

**Molecular and bioinformatic analyses of nematode-derived
proteins involved in plant-parasitism**

Sebastian Eves-van den Akker

Submitted in accordance with the requirements for the degree of Doctor of
Philosophy

The University of Leeds
School of Biology
Faculty of Biological Sciences

September 2014

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The Introduction and discussion of feeding tubes and plugs from Chapters 1 and 7 have formed part of a forum article publication currently in press:

S. Eves-van den Akker, C. J. Lilley, J. T. Jones and P. E. Urwin (2014). Plant-parasitic nematode feeding tubes and plugs: New perspectives on function. *Nematology* – In press.

The results from Chapter 3 have been compiled to form a publication:

S. Eves-van den Akker, C. J. Lilley, A. E. Ashcroft, J. R. Ault, J. Jones and P. E. Urwin (2014). The feeding tube of cyst nematodes: Characterisation of protein exclusion. *PLoS ONE* 9(1): e87289. doi:10.1371/journal.pone.0087289

The results from Chapters 4, 5, 6 and 7 have been compiled to form a publication:

S. Eves-van den Akker, C. J. Lilley, J. T. Jones and P. E. Urwin (2014). Identification and characterisation of a hyper-variable apoplastic effector gene family of the potato cyst nematodes. *PLoS Pathogens* 10(9): e1004391

This copy has been supplied on the understanding that it is copyright material and that no quotation from this thesis may be published without proper acknowledgement.

Acknowledgements:

I would like to thank both of my supervisors, Dr John T. Jones and Prof. Peter E. Urwin; each have been uniquely beneficial throughout the project and I have been very lucky to have both as supervisors.

In addition, I would like to thank Dr. Catherine Lilley for all of her help over the last few years. It is difficult to overstate the effect she has had on my work and my scientific approaches. It has been a genuine privilege working together and I hope it may continue in the future.

Finally, I would like to thank Les Brown and Eric Thomas.

This project was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and The James Hutton Institute. The James Hutton Institute received funding from the Scottish government.

Abstract:

Plant parasitic nematodes comprise several groups; the most economically damaging of these are the sedentary endoparasites. Sedentary endoparasitic nematodes modify host root tissues, using a suite of effector proteins to create and maintain a feeding site that is their sole source of nutrition. Ultrastructural studies of plant-nematode interactions, and in particular those of the feeding sites, have identified two key feeding structures; feeding tubes and feeding plugs. Sedentary plant-parasitic nematodes feed by withdrawing host cell assimilate from the feeding site through a feeding tube. The function, composition and molecular characteristics of both feeding tubes and feeding plugs are poorly characterised. It is hypothesised that the apparent selective uptake of certain proteins from the feeding site is attributed to feeding tube size exclusion.

A novel method is proposed to predict protein size based on protein database coordinates *in silico*. The validity of these predictions was tested using travelling wave ion mobility spectrometry – mass spectrometry, where predictions and measured values were within approximately 6% of each other. *In silico* predictions coupled with experimental techniques, such as mass spectrometry, analytical ultracentrifugation and protein electrophoresis, aimed to resolve seemingly-conflicting results of previous size exclusion experiments. Together these provided a pragmatic measurement of the upper limit for cyst nematode feeding tube size exclusion.

Putative feeding-structure genes were identified from the genome sequence of the potato cyst nematode *Globodera pallida* using a series of reasoned assumptions about their characteristics. As a result, several large gene families were identified, one of which displayed highly complex genomic variation within a population. Subsequent characterisation of these candidate genes informed their function. The expression of several candidates was demonstrated in tissues with the capacity to secrete proteins into the host, implicating their role in host-pathogen interactions. In addition, for the 444 gene family, the protein was detected in the apoplasm, between the anterior end of the nematode and the feeding site. *In planta* host induced gene silencing targeting 444s reduced nematode infection by > 50%; further supporting their important role in successful parasitism.

Feeding structure candidate genes were identified in *de novo* transcriptome assemblies of two related species (*Globodera rostochiensis* and *Rotylenchulus reniformis*). Differential expression analysis identified those candidates with congruent expression between species. 444 and 448 genes appear to be “core” effectors present in three genera of plant-parasitic nematodes that infect mono- and di-cotyledonous crop species.

Table of Contents

1. General Introduction	1
1.1. Nematoda.....	1
1.2. Nematode phylogeny.....	1
1.3. Plant parasitic nematodes	3
1.3.1. Root-knot nematodes.....	4
1.3.2. Cyst nematodes	5
1.3.3. The life cycle of cyst nematodes	5
1.3.4. Syncytia	6
1.3.5. The role of effectors in cyst nematode biotrophy	7
1.3.6. Nematode feeding structures	11
1.3.7. Feeding tubes	11
1.3.8. Feeding tube function.....	13
1.3.9. Feeding plugs.....	14
1.3.9.1. Feeding plug function	16
1.4. Project overview	17
2. General materials and methods	18
2.1. Bacterial Growth Media	18
2.1.1. Super Optimal Broth (SOB).....	18
2.1.2. Super Optimal broth with Catabolite repression (SOC).....	18
2.1.3. Luria-Bertani (LB).....	18
2.1.4. Luria-Bertani Agar (LB Agar).....	18
2.2. Plant growth media.....	19
2.2.1. Murashige and Skoog (liquid MS 20).....	19
2.2.2. Murashige and Skoog Agar (MS-A).....	19
2.3. Biological material	20
2.3.1. Maintenance of cyst nematode populations.....	20
2.3.2. Hatching of second-stage juveniles	20
2.3.3. Collection of parasitic-stage nematodes.....	20
2.4. Molecular Cloning.....	21
2.4.1. Polymerase Chain Reaction (PCR)	21
2.4.2. Restriction digestion of DNA.....	22
2.4.3. Agarose gel electrophoresis and gel extraction	22
2.4.4. DNA ligation	22
2.4.5. Plasmid DNA extraction from bacteria.....	22

2.4.6.	DNA extraction from plant and nematode tissue.....	23
2.4.7.	RNA extraction from plant and nematode samples.....	23
2.4.8.	cDNA synthesis.....	23
2.4.9.	Semi Quantitative Reverse Transcriptase - Polymerase Chain Reaction (SQRT-PCR)	24
2.4.10.	Preparation and transformation of chemically ultra- competent <i>Escherichia coli</i>	24
2.4.11.	Preparation and transformation of electro-competent <i>Escherichia coli</i>	25
2.4.12.	Preparation and transformation of competent <i>Agrobacterium tumefaciens / rhizogenes</i>	25
2.4.13.	Colony screening	26
2.5.	Protein expression, extraction and analysis	26
2.5.1.	Bacterial protein expression	26
2.5.2.	SDS-PolyAcrylamide Gel Electrophoresis (PAGE)	27
2.5.3.	Small-scale denaturing protein purification	27
2.5.4.	Large-scale protein expression and purification.....	28
2.5.5.	Removal of His-Tag from purified protein: Thrombin digestion	29
2.6.	Western blot	29
2.6.1.	Protein extraction from plant or animal tissue.....	29
2.6.2.	Transfer of protein to PVDF membrane.....	30
2.6.3.	Ponceau S red staining	30
2.6.4.	Colorimetric detection system	30
2.6.5.	Chemi-luminescent detection	31
2.7.	General primers.....	31
3.	Feeding tube size exclusion.....	32
3.1.	Introduction.....	32
3.1.1.	Feeding tubes and size exclusion.....	32
3.1.2.	Mass spectrometry as a way of measuring protein size.....	33
3.1.3.	Predicting protein size computationally.....	34
3.2.	Aims:	35
3.3.	Materials and methods	36
3.3.1.	Computational prediction of protein size – RotaMol.....	36
3.3.1.1.	Area Measurement.....	36
3.3.1.2.	Measurement in pixels and conversion to Angstroms.....	36

3.3.1.3.	Constructs for protein expression in <i>E. coli</i>	37
3.3.1.4.	Ion Mobility Spectrometry – Mass Spectrometry (IMS-MS).....	39
3.3.1.5.	Analytical Ultra Centrifugation (AUC).....	39
3.3.1.6.	Native - Polyacrylamide Gel Electrophoresis (PAGE).....	40
3.3.2.	Expression of fluorescent proteins in <i>Arabidopsis thaliana</i>	40
3.3.2.1.	Fluorescent protein fusion construct cloning.....	40
3.3.2.2.	Floral dip transformation of <i>Arabidopsis thaliana</i>	42
3.3.2.3.	Surface sterilisation of <i>Arabidopsis thaliana</i> seeds	42
3.3.2.4.	Selection of transformed seeds	42
3.3.2.5.	Selecting high expressing lines.....	43
3.3.2.6.	Surface sterilisation of juvenile nematodes.....	43
3.3.2.7.	Infecting <i>Arabidopsis thaliana</i> with juveniles in tissue culture.....	44
3.3.3.	Detecting uptake of fluorescent/fusion constructs.....	44
3.3.3.1.	Fluorescence microscopy	44
3.3.3.2.	Confocal Microscopy	44
3.4.	Results	45
3.4.1.	Computational protein size prediction.....	45
3.4.2.	Protein analyses.....	48
3.4.2.1.	Protein expression and removal of HIS tag.....	48
3.4.2.2.	Travelling wave ion mobility spectrometry – mass spectrometry	52
3.4.2.3.	Analytical Ultra Centrifugation.....	52
3.4.2.4.	Native PAGE	53
3.4.3.	<i>In planta</i> expression of fluorescent proteins	57
3.4.3.1.	Cloning of expression constructs.....	57
3.4.3.2.	Selection of transgenic <i>Arabidopsis thaliana</i>	59
3.4.3.3.	Selection of high expressing lines.....	59
3.4.4.	Uptake of fluorescent proteins.....	63
3.5.	Discussion	67
3.5.1.	Quality of computational size predictions - RotaMol	67
3.5.2.	Heterogeneity, quaternary structure, aggregation and their effects on feeding tube size exclusion.....	69
3.5.3.	Fluorescent proteins as a tool to test uptake	70

3.5.4.	Further resolution of cyst nematode feeding tube size exclusion.....	70
3.5.5.	Comparing nematode feeding tube size exclusion to other biological membranes.	71
4.	Identification of candidate feeding structure genes	74
4.1.	Introduction.....	74
4.1.1.	Effector identification strategies.....	74
4.2.	Aims:	76
4.3.	Materials and methods	77
4.3.1.	Data mining for candidate feeding structure genes.....	77
4.3.1.1.	Differential expression analysis	77
4.3.1.2.	Predicted secreted protein pipeline of novel proteins	77
4.3.2.	DNA and protein alignment	78
4.3.3.	3'RACE	78
4.3.4.	Single nematode DNA extractions.....	78
4.3.4.1.	Whole Genome Amplification (WGA).....	78
4.3.4.2.	NaOH digestion	78
4.3.4.3.	Chaos nucleic acid extraction buffer	79
4.3.5.	Tandem repeat analysis	79
4.3.6.	<i>In vitro</i> transcription.....	79
4.3.7.	Denaturing RNA gel electrophoresis	80
4.4.	Results	82
4.4.1.	Identification of genes that may encode components of the feeding structures	82
4.4.1.1.	Expression analysis.....	82
4.4.1.2.	Signal peptide and transmembrane prediction.....	82
4.4.1.3.	Similarity search	85
4.4.2.	Single copy genes.....	85
4.4.3.	Candidate 1078 misprediction	85
4.4.4.	Candidate 448 gene family.....	92
4.4.4.1.	Candidate 448 highly conserved motif	92
4.4.5.	Candidate 444 gene family.....	96
4.4.5.1.	Cloning from cDNA.....	96
4.4.5.2.	3'RACE.....	96
4.4.5.3.	Single nematode analysis.....	101
4.4.5.4.	Apparent non-canonical splicing	105

4.4.5.5.	444 gene family tree and transcript structure	107
4.4.5.6.	Tandem repeat analysis	107
4.4.5.7.	444 domain structure and modular organisation	113
4.5.	Discussion	115
4.5.1.	Quality of genome assembly	115
4.5.2.	Convergence of identification strategies	115
4.5.3.	Non-canonical splicing.	116
4.5.4.	The complex 444 multi-gene family	117
4.5.5.	Copy number variation of 444 sequences between individuals of the same population	118
4.5.6.	Inter-specific differences	118
5.	Characterisation of candidate feeding structure genes	121
5.1.	Introduction.....	121
5.1.1.	Validating effector identification.....	121
5.1.2.	Structures with the capacity to secrete proteins into the host	121
5.1.3.	Functional characterisation of effectors	122
5.1.3.1.	Localisation	122
5.1.3.2.	<i>In planta</i> interactions.....	124
5.1.3.3.	Loss of function knockdown.....	124
5.2.	Aims	126
5.3.	Materials and methods	127
5.3.1.	<i>In situ</i> hybridisation	127
5.3.2.	<i>In planta</i> RNA interference	129
5.3.3.	Acid fuchsin staining and nematode infection quantification.....	132
5.3.4.	Yeast-two-hybrid screens.....	132
5.3.5.	Antigenicity prediction	133
5.3.6.	Embedding, sectioning and immunochemistry.....	133
5.4.	Results	135
5.4.1.	Candidate 225.....	135
5.4.2.	Candidate 3453.....	137
5.4.2.1.	Candidate 3453 <i>in situ</i> hybridisation	137
5.4.2.2.	Candidate 3453 <i>in planta</i> RNA interference.....	137
5.4.3.	Candidate 176.....	143
5.4.3.1.	Candidate 176 <i>in situ</i> hybridisation	143

5.4.4.	Candidate 444 gene family	145
5.4.4.1.	Candidate 444 <i>in situ</i> hybridisation	145
5.4.4.2.	Yeast-two-hybrid analysis of candidates 444 and 176 148	
5.4.4.3.	Candidate 444 and 176 bacterial expression and mass spectrometry	148
5.4.4.4.	444 – Antigen design and immunolocalisation	150
5.4.4.5.	RNA interference of candidate 444	155
5.4.5.	Candidate 448 gene family	157
5.4.5.1.	Candidate 448 <i>in situ</i> hybridisation	157
5.4.5.2.	Candidate 448 RNA interference	159
5.4.5.3.	Candidate 448 antibody design	159
5.5.	Discussion	162
5.5.1.	Effector identification strategies.....	162
5.5.2.	The role of amphids during the feeding stages of <i>Globodera pallida</i>	162
5.5.3.	RNA interference.....	163
6.	Analysis of the transcriptomes of cyst nematodes and related species.....	167
6.1.	Introduction.....	167
6.1.1.	Related plant parasitic nematodes	167
6.1.1.1.	The root knot nematode <i>Meloidogyne incognita</i>	167
6.1.1.2.	The reniform nematode <i>Rotylenchulus reniformis</i>	167
6.1.1.3.	The cyst nematode <i>Globodera rostochiensis</i>	168
6.1.2.	Feeding structures in related plant parasitic nematodes ...	168
6.1.2.1.	Feeding tubes.....	168
6.1.2.2.	Feeding plugs.....	169
6.1.2.3.	Feeding structures in different species	169
6.1.3.	DNA Sequencing and next generation technology	171
6.1.3.1.	History of DNA sequencing.....	171
6.1.3.2.	Next Generation Sequencing (NGS).....	171
6.1.3.3.	RNAseq.....	172
6.2.	Aims	173
6.3.	Materials and methods	174
6.3.1.	Biological material	174
6.3.2.	Nematode collection and RNA extraction	174
6.3.2.1.	<i>G. rostochiensis</i> second stage juveniles	174

6.3.2.2.	<i>G. rostochiensis</i> sedentary female feeding females ...	174
6.3.2.3.	<i>R. reniformis</i> J2.....	174
6.3.2.4.	<i>R. reniformis</i> feeding stage	174
6.3.2.5.	RNA extraction and quality control.....	175
6.3.3.	Library construction, sequencing and quality control	175
6.3.4.	Assembly	176
6.3.4.1.	<i>In silico</i> normalisation	176
6.3.4.2.	Trinity assembly.....	176
6.3.4.3.	CLC bio assembly	177
6.3.5.	Mapping	177
6.3.5.1.	Mapping to genome - TopHat2	177
6.3.5.2.	Mapping to transcriptome - Bowtie2.....	177
6.3.6.	Differential expression.....	177
6.3.7.	Gene predictions	178
6.3.7.1.	Open reading frame prediction - Transdecoder.....	178
6.3.7.2.	CEGMA	178
6.3.8.	Bioinformatic comparisons	178
6.3.8.1.	BLAT	178
6.3.8.2.	BLAST.....	178
6.3.8.3.	Secreted protein pipeline	179
6.4.	Results	179
6.4.1.	RNA extractions and quality assessment	179
6.4.2.	RNAseq reads and quality control	180
6.4.3.	Assembled transcriptomes and quality assessment	186
6.4.3.1.	<i>R. reniformis</i> - Trinity.....	186
6.4.3.2.	<i>R. reniformis</i> – CLC	186
6.4.3.3.	<i>G. rostochiensis</i> – Trinity.....	186
6.4.3.4.	<i>G. rostochiensis</i> – CLC	187
6.4.3.5.	Removal of contamination	187
6.4.4.	Gene predictions and differential expression.....	190
6.4.4.1.	<i>R. reniformis</i> – within species gene cluster comparison	190
6.4.4.2.	<i>G. rostochiensis</i> – within species gene cluster comparison	193
6.4.5.	Comparing gene expression clusters between species	196
6.4.5.1.	Putative homologs between species.....	196

6.4.5.2. Functional categories in gene clusters between species.....	196
6.4.5.3. Functional categories in secreted proteins at sedentary female stage between species.....	196
6.4.6. Whole transcriptome comparison between <i>R. reniformis</i> and <i>G. rostochiensis</i>	202
6.4.6.1. BLAT	202
6.4.6.2. Reciprocal Best BLAST Hit (RBBH) analysis	202
6.4.7. Presence of <i>G. pallida</i> candidate feeding structure genes in the <i>R. reniformis</i> and <i>G. rostochiensis</i> transcriptomes	206
6.4.7.1. Gene 176.....	206
6.4.7.2. Gene 3453.....	207
6.4.7.3. 444 gene family	212
6.4.7.4. Candidate 444 diversity in <i>G. rostochiensis</i>	218
6.4.7.5. 448 gene family	220
6.4.8. Putative feeding tube components	223
6.4.8.1. <i>Rotylenchulus reniformis</i>	223
6.4.8.2. <i>Globodera rostochiensis</i>	223
6.5. Discussion.....	224
6.5.1. Transcriptome assembly and differential expression	224
6.5.1.1. Limitations of Trinity assembly.....	224
6.5.1.2. Normalisation and differential expression	225
6.5.2. Contamination removal and Horizontal gene transfer (HGT).....	226
6.5.3. Between species functional annotation comparison	226
6.5.4. Oxidoreductase activity	227
6.5.5. <i>G. pallida</i> candidate feeding structure genes in <i>G. rostochiensis</i> and <i>R. reniformis</i>	228
6.5.5.1. 448 gene family	228
6.5.5.2. 176 genes.....	229
6.5.5.3. 444 gene family	229
6.5.6. Putative orthologs between <i>R. reniformis</i> and <i>G. rostochiensis</i>	230
6.5.6.1. Proposed functions of super-conserved orthologs	230
7. General discussion	233
7.1. Feeding tube size exclusion	233
7.2. Feeding tube and feeding plug candidate genes	236

7.3. Success of the identification strategy	237
7.3.1. Tandem repeat proteins	239
7.3.2. Tandem repeated genes	245
8. References.....	247

List of Figures

Figure 1-1 The phylum Nematoda.	2
Figure 1-2 Schematic representation of a typical tylenchid nematode.....	10
Figure 1-3 Feeding tubes of plant parasitic nematodes.	12
Figure 1-4 Feeding plugs of plant parasitic nematodes.....	15
Figure 3-1 The number of pixels measured by RotaMol with varying degrees of Pixelskip for a single viewing angle of GFP.....	46
Figure 3-2 Empirically derived optimum values for RotaMol parameters.....	46
Figure 3-3 Comparison between RotaMol predictions and existing prediction methods.....	47
Figure 3-4 Elution Chromatograms for GFP and mRFP respectively.	49
Figure 3-5 Elution Chromatograms for Dual PI and a blank run.	50
Figure 3-6 Removing His tag from expressed proteins by thrombin digestion.....	51
Figure 3-7 ESI-TWIMS-MS drift plot of GFP, mRFP and Dual PI.....	54
Figure 3-8 Absorbance scans at a range of wavelengths before and after 5 minute 14,000g centrifugation.....	55
Figure 3-9 Analytical Ultra-Centrifugation (AUC) absorbance plots of mRFP, GFP and DUAL PI.....	55
Figure 3-10 Analysis of Dual PI multimers by native polyacrylamide gel electrophoresis.	56
Figure 3-11 Typical cloning procedure for plant transformation vectors using iLOV as an example.....	58
Figure 3-12 Selection of successfully transformed <i>Arabidopsis thaliana</i>	60
Figure 3-13 Selection of high expressing transgenic <i>Arabidopsis thaliana</i>	61
Figure 3-14 Detection of intact fusion constructs, expressed <i>in planta</i> , by western blot. plants (iLOV).	62
Figure 3-15 Fluorescence microscopy to detect uptake of fluorescent proteins by cyst nematodes.....	64
Figure 3-16 Confocal microscopy to detect uptake of mRFP.....	65
Figure 3-17 Confocal microscopy to detect uptake of DRONPA.....	66
Figure 3-18 GFP and mRFP external charge comparison.....	68
Figure 4-1 Expression analysis of candidate feeding structure genes.....	83
Figure 4-2 Typical positive output from signal peptide prediction using SignalP v4.1.....	84
Figure 4-3 Typical negative output from transmembrane domain prediction using TMHMM.....	84

Figure 4-4 Pipeline overview for identification of putative feeding structure components.	89
Figure 4-5 Transmembrane predictions on predicted and cloned 1078 sequences.	90
Figure 4-6 Alignment of mispredicted and cloned 1078 protein sequences.	91
Figure 4-7 Expression of all <i>G. pallida</i> 448 genes across the life cycle.	93
Figure 4-8 Phylogeny of 448 genes of <i>G. pallida</i> and <i>G. rostochiensis</i>	94
Figure 4-9 Motif common to all 448 sequences.	95
Figure 4-10 First 444 PCR amplification of cDNA.	98
Figure 4-11 3' Untranslated regions of 444 groups A, B and C identified by 3' RACE.	99
Figure 4-12 Schematic representation of 444 family gene structure, available primer pairs, and resultant PCR products on genomic DNA.	100
Figure 4-13 Whole genome amplification of single nematodes.	102
Figure 4-14 PCR on single nematode NaOH DNA extractions.	103
Figure 4-15 Single nematode Chaos DNA extraction.	104
Figure 4-16 Non-canonical apparent splicing of 444 genes.	106
Figure 4-17 Phylogenetic tree of genomic 444 DNA sequences.	109
Figure 4-18 Within subfamily deduced amino acid sequence from 444 genomic sequences.	110
Figure 4-19 Schematic representation of 444 gene family transcript structure.	111
Figure 4-20 Tandem repeat analysis of 444 genes.	112
Figure 4-21 Schematic representation of 444 domain organisations.	114
Figure 5-1 Preparation of probes for candidate 225 <i>in situ</i> hybridisation.	136
Figure 5-2 Candidate 225 <i>in situ</i> hybridisation.	136
Figure 5-3 Candidate 3453 <i>in situ</i> hybridisation probe preparation.	139
Figure 5-4 Candidate 3453 <i>in situ</i> hybridisation.	139
Figure 5-5 Typical transgenic hairy root transformation in tissue culture.	140
Figure 5-6 Semi-quantitative RT-PCR of candidate 3453 and GFP inverted repeat RNAi construct expression in hairy roots.	141
Figure 5-7 Candidate 3453 RNAi hairy root infections.	142
Figure 5-8 Candidate 176 <i>in situ</i> hybridisation probe.	144
Figure 5-9 Candidate 176 <i>in situ</i> hybridisation.	144
Figure 5-10 Generic 444 <i>in situ</i> hybridisation probe design.	146
Figure 5-11 Candidate 444 <i>in situ</i> hybridisation probe PCR.	146

