAAYSLSKAYAQNAINRVLEVLLKSTYFWNTKSQLLSGQNAQKNGKMAAIIVIDSMTQKNANVTIRTPA
QKVLMMQIELPSKVRPFSLRHQSAQQIDSLVQLFHHFSIKNRPQCTVDGRSVDTDFDIYTAPLSSNC
HSVLAKDCGSEEPQFVVLMMKVPQQQNNNNKSASTAEKMLVVITPKQQIECQPKTNTKGPQQRRLIC
KVNGQQVAGDEQCQGTSCVQFNPPQSDVTIRTEDIAVRFNKRKAWLKVSRFAKNTQCGLCGHYD
DSDDEEELRMANGKFASNLAEFHRSYALKDSELCSQQALDNFYSEQNGKQDSFLEQYNQLDNDSE
QNEAEGEALESDWQKEQFDGKLDSSSEHDEEDIGNNLSNISEPDGPSQPPPPAPKCCCCQEGE
DEQCQKCHTSKQCDVVCKGGRCAAEQPIWKATWPTWTKIQEQNHAVCFSTQPVKQCPSGAYS
SSVSSDYGSDEEEDSNDDEQQNEQKSKRVSVFCMERADPETRLLRLLRQGGQRVNVVSGHGASFT
ETVNEPQQCWHSRF

> *G. rostochiensis* 2396

MRTLVLAFVAAFLCLSAGRQYNSNPPARAGFLGFNPNSALKPFPQKNIEYSFLYNAQIASGLQSEEN
AANPQQKAVTRIQCLAKIQFVSEHGLQLQLQQCQAGQQNAQMSEAEVQPMGAFEQTTPSQTQKQ
QLQRPCQFSYNGVIERVQFQNDHVVWSKNFKKAVLNIIQLNLKRNNAAQLRGSDVRDEQEQQTEQ
DAWQAKQGKTFEIPEITLEGACQTTYTINKAWSGQQYNVTKTINFKKCQKIADVAYGFQTDQPQTQCA
QQCQYWAQQQNGQNLWPGQSDVNNAQFRDEQQNPCAKCDPKEVKENELDRSTVIRCRIQGEPE
QYALHGCELRSQYLYRNSKSAVGNGGAAMHTIVAANLQALDAKPMQQQQQQQIPASASKDETLMS
NEKENGVDVDFSTDGKDSFPAQVPKVQQATQALLKLMQNTQDKHKGIDQQVPVQQQRLVEMLRMCT
QQELKRVEQQCPGTAAQQKEQCKQIFVNSLAQCGTLNCVAELAKRIKNKQVSQATSVKALLALNQLP
APSDSILDELEQLCKSEAVNSQPTSKQSCWLTFGALVNEVCQHKTENAEQCVGETANCVQSGFHKK
GQCPTKQKQYKEAILDVYQTNECIYDKVAMSALGNAGLHNALTDLEKIKNPRESRVVRILAIEATR
LRVNAPWEVQNILLPVFLNTREQPQVRTAAAFANLLNTQPGPQVVDQLMYTMAKEPNKEVQSYVYRT
MKALAKSQNPSENKIAKHISALKMANMDEQQLRSSGKWQYSVFNAEEREQVFSFAQAVSSRSAL
PAYAHAQLDSHFNDQYQLNNIKVFMQLQQAENWYGNLADRFIKQKHTRAQRNNQNSQRLHNIYE
ALGIKNRRADFGTELDSDKEFDKNQQNIKPFAMLCCLRSNNVDKAFLALDEHTMEEQTRKILNGEEKAN
WALLFGGQQQLNTAIVQNLNEKEAKVPTSIGVPLRILNSMPVLANIEGQIKIKVESARLTSPGLIQFDVN
AKASGRLAQIQKMEWLTPFAASGVECVRSQVNLPLQGELNANSDDQGISLAKLPQQKTDLASAHS
PITYIAKVDPKTNLVREPRQILTIHDPKLEQQQHEVNVVQGGQKNLAMPLHFRGHFWPAEPTSYQQFA
QLLLATENAVHVHYQPSRDSPREVIFRVTGEQFQQVAQHEKPDMLRKFFAEEFHNAYSNAEQCQ
QGGDCEQCSKFEQQQKEDAEIEGQEQRQMLNFIGAYKPREQYKHSLKLIQNTNGGAKDAKAQ
CDIQAQCDEQMNYCKVQLKAERPPIANENMKWTLNAQSHLVMPETAASIEQLQKQPKQLKQFIAQAE
AQWGHQGGQPKQHVNIHMHGTQAHTQQVRAVLQQQNKRLISAQQQRNFQRRSAFLNKFQIEAAYSL
KSNAQNAINRVLEVLLKSTYFWNTKSQMLLSGQNAQKNGKMAATILIDPMTQKNANVTIRTPAQQVRM

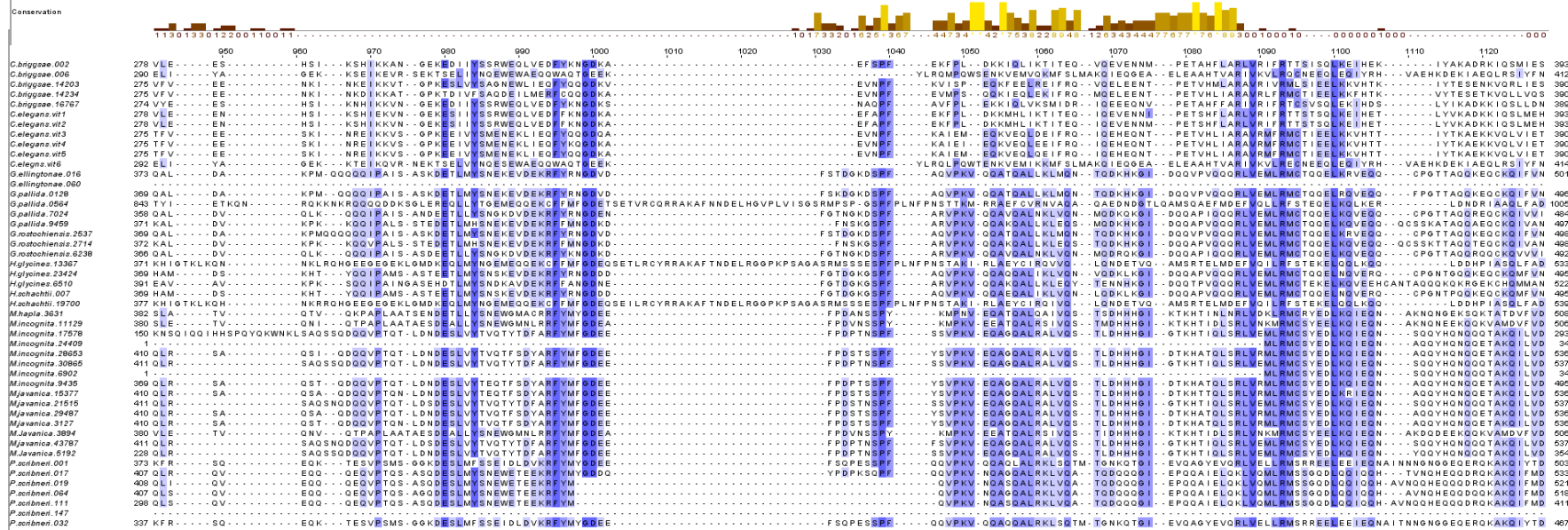
QQIELPSKVRPFSLIRQQSARQTNSFSQLFHNFNINNRPQCTVDGRNVNTFDDLIYTAPLSSNCHSVL
AKDCGSDQPQFVVLMMKKVPQQQNNKKSANSEKVLKVITPEHQIECQPKTNANSPRQLICRVDGKQI
TGDEQCQDNYCVQFNNPQQNDLTVQTEGIAVRFNNGRAWLKISQAYKNTQCGLCGHYDDSDDQEE
ELRMANGQTTSNLAEFHRSYVLEDSEQCSQQVLNNFYNEQSSEQDTYYQQYNQQFDNNSEQEKNE
AEDEDQQDQVDWQKEQFDGQFNSLEEDNDEDNVNRPLEIIPNISAQKCCCCCDEDEDGQCQQNCQDN
DECQEKCDQCEQSPPPQCCCCCQEGDNQCQEKQTSKQCDANCKDDRCAEQLAVWPTWTKIHEE
SHAVCFSTRPVKQCPQGAQAYSSLTNDYATGVEQKSKKVNFCMERSEPETRLLSLLRRGQRVVD
VNARGASFTQNVSEPQQCWRNRY