Figure 5-12 5-13 Candidate 444 <i>in situ</i> hybridisation.	147
Figure 5-14 444 and 176 Yeast-two-hybrid analysis.	149
Figure 5-15 444 antigenic regions for peptide design.	151
Figure 5-16 444 Immunolocalisation - Peptide design and antiserum quality control.	152
Figure 5-17 444 antiserum specificity control.	153
Figure 5-18 Immunolocalisation of 444 proteins in infected root tissue.	154
Figure 5-19 Semi-quantitative RT-PCR of 444-A-IR hairy roots.	156
Figure 5-20 <i>In planta</i> RNAi of 444 genes.	156
Figure 5-21 Candidate 448 <i>in situ</i> hybridisation probe.	158
Figure 5-22 Candidate 448 <i>in situ</i> hybridisation.	158
Figure 5-23 448 triple inverted repeat construct.	160
Figure 5-24 448 antigen design.	161
Figure 6-1 Phylogenetic distribution and morphology of feeding tubes and feeding plugs.	170
Figure 6-2 RNA extraction quality control.	182
Figure 6-3 FastQC report of GC content of <i>G. rostochiensis</i> sedentary female reads identifies putative contamination.	184
Figure 6-4 Length distribution of reads following quality score trimming. ...	185
Figure 6-5 Comparison of contig size between assembly methods.	188
Figure 6-6 Analysis of GC content of transcripts removed from <i>G. rostochiensis</i> assembly.	189
Figure 6-7 <i>R. reniformis</i> gene expression clusters.	191
Figure 6-8 Blast2GO cellular component analysis of differentially expressed gene clusters for <i>R. reniformis</i>.	192
Figure 6-9 <i>G. rostochiensis</i> gene expression clusters.	194
Figure 6-10 Blast2GO cellular component analysis of differentially expressed clusters for <i>G. rostochiensis</i>.	195
Figure 6-11 Gene ontology terms for secreted proteins at sedentary female for <i>G. rostochiensis</i> and <i>R. reniformis</i>.	198
Figure 6-12 Blast2GO comparison between secreted proteins up-regulated at sedentary female for <i>R. reniformis</i> and <i>G. rostochiensis</i>.	199
Figure 6-13 Expression of glutathione synthetase genes in <i>G. rostochiensis</i>.	200
Figure 6-14 Expression of glutathione synthetase genes in <i>R. reniformis</i>.	201
Figure 6-15 Frequency distribution of percentage identity of putative homologs between <i>R. reniformis</i> and <i>G. rostochiensis</i>.	204

Figure 6-16 Comparison of expression between species of two example highly conserved homologs.....	205
Figure 6-17 Phylogenetic tree of 176 genes in cyst nematodes.....	208
Figure 6-18 Comparison of expression of 176-like sequences in <i>G. rostochiensis</i> and <i>G. pallida</i>	209
Figure 6-19 Phylogenetic tree of 3453 genes in cyst and reniform nematodes.....	210
Figure 6-20 Comparison of expression of 3453-like genes in <i>G. rostochiensis</i> , <i>R. reniformis</i> and <i>G. pallida</i>	211
Figure 6-21 Phylogenetic analysis of all 444 sequences from <i>G. pallida</i> , <i>G. rostochiensis</i> , <i>R. reniformis</i> and <i>H. glycines</i>	213
Figure 6-22 Protein alignment of 444-B like sequences from <i>G. pallida</i> , <i>G. rostochiensis</i> and <i>R. reniformis</i>	217
Figure 6-23 Comparison of expression of 444 genes in <i>G. rostochiensis</i> , <i>R. reniformis</i> and <i>G. pallida</i>	217
Figure 6-24 Comparison of candidate 444 genes in <i>G. pallida</i> and <i>G. rostochiensis</i>	219
Figure 6-25 Phylogenetic tree of 448 proteins in <i>G. pallida</i> , <i>G. rostochiensis</i> and <i>R. reniformis</i>	221
Figure 6-26 Comparison of 448 gene expression in <i>G. rostochiensis</i> , <i>R. reniformis</i> and <i>G. pallida</i>	222
Figure 7-1 Predicted structure of the leucine rich repeat domain of plant resistance proteins.	241
Figure 7-2 Transcription activator like effectors DNA-bound structure.....	243

List of Tables:

Table 2-1 General primers common to multiple Chapters	31
Table 3-1 Summary of primers to cloning Dual PI, mRFP and GFP for expression in <i>E. coli</i> and subsequent size measurement	38
Table 3-2 Primers to clone fluorescent protein fusion construct for plant transformation vectors	41
Table 3-3 Summary of molecular mass, size, and uptake of mRFP, GFP and Dual PI	54
Table 3-4 Summary of fluorescent protein fusion constructs and previously reported uptake by cyst nematodes	57
Table 4-1 Primer sequences, annealing temperatures and expected product sizes where applicable	81
Table 4-2 Summary of candidate feeding structure genes identified including expression at each life stage and BLAST hits	86
Table 5-1 <i>In situ</i> hybridisation primers, annealing temperatures and expected amplicon size	128
Table 5-2 RNAi construct primers, expected annealing temperatures and product sizes	131
Table 6-3 Summary of quality control data for RNA samples	181
Table 6-4 Summary of raw and trimmed reads	183
Table 6-5 Transcriptome assembly quality assessment	189

List of abbreviations:

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
dpi	Days post infection
DRONPA	Reversibly switchable photoactivatable fluorescent protein
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tag
GFP	Green fluorescent protein
<i>g</i>	Gravitational force
g	Grams
h	Hours
iLOV	Fluorescent protein derived from light, oxygen or voltage-sensing domain of the plant blue light receptor, phototropin.
J2	Juvenile nematode (second stage)
m	Metre
mRFP	Monomeric Red fluorescent protein
mRNA	Messenger RNA
MS	Murashige and Skoog
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PVDF	Polyvinylidene fluoride
qPCR	Quantitative real-time PCR
RKN	Root-knot nematode
RNAi	Ribonucleic acid interference
RPM	Rotations per minute
s	Seconds
TMHMM	Trans membrane hidden Markov model
μ	Micro
Y2H	Yeast-two hybrid

1. General Introduction

1.1. Nematoda

Nematodes are unsegmented, bilaterally symmetrical roundworms, with over 23,000 species identified to date. Although it is difficult to estimate the total number of nematode species accurately, due to the fact that many species live underground and in river and sea beds, it is suggested that they extend into the millions (Blaxter, 2011). There is almost no characterised fossil record for nematodes as many are microscopically small and all are soft bodied. The earliest nematode fossil on record dates from approximately 400 million years ago (Poinar et al., 2008). However, the phylum Nematoda is hypothesised to have originated during the Cambrian explosion that occurred approximately 550 – 600 million years ago (van Megen et al., 2009). Nematodes are one of the most speciose groups of all animals, and are present in virtually all marine, freshwater and terrestrial environments (van Megen et al., 2009). In marine environments nematodes outnumber every other animal many fold (Blaxter, 2011, Creer et al., 2010). Although body morphology is generally conserved, with the majority showing a vermiform body plan, there is a wide range of body sizes spanning 4 orders of magnitude from less than 1 mm to several metres (Blaxter, 2011). Nematodes can be free-living, plant-, or animal-parasites, predators or necrotrophs.

1.2. Nematode phylogeny

Nematodes diverged before arthropods in the Animalia branch of the tree of life. The most comprehensive and recent phylogenetic analysis of nematodes is based on the Small Subunit ribosomal DNA (SSU rDNA) sequence of over 1200 species has resolved the phylum Nematoda into 12 clades highlighted in Figure 1-1 (van Megen et al., 2009). This subdivision is broadly consistent with previous work in which nematodes are divided into three main groups, the Enoplia, Dorylaimia, and Chromadoria (Figure 1-1). Despite the heavy bias of modern phylogenies towards economically important, and typically parasitic, species the Enoplia contain almost exclusively free-living nematodes. Insect, animal or plant parasites are found in both the Dorylaimia and Chromadoria, the latter being the most well studied. Chromadoria contain the most economically devastating plant parasitic nematodes (Tylenchina), animal parasitic nematodes of global importance, and by far the most well studied nematode; the model organism *Caenorhabditis elegans*.

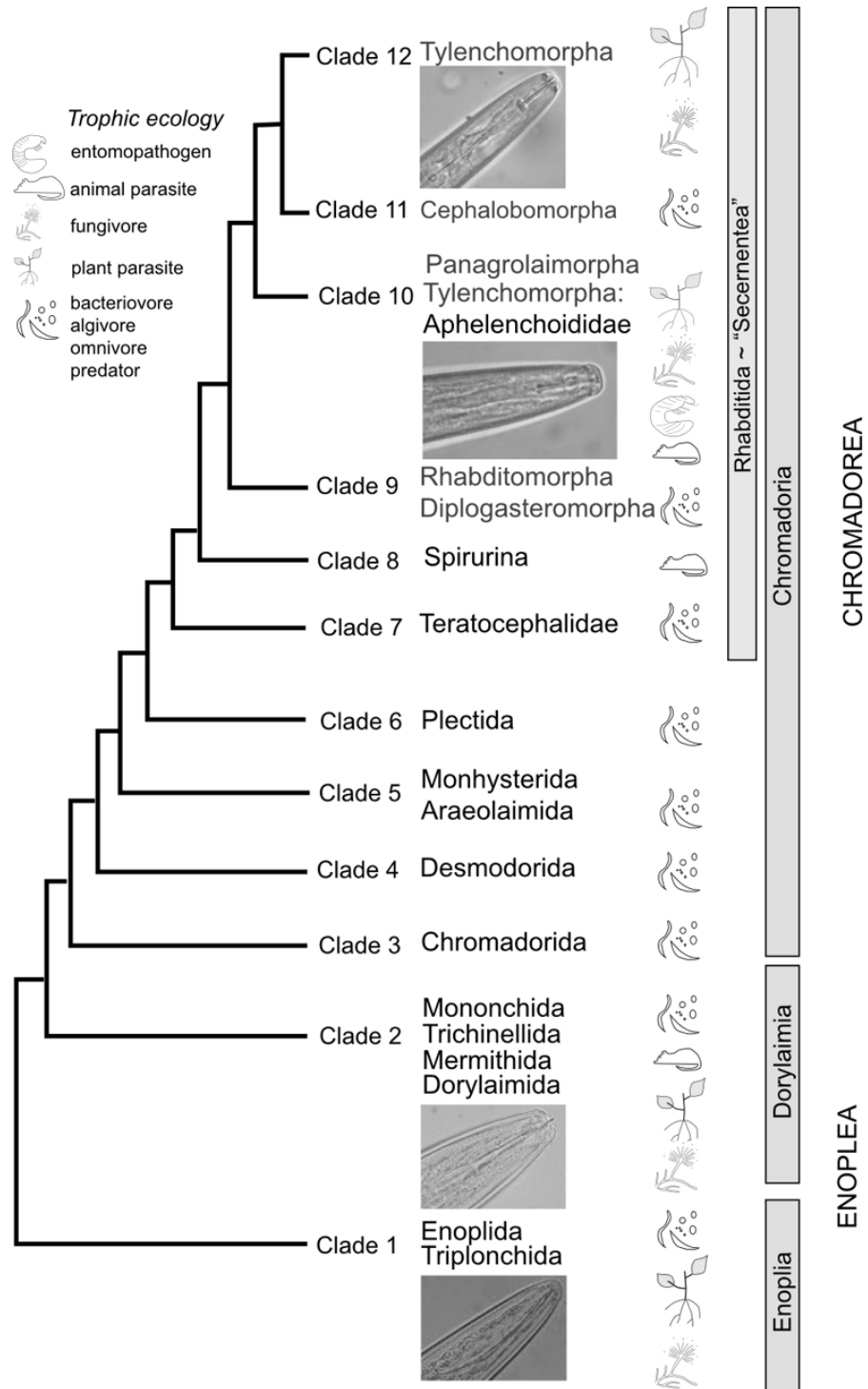


Figure 1-1 **The phylum Nematoda.** Schematic representation of small subunit ribosomal DNA phylogeny of nematodes from Jones et al. (Jones et al., 2013) can be divided into three main groups; Dorylaimia, Enoplia and Chromadorea. The most recent phylogeny of 1200 species has further subdivided the phylogeny into 12 clades (van Megen et al., 2009). Various feeding modes are indicated by small icons. Representatives of each clade containing plant-parasitic nematodes are shown with images. In each of these clades the stylet appears to have evolved independently.

1.3. Plant parasitic nematodes

It has been suggested that at least one species of parasitic nematode has been identified for almost every plant and animal species on the planet (Blaxter, 2011). Although parasitic nematodes constitute the overwhelming minority of nematode species, they are amongst the most widely studied groups due to their economic and medical importance (van Megen et al., 2009). Plant parasitic nematodes have been identified in four of the 12 clades proposed by van Megen *et al.* (van Megen et al., 2009). Clades 1 and 2 contain migratory ectoparasitic species including *Trichodorus* (Clade 1), *Xiphinema* and *Longidorus* (Clade 2). Although Clade 10 is composed mainly of fungal feeding nematodes, it also includes two plant parasitic species of *Bursaphelenchus* and plant-parasitic (foliar) *Aphelenchoides* species (e.g. *A. fragariae*). However, the majority of plant parasitic species, including the most economically important nematodes, are found within Clade 12.

Plant-parasitism by nematodes encompasses a diverse array of parasitism strategies that can be subdivided into two groups, endo-parasitism and ecto-parasitism. Those nematodes that are ecto-parasites spend their entire lifecycle, including the parasitic stages, outside their host (e.g. *Trichodorus similis*). They typically feed on root epidermal cells as they migrate in the soil. Those that are endo-parasites are hypothesised to be more highly adapted than ectoparasites, spending at least part of their life cycle, and in some cases their entire life, within the host (Jones et al., 2013). These nematodes are often highly destructive.

Endo-parasites can be further subdivided into the migratory and sedentary species. Both lifestyles are found in Clade 12. The former feed as they migrate within host tissues, causing tissue damage, often resulting in necrotic lesions. These nematodes form no biotrophic interactions. In addition, Clade 12 includes two main groups of sedentary endoparasitic nematodes, the cyst nematodes (CN) and the root-knot nematodes (RKN). In both cases, the nematodes invade the root, and modify host root tissue to create a unique and highly metabolically active nutrient sink from which they feed for the duration of their life (Kyndt et al., 2013).

All sedentary endo-parasites in Clade 12 form obligate biotrophic interactions. This remarkable dependency on the host, and the ability to induce feeding sites, appears to have evolved on at least two separate occasions in Clade 12 (the RKNs and the CNs). However, given the range and diversity of nematode feeding sites within this clade, this is probably an underestimate. Although the life cycles of the RKNs and

the CNs are quite different, both spend the majority of their non-dormant life within the host, and both create a feeding site that is their sole source of external nutrition.

1.3.1. Root-knot nematodes

Root-knot nematodes are the most widespread sedentary endoparasitic nematodes. Some species are remarkable in that despite the close and presumably highly specialised biotrophic interaction, they are capable of infecting thousands of plant species (Trudgill, 1997). The life cycle of a root knot nematode, such as *Meloidogyne incognita*, starts when the second stage juveniles (J2) hatch from eggs in the soil. It is hypothesised that J2s may find their host by sensing gradients of ions, CO₂, pH or amino acids in the rhizosphere of growing plants (Perry, 1996). Once a host is located, they enter the root, often at the zone of elongation (Wyss and Grundler, 1992). By using a combination of mechanical disruption and degradation of cell wall polysaccharides by enzymatic secretions, they migrate intercellularly (between cells). The J2s initially move towards the growing root tip and then change direction at the root apical meristem prior to migrating towards the differentiating vascular cylinder (Wyss and Grundler, 1992). Once they reach pro-vascular cells the J2s stop migrating and initiate the formation of a feeding site. Each root knot nematode induces the formation of several modified cells, termed giant cells, from which they feed. Giant cells are created from individual parenchymal cells, by induction of multiple rounds of mitosis in the absence of cytokinesis. Normal cell division appears to be arrested at the formation of the cell plate in the giant cells, and vesicles that would normally fuse to form the cell plate instead disperse resulting in a binucleate cell. This process of mitosis in the absence of cytokinesis is repeated several times resulting in a mature giant cell that may contain up to 32 nuclei (Gheysen and Fenoll, 2002). The cell wall between two adjacent giant cells is characterised by well-developed ingrowths (Jones et al., 1974). A large increase in the number of plasmodesmata between adjacent giant cells is also observed suggesting that *de novo* plasmodesmata formation may have occurred (Jones et al., 1974). Proliferation of cortical and vascular tissue surrounding the giant cells results in a large gall that gives rise to the common name "root-knot nematode".

Feeding by root-knot nematodes occurs in phases, with pumping of the metacarpal bulb and subsequent withdrawal of host cell assimilate through the stylet. Root-knot nematodes induce several giant cells and feed on these sequentially, resulting in a characteristic pattern of head movement and stylet insertion (Wyss and Grundler, 1992). After formation of the giant cells, root-knot nematodes undergo three moults while feeding is arrested. Feeding is then resumed following the final moult to

female. Females feed for several weeks as they mature and swell. Females deposit eggs in a gelatinous matrix inside, or on the surface of, the gall.

1.3.2. Cyst nematodes

Cyst nematodes (including *Globodera* and *Heterodera* species) often have a highly restricted host range that is reflected in their common names (e.g. potato cyst nematode, soybean cyst nematode). In the UK, the potato cyst nematode is the most economically important plant parasitic nematode, causing losses to UK agriculture in excess of £50 million per year (UK Potato Council, 2013). Potato cyst nematodes (PCN) are thought to have originated from South America (Plantard et al., 2008, Grenier et al., 2010). It is hypothesised that PCN were introduced to Europe in the 19th Century along with potato material used for resistance breeding against the late blight pathogen *Phytophthora infestans* (Evans et al., 1975). Several introductions to Europe gave rise to the current populations (Hoolahan et al., 2012, Papayiannis et al., 2013, Picard et al., 2004). Two species of potato cyst nematode are present in the UK; *Globodera pallida* and *Globodera rostochiensis*. The H1 resistance gene, now present in many commercially viable cultivars (e.g. Maris Piper), provides effective control against the most prevalent European *G. rostochiensis* pathotype; Ro1. Widespread use throughout the UK has successfully suppressed this species, but has inadvertently led to an increase in prevalence of *G. pallida*; now present in 64% of potato growing fields (Minnis et al., 2002). No commercially viable resistant varieties are available for *G. pallida*. Current control measures rely on integrated pest management strategies of crop rotation, nematicides and trap crops. However, DEFRA's potato cyst nematode management guidelines suggest that growing potatoes more frequently than once every 6 years in infected fields provides no control against the pathogen.

1.3.3. The life cycle of cyst nematodes

The second stage juvenile (J2) of a typical cyst nematode, like that of RKNs, is the infective stage. The J2s hatch from eggs within cysts in response to the presence of host root diffusates. This ensures that the host-specific nematodes only hatch in the presence of a plant that they can infect, in contrast to the J2 of RKN that tend to hatch when conditions (temperature) are favourable for plant growth (Devine et al., 1996). It is generally accepted that chemosensory sensilla in the main sense organs, the amphids, mediate detection of gradients of either ions, CO₂, pH or amino acids in the rhizosphere of growing plants (Perry, 1996). Once they reach their host the J2s use a rapid thrusting action of their needle-like stylet, coupled with the secretion of cell wall degrading enzymes, to migrate intra-cellularly (through

cells), towards the vascular tissue, causing damage to the root tissue as they do so (Atkinson et al., 1995). This thrusting action of the stylet becomes more delicate and exploratory when the nematode is choosing an initial cell which will eventually become the highly metabolically active feeding site, known as a syncytium. *In vivo* observations of *Heterodera schachtii* feeding on *Arabidopsis thaliana* indicate that this initial syncytial cell (ISC) is selected based on the response from the plant to the insertion of the nematode stylet (e.g. the protoplast retains its shape (Urwin et al., 1997b) and the lack of callose-like depositions (Wyss and Grundler, 1992)). Once the ISC has been selected the nematode becomes sedentary. Observations show that feeding occurs in three phases (Wyss and Grundler, 1992). In Phase I, often the longest of the three, the stylet is inserted into the feeding site across the cell wall and contacts the plasma membrane. Cell contents are then ingested by continuous and rapid metacarpal bulb pumping. In Phase II the stylet is retracted before later being reinserted into the syncytium. Phase III is characterised by a continuous forward movement of secretory granules through the stylet into the syncytium (Wyss, 1992, Wyss and Grundler, 1992). It has been suggested that cyst nematodes are capable of withdrawing 4 times their body volume of host cell cytoplasm per day (Sobczak et al., 1997). During the following 1-2 weeks, feeding nematodes develop into either males or females. Sex determination in cyst nematodes is unclear, although it is thought that the size and 'quality' of the syncytium plays a key role (Sobczak and Golinowski, 2011). Syncytia resulting in poor nutrition for the nematode are observed to more frequently give rise to males. Males arrest feeding, moult within the cuticle and regain motility to become migratory once more. Females swell as they feed for a number of weeks. Fertilised females produce eggs within their body and their cuticle tans as a result of polyphenol oxidase activity to form a tough cyst that encloses the eggs. Cysts become detached from roots and can remain viable in the soil for decades.

1.3.4. Syncytia

Although root-knot nematode giant cells and cyst nematode syncytia are morphologically and developmentally distinct, the two structures show similar cellular features including reduced vacuoles and extensive proliferation of the smooth endoplasmic reticulum (SER), ribosomes, mitochondria and plastids (Gheysen and Fenoll, 2002, Gray et al., 1994, Bleve-Zacheo et al., 1997). In the early stages of syncytium induction, plasmodesmata between the ISC and the surrounding cells become enlarged and the walls between the cells partially degrade. Eventually the protoplasts of the ISC and its surrounding cells become fused to generate the multinucleate syncytium. This process is repeated so that additional layers of cells become incorporated into the syncytium. Prior to

incorporation adjacent cells appear to undergo mitosis (de Almeida Engler et al., 1999). The detection of CYCB1;1 transcripts inside the syncytium has led to the hypothesis that a shortened cell cycle (endoreduplication) until at least phase G2 is repeatedly stimulated to contribute to the greatly enlarged nuclei (de Almeida Engler and Gheysen, 2013, de Almeida Engler et al., 1999). Syncytia are also characterised by a reduced central vacuole and cytoplasm enriched in subcellular organelles and are highly metabolically active (Wyss and Grundler, 1992, Sobczak et al., 1999). The first cell wall openings between the ISC and its neighbouring cells are visible six hours after selection of the initial syncytial cell (Grundler et al., 1998), and a mature syncytium, composed of approximately 200 cells, reaches its maximum size ten days after infection (Urwin et al., 1997a). Once formed the syncytium is the sole source of nutrition for the remainder of the nematode's life.

Formation of the syncytium involves remarkable changes in gene expression with genes often differentially regulated by several orders of magnitude (Bauters et al., 2014). Moreover, a microarray analysis of syncytia induced by *Heterodera schachtii* in *Arabidopsis* showed that 18.4% of the 21,138 *Arabidopsis* genes tested were up regulated and 15.8% were down regulated when compared to uninfected controls (Szakasits et al., 2009). These profound changes in gene expression, and the consequent structural changes that occur during syncytium development are attributed to the actions of nematode effectors (Haegeman et al., 2012).

1.3.5. The role of effectors in cyst nematode biotrophy

Effectors play a critical role in the interactions of many plant pathogens, including nematodes, with their hosts. Although there are many widely varying definitions of effectors, for the purpose of this thesis the broad definition of effectors is used: Proteins (or other factors) secreted by a pathogen into its host in order to manipulate the host to the advantage of the pathogen.

Although nematodes have numerous tissues with the capacity to secrete proteins into the host (amphids, hypodermis, vulva), nematode effectors are classically hypothesised to originate from the oesophageal gland cells; two subventral and one dorsal (Figure 1-2). Based on observations of structural changes throughout the life cycle, the subventral gland cells are hypothesised to secrete effectors involved in invasion while the dorsal gland cell secretions are hypothesised to be involved in induction and maintenance of the feeding site (Hussey, 1989, Abad et al., 2008, Davis et al., 2004).

Effectors clearly play a key role in several aspects of the biology of cyst nematodes and particularly their interactions with their hosts. A variety of genes encoding cell wall degrading enzymes have been identified from several plant-parasitic nematode species including cyst and root-knot nematodes. These genes have high homology to bacterial sequences, are absent from most other animals and are therefore likely to have been acquired by Horizontal Gene Transfer (HGT) (Haegeman et al., 2011, Danchin et al., 2010). For example, a well characterised functional GH28 polygalacturonase has been identified in RKN and is expressed in the subventral gland cells (Jaubert et al., 2002). These proteins are likely to be involved in the breakdown of cell walls to aid the nematode in migration.

Cyst nematodes spend a significant proportion of their life within the host plant. Like other plant pathogens, cyst nematodes have evolved methods of suppressing host defences (Postma et al., 2012, Jaouannet et al., 2013). For example, a secreted SPRY domain-containing protein from the Clade 12 cyst nematode *Globodera rostochiensis* is able to suppress disease resistance responses mediated by the most common class of plant resistance genes; Coiled-coil Nucleotide-binding Leucine-rich repeat receptors (Postma et al., 2012). In addition, the 30C02 effector of the cyst nematode *Heterodera glycines* is able to interact with a plant β -1,3-endoglucanase as confirmed both in Yeast-two-Hybrid (Y2H) and *in planta* analyses (Hamamouch et al., 2012). Transgenic over-expression of 30C02 in *Arabidopsis thaliana* results in increased susceptibility to nematode infection. Additionally, either *in planta* RNA interference targeting 30C02 or a T-DNA insertion mutant defective in the host target result in reduced susceptibility. While β -1,3-endoglucanases are often involved in plant defence, it is hypothesised that the physical interaction with 30C02 may reduce callose deposition around the stylet thus suppressing the host defences. Interestingly these susceptibility phenotypes are cyst nematode specific, and no effect was observed with the root-knot nematode *M. incognita*.

Effectors also need to be produced that induce and maintain the feeding site, although due to the technical difficulties in working with an obligate biotroph, little information is available for these. However, it is known that plant-parasitic nematodes have evolved short proteins/peptides that mimic endogenous plant peptide hormones. The CLAVATA-like (CLE peptides) of the cyst nematode *Heterodera schachtii* are apparently secreted into syncytia and subsequently transported to the apoplast by plant mechanisms (Replogle et al., 2011). Since endogenous plant-CLEs carry out a range of functions, the precise role of the nematode peptides in parasitism is unclear, although they appear to be functional *in*

planta. A *Heterodera glycines* CLE peptide can complement an *Arabidopsis thaliana* CLE mutant (Wang et al., 2005) and over-expression of a nematode derived CLE peptide in *A. thaliana* results in a 'wuschel-like' phenotype similar to that caused by over expression of endogenous CLE peptides (Wang et al., 2011). These signalling peptides are some of the best characterised effectors in plant parasitic nematodes. The CLE-like peptides have been identified in both cyst and root-knot nematodes and are one of the few features that are shared between these two groups, leading to a suggestion that they play a fundamental role in the development of two morphologically diverse feeding sites. In addition, effector 19C07 from *Heterodera* species (*H. glycines* and *H. schachtii*), expressed in the dorsal gland cell of parasitic stage J2/J3 is hypothesised to be involved in feeding site development (Lee et al., 2011). This effector specifically interacts with the *A. thaliana* auxin influx transporter LAX3. It is proposed that this interaction increases LAX3-dependent auxin influx and may provide a mechanism to stimulate cell wall hydrolysis for syncytium development.

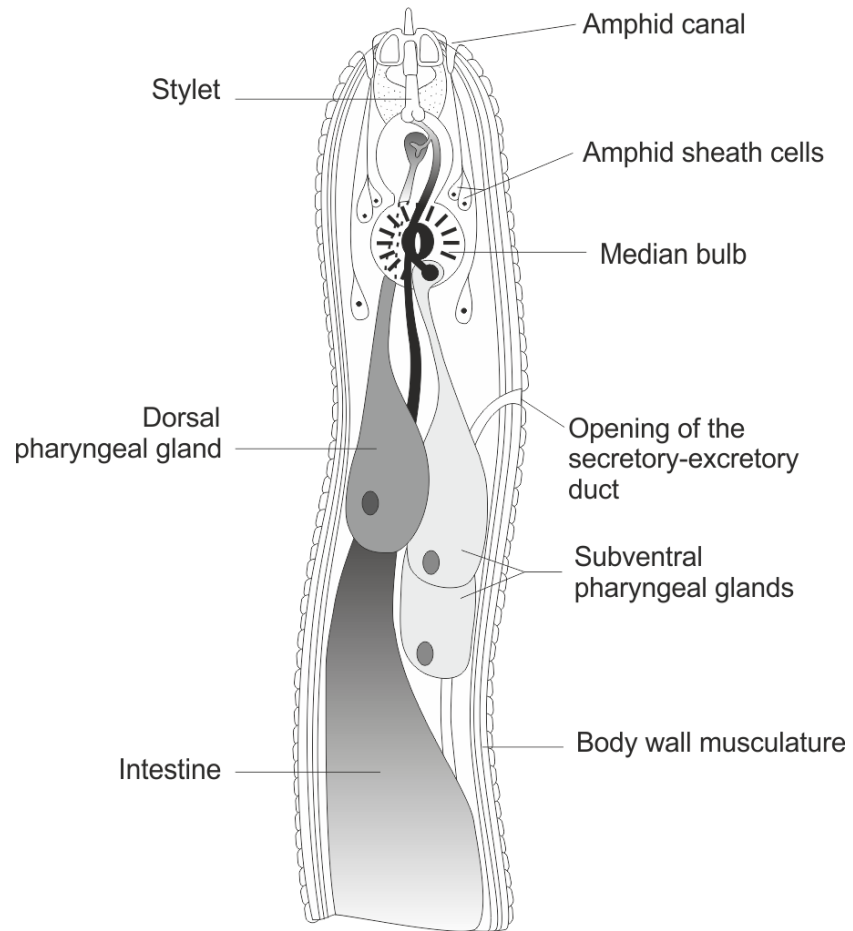


Figure 1-2 **Schematic representation of a typical tylenchid nematode.** Effectors have been identified that are secreted from the pharyngeal gland cells (dorsal and subventral) via the stylet, and directly into the apoplasm from either the hypodermis or the amphid openings.

1.3.6. Nematode feeding structures

Electron microscopy studies have identified two key feeding structures produced by the nematode that are directly associated with the plant-nematode interface, the feeding tube and the feeding plug (Holtmann et al., 2000, Endo, 1978, Rebois, 1980, Razak and Evans, 1976, Rebois et al., 1975). Feeding tubes are produced inside the feeding site and are thought to act as molecular sieves (Eves-van den Akker et al., 2014b). Feeding plugs appear as electron dense material intimately associated with the plant-nematode interface and are located in the apoplast between the anterior end of the nematode and the feeding site (Endo, 1978).

1.3.7. Feeding tubes

Feeding tubes appear to be specifically associated with plant-parasitic nematodes and have been described for species in Clade 1 and two groups of nematodes in Clade 12, suggesting an essential role in plant parasitism. However, the feeding tubes of migratory ectoparasites, such as *Trichodorus similis* (Clade 1) form around the stylet as it is inserted into the host cell. These feeding tubes can form without interaction with the host cell cytoplasm, bridge the cell wall and are open ended. The stylet of *T. similis* is completely solid at the anterior end (Wyss et al., 1979), and so it is impossible for it alone to function as a means to withdraw host cell cytoplasm. It is therefore hypothesised that the feeding tube of the Clade 1 plant-parasitic nematodes, with its open distal end, represents the true functional analogue for withdrawing host cell assimilates (Eves-van den Akker et al., 2014b). The name of these structures is somewhat misleading as they seem to have a different functional role from those produced in other nematode species.

Two independently evolved groups of obligate biotrophs within Clade 12 (RKN and cyst/reniform nematodes) both produce intracellular feeding tubes. Morphological comparison of the feeding tubes of the two groups suggests they are structurally different (Figure 1-3) (Hussey and Mims, 1991, Rebois, 1980, Wyss, 1992). Taken together, these findings suggest that feeding tubes within Clade 12 represent an excellent example of convergent evolution.

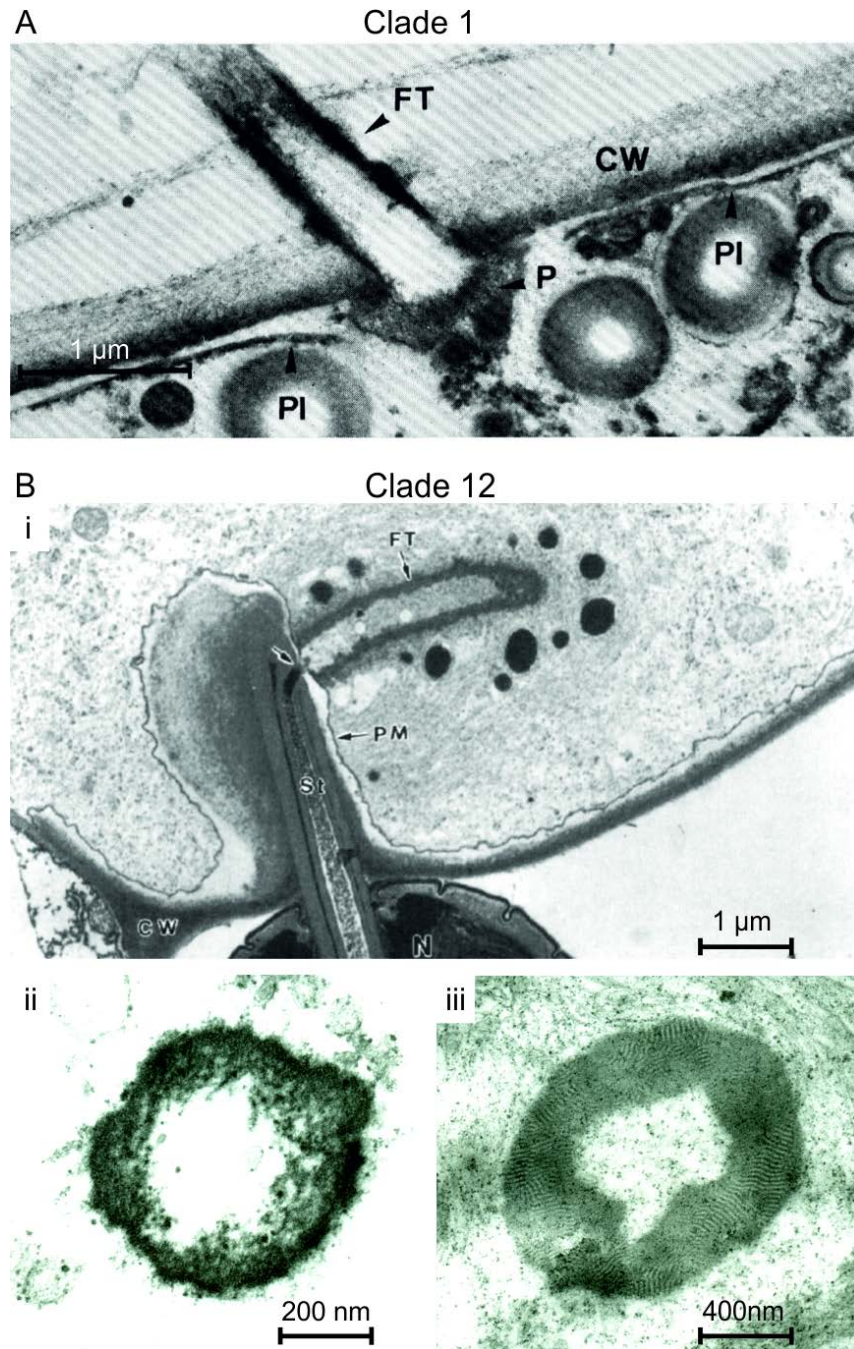


Figure 1-3 **Feeding tubes of plant parasitic nematodes.** A) Feeding tubes (FT) of the Clade 1 plant parasitic nematode *Trichodorus similis* appear to bridge the cell wall (CW) and disrupt the plasmalemma (PI) (modified from (Wyss et al., 1979)). Congealed cytoplasm can be seen where the feeding plug (P) contacts the host cell cytoplasm. B) Feeding tubes of Clade 12 plant parasites form inside the host cell cytoplasm, at the end of the stylet orifice (S), and contact the plasma membrane (PM) exemplified by *Rotylenchulus reniformis* (B-i modified from (Rebois, 1980)). Within the Clade 12 plant parasites, two morphologically distinct feeding tubes exist as seen from cross sections: the electron dense mesh of the cyst nematode feeding tube (B-ii), and the regular lattice structure of the root-knot nematode feeding tube (B-iii). No cross sections are available for *T. similis*, although longitudinal sections do not indicate a regular lattice structure.

Feeding tubes of Clade 12 plant-parasitic nematodes are blind ended tubes that form from nematode secretions within the feeding site at the end of the stylet orifice. Although no components of feeding tubes have been identified, they are hypothesised to be of nematode origin. Different nematode species infecting the same plant produce feeding tubes characteristic of their species. Similarly, nematodes that infect different plant species produce a similar feeding tube in each host (Hussey and Mims, 1991). *In vivo* observations suggest that feeding tube components originate from the pharyngeal gland cells although no direct evidence has yet been provided (Grundler and Böckenhoff, 1997).

Feeding tubes have been hypothesised to assemble in response to an unknown interaction with the host cell cytoplasm (Rumpfenhorst, 1984, Berg et al., 2009, Sobczak and Golinowski, 2009, Sobczak et al., 1999) as inducing sedentary parasitic nematodes to secrete pharyngeal gland contents *in vitro* does not result in the formation of feeding tubes (Hussey and Mims, 1991). A new feeding tube is formed before each bout of feeding, and is detached from the stylet orifice and discarded after feeding and stylet withdrawal. Discarded feeding tubes persist for up to 14 days post production, and accumulate within the feeding site (Hussey and Mims, 1991). Everything the nematode ingests during its juvenile and adult life must pass through the walls of the feeding tube. Intracellular feeding tubes are often associated with whorls of membrane systems, with an unknown function that resemble smooth endoplasmic reticulum (Rebois, 1980, Sobczak et al., 1999, Grundler and Böckenhoff, 1997). It has been hypothesised that the elaborate membrane systems associated with feeding tubes may provide direct synthesis and transport of nutrients into the tube or in some way facilitate feeding tube formation (Grundler and Böckenhoff, 1997) as they still form around, or remain attached to, discarded feeding tubes (Hussey and Mims, 1991).

1.3.8. Feeding tube function

Despite the morphological differences between them, the feeding tubes of the various sedentary endoparasitic nematodes probably share a common function that may be vital to parasitism of plants. It is likely that feeding tubes facilitate uptake of soluble assimilates from the feeding cells (Hussey and Mims, 1991). It has been hypothesised that the feeding tube may act as a molecular sieve, excluding large debris from being taken up by the nematode during bouts of feeding. Exclusion of some macromolecules rather than others, on the basis of size, may confer several benefits to the nematode. The size exclusion hypothesis states that the feeding tube excludes large protein complexes or small organelles that may cause partial or

total blockage of the stylet. Blocking of the stylet would undoubtedly be fatal as it is essential to the uptake of nutrition throughout the feeding stages of the nematode. Moreover, effector proteins involved in the maintenance of the feeding site are delivered into the plant via the stylet. The size exclusion hypothesis is supported by the fact that a new feeding tube is synthesised before each bout of feeding. This implies that an aspect of the feeding tube reduces in efficiency over time and requires replacement. The proposed molecular sieve may become blocked and a new one must be synthesized. Feeding tubes are reported to persist in the cytoplasm up to 14 days post production (Hussey and Mims, 1991). Degradation of the feeding tube is therefore unlikely to underlie the need for their replacement.

1.3.9. Feeding plugs

Feeding plugs have been reported in a range of cyst nematodes (Sobczak et al., 1999, Holtmann et al., 2000, Endo, 1978) and a similar structure, variously referred to as a feeding plug or feeding peg, has also been described for *Rotylenchulus reniformis* (Razak and Evans, 1976, Rebois, 1980). *R. reniformis* is not a cyst nematode, but is the closest known relative (van Megen et al., 2009) and also induces the formation of syncytia similar in ontogeny to those of cyst nematodes (Agudelo et al., 2005, Rebois et al., 1975). This is probably reflective of a common sedentary plant parasitic ancestor. Two previous studies showed the feeding plugs of cyst nematodes to be continuous with the amphid openings of sedentary females (Endo, 1978, Sobczak et al., 1999), one of which concluded that the feeding plug originates from the amphidial canal (Endo, 1978). Both are of the view that despite feeding plugs forming in the host apoplasm, they are of nematode origin. The feeding pegs of *R. reniformis* are more difficult to distinguish from cell wall material (Figure 1-4) and may be derived mainly, although not exclusively, from callose-like plant material (Rebois, 1980, Rebois et al., 1975). In Clade 1, feeding plugs have been described for *T. similis*, although these appear to be very similar in composition to, and contiguous with, their feeding tubes (Wyss et al., 1979) (Figure 1-4). Feeding plugs have not been identified in electron microscopy studies of root-knot nematode giant cells (Hussey and Mims, 1991). Feeding plugs of the Clade 12 nematodes are produced early in the infection process (between 12 – 24 hours post infection), and persist throughout parasitism (Sobczak et al., 1999). They also appear to increase in size during infection (Sobczak et al., 1999, Rebois, 1980), suggesting both continual renewal and an important function.

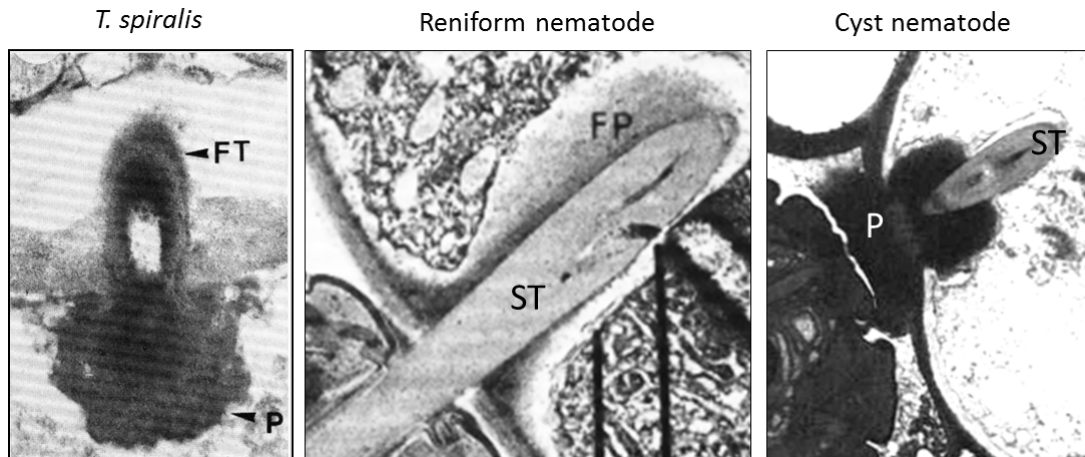


Figure 1-4 **Feeding plugs of plant parasitic nematodes.** For *T. similis* the feeding plug (P) forms inside the host cell adjacent to the cell wall, but attached to the feeding tube (FT). For reniform nematodes the feeding plug (FP) forms around the stylet (ST) inside the host cell and is difficult to distinguish from cell wall material. Cyst nematode feeding plugs (P) cross the cell wall and encase the stylet (ST).

1.3.9.1. Feeding plug function

Despite differences in morphology, structure, and even origin, all feeding plugs may be functionally analogous. It has been suggested that feeding plugs may play a role in anchoring the nematode in place throughout development (Endo, 1978). Dense pathogen-derived fibrous material in close association with the host-parasite interface has been observed in animal parasitic nematodes (Simpson and Lee, 1976) and fungi (Aist, 1976) and been linked to attachment. Each root-knot nematode induces multiple giant cells and feeds from them in turn throughout development inside the plant (Wyss and Grundler, 1992). By contrast, each cyst or reniform nematode induces, and feeds from, a single syncytium. Cyst nematodes undergo three moults within the host root, and during each of these moults the stylet is replaced (Wyss, 1992). Although root-knot nematodes also undergo moults, unlike cyst nematodes they do not feed between moults so may therefore not reinsert the stylet until adult. Adhesion of the cyst and reniform nematodes to the single syncytium from which they feed is likely to be beneficial. However, root-knot nematodes feed from multiple different giant cells, and so adhesion to one is likely to be detrimental. Although the feeding plugs of *T. similis* appear to be of similar material to that of the feeding tube of *T. similis*, and can form without contact with host cell cytoplasm (Wyss et al., 1979) they may be functionally similar to feeding plugs of Clade 12 plant-parasitic nematodes. If the feeding tube of *T. similis* is functionally analogous to the stylet of other plant-parasitic nematodes, feeding plug formation at the base of the feeding tube inside the host cell, irrespective of its composition, could also have a role in adhesion. Feeding plugs would not be required for any of the migratory endoparasites of Clade 12 or the migratory ectoparasites of Clade 2, and have not been described. Feeding plugs of cyst nematodes appear to be nematode derived but distinct from the feeding tube. Since the identification of feeding plugs in the 1970s (Endo, 1978, Razak and Evans, 1976), very few studies have been published on the subject, primarily due to the difficulties in working with sedentary endoparasites.

1.4. Project overview

The aims of this project were to characterise nematode derived proteins involved in plant-parasitism with a direct focus on the feeding tube and feeding plug. Due to the clear functional differences, the terms “feeding tube or plug” from this point onwards will apply to Clade 12 plant-parasitic nematodes only. As molecular sieving characteristics for the feeding tube have been previously described, this will be studied in detail by first reviewing the reported data in the literature. Attempts will then be made to provide further resolution, and a more pragmatic view on, feeding tube size exclusion. Using a combination of *in silico* measures of protein size coupled with experimental evaluation of protein heterogeneity in the liquid, solid and gas phase we aim to further characterise feeding tube size exclusion. Secondly, candidate feeding structure genes will be identified from the genome and accompanying transcriptome data of the cyst nematode *Globodera pallida*. Using molecular biology techniques, candidate genes will be characterised to determine if feeding tube or plug components have been identified. Finally, putative feeding structure genes will then be identified in strategically-chosen closely-related species by bioinformatic analysis. Whole transcriptomes will be generated for two species, and candidate feeding structure gene homologues will be identified.

2. General materials and methods

All routine chemicals and reagents were supplied from either Sigma Aldrich (Dorset, United Kingdom) or Fisher Scientific (Loughborough, United Kingdom) unless specified otherwise.