Supplementary File 3.3: Full sequence alignment of putative *vit* genes.



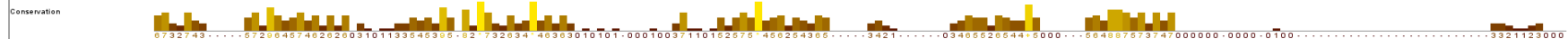
[illegible]

Conservation



[illegible]

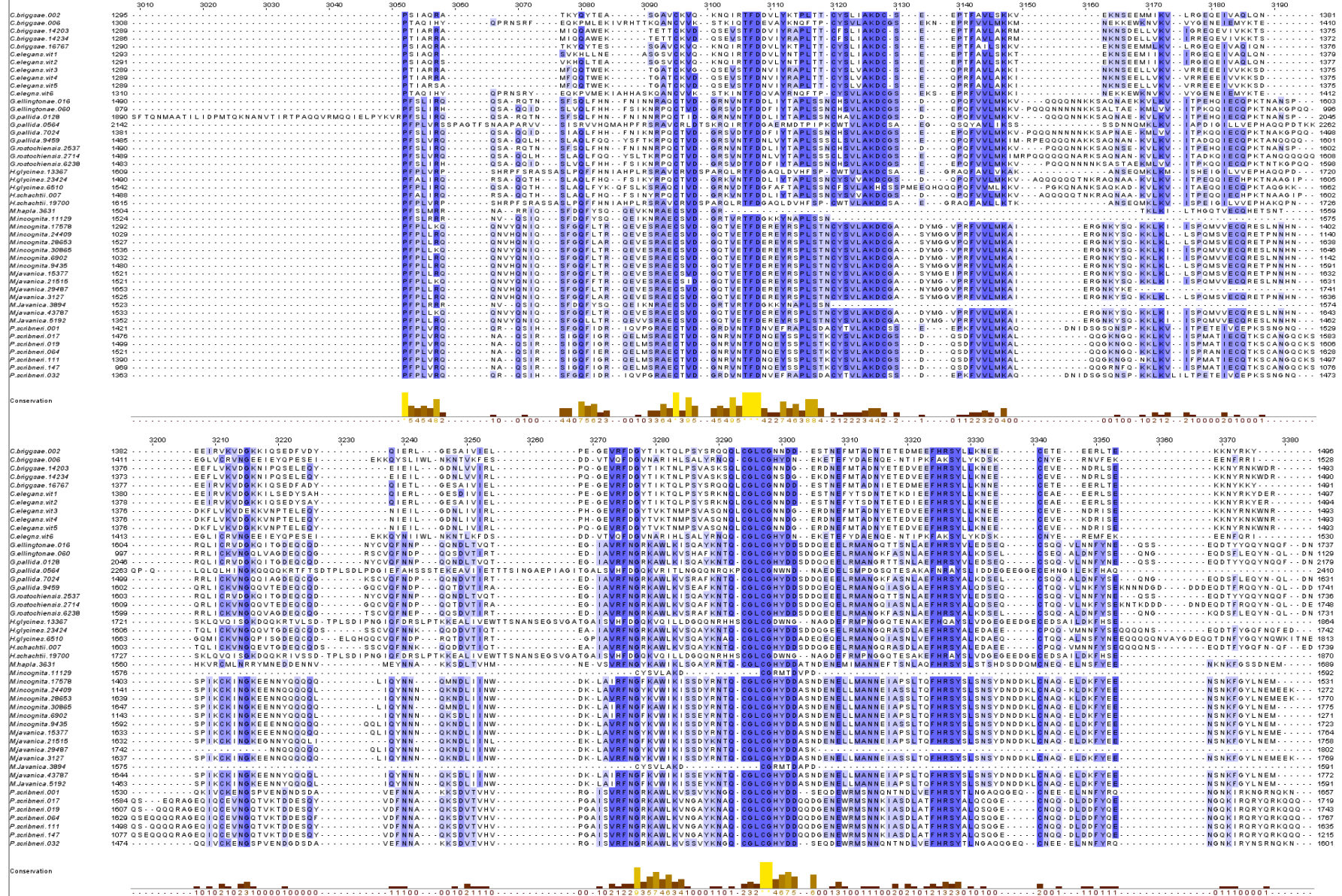


[illegible]

	2070	2080	2090	2100	2110	2120	2130	2140	2150	2160	2170	2180	2190	2200	2210	2220	2230	2240	2250
<i>C.biggusae</i> .002	954	S -	KFEYIEA	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.biggusae</i> .006	985	E -	AROYPEP	VEMTLMLRNDQLQSLD	HEYKOF	RVGOTLNR	TSGRV	POWIV	DNTEVLEP	HKIEKYAMLEDMYVNMKMN				VEKEYKHTGKR			YLESEPEYDE		EEHROITK
<i>C.biggusae</i> .14003	965	Q -	ETQYIEA	ERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					VEKEYKHTGKR			YLESEPEYDE		EEHROITK
<i>C.biggusae</i> .14234	940	-	NDQYIEA	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.biggusae</i> .16767	940	A -	KHNVDSD	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	952	S -	KFEYIEA	ERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	952	S -	KFEYIEA	ERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	950	Q -	DSEYVET	EERTVVAQWQKTOEVE	RTENYLLEFSTRONIL	ROHSVENWLLAEOFTVEN	KNRBAEFVFMRTASPLEKTPLE	I	KISNEI					FEKEIELEQERS					ENNRBF5SK
<i>C.elegans</i> .v12	95																		