2.1. Bacterial Growth Media

2.1.1. Super Optimal Broth (SOB)

2 % w/v Tryptone

0.5 % w/v Yeast extract

10 mM NaCl

2.5 mM KCl

10 mM MgSO₄

10 mM MgCl₂

2.1.2. Super Optimal broth with Catabolite repression (SOC)

As above with the addition of glucose to 20 mM

2.1.3. Luria-Bertani (LB)

10 g/l Tryptone

5 g/l NaCl

5 g/l Yeast extract

2.1.4. Luria-Bertani Agar (LB Agar)

As above with the addition of 1% Agar (w/v)

2.2. Plant growth media

2.2.1. Murashige and Skoog (liquid MS 20)

4.3 g/l Murashige and Skoog medium (with vitamins)

20 g/l sucrose

pH 5.3 - 5.6

(where specified ½ MS will refer to 2.15 g/l Murashige and Skoog medium (with vitamins) and 10 g/l sucrose at pH 5.3-5.6)

2.2.2. Murashige and Skoog Agar (MS-A)

As above with the addition of 2 – 4 g/l plant agar depending on application

2.3. Biological material

2.3.1. Maintenance of cyst nematode populations

Cysts of *Globodera pallida*, *G. rostochiensis* and *Heterodera schachtii* were stored at 4 °C. Independent stocks were maintained for *Globodera* species by growing susceptible potato cultivar Désirée in soil infected with approximately 50 eggs/g. *Heterodera schachtii* were maintained by planting 4 week old cabbage plants (*Brassica oleracea* var. Primo II) in 50:50 sand loam mix containing approximately 20 eggs/g. For both *Globodera* and *Heterodera* species plants were allowed to grow in infected soil for approximately 2-3 months before the above ground parts of the plants were removed. Damp soil containing *H. schachtii* cysts and dry soil containing cysts of *Globodera* species were stored at 4 °C. Cysts were collected by resuspending the 50:50 sand loam mixture in three volumes of water, the silt was allowed to settle leaving the floating cysts on the surface. Cysts were concentrated using a 300 µm sieve and collected into a 1.5 ml microfuge tube for storage.

2.3.2. Hatching of second-stage juveniles

Second-stage juveniles (J2) were hatched from cysts following diapause at 4 °C. Cysts were put onto a > 30 µm mesh, and soaked at 20 °C in potato root diffusate or 2 mM ZnCl₂, for *Globodera* and *Heterodera* species respectively. Potato root diffusate was prepared by soaking 3 week old potato plant roots in water at 4 °C for 16 hours (80 g/L). Diffusate was filter sterilised prior to use. J2s that passed through the mesh were collected every 3-4 days, for a period of up to 2 weeks, and stored in 15 ml polystyrene tubes (Starstedt, Leicester). J2s were then centrifuged at 3000 g for 10 minutes, and the pellet was resuspended in sterile tap water. J2s were then stored at 10 °C for up to 2 weeks prior to use.

2.3.3. Collection of parasitic-stage nematodes

Potato tubers stored at 4 °C were placed at room temperature until chits 0.5 – 1 cm had formed. At this stage, whole potatoes were planted in soil containing 80 eggs/g and allowed to grow for 1-2 weeks longer than the age of sample needed. For example 14 days post infection nematodes would be collected 21-28 days post planting, where the most numerous stage of nematode would be that of 14 days post infection. Alternatively, where greater synchrony of infection was required, individual chits would be removed from tubers and planted in root trainers (Haxnicks, Oxford, United Kingdom). Following an initial growth phase of the potato plants, J2 nematodes could be pipetted directly onto the roots of the plants. Typically a concentration of 1 juvenile per µl of water was used, infecting 2 - 3000

nematodes across the entire root system per plant. For this method nematodes were collected on the date 1-2 days longer than the age of sample needed. For example 14 days post infection nematodes would be collected 15 - 16 days post inoculation. In this case the most developed nematodes would be collected.

Plants were not watered 2 days prior to collection of feeding stage nematodes. Roots were released from dry soil by gentle agitation and washed in tap water to remove remaining soil. Clean root tissue was cut into approximately 2 cm sections and subjected to a short < 1 second burst in a standard food blender. The contents were then passed over a series of sieves (1 mm, 500 µm, 250 µm, 150 µm, 63 µm and 25 µm mesh size). To collect nematodes at a particular life stage the debris was collected onto a sieve with an appropriate mesh size prior to being washed into a 50ml Falcon tube containing 40% sucrose. This was centrifuged at 4000 g for 5 minutes, the supernatant was then passed back over the corresponding sieve, the sucrose solution washed off with tap water, and the nematodes collected into a glass beaker. Individual nematodes were then collected into watch glasses by pipetting.

2.4. Molecular Cloning

2.4.1. Polymerase Chain Reaction (PCR)

PCR was carried out for sequences of interest with the relevant primers using the manufacturer's instructions included with BioTaq Red DNA polymerase (Bioline, London). Each PCR reaction contained 1.5 units of Taq DNA polymerase, 1.5 mM MgCl₂, 1x NH₄ reaction buffer, 0.2 mM dNTPs, 0.5 µM of each relevant primer. PCR cycling conditions were 5 minutes of 95 °C denaturing phase, followed by 30 cycles of 95 °C for 1 minute, relevant annealing temperature for 30 seconds and extension at 72 °C for 1 minute per kb of expected amplicon. A final extension step at 72 °C for 5 minutes was added after all cycles were complete. Where downstream processes required purified DNA the PCR reaction was cleaned using the QIAquick PCR Purification Kit (Qiagen, Manchester, United Kingdom) following the manufacturer's instructions.

In the cases where a proof reading enzyme was required, Phusion High-fidelity DNA polymerase (New England Biolabs, Ipswich, United Kingdom) was used. Each PCR reaction contained 1x Phusion HF Buffer, 0.2 mM dNTPs, 1 µM of each relevant primer and 1 unit of Phusion DNA polymerase per 50 µl reaction. PCR cycling conditions were an initial denaturation step of 98 °C for 30 seconds, followed by 30-35 cycles of 98 °C for 10 seconds, the relevant annealing

temperature for 30 seconds, extension at 72 °C for 30 seconds per kilobase, ending with an extension of 72 °C for 5 minutes. In the cases where the PCR product would be used for subsequent TA cloning (Section 2.4.4), after the PCR had finished, an additional 1 unit of BioTaq Red DNA polymerase was added and the PCR reaction incubated at 72 °C for a further 10 minutes, followed by immediate purification using QIAquick PCR purification Kit (Qiagen).

2.4.2. Restriction digestion of DNA

Restriction enzyme (RE) digestions were typically carried out in a 20 µl volume with a final concentration of 1 x relevant NEBuffer (New England Biolabs) containing 20 units of the relevant RE and 1 µg of DNA. Where multiple restriction enzymes were used in the same reaction, a reaction buffer that indicated 100 % activity for each enzyme was used. Digestion was carried out at the relevant digestion temperature (usually 37 °C) for 3 hours. Complete digestion was confirmed by agarose gel electrophoresis.

2.4.3. Agarose gel electrophoresis and gel extraction

To determine the size of the DNA sample of interest, approximately 10 % of the reaction (PCR or RE-digestion) was analysed by agarose gel electrophoresis. 1.8 % w/v agarose (Invitrogen) was added to TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) and dissolved by microwaving on full power for 2 minutes. DNA was visualised by addition of GelRed (Cambridge Bioscience, Cambridge, UK) to the agarose, before it set, at a concentration of 1:20,000. The gel was electrophoresed at 100 volts for sufficient time to resolve the relevant bands. Where necessary the relevant DNA fragment was recovered from the gel using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions.

2.4.4. DNA ligation

Linear DNA fragments suitable for ligation (corresponding T-A overhangs or matching restriction digest "sticky-ends") were combined at a 3:1, insert:vector molar ratio. A final reaction volume of 10 µl of 1x T4 DNA ligase buffer containing 1 unit of T4 DNA ligase (Invitrogen, Paisley, United Kingdom) was incubated at 16 °C for 16 hours.

2.4.5. Plasmid DNA extraction from bacteria

Positive colonies as identified by PCR were used to inoculate 5 ml of LB medium containing the relevant antibiotics. This inoculated culture was incubated over night

at 37 °C. Three millilitres of culture were used in the Qiagen Mini Prep Kit (Qiagen) following the manufacturer's instructions.

2.4.6. DNA extraction from plant and nematode tissue

Leaf, root or nematode tissue samples were frozen at the point of collection in liquid nitrogen. Frozen tissues were ground using a cooled micro pestle, the resultant powder was resuspended in 700 µl of DNA extraction buffer (0.1 M Tris (pH 8.0), 0.5 M NaCl, 50 mM EDTA), vortexed and incubated at 65 °C for 15 minutes. 500 µl of phenol:chloroform:isoamylalcohol (25:24:1) were added to each sample, vortexed to mix, and centrifuged at 12,000g for 10 minutes. The upper aqueous phase was added to a separate tube containing 500 µl of chloroform-isoamylalcohol (24:1) and the vortex and centrifugation steps were repeated. The top phase was then added to a separate tube containing 500 µl isopropanol, mixed by inversion and centrifuged at 12,000g for 10 minutes. The supernatant was discarded and the pellet washed with 500 µl of 70 % EtOH. The pellet was collected by centrifugation at 12,000g for 2 minutes, followed by removal of all the supernatant. The pellet was allowed to air dry (until clear) and then resuspended in an appropriate volume of sterile deionised water containing 1 µl DNase-free RNase (Qiagen) and incubated at 65 °C for 15 minutes. The resultant purified DNA was stored at -20 °C for later use.

2.4.7. RNA extraction from plant and nematode samples.

Samples were flash frozen in liquid nitrogen upon collection. An RNeasy mini kit (Qiagen) was used to extract RNA from frozen samples following the manufacturer's instructions. For extraction from nematode samples, the manufacturer's instructions for animal tissues were followed. In both cases the optional on-column DNase I digestion (Qiagen) was carried out.

2.4.8. cDNA synthesis

First strand cDNA synthesis was carried out on column purified RNA samples. Between 100 ng and 1 µg of total RNA was used in each cDNA synthesis reaction. Superscript II Reverse transcriptase (Invitrogen) was used according to the manufacturer's instructions, using oligo(dT) primers (Bioline). The optional denaturation step (65 °C) was routinely used.

2.4.9. Semi Quantitative Reverse Transcriptase - Polymerase Chain Reaction (SQRT-PCR)

RNA was extracted from frozen tissue samples and cDNA synthesised according to sections 2.4.7 and 2.4.8 respectively. When comparing between samples, equal quantities of RNA were used in the cDNA synthesis reactions. Gene-specific primers were then used to amplify the gene of interest from the cDNA at typically between 25-28 cycles. A number of cycles is decided upon for a given gene based on a comparison between samples. If bands amplified from all samples are very, and crucially equally, intense, this is not considered to be quantitative and so the number of cycles is reduced. At such a reduced number of cycles, it is assumed that the determining factor in the amount of product generated is the number of mRNA copies that were present in the total RNA extract (assuming cDNA synthesis is non-selective). The relative brightness of each band was estimated relative to a control “housekeeper” gene. For both plant and nematode samples the housekeeper gene used was Elongation Factor 1 α (EF1 α).

2.4.10. Preparation and transformation of chemically ultra-competent *Escherichia coli*

To prepare ultra-competent *Escherichia coli* cells 10-12 large colonies were used to inoculate 250 ml of sterile SOC medium in a 1 L flask. Cells were allowed to grow at 19°C with shaking at 200 rpm until they reached OD_{600nm} = 0.5 (between 24-36 hours). Once the desired optical density was reached, the cells were cooled on ice for 10 minutes, followed by centrifugation at < 100g for 10 minutes at 4°C. The cell pellet was gently resuspended in 80 ml ice-cold TB (10 mM PIPES, 15 mM CaCl₂, 250 mM KCL) and stored on ice for 10 minutes. The cells were centrifuged a second time using the same conditions, followed by resuspending in 20 ml ice-cold TB with the addition of 1.4 ml DMSO. 100 – 200 μ l of cells were aliquoted into individual 1.5 ml microfuge tubes, flash frozen in liquid nitrogen and stored at -80°C.

To transform ultra-competent cells, an aliquot per sample transformation was thawed on ice. 1-2 μ l of plasmid or ligation mix was added to thawed cells and incubated on ice for 5 minutes. Typically 10 μ l, 30 μ l and the remaining volume of cells was spread onto individual pre-warmed LB agar plates containing the relevant antibiotic concentration without a growth phase. Plates were incubated over night at 37°C for colonies to grow.

2.4.11. Preparation and transformation of electro-competent *Escherichia coli*

A single colony of *E. coli* was used to inoculate 5 ml of SOC medium. The culture was incubated at 37 °C with shaking at 200 rpm for 16 hours. 1 ml of this culture was used to inoculate 200 ml of SOC medium, which were incubated at 37 °C (200 rpm) until OD_{600nm} was between 0.6-0.9 indicating log phase growth. The cell suspension was centrifuged at 4 °C for 15 minutes at 2,500g, the supernatant was discarded and the pellet re-suspended in 1 volume of ice-cold 10 % v/v glycerol. Following a 5 minute incubation on ice, the centrifugation was step repeated, and the pellet re-suspended in 0.5 volumes of 10 % ice cold glycerol. The cells were incubated on ice for 5 minutes, pelleted by centrifugation a third time, and re-suspended in 0.1 volumes of ice cold 10 % glycerol. Following a final 5 minute incubation on ice the cells were centrifuged again, and the pellet re-suspended in 0.01 volumes of 10 % glycerol. The cells were divided into 90 µl aliquots in pre-chilled 1.5 ml microfuge tubes, flash frozen in liquid nitrogen and stored at -80 °C for future use.

Prior to transformation, one aliquot per event was thawed on ice for 10 minutes, LB agar plates containing the relevant antibiotics were pre-warmed at 37 °C and the electroporation cuvettes were chilled on ice. To transform, 1-2 µl of vector or ligation mix was added to each aliquot and stirred gently with the pipette tip. The cells were transferred to the chilled electroporation cuvette, the cuvette wiped dry and loaded into the electroporator (BioRad MicroPulser). Following a single pulse of 1500 Volts for 6.1 ms, 500 µl of room temperature SOC medium were added to the cells, and transferred to a 1.5 ml microfuge tube. To allow expression of the antibiotic resistance, cells were grown at 37 °C for 1 hour at 200 rpm. Typically 10 µl, 30 µl and 100 µl of cells were plated onto three individual plates and the remaining cells stored at 4 °C for later use if necessary.

2.4.12. Preparation and transformation of competent *Agrobacterium tumefaciens / rhizogenes*

Five millilitres of LB medium containing rifampicin (50 µg/ml) were inoculated with a single colony of the desired *Agrobacterium* strain, and incubated with shaking at 200 rpm for 16 hours at 28 °C. 2 ml of this culture was used to inoculate 50 ml of LB medium rifampicin (50 µg/ml) which was replaced to the incubator for a further 5 hours until the culture OD_{600nm} measured between 0.5 and 1 using a bench top spectrophotometer (Colorimeter model 45, Fisher). The culture was chilled on ice, centrifuged for 5 minutes at 4000g at 4 °C, and the supernatant discarded. The

pellet was resuspended in 1 ml ice cold 20 mM CaCl₂ by gentle shaking, and divided into aliquots of 1-200 µl prior to being frozen in liquid nitrogen.

To transform the *Agrobacterium*, an aliquot was thawed on ice and 10 – 100 ng of purified plasmid were added and mixed gently by stirring with the pipette. The cells were frozen again in liquid nitrogen, and placed in a 37 °C water bath for 5 minutes. 1 ml of LB media was added, the cells transferred to a 15 ml polypropylene tube (Grenier bio-one, Frickenhausen, Germany), and incubated at 28 °C for at least 2 hours with shaking (200 rpm). Three quantities, typically 50 µl, 250 µl and the remaining sample, were spread onto LB medium plates containing 50 µg/ml rifampicin and the relevant selection antibiotic. The plates were incubated upside down at 28 °C for 2-3 days, visible colonies were screened by PCR to confirm presence of desired plasmid.

2.4.13. Colony screening

Colonies were screened for presence of desired gene by PCR. A single colony was touched with a p200 tip. This tip was then inserted into a 1.5 ml microfuge tube containing 100 µl H₂O and mixed by pipetting up and down. 2 µl of this suspension were used in a subsequent PCR reaction as template, using the relevant primer pair. PCR was carried out as described in section 2.4.1 with the addition of a 5 minute 99 °C incubation before any cycling.

2.5. Protein expression, extraction and analysis

2.5.1. Bacterial protein expression

Relevant coding regions were cloned into the inducible bacterial expression vector PET28(b). Correct constructs, identified by RE digest and subsequent sequencing, were transformed into the expression strain of *E. coli*, BL21 DE3-RIL. Positive colonies, screened by PCR using gene specific primers, were initially tested for expression on a small scale prior to protein purification. A single colony was inoculated into 2 ml of LB medium containing the relevant concentration of antibiotics (in this case 50 µg/ml kanamycin and 24 µg/ml chloramphenicol) and incubated with shaking at 37 °C for 3 hours. 200 µl of each culture were removed and centrifuged at 12,000g for 1 minute. The pellet was resuspended thoroughly by vortexing in 100 µl 1x SDS-PAGE loading buffer (final concentration 0.05% Bromophenol blue, 10% glycerol, 1% SDS, 0.05 M Tris-Cl pH 6.8, and the addition of 1 µl of β-mercaptoethanol per sample) and kept for later use as an un-induced control. Protein expression in the remaining culture was induced by addition of IPTG to a final concentration of 1 mM and returned to the shaking incubator for a

further 3 hours at 37 °C. As a negative control 100 µl of culture were removed prior to induction of protein expression (to approximately account for growth of cells), centrifuged at 12,000g for 1 minute, and the pellet resuspended thoroughly by vortexing in 100 µl 1x SDS-PAGE loading buffer. 10 µl of un-induced and induced cell extracts were electrophoresed in parallel on an SDS-PAGE gel to confirm expression of a protein of the expected molecular weight.

2.5.2. SDS-PolyAcrylamide Gel Electrophoresis (PAGE)

SDS-PAGE gels were prepared in two parts, a main gel, and a stacking gel, for each the ratio of acrylamide:bis-acrylamide was 37.5:1. The main gel consisted of 12% acrylamide (w/v), 0.32 % bis-acrylamide (w/v), 0.3125 M Tris pH 8.8, 0.08% SDS (w/v), 0.02% ammonium persulphate (w/v), and 0.05% TEMED. The main gel was poured and allowed to set, before the stacking gel was prepared. The stacking gel was prepared in the same manner as the main gel except with a final concentration of 4% acrylamide (w/v), 0.1 % bis-acrylamide (w/v), and 0.3125 M Tris pH 6.8. The comb containing the wells was inserted in the stacking gel before it set. Protein samples were boiled in 1 x SDS-PAGE loading buffer for 10 minutes, prior to loading. The gel was electrophoresed at 200 V in 1 x running buffer (3 g Tris, 14.4 g glycine and 1 g SDS per litre) until the dye front reached the end of the gel. Protein was stained by incubation at room temperature for 1 hour in Coomassie stain (1 g Coomassie Brilliant Blue, 400 ml MeOH, 100 ml acetic acid and 500 ml H₂O). To visualise the protein, the gel was de-stained for a minimum of 1 hour at room temperature (400 ml MeOH, 100 ml acetic acid, 500 ml H₂O), or until bands were clearly visible.

2.5.3. Small-scale denaturing protein purification

Prior to large scale purification, His-tagged proteins expressed in bacteria were subject to small scale denaturing purification using Ni-NTA resin (Qiagen). A single colony was used to inoculate 5 ml of LB medium containing the relevant antibiotics, which were incubated at 37 °C over night with shaking at 200 rpm. Five microliters of this initial culture were used to inoculate a fresh 5 ml of LB medium containing the relevant antibiotics, which were returned to the incubator for 3 hours. At this stage IPTG was added to a concentration of 1 mM to induce protein expression, and the culture was returned to the incubator for a further 3 hours. To extract protein, 1 ml of culture was pelleted in a 1.5 ml microfuge tube by centrifugation at 12,000g for 1 minute. The pellet was resuspended in 200 µl of lysis buffer B (8 M urea, 0.1 M NaH₂PO₄, 0.01 M Tris HCl pH 8.0), vortexed to lyse cells, and the remaining cell debris was pelleted by centrifugation at 12,000g for 10 minutes. The

supernatant was removed to a fresh tube containing 50 μ l of 50 % Ni-NTA resin slurry, and mixed at room temperature for 30 minutes. The resin was pelleted by centrifugation at 12,000g for 1 minute, and washed 3 times in wash buffer C (8 M urea, 0.1 M NaH_2PO_4 , 0.01 M Tris HCl pH 6.3). Protein was eluted by incubating the resin at room temperature for 2 minutes in 50 μ l of wash buffer C containing 100 mM EDTA to chelate the Ni^{2+} ions from the resin. The resin was pelleted by centrifugation at 12,000g for 15 seconds, and the supernatant containing purified protein was removed to a fresh tube, and tested for the presence of the expected protein by SDS-PAGE.

2.5.4. Large-scale protein expression and purification

Having identified bacterial clones expressing the protein of interest in a translational fusion to a His tag, a single colony/scrape of a glycerol culture was added to 5 ml of LB medium, containing the relevant antibiotics, and incubated at 37 °C for 16 hours at 200 rpm. The entire 5 ml culture was used to inoculate 500 ml of LB medium in a 2 L conical flask containing the same final concentration of relevant antibiotics. This culture was incubated at 37 °C and 200 rpm until an optical density of 0.5-0.8 at $A_{600\text{nm}}$ was reached. IPTG was then added to a final concentration of 1 mM and the culture was returned to the incubator for a further 3 hours (or appropriate temperature and time depending on protein being expressed). Following induction the culture was cooled on ice for 10-30 minutes, and centrifuged at 4000g for 30 minutes at 4 °C in a pre-cooled centrifuge. The supernatant was removed and the cell pellet stored at -20 °C prior to protein purification.

Frozen cell pellets were resuspended in 30 ml resuspension buffer (0.3 M sodium phosphate pH 8.0, 0.5 M NaCl, 5 % glycerol v/v, 0.1 mM EDTA and 10 mM imidazole) by vortexing thoroughly. PMSF was added as a proteinase inhibitor to a final concentration of 0.1 mM, followed by lysozyme and Triton-X100 to final concentrations of 0.2 mg/ml and 0.1 % respectively. The suspension was incubated with gentle swirling at room temperature for 5 minutes or until the cell suspension became viscous. At this point 90 μ l of 1 M MgCl_2 were added along with 100 μ l of DNase at 1 mg/ml (stored in 10 mM Tris.HCl pH 7.9, NaCl 150 mM), and incubated again at room temperature for 5 minutes until the solution returned to its original viscosity. The supernatant was centrifuged at 15,000g for 30 minutes at < 5 °C in a pre-cooled centrifuge. The clear supernatant was decanted to a new polypropylene tube and stored on ice prior to nickel affinity purification.

A 1 ml HIS-Trap Fast Flow Nickel Tag Affinity (FF-NTA) column was loaded on an ÄKTA explorer instrument (GE Healthcare, Buckinghamshire, UK), and equilibrated

first in buffer His B (20 mM sodium phosphate pH 8.0, 0.5 M NaCl, 5 % v/v glycerol, 500 mM imidazole, 0.1 mM EDTA, 5 mM 2-mercaptoethanol) and subsequently in His A (20 mM sodium phosphate pH 8.0, 0.5 M NaCl, 5 % v/v glycerol, 10 mM imidazole, 0.1 mM EDTA, 5 mM 2-mercaptoethanol). The soluble total protein extract was passed across the NTA column at a flow rate of 1 ml/minute. Bound proteins were eluted using a gradient of increasing imidazole by altering the percentage of His B from 0% to 100% over 20 column volumes. Real time absorbance at 280nm and 230nm was collected during elution to indicate proteins eluting from the column. 2 ml fractions were collected and those containing the protein of interest were confirmed by SDS-PAGE. Pure fractions were pooled and dialysed against 150 mM NaCl/10 mM Tris pH 7.5 using a 3 kDa spectra/por membrane (Spectrum, Breda, Netherlands) for 16 hours at 4 °C with two buffer changes. Dialysed proteins were subsequently stored at 4 °C for future analyses. The absorbance maximum of each protein sample at 280 nm was recorded. The amino acid sequence of each protein was analysed using 'Protparam' (web.expasy.org/protparam/ 14-10-11) to obtain an extinction coefficient. The concentration of each sample was then determined using a derivative of the Beer-Lambert Law, $A / E = C$. Where:

A = absorbance at 280 nm, E = extinction coefficient and C = concentration in mg/ml.

2.5.5. Removal of His-Tag from purified protein: Thrombin digestion

Proteins expressed using the PET28 system contained a thrombin cleavage site between the His-Tag and start of the sequence. The His-tag was removed after purification and prior to further analysis. Proteins of interested were incubated for 16 hours at room temperature at a concentration of 1:10 w/w of thrombin. An aliquot of each digest was analysed by SDS-PAGE alongside non-digested protein to confirm complete digestion by a drop in molecular weight of approximately 2 kDa.

2.6. Western blot

2.6.1. Protein extraction from plant or animal tissue

Tissue samples were flash frozen in liquid nitrogen upon collection, prior to protein extraction, samples were stored at -80 °C. Samples were thoroughly ground with a cooled micro-pestle and the resultant powder resuspended in an appropriate volume of phosphate buffered saline (PBS: 137 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, 1.5 mM KH₂PO₄, pH 7.4) with the addition of either PMSF to a final

concentration of 0.1 mM or proteinase inhibitor cocktail tablets (Roche). The sample was centrifuged at 12,000g for 3 minutes, the supernatant containing the soluble protein was removed to a new chilled microfuge tube, and the centrifugation process repeated until the supernatant was clear. The protein extract was quantified on a Nanodrop spectrophotometer ND-1000 (Thermo) at A_{280} .

2.6.2. Transfer of protein to PVDF membrane

Following SDS-PAGE, protein was transferred from the acrylamide gel to an Immobilon-P PVDF membrane (Millipore, Billerica, USA) prior to either colorimetric or chemo-luminescent detection. 2 pieces of 3MM filter paper (Whatman) were cut to 12 cm x 8 cm, and soaked, along with nylon pads of the same size and the polyacrylamide gel, in transfer buffer (25 mM Tris, 190 mM glycine and 20 % v/v methanol) for 5 minutes. One piece of PVDF membrane, cut to 9 cm x 6 cm, was hydrated in methanol for 10 seconds before being soaked in transfer buffer. The transfer cassette was assembled as follows; nylon pad, 3MM paper, gel, membrane, 3MM paper, nylon pad. The cassette was closed, and protein was transferred to the membrane at 100 v for 1 hour. After transfer the membrane was used directly in either colorimetric or chemo-luminescent detection protocols.

2.6.3. Ponceau S red staining

The efficiency of the transfer process can be tested by directly staining the membrane after transfer with the reversible protein stain Ponceau Red. A solution of 0.5 % w/v Ponceau red, 1 % v/v acetic acid was prepared. The membrane was incubated in the stain for 1 - 2 minutes until bands were visible, the red stain can then be washed off using distilled water prior to chemo-luminescent or colorimetric detection.

2.6.4. Colorimetric detection system

Following transfer, the membrane was rinsed for 5 minutes in 1 x PBS. The blot was then placed into blocking solution (PBS/0.05 % Tween-20, 5 % w/v non-fat milk powder) on a shaking table for 1 hour at room temperature. The blocking solution was then replaced with the primary antibody solution at the relevant concentration diluted in PBS/0.5 % Tween-20 with 0.5 % non-fat milk powder for a minimum of 1 hour at room temperature or overnight at 4°C. The membrane was then washed 6 x 10 minutes in PBS/0.5 % Tween-20 at room temperature, prior to incubation for 1 hour at room temperature in the relevant alkaline phosphatase-conjugated secondary antibody diluted 1:2000 in PBS/0.5 % Tween-20 with 0.5 % milk powder. The membrane was washed a further 6 x 10 minutes in PBS/0.5 % Tween-20 at

room temperature. The substrate tablet (Sigma Fast Tablets BCIP/NBT) was dissolved in 10 ml of water, applied to the protein side of the membrane and incubated in the dark at room temperature until bands were sufficiently developed. The membranes were then washed in H₂O to stop the reaction and allowed to dry on 3MM paper.

2.6.5. Chemi-luminescent detection

Following transfer, the membrane was placed in blocking solution (5 % non-fat dry milk powder w/v in PBST (PBS with the addition of Tween 20 to 0.1 % v/v) for a minimum of 2 hours at room temperature or overnight at 4 °C. The blocking solution was replaced by primary (1^o) antibody, diluted to the optimised concentration in PBST (typically between 1:10,000 and 1:20,000), for 1 hour at room temperature. The membrane was rinsed in PBST followed by three 20 minute wash steps on a shaking table at room temperature. The membrane was incubated with the relevant secondary antibody (2^o) diluted 1:50,000 in PBST at room temperature for 1 hour. Following this the membrane was washed briefly in PBST followed by three 20 minute washes and a final 5 minute wash in PBS at room temperature.

Chemi-luminescent detection was carried out using the SuperSignal West Pico reagents (Thermo Scientific, Waltham USA). Reagents (Luminol/Enhancer and Stable Peroxide solutions) were removed from 4 °C, allowed to equilibrate to room temperature and mixed at a ratio of 1:1. The detection solution was continually washed over the membrane using a pipette for 5 minutes. The membrane was wrapped in Saran wrap, and placed protein side up in an X-ray film cassette. Detection of protein was achieved by exposing the membrane to autoradiography film for between 3 seconds to 30 minutes depending on the level of detection required. The film was developed using a Compact x4 developer according to the manufacturer's instructions (XoGraph, Gloucestershire, UK).

2.7. General primers

Table 2-1 **General primers common to multiple Chapters**

Primer name	Seq 5' - 3'	TM
M13_F	GTAAAACGACGGCCAGT	55
M13_R	GTTTCCAGTCACGAC	55
EF1 α _ST_F	GAT TGG AAA CGG ATA TGC TC	55
EF1 α _ST_R	CGA CAG CAA CAG TTT GCC TC	55

3. Feeding tube size exclusion

3.1. Introduction

3.1.1. Feeding tubes and size exclusion

Feeding tubes are produced by several different biotrophic nematodes including cyst nematodes, root knot nematodes and the reniform nematodes. It has been hypothesised that sedentary endoparasitism has evolved independently in the cyst and root knot nematodes (van Megen et al., 2009). The close phylogenetic proximity of the reniform nematodes to the cyst nematodes, and the similar biology, suggest they share a common sedentary parasitic ancestor.

Morphological studies show that feeding tubes differ between nematode groups. Under electron microscopy the feeding tubes of root-knot nematodes appear to have a regular crystalline structure, however the feeding tubes of cyst and reniform nematodes both have an electron-dense irregular structure. Taken together these data support the independent origins of sedentary parasitism, and the convergent evolution of feeding tubes implies an important role for successful biotrophic interactions (Sobczak et al., 1999, Rumpfenhorst, 1984, Hussey and Mims, 1991).

Everything the nematodes ingest must pass through the walls of the feeding tube. It has been hypothesised that the feeding tube may act as a molecular sieve that excludes organelles and very large macromolecules, which may otherwise cause partial or total blockage of the stylet (Razak and Evans, 1976). Blocking of the stylet would undoubtedly be fatal as it is essential to the uptake of nutrients throughout the feeding stages of the nematode. A new feeding tube is formed before each bout of ingestion, so can be replaced if blocked. It is also possible that the feeding tube prevents nematode feeding activity from causing fatal damage to the feeding structure. Nematodes can only induce one feeding site and therefore need to keep this structure alive for the duration of the life cycle.

Determining the size exclusion limit of the feeding tube has proved challenging. In previous studies, host plants expressing reporter proteins have been infected with parasitic nematodes (Urwin et al., 1997b, McCarter et al., 2010, Valentine et al., 2007, Goverse et al., 1998), or fluorescently labelled-dextran have been injected directly into feeding sites (Bockenhoff and Grundler, 1994). Reporter molecules that are detected in the nematode digestive system can clearly pass through the feeding

tube. For cyst nematode feeding tubes, there is some ambiguity in these size exclusion experiments, highlighted by a series of seemingly conflicting results. Dextrans of 20 kDa but not 40 kDa were detected within nematodes (Bockenhoff and Grundler, 1994). Similarly uptake of an 11 kDa single cystatin (Urwin et al., 1998), and a 24.5 kDa monomeric red fluorescent protein (mRFP) (Valentine et al., 2007) has been demonstrated. However, although there was no evidence for ingestion of a 22 kDa double proteinase inhibitor fusion (Dual PI) (Urwin et al., 1998), the heavier 26.9 kDa green fluorescent protein (GFP) was observed to be taken up in one study (Goverse et al., 1998), but not in another (Urwin et al., 1997b). In experiments to date molecular mass has been used as an indicator of protein size. These apparently contradictory results suggest that a more pragmatic measure of protein size needs to be applied to feeding tube exclusion.

3.1.2. Mass spectrometry as a way of measuring protein size

Here we define protein size as the rotationally averaged cross section of a protein. In recent years, an emphasis has been placed on elucidating protein structure and its link to function. Presently, the two main methods used for determining protein structure are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Both give information about protein shape, but do not intrinsically tell us anything about size. Protein size can, however, be measured using ion mobility spectrometry (IMS).

In conventional IMS the size of a molecule is determined by the accurate measurement of its drift time through an IMS drift tube of known length (Mesleh et al., 1996). The drift tube is filled with a neutral buffer gas of known pressure. The movement of an ion through the tube, under the influence of a low electric field, is inversely proportional to its cross-sectional area and proportional to the number of charges it carries. IMS coupled with soft ionisation, such as electrospray ionisation (ESI), and mass spectrometry (MS) allows rapid determination of both a macromolecule's size and its mass in its native state. This can be achieved in a single experiment using a much lower amount of material than required for X-ray crystallography or NMR. One type of IMS in common usage coupled with MS is travelling wave IMS (TWIMS) (Giles et al., 2004, Pringle et al., 2007). In these devices the ion mobility separation occurs in a stacked-ring ion guide that contains the neutral buffer gas. A direct current is applied to the rings to radially confine the ions and a series of transient voltage pulses that create the travelling wave are superimposed on this. As molecules traverse the drift cell under the influence of the

travelling wave, they interact with the neutral buffer gas. The frequency of these interactions, due to their size, will determine if the molecule travels along with the wave or 'falls back' over the wave leading to a longer drift time. The protein size for a given mass to charge ratio is then calculated. This measure of size is known as the temperature-dependent, rotationally-averaged, Collision Cross Sectional area (CCS). In conventional IMS this can be calculated directly from the ion's drift time. However, for TWIMS the relationship between CCS and drift time is not linear and a calibration of the device must be performed with standards of known CCS (Mack, 1925, Smith et al., 2009).

3.1.3. Predicting protein size computationally

It is widely accepted that there is a relationship between the CCS of a protein and its structure, as solved by NMR or X-ray crystallography. Various attempts have been made to computationally predict the CCS of proteins based on these structures (Shvartsburg et al., 2007, Shvartsburg and Jarrold, 1996, Clemmer and Jarrold, 1997, Mesleh et al., 1996). The main assumption used in these predictions is that protein conformations in the gas phase are comparable to those in crystals or solutions and, more broadly, that both of these are analogous to conformations *in vivo*. Although this assumption is widely accepted, there are reports of protein complexes collapsing when ionised into the gas phase (Hogan et al., 2011).

Of these prediction methods, the most simple is the Projection Approximation approach (PA). This calculates the CCS by averaging the area of a 2D projection of a protein over a range of viewing angles (Mack, 1925). The reliability of PA has been questioned (Shvartsburg et al., 2007), most notably because it fails to take into account the buffer gas in its predictions (Shelimov et al., 1997, Jurneczko and Barran, 2011). It has been suggested that the PA approach will underestimate CCS for large concave molecules (Shvartsburg et al., 2007). However, PA predictions have been shown to be the closest to experimental measures for predicting CCS (van Duijn et al., 2009), even for concave molecules such as ubiquitin (Smith et al., 2009). This chapter introduces a new program, "RotaMol", that can be used to predict protein size based on information from either X-ray crystallography or NMR, in a similar manner to the PA approach, using the solvent accessible surface of the protein. Having validated its agreement with experimentally determined values, RotaMol predictions are used in conjunction with experimental methods of determining protein heterogeneity and size, to resolve conflicting data from past nematode feeding tube size exclusion experiments.

3.2. Aims:

Characterise the size exclusion limit of the cyst nematode feeding tube.

- Design a method to calculate protein size *in silico*.
- Validate the predictions experimentally.
- Provide better resolution of feeding tube exclusion, using a combination of protein size predictions and measurements, to design a range of different sized fluorescent proteins to test for uptake *in vivo*.

3.3. Materials and methods

3.3.1. Computational prediction of protein size – RotaMol

3.3.1.1. Area Measurement

The Protein Data Base (PDB) file of the protein of interest was loaded into the protein modelling program PyMol. This generated a graphical representation of the protein. The 1.4 Å solvent accessible surface was then loaded onto the protein using PyMol's built in surface function. The area of the two dimensional face, or 'viewing angle' that is presented to the user was then recorded by counting the number of pixels which make up the protein within the PyMol viewer. The protein was then rotated by a user defined angle of rotation (θ). After each rotation a new 'viewing angle' was measured. The average of all the viewing angles makes up the Collision Cross Section (CCS) or size of the protein. θ is limited by the program in that $180/\theta$ must result in an integer. Where this is not the case, the closest angle for which this is true is automatically used.

3.3.1.2. Measurement in pixels and conversion to Angstroms

Measuring every pixel in the protein viewer would be computationally expensive and unnecessary, as such, a new term 'Pixelskip' was defined. A Pixelskip of 5 will measure 1 in every 5 pixels, the pixels in between will be ignored. This builds up a coarse grained image of the protein. For the same viewing angle, a Pixelskip of 5 will measure 2307 pixels, where a Pixelskip of 10 will measure 578 pixels. The resultant area in pixels is multiplied by the pixel skip for X and Y.

For example: $2307 * 5 * 5 = 57675$

Similarly: $578 * 10 * 10 = 57800$

To convert this measure of area from pixels to Angstroms, the size of the protein viewer window in PyMol can be defined at discrete Angstrom values using the built in zoom function. Measuring the size of the window in pixels and dividing it by the size of the window in Angstroms, gives the number of pixels for one Angstrom. Full documentation, .exe and source code are available for download (<http://code.google.com/p/rotamol/>).

3.3.1.3. **Constructs for protein expression in *E. coli***

Primers were designed to amplify the desired coding sequences of mRFP, GFP and Dual PI with the addition of relevant restriction enzyme sites for cloning. Each PCR product was amplified from existing plasmid templates using Phusion polymerase (New England Biolabs) according to section 2.4.1 using the relevant primers detailed in Table 3-1. Following addition of 3' A overhangs the PCR product was purified immediately with a Qiaquick PCR purification kit (Qiagen, Manchester, UK) following the manufacturer's instructions. Purified PCR product was cloned into the pGEM-T Easy vector (Promega) following the manufacturer's recommendations and clones confirmed by sequencing. Two micrograms of pGEM-T Easy plasmid containing the gene of interest were digested with NdeI and BamHI, the released gene fragments were gel purified (QIAquick Gel Extraction Kit; Qiagen), and ligated into digested pPET28b vector. Positive constructs, identified by restriction enzyme digestion, were transformed into the expression strain of *E. coli*, BL21 DE3-RIL. Protein, extracted as described in section 2.5 with the His purification tag removed as described in section 2.5.5, was used in the following analyses.

Table 3-1 Summary of primers to cloning Dual PI, mRFP and GFP for expression in *E. coli* and subsequent size measurement.

Primer name	Seq 5' - 3'	TM	Additional restriction enzyme site
Dual_to_PET_F	ACA <u>CATATG</u> ATGTCATCAGACGGAGGACC	58	Nde I
Dual_to_PET_R	TGT <u>GGATCC</u> TTA ² CTCATCATCTTCATCC	46	Bam HI
mRFP_to_PET_F	ACA <u>CATATG</u> ATGGCCTCCTCCGAGGAC	61	Nde I
mRFP_to_PET_R	TGT <u>GGATCC</u> CTAGGCGCCGGTGGAGTG	63	Bam HI
GFP_to_PET_F	ACA <u>CATATG</u> ATG AGT AAAGGAGAAGA ² ACTTTTC	53	Nde I
GFP_to_PET_R	TGT <u>GGATCC</u> CTATTTGTATAGTTCATCCATGC	52	Bam HI

3.3.1.4. **Ion Mobility Spectrometry – Mass Spectrometry (IMS-MS)**

Protein samples for mass spectrometry were concentrated and the buffer switched to 50 mM ammonium acetate, using Amicon Ultra 3K spin columns (Milipore, Billerica, MA, USA). Ion mobility spectrometry-Mass spectrometry samples were analysed by Z-spray nanoelectrospray ionisation (nanoESI) MS using a quadrupole-IMS-orthogonal time-of-flight mass spectrometer (Synapt HDMS, Waters UK Ltd., Manchester, U.K.) with gold/palladium coated nanoESI tips prepared in-house. The instrument was operated in positive nanoESI-ion mobility spectrometry-TOF mode using a capillary voltage of 1.5 kV and cone voltage of 8 V. The source and desolvation temperatures were set at 80 °C and 150 °C, respectively. The nanoESI gas pressure was 0.145 psi, the source backing pressure was 0.034 psi, the trap and transfer argon gas pressures were 1.8×10^{-5} psi and the IMS cell nitrogen gas pressure was 7.1×10^{-3} psi. The trap collision energy was 21.8 V, the transfer collision energy was 4.0 V and a trap bias of 22.8 V was used. The IMS travelling wave speed was 225 m/s and the wave height was 4.9 V. Mass calibration was performed by a separate injection of aqueous sodium iodide at a concentration of 2 µg/µl. The IMS drift cell calibration was performed by separate injection of the denatured protein standards myoglobin, cytochrome c and ubiquitin at a concentration of 10 µM in acetonitrile/water/formic acid (50/49/1; v/v/v). Reduced CCSs (Ω') were calculated from published cross-sections determined using conventional ion mobility measurements (www.indiana.edu/~clemmer/Research/Cross%20Section%20Database/Proteins/protein_cs.htm) and were plotted against measured drift times (tD). An allometric $y = Ax^B$ fit was applied to the data. Experimental cross-sections were determined after separate infusion of the analytes and measurement of the drift time centroid for the lowest charge state ions. Data processing was performed using the MassLynx v4.1 suite of software supplied with the mass spectrometer.

3.3.1.5. **Analytical Ultra Centrifugation (AUC)**

Prior to Analytical Ultra-centrifugation (AUC) samples of purified protein in PBS buffer were adjusted to an absorbance between 0.1 and 1 (280 nm) and subjected to centrifugation at 8,000g for 2 minutes to remove any insoluble material. Samples were centrifuged at 200,000 rpm using an Optima XL-I Analytical Ultra centrifuge rotor (Beckman) at 20 °C. Scans were taken at 280 nm every 5 minutes for 100 scans per sample and results were analysed using SEDFIT V 12.44.

3.3.1.6. **Native - Polyacrylamide Gel Electrophoresis (PAGE)**

A native (non-denaturing) polyacrylamide gel was prepared in two sections. A 10% resolving gel was prepared by combining 3.4 ml of Acrylamide/Bis-acrylamide (30%/0.8% w/v) (Severn Biotech, Kidderminster, UK), 6.49 ml 0.375 M Tris-HCL (pH 8.8), 100 µl 10% (w/v) ammonium persulfate, and 10 µl TEMED. Once set, the stacking gel was prepared by combining 0.67 ml Acrylamide/Bis-acrylamide (30%/0.8% w/v), 4.275 ml 0.375 M Tris-HCL (pH 8.8), 50 µl 10% (w/v) ammonium persulfate and 5 µl TEMED. Once both gels had set, the samples were mixed with 2x sample buffer (62.5 mM Tris-HCL (pH 6.8), 25% glycerol, 1% bromophenol blue) at room temperature and loaded into each well. The gel was electrophoresed at 100 V in running buffer (25 mM Tris, 192 mM glycine) for 2 hours, or until desired resolution had been reached. The protein was visualised using the Coomassie-blue staining protocol described in section 2.5.2.

3.3.2. **Expression of fluorescent proteins in *Arabidopsis thaliana***

3.3.2.1. **Fluorescent protein fusion construct cloning**

A variety of different sized and coloured fluorescent proteins were available in the laboratory prior to the project: Green Fluorescent Protein (GFP), monomeric Red Fluorescent Protein (mRFP), iLOV and DRONPA. Both DRONPA and GFP were already present in the binary vectors pBI121 and pGRAB respectively. Protein coding sequences not present in binary vectors were amplified from the relevant plasmids using the corresponding primers, described in Table 3-2, to add restriction enzyme recognition sites. All restriction enzyme digests and cloning procedures were carried out as described in detail in section 2.4. All amplicons were sub-cloned by TA cloning to pGEM-T Easy vector for sequencing. For single proteins (iLOV, mRFP, DRONPA and GFP) sequences without errors were subsequently cloned to the plant transformation binary vector pBI121 by standard restriction digest cloning. For fusion constructs, various combinations of iLOV and mRFP were cloned in frame as translational fusions in pBI121. Restriction enzyme sites in the binary vector pBI121 occur in the order BamHI, KpnI and SacI. By including combinations of these restriction sites on either the forward or reverse oligonucleotide primers (Table 3-2), with or without stop codons, iLOV-iLOV, iLOV-mRFP, mRFP-iLOV, and mRFP-mRFP could be generated. Having generated fusion constructs, and confirmed translational fusion by sequencing, binary vectors were transformed into *Agrobacterium tumefaciens* strain GV3101 as described in section 2.4.12.

Table 3-2 Primers to clone fluorescent protein fusion construct for plant transformation vectors

Primer name	Seq 5' - 3'	TM	Additional restriction enzyme site
DRONPA_F	ATGGTGAGTGTGATTAACC	50	NA
DRONPA_R	TACTTGGCCTGCCTCGGCAG	67	NA
mRFP_F	ATGGCCTCCTCCGAGGAC	61	NA
mRFP_R_STOP	CTAGGCGCCGGTGGAGTG	63	NA
mRFP_BAM_F	ACA <u>GGATCC</u> ATGGCCTCCTCCGAGGAC	61	BamHI
mRFP_KPN_R_NO_STOP	ACA <u>GGTACC</u> GCGCCGGTGGAGTG	61	KpnI
mRFP_KPN_F	ACA <u>GGTACC</u> ATGGCCTCCTCCGAGGAC	61	KpnI
mRFP_SAC_R_STOP	ACA <u>GAGCTC</u> CTAGGCGCCGGTGGAGTG	63	SacI
iLOV_F	ATGGCTAGCATAGAGAAGAATTTTC	56	NA
iLOV_R_STOP	CTATACATGATCACTTCCATCGAGCTG	63	NA
iLOV_BAM_F	ACA <u>GGATCC</u> ATGGCTAGCATAGAGAAGAATTTTC	56	BamHI
iLOV_KPN_R_NO_STOP	ACA <u>GGTACC</u> TACATGATCACTTCCATCGAGCTG	62	KpnI
iLOV_KPN_F	ACA <u>GGTACC</u> ATGGCTAGCATAGAGAAGAATTTTC	56	KpnI
iLOV_SAC_R_STOP	ACA <u>GAGCTC</u> CTATACATGATCACTTCCATCGAGCTG	63	SacI
GFP_F	ATGAGTAAAGGAGAAGAACTTTTC	53	NA
GFP_R_STOP	CTATTTGTATAGTTCATCCATGC	52	NA

3.3.2.2. **Floral dip transformation of *Arabidopsis thaliana***

Eight *Arabidopsis thaliana* plants were used for each construct. Plants were grown in the glass house typically at 20 °C under long day conditions, until several unopened flower buds could be seen.

Twenty millilitres of LB medium containing the relevant antibiotics were inoculated with a single colony, or a scrape from a glycerol stock, of *A. tumefaciens* strain GV3101 containing the relevant binary vector. The culture was incubated with shaking (200 rpm) at 28 °C overnight. The entire 20 ml culture was used to inoculate 200 ml of LB medium containing the same concentration of the relevant antibiotics. The 200 ml culture was incubated with shaking (200 rpm) at 28 °C for approximately 4-5 hours until OD at A₆₀₀ was 0.5-0.8. The bacteria were pelleted by centrifugation at 4000g for 10 minutes and resuspended in 200 ml of 5% sucrose containing 100 µl Silwet (LEHLE seeds, USA). *A. thaliana* inflorescences were submerged in the solution for 2 minutes with occasional agitation, removed and the plants placed under a propagator lid for 24 hours. Plants were then allowed to continue growing as normal and the seed was collected and screened for successful transformation events.