	2830	2940	2950	2960	2970	2980	2990	3000
C.briggsae.002	.	.	.	.	.	.	.	.
C.briggsae.006	.	.	.	.	.	.	.	.
C.briggsae.14203	.	.	.	.	.	.	.	.
C.briggsae.14234	.	.	.	.	.	.	.	.
C.briggsae.16797	.	.	.	.	.	.	.	.
C.elegans.wt1	.	.	.	.	.	.	.	.
C.elegans.wt2	.	.	.	.	.	.	.	.
C.elegans.wt3	.	.	.	.	.	.	.	.
C.elegans.wt4	.	.	.	.	.	.	.	.
C.elegans.wt5	.	.	.	.	.	.	.	.
C.elegans.wt6	.	.	.	.	.	.	.	.
G.ellingtonae.016	.	.	.	.	.	.	.	.
G.ellingtonae.060	.	.	.	.	.	.	.	.
G.ellingtonae.064	1702	QF DgGf HSL EDDDED TVNPL E I	S DI SAQK GCGCCD NEDGG CCHCNDDE CGEK DDGCEGSP PCCCCCEE GSG CGEKG GT SKGCDANCKDGRC AELAVMP TWTKT KI NEESHAVGF STRPV KGC PGAGAYSS L THDYATOGVEON KSKVNFV CMERSEPET RRLLS LL RRGRVRVDYNARA G	1899				
G.pallida.0564	.	.	.	.	.	.	.	.
G.pallida.7024	.	.	.	.	.	.	.	.
G.pallida.9459	.	.	.	.	.	.	.	.
G.rustochiensis.2537	.	.	.	.	.	.	.	.
G.rustochiensis.2714	.	.	.	.	.	.	.	.
G.rustochiensis.6238	.	.	.	.	.	.	.	.
H.glycines.1367	.	.	.	.	.	.	.	.
H.glycines.23424	.	.	.	.	.	.	.	.
H.glycines.6510	.	.	.	.	.	.	.	.
H.schachti.007	.	.	.	.	.	.	.	.
H.schachti.19700	.	.	.	.	.	.	.	.
M.hapla.3631	.	.	.	.	.	.	.	.
M.incognita.11129	.	.	.	.	.	.	.	.
M.incognita.17578	.	.	.	.	.	.	.	.
M.incognita.24409	.	.	.	.	.	.	.	.
M.incognita.26853	.	.	.	.	.	.	.	.
M.incognita.30885	.	.	.	.	.	.	.	.
M.incognita.6902	.	.	.	.	.	.	.	.
M.incognita.9436	.	.	.	.	.	.	.	.
M.javanica.15377	.	.	.	.	.	.	.	.
M.javanica.21615	.	.	.	.	.	.	.	.
M.javanica.29487	.	.	.	.	.	.	.	.
M.javanica.3127	.	.	.	.	.	.	.	.
M.javanica.3894	.	.	.	.	.	.	.	.
M.javanica.43787	.	.	.	.	.	.	.	.
M.javanica.5162	.	.	.	.	.	.	.	.
P.acitbneri.001	.	.	.	.	.	.	.	.
P.acitbneri.017	.	.	.	.	.	.	.	.
P.acitbneri.018	.	.	.	.	.	.	.	.
P.acitbneri.064	.	.	.	.	.	.	.	.
P.acitbneri.111	.	.	.	.	.	.	.	.
P.acitbneri.147	.	.	.	.	.	.	.	.
P.acitbneri.032	.	.	.	.	.	.	.	.
Conservation	.	.	.	.	.	.	.	.



Conservation

Conservation

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

1540 1550 1560 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680 1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280 2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520 2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640 2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880 2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000 3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120 3130 3140 3150 3160 3170 3180 3190 3200 3210 3220 3230 3240 3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350 3360 3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480 3490 3500 3510 3520 3530 3540 3550 3560 3570

Conservation

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

[illegible]

Supplementary material 7.1a: Sequence alignment of reference sequence of pJET 1.2 plasmid with *Heterodera schachtii* genomic DNA of the region containing the *vit* gene Hsc_gene_6726 and sequences derived from the use of the forward primer in Sanger sequencing of samples: H1, H2, H3, I2 and I5. Blue arrows indicate PCR forward primer sequence from initial PCR reactions.