3.3.2.3. **Surface sterilisation of *Arabidopsis thaliana* seeds**

Prior to use in sterile tissue culture, *Arabidopsis thaliana* seeds were surface sterilised by soaking in 20% v/v household bleach for 20 minutes at room temperature on a rotational mixer. All work following bleach treatment was carried out in a laminar flow hood to prevent re-contamination. The seeds were then washed 5 times in sterile H₂O by centrifugation at 3000 g for 30 seconds to remove all traces of bleach.

3.3.2.4. **Selection of transformed seeds**

Successfully transformed seeds were identified in tissue culture using kanamycin selection. T1 seeds were allowed to germinate on ½ MS 10 plates containing 50 µg/ml kanamycin. Successfully transformed seeds were able to progress to the first 4 true leaf stage on this media, at which point the T1 plants were transferred to individual ½ MS 10 plates. A single leaf was taken, flash frozen on removal, and used for DNA extraction (Section 2.4.6) and subsequent PCR using gene specific primers (Section 2.4.1) to confirm the presence of the transgene.

In some cases the binary vector used for transformation conferred DL-Phosphinothricin (BASTA – Duchefa Biochemie, Netherlands) resistance.

Transformants of these lines were identified following growth in soil. T1 Seeds were thinly spread on a 30 cm by 10 cm tray of compost. Once seeds had germinated, BASTA was sprayed directly onto the surface of the leaves, once per week for three weeks at a concentration of 0.75 mg/L in sterile distilled water containing 0.04% Silwet (LEHLE seeds). Successfully transformed plants, identified as those that had survived this treatment, were transferred to individual pots, and a single leaf taken, flash frozen on removal, and used for DNA extraction (Section 2.4.6) and subsequent PCR using gene specific primers (Section 2.4.1) to confirm the presence of the transgene.

3.3.2.5. **Selecting high expressing lines**

For all lines containing the desired transgene, the highest expressing lines were identified either by SQRT-PCR (Section 2.4.9) with the relevant gene specific primers, or by qualitatively assessing the brightness of relevant fluorescent protein (below).

Ten to twelve transgenic *A. thaliana* seeds of each line were grown in sterile tissue culture on ½ MS 10 media. After 2 weeks roots were sufficiently developed to assess the expression levels of fluorescent proteins on the basis of the strength of the fluorescent protein signal. Bright field and fluorescent images were acquired for each plant using a Leica Leitz DMRB binocular microscope (Leica microsystems, Milton Keynes, UK). Fluorescent images were converted to grey scale, inverted and the number of non-white pixels counted. As this count will depend on the total amount of root present to start with (regardless of brightness) the total root area was estimated in a similar manner, in this case counting every pixel that was not background colour. This gave a relative brightness per area of root tissue that was used to prioritise the highest expressing lines. Pixel counts were carried out using a custom script.

3.3.2.6. **Surface sterilisation of juvenile nematodes**

Prior to use in tissue culture, hatched second stage juvenile nematodes (J2s - section 2.3.2) were pelleted in 1.5 ml microfuge tubes (Maxymum Recovery™ – Axygen, Amsterdam, NL). J2s were resuspended in an appropriate volume of hexadecyltrimethylammonium bromide (CTAB - 0.5 mg/ml – Sigma) containing 0.1 % v/v chlorhexidine digluconate (Sigma) and 0.01% v/v Tween-20. J2s were sterilised in this solution for 25 minutes at room temperature on a rotator in a laminar flow hood and were subsequently suspended in sterile distilled water at a concentration of approximately 1 nematode per µl.

3.3.2.7. Infecting *Arabidopsis thaliana* with juveniles in tissue culture

Arabidopsis thaliana grown on vertical square ½ MS 10 medium plates (10 g/L agar) were infected 2 – 3 weeks post germination. Typically 5 plants were grown on each 12 cm square plate (Sterilin, Newport UK), with three infection points per plant. Thirty surface sterilised J2s were applied to each infection point, after which a single square (0.5 cm) of sterile GF/A paper (Whatman) was placed on top of the infection site. The plates were sealed with micropore tape and incubated at 20 °C under long day conditions. GF/A squares were removed in a laminar flow hood 2 days later and the plates were re-sealed and returned to the incubator.

3.3.3. Detecting uptake of fluorescent/fusion constructs

3.3.3.1. Fluorescence microscopy

Ten to twelve transgenic *A. thaliana* seeds of each high expressing line were grown in sterile tissue culture on ½ MS 10 media. Two weeks post sowing roots were sufficiently developed to infect with J2 nematodes (3.2.1.5). Successfully infecting nematodes were viewed *in situ* under a Leica Leitz DMRB binocular microscope to visualise presence of the fluorescent protein in the digestive system of the feeding female.

3.3.3.2. Confocal Microscopy

Transgenic plants expressing fluorescent protein constructs were grown as described above (3.2.1.3) and infected as described in section 3.2.1.5. Sections of root containing syncytia and feeding nematodes were removed and placed on a microscope slide under a cover slip. In some cases cover slips were placed over the agar plate and nematodes were visualised *in situ*. Confocal microscopy was carried out using an axioCam HPC camera (Carl Zeiss Microimaging, New York, USA) mounted on a Zeiss LSM710 confocal microscope (Carl Zeiss Microimaging). Images were captured and processed using the software package Zen 2011 version 7.0 (Carl Zeiss Microimaging). Spectra for mRFP were obtained from either mRFP expressing plants without nematodes or purified mRFP protein.

3.4. Results

3.4.1. Computational protein size prediction

RotaMol calculates the area of an image over a range of different viewing angles. In addition to analyte size, there are two variables that will affect the result, the resolution of the measurement (Pixelskip – Figure 3-1) and the number of viewing angles taken (inversely proportional to the angle of rotation, θ). For the Pixelskip analyses, a θ of 30° was used on the PDB model of mRFP. Pixelskip was altered in a stepwise manner for every integer between 3 and 100. Figure 3-2 shows the area of the protein in Angstroms² (\AA^2) plotted against Pixelskip. Increasing Pixelskip from 3 to 40 resulted in no appreciable decrease in accuracy. The pattern of regular peaks and troughs above a Pixelskip of 40 is characteristic of all proteins measured ($n=20$), highlighting the inaccuracy associated with use of high Pixelskip values. The optimum value for Pixelskip was defined as 20, as it provided the highest accuracy for the shortest computational time.

For the θ analyses, a Pixelskip of 20 was used on luteinising hormone-releasing hormone (LHRH, PDB code; 1YY2). Decreasing θ results in more viewing angles being analysed. The angle of rotation θ was increased from 1° to 180° , corresponding to a decrease in the number of viewing angles from 32,400 to 1 respectively. Figure 3-2ii shows the average area of LHRH plotted against the number of viewing angles used to make the average. There was no appreciable difference between the average area predicted using 32 viewing angles, and that of 32,400. The optimum value for θ was therefore defined as 30° . Using these newly defined default parameters, predictions made by RotaMol were compared to the existing prediction methods, MOBCAL Exact Hard Sphere Scattering (EHSS) and projection approximation (PA) (Figure 3-3). Trajectory method was not included as it was not published for all proteins in the comparison. In the majority of cases (70%) the predictions made by RotaMol lay between PA and EHSS.

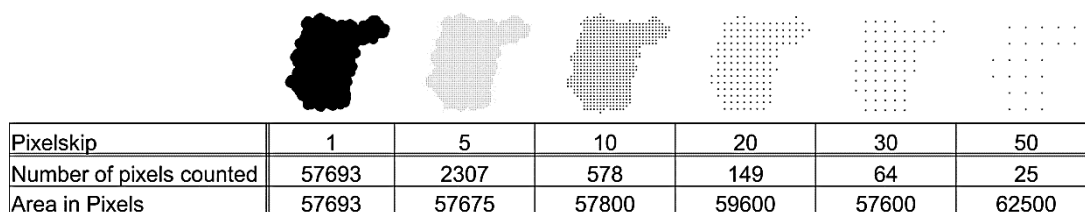


Figure 3-1 **The number of pixels measured by RotaMol with varying degrees of Pixelskip for a single viewing angle of GFP.** The images represent every pixel analysed at varying Pixelskips, at a Pixelskip of 1 every pixel is measured, at a Pixelskip of 50, 1 in every 50 pixels are measured. The number of pixels is then multiplied by the Pixelskip for X and Y.

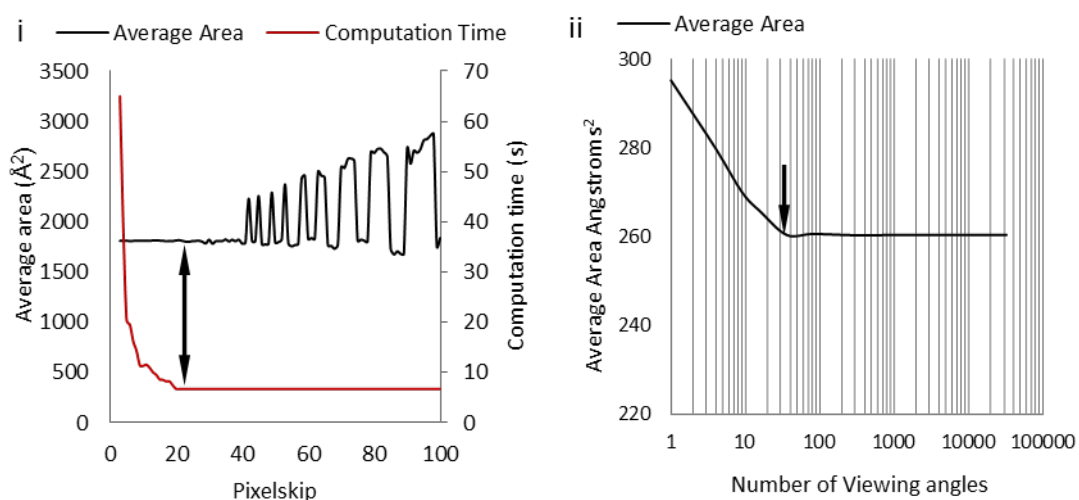


Figure 3-2 **Empirically derived optimum values for RotaMol parameters.** i)

RotaMol analysis of mRFP at $\theta = 30^\circ$ varying Pixelskip in a stepwise manner from 3 to 100. Average area (black) in Angstroms² plotted against Pixelskip. Computation time (red) plotted on left axis in seconds, shows exponential increase with decreasing Pixelskip. The optimum value for Pixelskip was defined as 20, as it has the highest accuracy with the shortest computational time (arrow). ii) RotaMol analysis of Luteinising hormone-releasing hormone (LHRH) using a Pixelskip of 20 and varying the number of viewing angles from 1 to 32,400 ($\theta = 180^\circ$ to 1°). The optimum value for the number of viewing angles is defined as 32 ($\theta = 30^\circ$) as it has no appreciable difference when compared to 32,400 viewing angles ($\theta=1^\circ$).

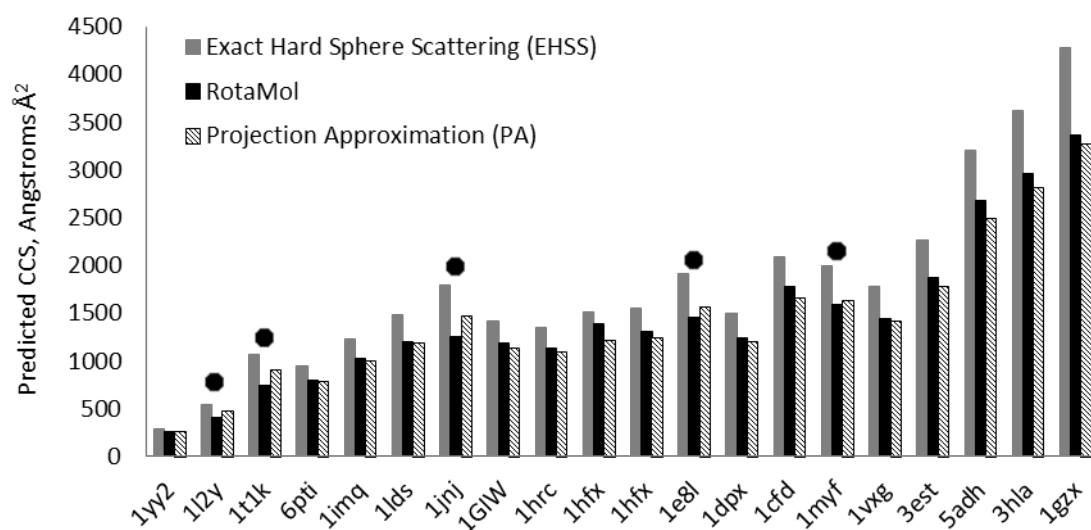


Figure 3-3 **Comparison between RotaMol predictions and existing prediction methods.** Predictions shown for MobCal Exact Hard Sphere Scattering (EHSS) in grey bars, and Projection Approximation (PA) in striped bars, compared to RotaMol, in black bars, for a range of proteins (PDB code given) from published sources. For the majority of predictions (70 %) RotaMol lies between that of PA and EHSS (● indicates the cases where RotaMol predictions are not between PA and EHSS) [16, 21, 22].

3.4.2. Protein analyses

3.4.2.1. Protein expression and removal of HIS tag

Having determined the optimum parameters for RotaMol, and compared its predictions to existing methods, the accuracy of its predictions was tested using three proteins of particular interest to cyst nematode feeding tube size exclusion, mRFP, GFP and Dual PI. Each protein was expressed in *E. coli*, and purified using nickel affinity chromatography. Figures 3-4 and 3-5 show the elution chromatograms for the three proteins. For GFP (Figure 3-4i), A single peak of protein was eluted from the column as confirmed by absorbance at A_{280} and A_{230} , reflecting aromatic rings and peptide bonds respectively. For mRFP (Figure 3-4ii) a peak of protein was eluted from the column as confirmed by absorbance at both A_{280} and A_{605} , in the case of the latter, detecting the colour of the expressed protein. For mRFP and GFP, eluted fractions corresponding to absorbance peaks were collected and confirmed visually to contain functioning fluorescent proteins. In the case of Dual PI, multiple elution peaks can be seen that, as confirmed by SDS-PAGE, all correspond to the protein of interest (Figure 3-5). As the extraction was carried out under native conditions the presence of multiple peaks may indicate the presence of oligomers.

Prior to further protein analysis, the HIS tag was removed from all three expressed proteins by thrombin digestion. Figure 3-6 shows a drop in molecular mass of approximately 2 kDa corresponding to the loss of the His tag. It is interesting to note that, despite the denaturing conditions of the SDS-PAGE, Dual PI does not migrate as a single band. Importantly, all bands identified correspond to Dual PI, rather than contaminating bacterial protein, because all experience the drop in molecular mass as a result of losing the His tag. There appear to be degradation products present for mRFP as the sum of the molecular mass of the two lower bands corresponds to that of the upper band. In addition, one of the bottom two bands experiences a 2 kDa shift in molecular mass corresponding to the HIS tag removal. All subsequent mass spectrometry experiments only analysed proteins of a mass corresponding to the non-degraded form of mRFP.

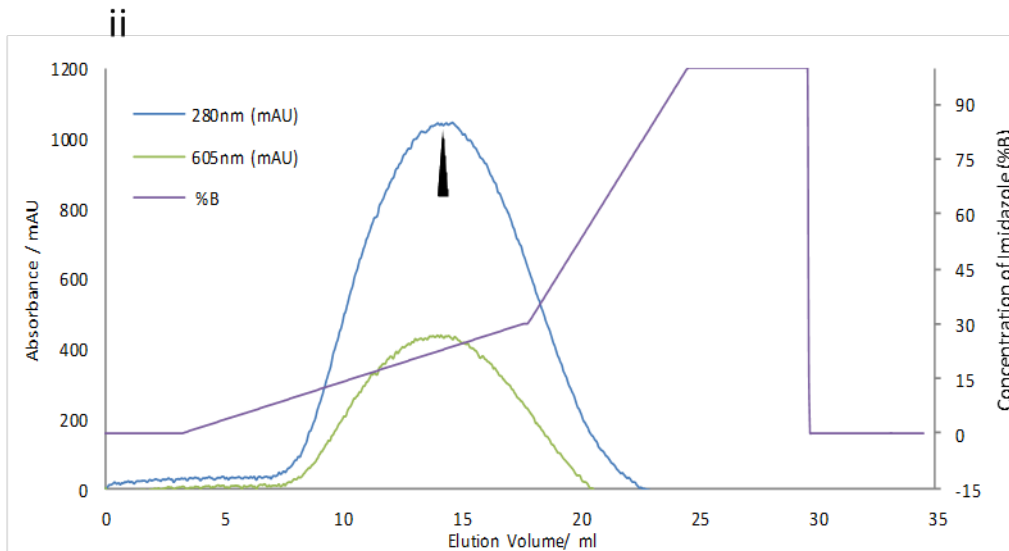
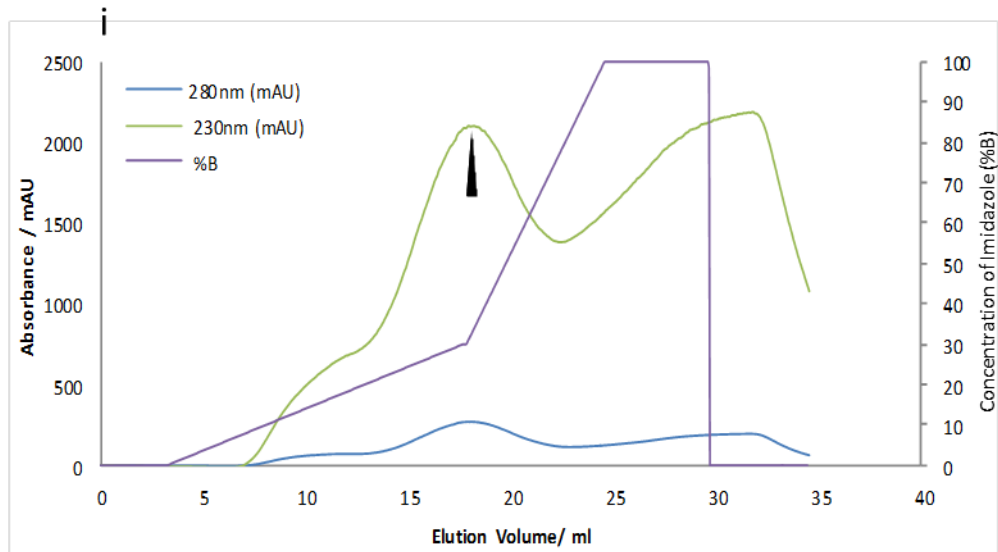


Figure 3-4 Elution Chromatograms for GFP and mRFP respectively. Protein eluted from the His-Trap column by increasing concentration of imidazole (purple). i) Absorbance measured at 280 nm (blue) and 230 nm (green), arrow indicates peak that corresponds to GFP. ii) Absorbance measured at 280 nm (blue) and 605 nm (green), arrow indicates peak corresponding to mRFP. In both cases the peaks predicted to be mRFP and GFP respectively were confirmed visually for GFP (left) and mRFP (right) (iii).

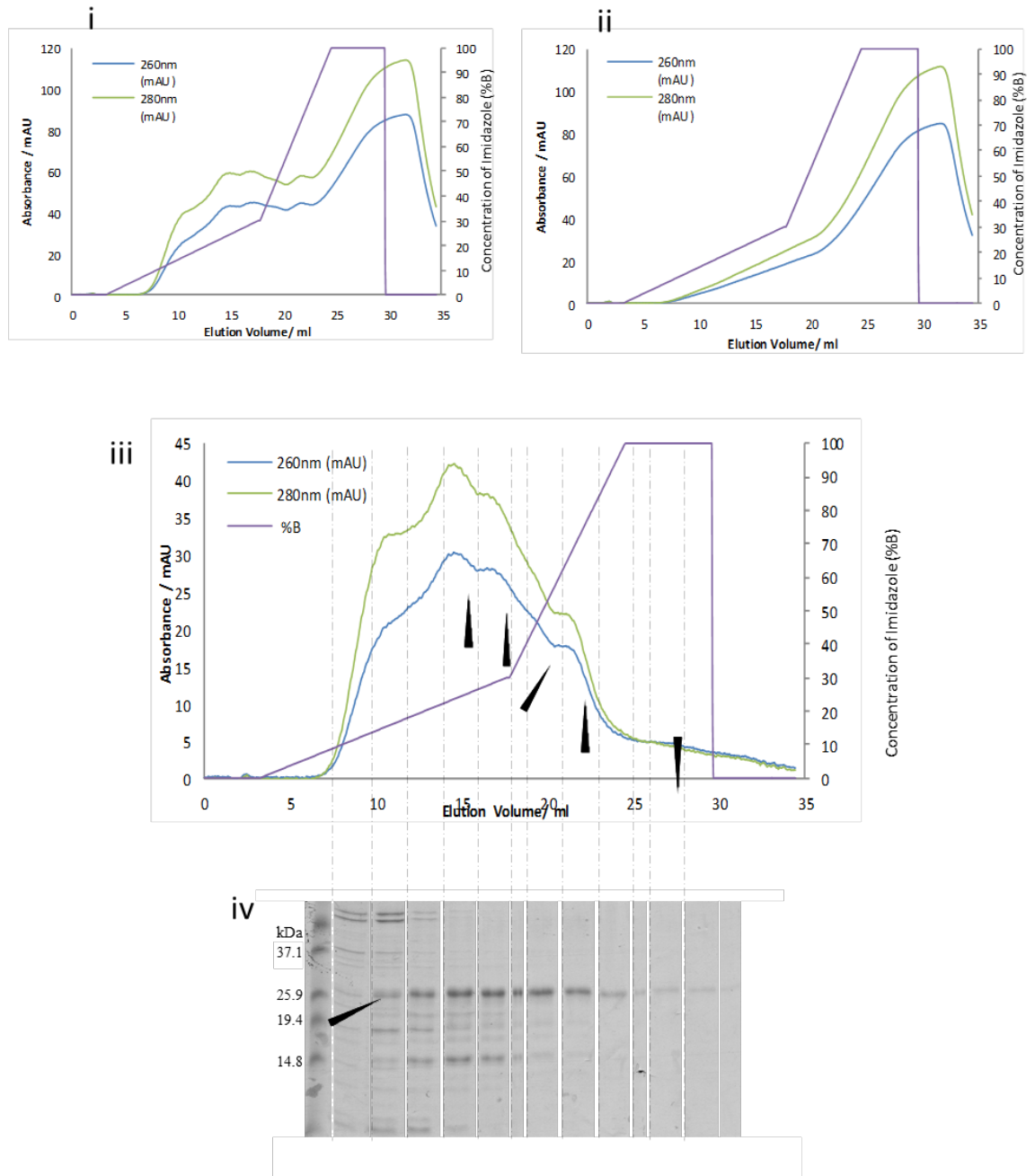


Figure 3-5 **Elution Chromatograms for Dual PI and a blank run.** Protein eluted from the His-Trap column by increasing concentration of imidazole (purple). **i)** Absorbance measured at 260 nm (blue) and 280 nm (green). **ii)** Absorbance measured at 260 nm (blue) and 280 nm (green). **iii)** The result of subtracting negative (**ii**) from Dual PI elution data (**i**), arrows indicate multiple peaks that correspond to Dual PI, as confirmed by analysing an aliquot of each fraction on SDS-PAGE (**iv**).