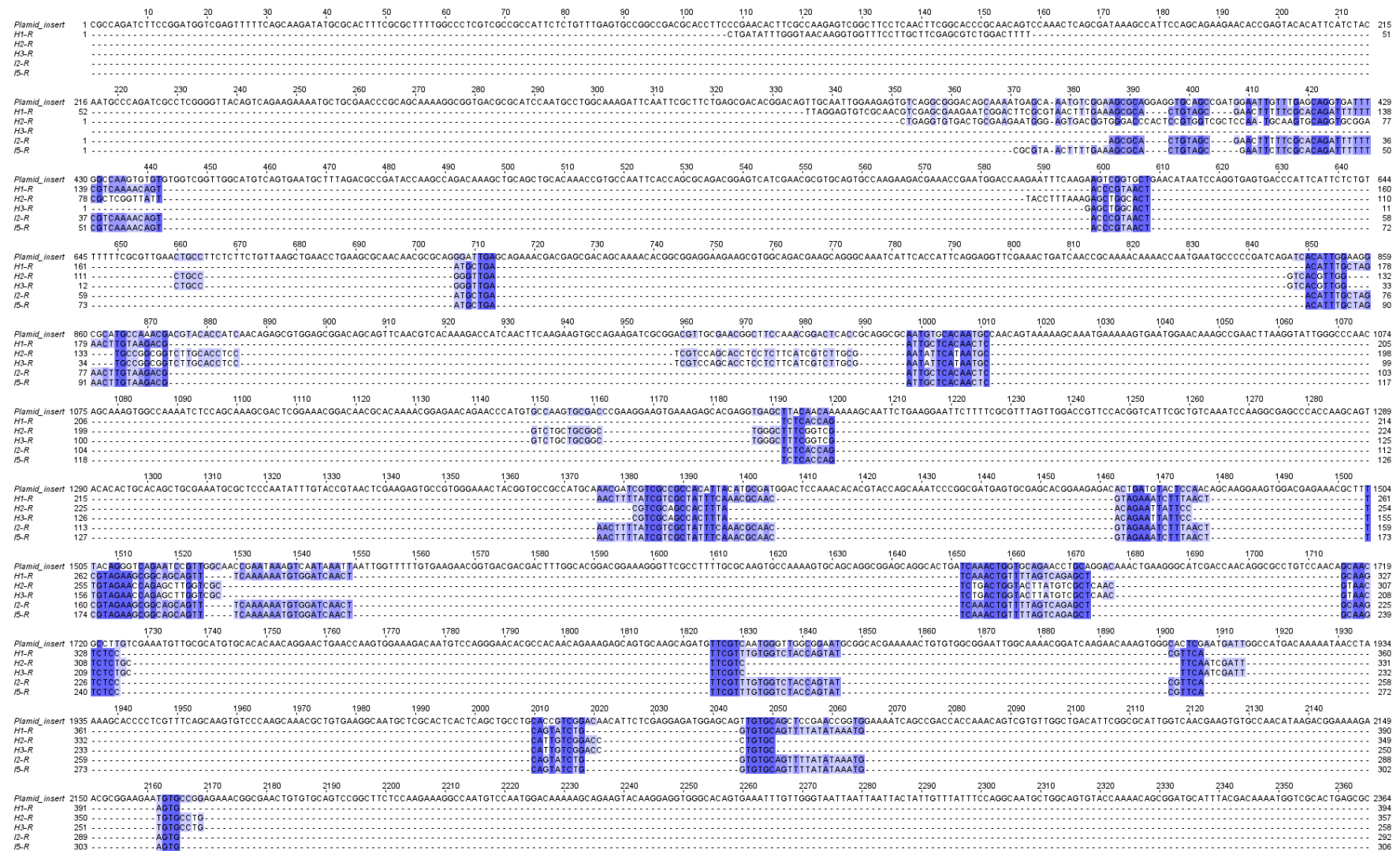


[illegible]

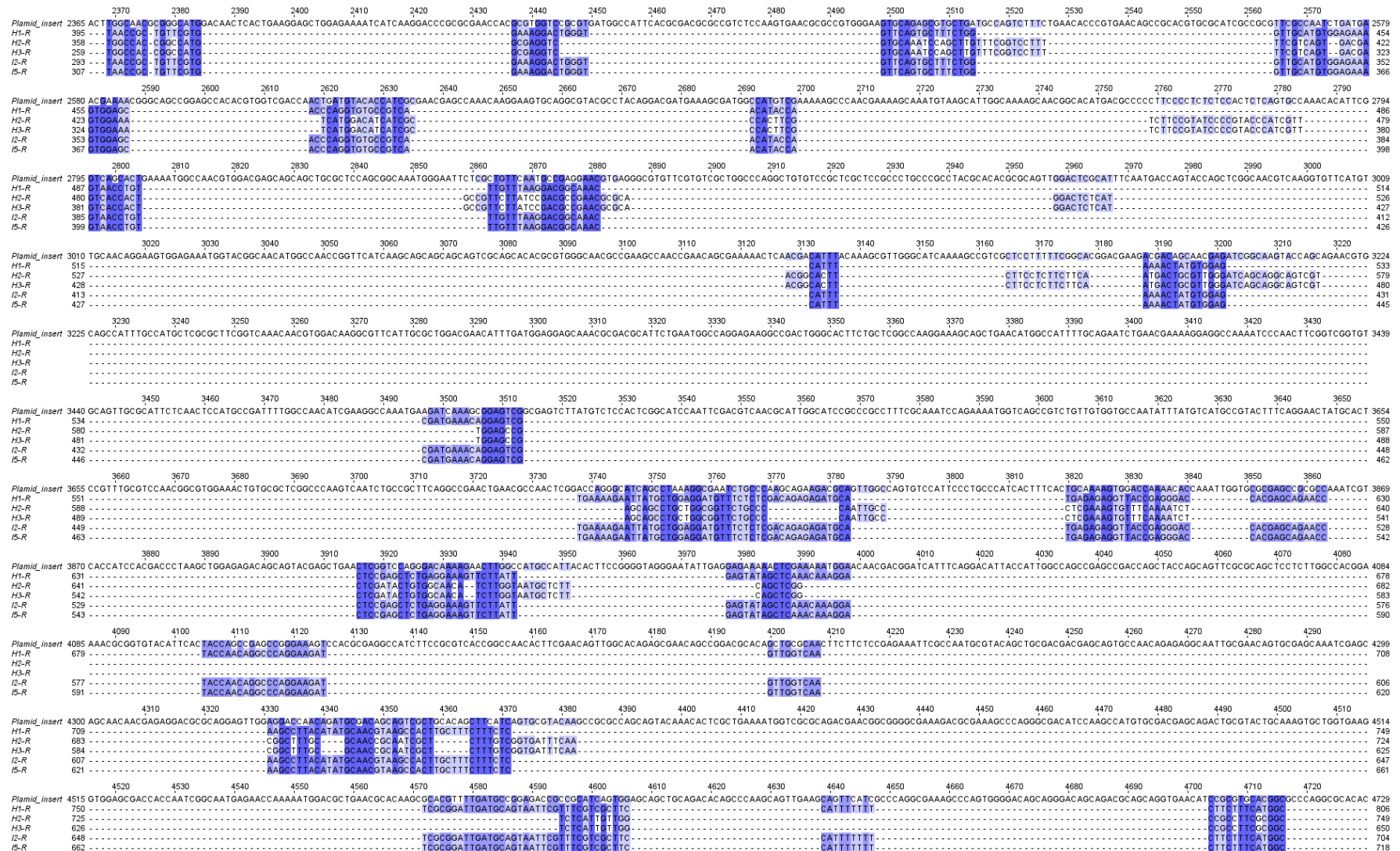
Supplementary material 7.1c: Sequence alignment of reference sequence of pJET 1.2 plasmid with *Heterodera schachtii* genomic DNA of the region containing the *vit* gene Hsc_gene_6726 and sequences derived from the use of the forward primer in Sanger sequencing of samples: H1, H2, H3, I2 and I5.



Supplementary material 7.1d: Sequence alignment of reference sequence of pJET 1.2 plasmid with *Heterodera schachtii* genomic DNA of the region containing the *vit* gene Hsc_gene_6726 and sequences derived from the use of the reverse primer in Sanger sequencing of samples: H1, H2, H3, I2 and I5.



Supplementary material 7.1e: Sequence alignment of reference sequence of pJET 1.2 plasmid with *Heterodera schachtii* genomic DNA of the region containing the *vit* gene Hsc_gene_6726 and sequences derived from the use of the reverse primer in Sanger sequencing of samples: H1, H2, H3, I2 and I5.



Supplementary material 7.1f: Sequence alignment of reference sequence of pJET 1.2 plasmid with *Heterodera schachtii* genomic DNA of the region containing the *vit* gene Hsc_gene_6726 and sequences derived from the use of the reverse primer in Sanger sequencing of samples: H1, H2, H3, I2 and I5.