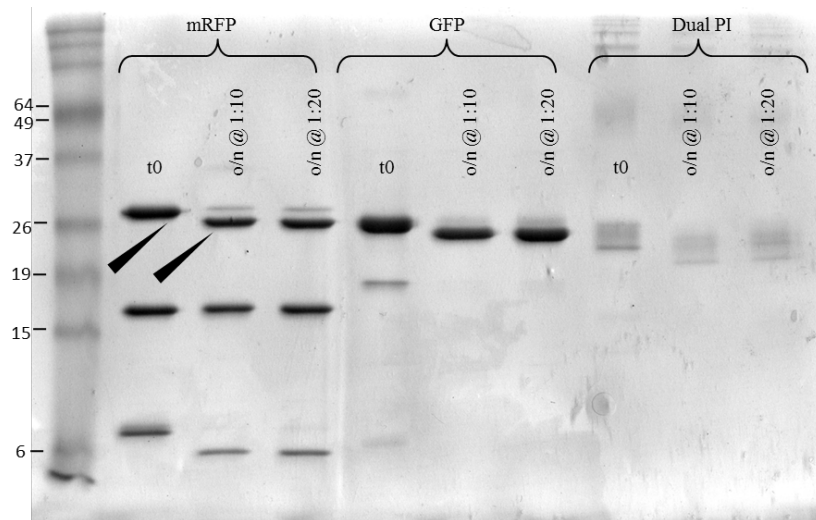


Figure 3-6 Removing His tag from expressed proteins by thrombin digestion.

Thrombin specifically cleaves the His tag from expressed proteins as seen by a shift in apparent molecular mass of approximately 2 kDa on SDS-PAGE (example indicated by arrows). The protein degradation products for mRFP are present before thrombin digestion, one of which experiences a drop of 2 kDa following digestion suggesting that the bands correspond to two halves of mRFP. Thrombin digestion appears to be sufficiently complete at a ratio of 1:10 when incubated with recombinant proteins overnight at room temperature. All bands in the Dual PI lane experience a shift in apparent molecular mass, suggesting they all correspond to Dual PI in various conformations.

3.4.2.2. Travelling wave ion mobility spectrometry – mass spectrometry

A comparison was made between predicted CCS derived using RotaMol, and those measured experimentally using ESI-Travelling Wave Ion Mobility Spectroscopy – MS (TWIMS-MS). Equal quantities of purified mRFP, GFP and Dual PI in 50 mM ammonium acetate were analysed by ESI-TWIMS-MS. Figures 3-7i, ii and iii show the individual spectra and corresponding drift plots for GFP, mRFP and Dual PI respectively. For GFP (Figure 3-7i) a strong signal was detected for the monomeric species at molecular mass 27,334 Da and a signal corresponding to a dimer at 54,627 Da. mRFP was present as a monomeric form of molecular mass 25,786 Da and a dimeric form in low abundance at 51,575 Da (Figure 3-7ii) For Dual PI a single dimeric species could be detected, but in such low abundance that its mass or size could not be accurately determined. Two monomeric forms were identified corresponding to a folded and an unfolded conformer both of 21,690 Da (Figure 3-7iii). This analysis was repeated with higher concentrations of Dual PI on multiple occasions with the same consistent pattern. Measured size of analytes was calculated by comparison to known standards, predictions were made with RotaMol (Pixelskip = 20, $\theta = 30^\circ$). Table 3-3 shows the predicted size and measured size for both mRFP and GFP and the measured size for Dual PI. No size prediction is available for Dual PI in either conformation as there is no solved structure. Predictions for mRFP and GFP were 94.4% and 96.5% accurate respectively.

3.4.2.3. Analytical Ultra Centrifugation

Analytical Ultra Centrifugation (AUC) characterises a protein's hydrodynamic properties in the liquid phase, determining the sedimentation coefficient of a protein/protein complex giving information about hydrodynamic shape and heterogeneity without the need for interaction with a matrix or ionisation into the gas phase (Lebowitz et al., 2002). Prior to AUC analysis the proteins were subjected to a short centrifugation at 14,000g. Their absorbance at 280 nm was measured before and after the centrifugation. A large reduction in absorbance would signify that the protein is largely insoluble. As no appreciable reduction was noted for any of the three scans it was concluded that all protein samples were in solution (Figure 3-8). The AUC analysis confirmed that mRFP was present as a single monomeric species whilst both monomeric and dimeric species of GFP were present. A monomeric species was detected for Dual PI, although in low abundance. The majority of the protein (>88%) was present in a large heterogeneous distribution of multimers (Figure 3-9), the molecular mass of which can be estimated to be a range of oligomers formed from between 4 and 12 monomers.

3.4.2.4. **Native PAGE**

The observed heterogeneity of Dual PI was further analysed, in the solid state, using Coomassie stained native-PAGE. Figure 3-10 shows Dual PI migrating as multiple bands under native conditions. For comparison it was electrophoresed alongside mRFP, which migrated as a single monomeric species.

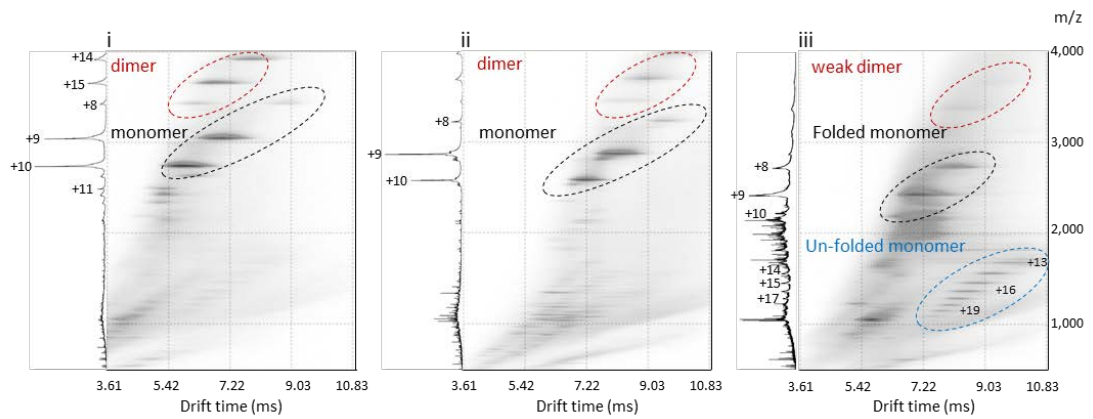


Figure 3-7 **ESI-TWIMS-MS drift plot of GFP, mRFP and Dual PI.** For each, mass spectra are represented on the left and on the right the corresponding drift plot. **i)** GFP drift plot shows both monomeric species of 27,334 Da and dimeric species at 54,627 Da (red). **ii)** mRFP drift plot shows a monomeric species at 25,786 Da and a less intense dimeric species at 51,575 Da (red). **iii)** Dual PI drift plot shows a single weak dimeric species below the measurement threshold, and two monomeric species, both at 21,690 Da, presumed to be folded and unfolded variants.

Table 3-3 **Summary of molecular mass, size, and uptake of mRFP, GFP and Dual PI**

Protein Name	Ingested by nematode	Molecular mass (Da)	Predicted Size (\AA^2)	Measured Size (\AA^2)	Predicted compared to measured (%)
Monomeric Red Fluorescent Protein (mRFP)	Yes	25,786	1,759	1,864	94.4
Green Fluorescent Protein (GFP)	Conflicting results	27,334	1,913	1,983	96.5
Dual Proteinase Inhibitor (Dual PI) - folded	No	21,690	N/A	1,686	N/A
Dual Proteinase Inhibitor (Dual PI) - un-folded	No	21,690	N/A	3,203	N/A

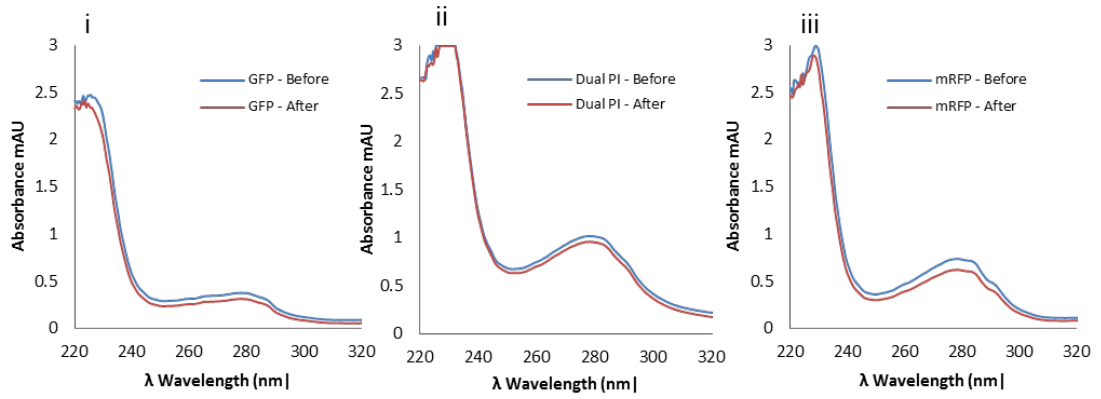


Figure 3-8 **Absorbance scans at a range of wavelengths before and after 5 minute 14,000g centrifugation.** All 3 scans show a small drop in absorbance following centrifugation. **i)** GFP shows a final absorbance at 280 nm of 0.315 mAU **ii)** DUAL PI shows a final absorbance at 280 nm of 0.955 mAU. **iii)** mRFP shows a final absorbance at 280 nm of 0.621 mAU. The absorbance of all samples was between the required 0.1 and 1 mAU

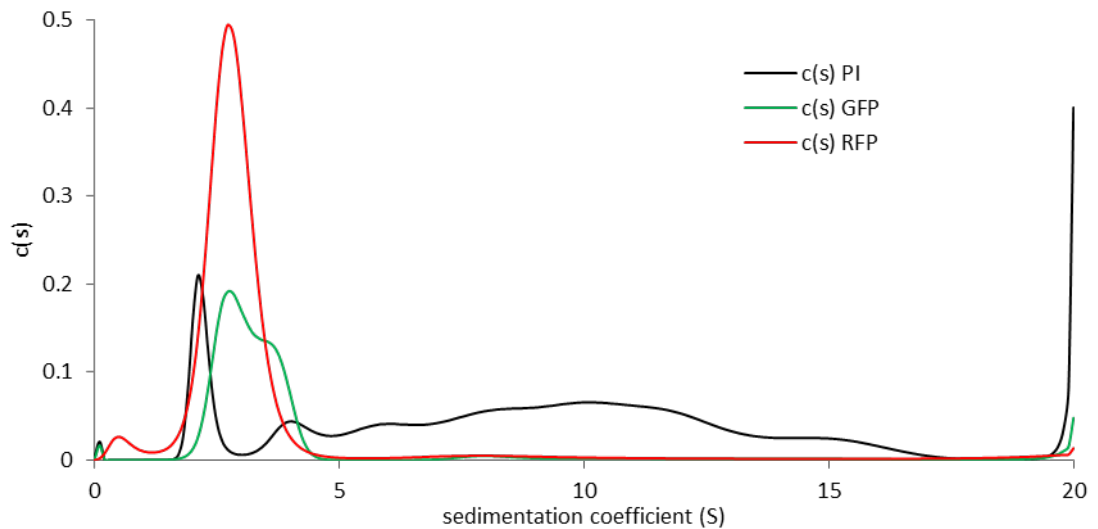


Figure 3-9 **Analytical Ultra-Centrifugation (AUC) absorbance plots of mRFP, GFP and DUAL PI.** For all samples, sedimentation coefficient is plotted against concentration distribution $c(s)$. A Single monomeric species was detected for mRFP (red), both monomeric and dimeric species detected for GFP (green). A monomeric form and large heterogeneous multimers estimated at >150 kDa were detected for Dual PI (black).

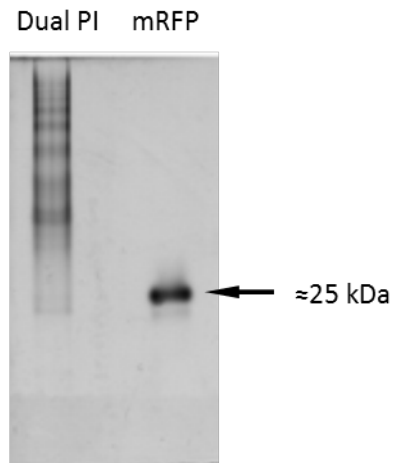


Figure 3-10 **Analysis of Dual PI multimers by native polyacrylamide gel electrophoresis.** Dual PI, when electrophoresed under native conditions, migrates as multiple discrete bands suggesting the presence of multimers. In comparison mRFP electrophoresed under the same conditions migrates as a single monomeric species.

3.4.3. *In planta* expression of fluorescent proteins

A range of different sized proteins were designed to span the expected size exclusion range of the cyst nematode feeding tube. All of these proteins, summarised in Table 3-4, were fluorescent proteins in order to enable detection by microscopy. The sizes of proteins that were monomeric, or that were similar to another protein for which the structure has been solved, were predicted by RotaMol. Following cloning of each coding region into a plant transformation vector, each construct was transformed into *Arabidopsis thaliana* for uptake experiments.

Table 3-4 **Summary of fluorescent protein fusion constructs and previously reported uptake by cyst nematodes.**

Construct name	Molecular mass (kDa)	Predicted size (\AA^2)	Binary vector - plant selection marker	Promoter <i>in planta</i>	Exclusion by feeding tube
iLOV	13	1214	pBI121 - Kanamycin	35S	Unknown
mRFP	25.4	1768	pBI121 - Kanamycin	35S	Not-excluded
mRFP++	27.042	N/A	pK7WGR - Kanamycin	35S	Unknown
GFP	26.89	1946	pBI121 - Kanamycin	35S	Conflicting results
iLOV-iLOV	26.05	N/A	pBI121 - Kanamycin	35S	Unknown
iLOV-mRFP	38.4	N/A	pBI121 - Kanamycin	35S	Unknown
mRFP-iLOV	38.4	N/A	pBI121 - Kanamycin	35S	Unknown
mRFP-mRFP	52.4	N/A	pBI121 - Kanamycin	35S	Excluded
DRONPA	Tetramer >100	>4660	pGRAB - BASTA	35S	Unknown

3.4.3.1. Cloning of expression constructs

A typical cloning procedure, using iLOV as an example, involved the following: The protein coding sequence of interest was amplified by PCR using gene specific primers with the addition of the relevant restriction enzymes for cloning (Figure 3-11i). Amplified products were sub cloned to pGEM T-Easy vector by TA cloning and sequenced to ensure that no PCR errors had been introduced (Figure 3-11ii). Correct sequences were transferred by standard restriction digest cloning to the relevant plant binary vector, and transformants were identified by colony PCR (Figure 3-11iii). Generation of the desired clones was confirmed initially by restriction digest (Figure 3-11iv) and subsequent sequencing and alignment to the original sequence (Figure 3-11v).

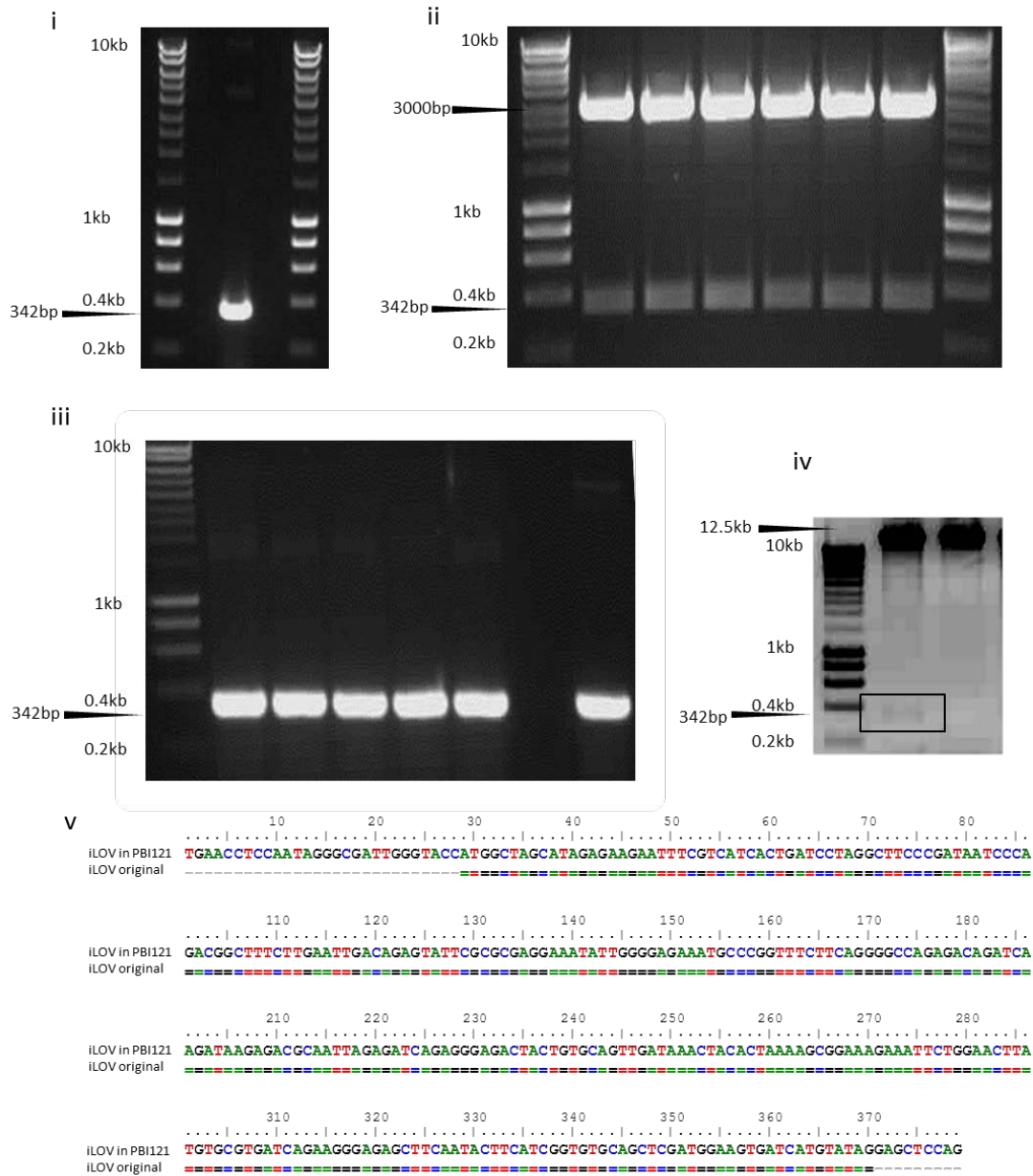


Figure 3-11 Typical cloning procedure for plant transformation vectors using iLOV as an example. i) Initial PCR with gene specific primers containing the addition of 5' NheI and 3' BamHI restriction enzyme sites, with an expected product of 342 bp (arrow). ii) Sub cloned to pGEM-T easy vector and digested with NheI and BamHI to confirm the presence of the appropriate insert and restriction sites. iii) Colony PCR in plant transformation vector pBI121, lanes from left to right show 5 individual colonies, no template control and positive control. iv) Digested out of pBI121 with NheI and BamHI to confirm insert. v) Plasmid used to generate the first lane in iv was sequenced to confirm the insert aligns to the expected sequence with zero miss-match (identity shown as an "=").

3.4.3.2. Selection of transgenic *Arabidopsis thaliana*

Successfully transformed *Arabidopsis thaliana* were identified either by treating T1 seedlings with BASTA or plating T1 seeds onto selection media where appropriate. Figure 3-12 shows a typical result from the selection process. Once these plants were established, DNA was extracted and PCR carried out with gene specific primers. The BASTA treatment had a false positive rate of 18% as confirmed by gene specific PCR. Kanamycin selection, confirmed in the same manner, had a false positive rate of < 1%.

3.4.3.3. Selection of high expressing lines

Having identified a number of lines for each construct containing the relevant transgene, the expression of the transgenes was analysed. In the case of fluorescent proteins that could be visualised an initial qualitative approach was taken using fluorescence microscopy. Figure 3-13 shows comparison between wild type (Figure 3-13A) and transgenic plants expressing mRFP (Figure 3-13B). This expression was quantified by transforming the image to grey scale (Figure 3-13iii), inverting (Figure 3-13iv) and counting the number of non-white pixels. All fusion constructs containing exclusively iLOV proteins could not be visualised. In these cases, semi-quantitative RT-PCR was used (Figure 3-13C). High expressing lines were identified by comparison of the intensity of a product obtained using gene specific primers with that from a control gene (elongation factor). For fusion constructs that contained one fluorescent protein fused to iLOV, the expression of the construct was quantified using the visible fusion protein only. In both cases the fusion proteins were confirmed to be intact *in planta* by western blot analysis (Figure 3-14).

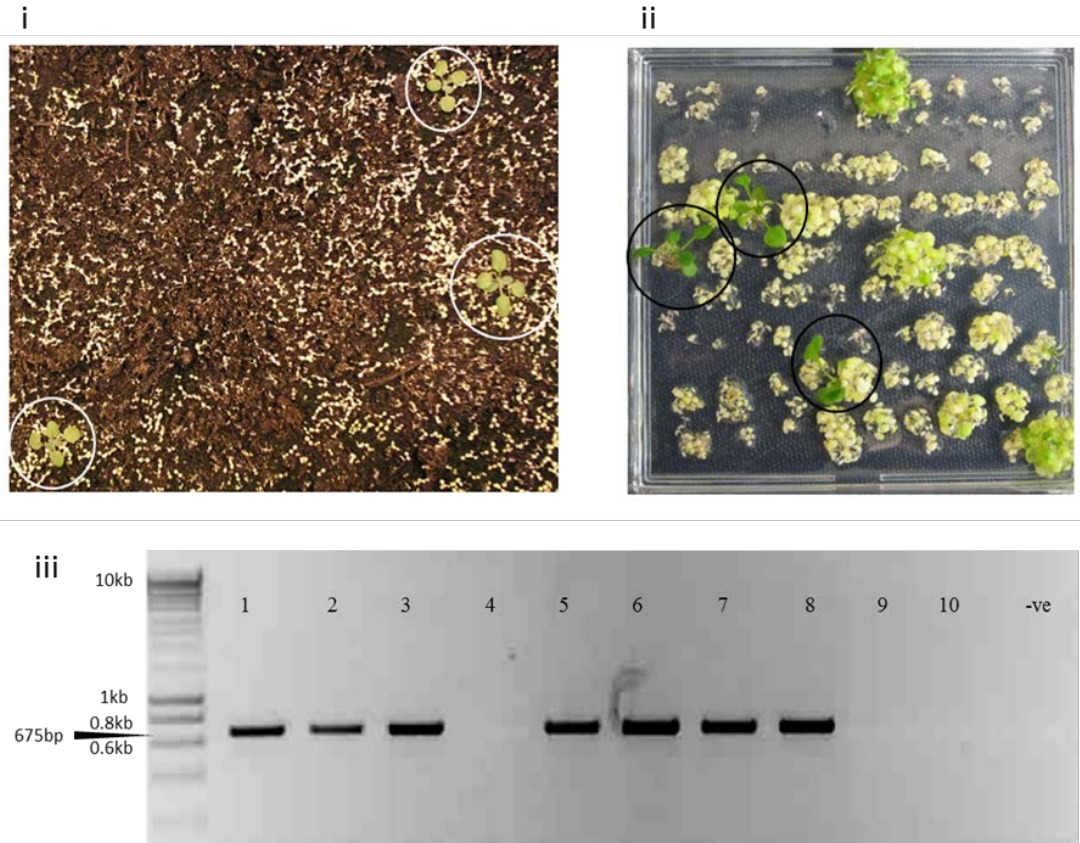


Figure 3-12 **Selection of successfully transformed *Arabidopsis thaliana*.** i and ii) Selection of positive transformants (circled) by BASTA selection and kanamycin selection respectively. iii) Subsequent PCR on DNA extracted from DRONPA transformed *A. thaliana* that passed BASTA selection. In this case lines 4, 9 and 10 were negative for the transgene.