[illegible]

Supplementary material 7.2a: Sequence alignment of reference sequence of pJET 1.2 plasmid with *Heterodera schachtii* genomic DNA of the region containing the *vit* gene Hsc_gene_6726 and a consensus sequence of cloned DNA from plasmid constructs H1, H2, H3, I2 and I5.

Reference	1	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220		
Consensus	1	CAGT	CG	CGA	CGAC	CT	CT	CG	CG	CT	CT	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	177
Reference	215	230	240	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440	437	
Consensus	178	TAC	ATC	TAT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	437	
Reference	438	460	470	480	490	500	510	520	530	540	550	560	570	580	590	600	610	620	630	640	650	660	655		
Consensus	314	T	GAT	TGG	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	655		
Reference	656	670	680	690	700	710	720	730	740	750	760	770	780	790	800	810	820	830	840	850	860	870	880	890	874
Consensus	456	T	GAAT	TGG	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	874	
Reference	875	900	910	920	930	940	950	960	970	980	990	1000	1010	1020	1030	1040	1050	1060	1070	1080	1090	1100	1110	1096	
Consensus	573	ATG	T	GAAT	TGG	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	1096	
Reference	1097	1120	1130	1140	1150	1160	1170	1180	1190	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300	1310	1320	1330	1316	
Consensus	707	CG	TAT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	1316	
Reference	1317	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560	1533
Consensus	847	CG	TAT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	1533	
Reference	1534	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700	1710	1720	1730	1740	1750	1760	1770	1780	1755	
Consensus	1015	TC	TG	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	1755	
Reference	1756	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	1968	
Consensus	1174	GT	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	1968	
Reference	1969	2010	2020	2030	2040	2050	2060	2070	2080	2090	2100	2110	2120	2130	2140	2150	2160	2170	2180	2190	2200	2210	2220	2178	
Consensus	1319	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	2178	
Reference	2179	2240	2250	2260	2270	2280	2290	2300	2310	2320	2330	2340	2350	2360	2370	2380	2390	2400	2410	2420	2430	2440	2450	2400	
Consensus	1510	TC	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	2400	
Reference	2401	2460	2470	2480	2490	2500	2510	2520	2530	2540	2550	2560	2570	2580	2590	2600	2610	2620	2630	2640	2650	2660	2670	2603	
Consensus	1660	AT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	2603	
Reference	2604	2680	2690	2700	2710	2720	2730	2740	2750	2760	2770	2780	2790	2800	2810	2820	2830	2840	2850	2860	2870	2880	2890	2813	
Consensus	1845	GT	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	2813	
Reference	2814	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000	3010	3020	3030	3040	3050	3060	3070	3080	3090	3100	3110	3120	3036	
Consensus	2007	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	3036	
Reference	3037	3130	3140	3150	3160	3170	3180	3190	3200	3210	3220	3230	3240	3250	3260	3270	3280	3290	3300	3310	3320	3330	3340	3243	
Consensus	2111	CG	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	3243	
Reference	3244	3350	3360	3370	3380	3390	3400	3410	3420	3430	3440	3450	3460	3470	3480	3490	3500	3510	3520	3530	3540	3550	3560	3466	
Consensus	2276	AT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	3466	
Reference	3467	3570	3580	3590	3600	3610	3620	3630	3640	3650	3660	3670	3680	3690	3700	3710	3720	3730	3740	3750	3760	3770	3780	3790	3688
Consensus	2403	CT	CA	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	3688	

Supplementary material 7.2b: Sequence alignment of reference sequence of pJET 1.2 plasmid with *Heterodera schachtii* genomic DNA of the region containing the *vit* gene Hsc_gene_6726 and a consensus sequence of cloned DNA from plasmid constructs H1, H2, H3, I2 and I5.

[illegible]

Supplementary material 7.2c: Sequence alignment of reference sequence of pJET 1.2 plasmid with *Heterodera schachtii* genomic DNA of the region containing the *vit* gene Hsc_gene_6726 and a consensus sequence of cloned DNA from plasmid constructs H1, H2, H3, I2 and I5.