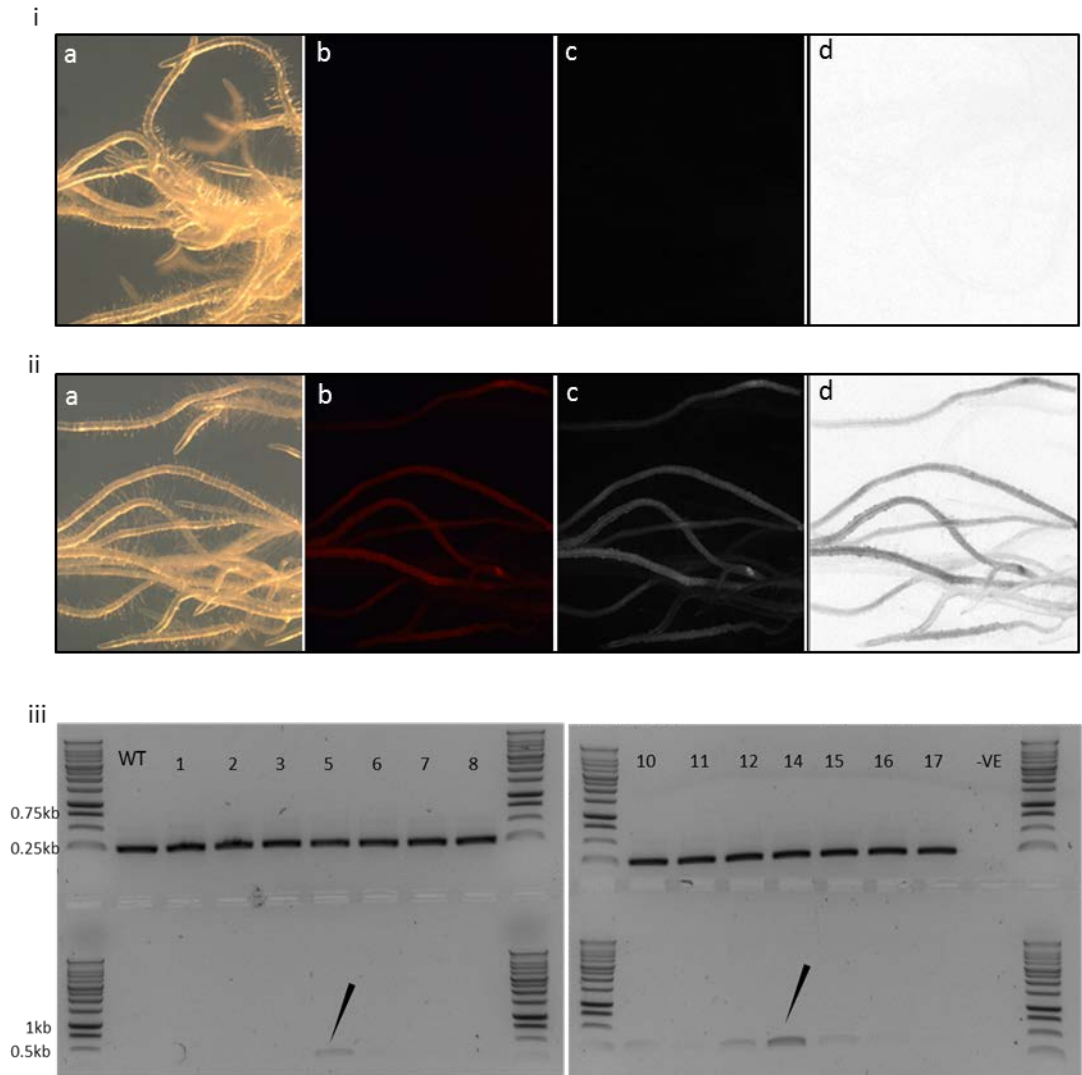


Figure 3-13 **Selection of high expressing transgenic *Arabidopsis thaliana*.** For fluorescent proteins that could be visualised, high expressing lines were selected qualitatively. **i** and **ii** show 4 images of WT and transgenic lines respectively. Comparing bright field (**a**), and fluorescence (**b**) reveals expression levels. Expression was quantified by transforming the image to grey scale (**c**), inverting (**d**) and counting the number of non-white pixels. A similar counting procedure was carried out for bright field images to normalised for initial root quantities. For constructs that could not be visualised, semi-quantitative RT-PCR was carried out (**iii**). High expressing lines (arrow) were identified by comparing quantity of gene specific amplicon (bottom) to elongation factor amplicon (top).

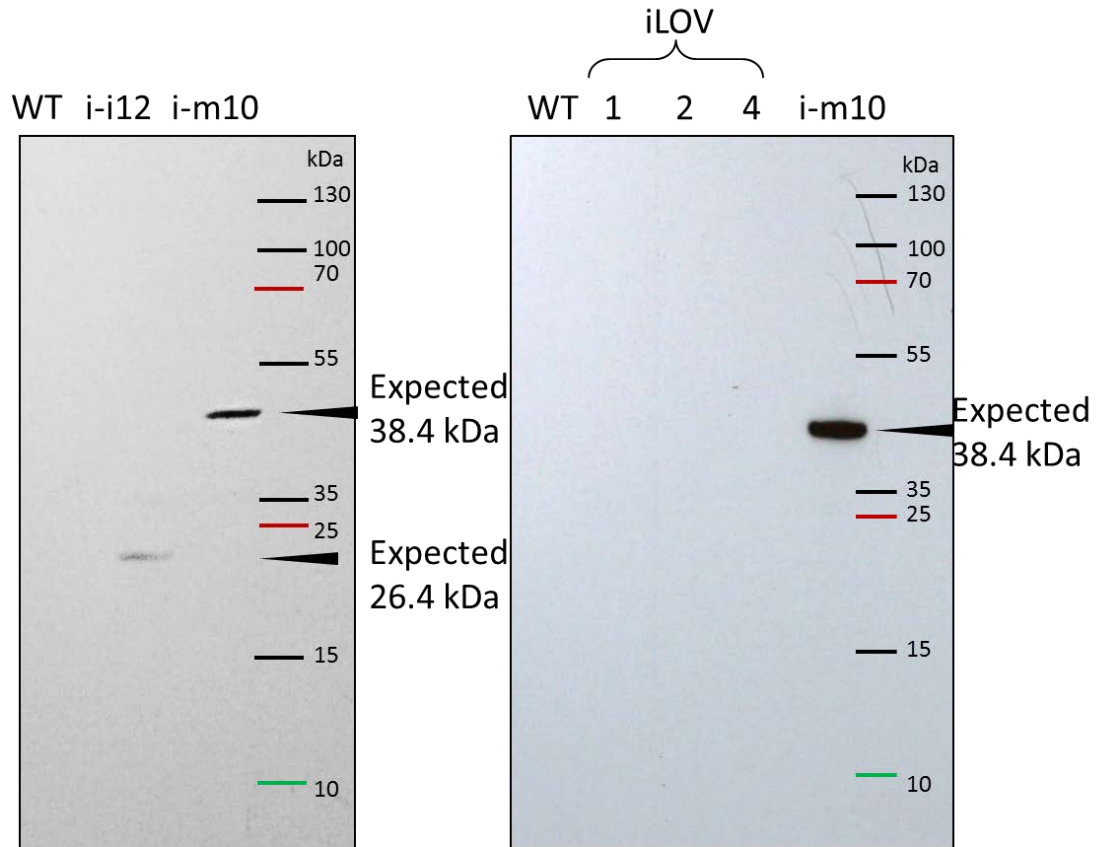


Figure 3-14 **Detection of intact fusion constructs, expressed *in planta*, by western blot.** Confirming detection of intact iLOV-mRFP fusion (i-m10) and iLOV-iLOV fusion (i-i12) in protein extracted from transgenic plants expressing the relevant fusion construct. No protein is detected in the wild type plants (WT) or the single iLOV transgenic plants (iLOV).

3.4.4. Uptake of fluorescent proteins

Although it was expected that all fluorescent proteins would be detectable, as described this was in fact not the case. For those that could be visualised, *H. schachtii* infecting transgenic *A. thaliana* lines expressing the fluorescent constructs were visualised initially using fluorescence microscopy. This however had insufficient resolution and brightness to definitively detect uptake, and so confocal microscopy was used for all uptake experiments (Figure 3-15). mRFP fluorescence can be seen in discrete bodies within the gut of the nematode (Figure 3-16). However similar patterns of fluorescence are also seen for male, and juvenile nematodes that do not feed. It is therefore assumed that this is, in fact, auto fluorescence from lipid droplets.

In no combination of fluorescent constructs tested could we replicate uptake results similar to those published. Spectral un-mixing was carried out and yet still even the exact spectrum of mRFP was visible in these lipid droplets of non-feeding nematodes. Juvenile nematodes prior to infection also had similar auto-fluorescent patterns. Taken together these results suggest that confocal microscopy of uptake is an inadequate tool for testing feeding tube size exclusion.

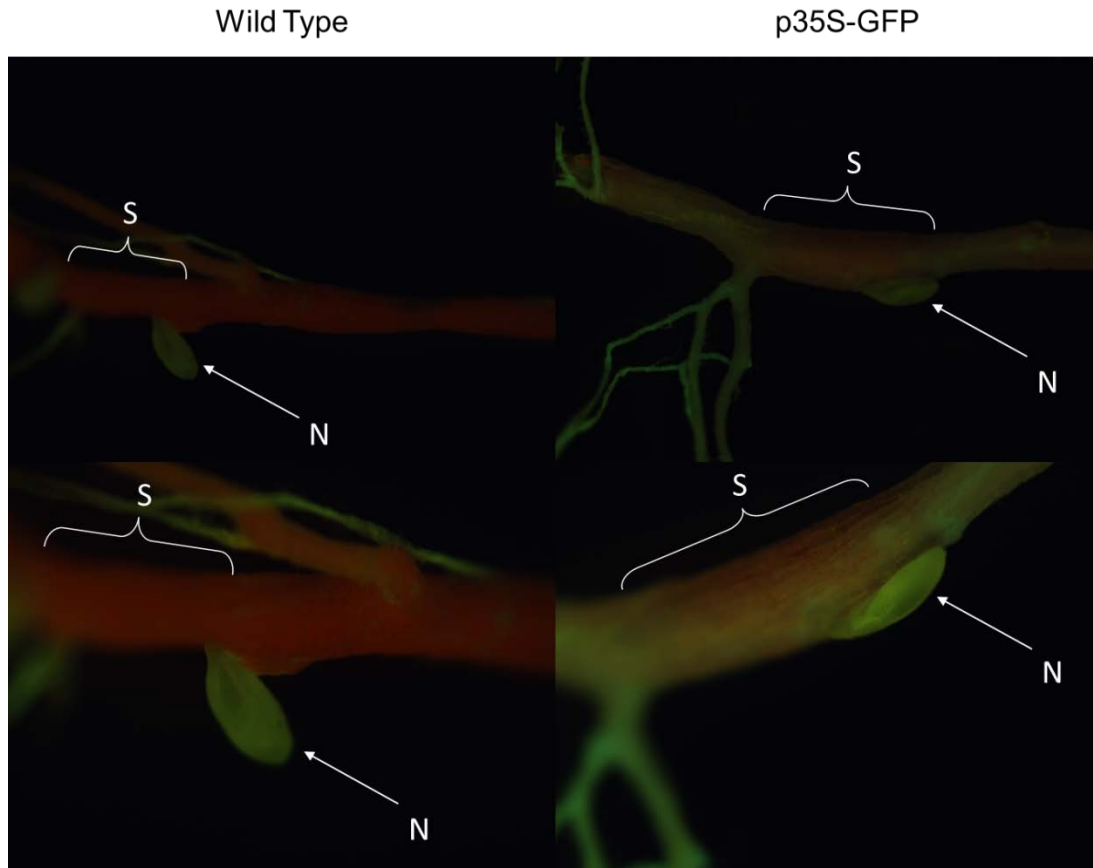


Figure 3-15 **Fluorescence microscopy to detect uptake of fluorescent proteins by cyst nematodes.** Wild type (left) and GFP (right) *A. thaliana* roots can be seen infected with the cyst nematode *H. schachtii* (N) with the syncytial region highlighted (S). Although the roots of the GFP expressing plant are “more green” than the wild type, there is insufficient brightness and resolution to determine uptake by the nematode conclusively.

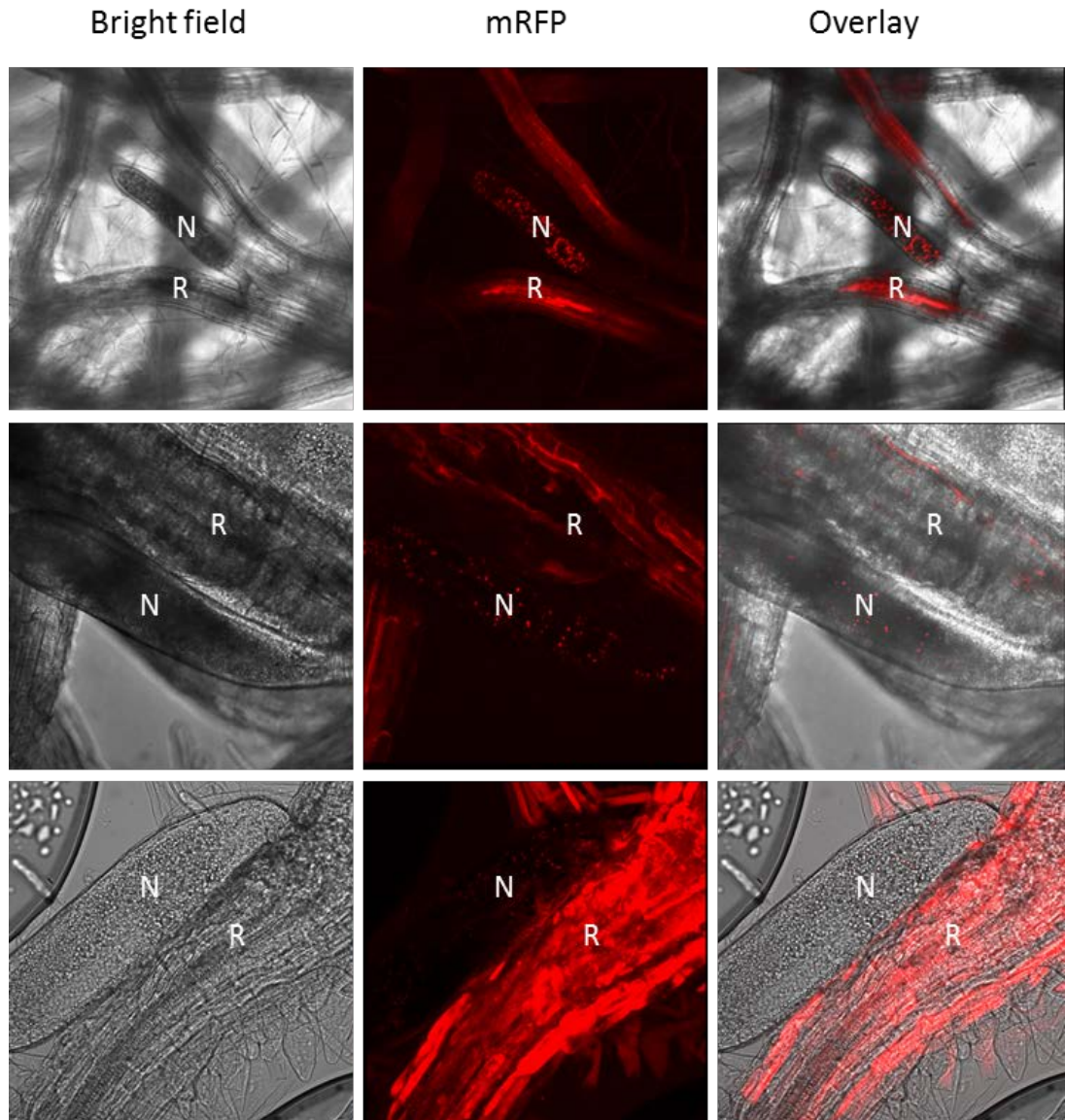


Figure 3-16 **Confocal microscopy to detect uptake of mRFP.** Confocal microscopy of transgenic *A. thaliana* expressing construct mRFP⁺⁺. Bright field, fluorescence, and overlay can be seen for three nematodes (N) feeding on roots (R). Red fluorescence can be seen in discrete bodies within the nematodes.

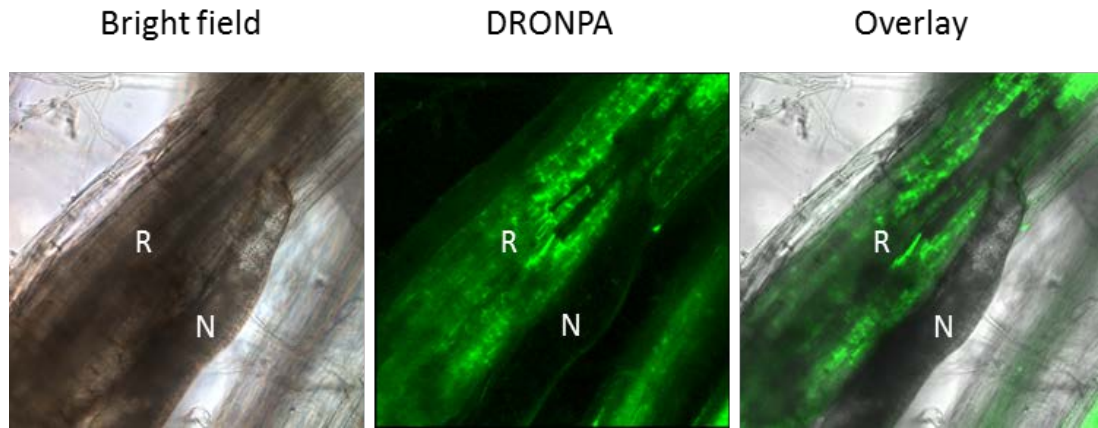


Figure 3-17 **Confocal microscopy to detect uptake of DRONPA.** Confocal microscopy of transgenic *A. thaliana* expressing construct DRONPA. Bright field, fluorescence, and overlay can be seen for a nematode (N) feeding on roots (R). No green fluorescence can be seen within the nematode.

3.5. Discussion

Nematode feeding tube size exclusion has remained unclear for a number of years. In previous studies molecular mass has been used as an indicator of protein size. In this study we aimed to clarify the conflicting results between previous size exclusion experiments by assessing the size and heterogeneity of the proteins tested. We then aimed to further resolve the size exclusion limit by designing a range of proteins to span the size range previously tested.

3.5.1. Quality of computational size predictions - RotaMol

For the two measured proteins, mRFP and GFP, the CCS predictions made by RotaMol were 94.4% and 96.5% of the measured values respectively. However, these need to be put in the context of other prediction methods. Of the available prediction methods, MOBCAL is the most widely used (Shvartsburg and Jarrold, 1996, Shvartsburg et al., 1998, Mesleh et al., 1996). It has been noted on a number of occasions that the MOBCAL Projection Approximation (PA) approach has a tendency to underestimate protein size measurement (Shvartsburg et al., 2007, Smith et al., 2009, Jurneczko and Barran, 2011), whereas the MOBCAL Exact Hard Sphere Scattering (EHSS) and MOBCAL Trajectory Method (TM) have a tendency to overestimate protein size measurement (Smith et al., 2009). For a series of published protein size predictions and measurements, ranging from 1,000 to 64,000 Da ($n=20$) (Shelimov et al., 1997, Valentine et al., 1997, Hopper and Oldham, 2009), 70% of predictions by RotaMol fell between the MOBCAL PA and EHSS values (TM was not published for all proteins and so was not included in the comparison). RotaMol differs from the traditional PA approach in that it will measure any surface loaded onto the protein in PyMol, the cases described herein used the 1.4 Å solvent accessible surface. A common criticism of the PA approach is its inability to take into account electrostatic forces (Jurneczko and Barran, 2011), however in place of the 1.4 Å solvent accessible surface, a projected image of external charge can be loaded onto a protein model in PyMol (Figure 3-18). This can then be used for downstream RotaMol analyses to give a combined measure of shape and charge. It is interesting to note that there was no difference between the average area made from 32 viewing angles and that of 32,400, despite the inherent asymmetry of the protein in question (LHRH - Figure 3-2).

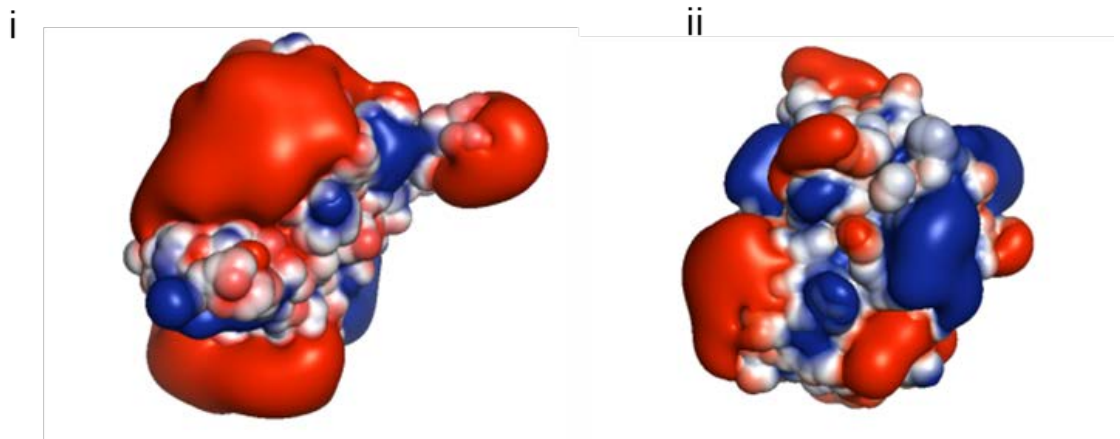


Figure 3-18 **GFP and mRFP external charge comparison.** Electrostatic potentials, negative (red) and positive (blue), are shown for GFP (i) and mRFP (ii) generated using PyMOL and the APBS tools plugin.

3.5.2. Heterogeneity, quaternary structure, aggregation and their effects on feeding tube size exclusion

Assessing the conformations and heterogeneity of proteins used in previous size exclusion experiments has addressed the lack of correlation between molecular mass and observed passage through the feeding tube. In the literature, all cyst nematodes tested have been able to uptake mRFP. Weak dimer formation was noted in the mass spectrum (Figure 3-7ii), however no such formation was observed in the AUC (Figure 3-9). ESI-MS is considerably more sensitive than AUC, and it has been known that oligomers can be detected in the former rather than the latter for the same sample (Woods et al., 2011). Moreover, as the AUC analysis is performed in the liquid phase, the conformation of mRFP here is expected to be more analogous to that *in vivo*. GFP was noted to form a dimer in the mass spectrum, whilst in AUC roughly half of the GFP was present as a dimer. The GFP dimer was measured at 160% the size of GFP monomer (Table 3-3), assuming a similar ratio *in vivo*, detecting the remaining GFP being taken up by the nematode would be challenging. In general, it is known that protein concentration is directly proportional to aggregation under quiescent conditions (Treuheit et al., 2002, Fields et al., 1992), and this is not the first report of GFP forming dimers (Zacharias et al., 2002), highlighting the care needed when selecting reporter proteins.

Dual PI was one of the least massive proteins tested for exclusion and yet has never been detected to pass through the cyst nematode feeding tube. The mass spectrum shows weak dimer formation, but more importantly two different monomeric states are present, folded and unfolded (Figure 3-7). Not only does the unfolded state have a considerably larger measured size (nearly twice that of the folded), it may also play a role in aggregation (Roberts, 2007). The AUC revealed multiple different oligomers of dual PI. These were in the mass range of 4-12 times that of the monomer. From this study it is unclear if the oligomers are formed from the folded, unfolded, a partially folded intermediate or a combination of all three forms (Roberts, 2007, Raso et al., 2005, Kiese et al., 2008). These aggregates were not present in the mass spectrum, suggesting an inability to ionise some of the aggregates into the gas phase, highlighting a potential limitation of ESI-MS in assessing protein interactions in this case. AUC and Native-PAGE both highlighted the aggregates, the former suggesting that the majority of the protein (>88%) was present in a large heterogeneous distribution. This estimate of 88% is, if anything, conservative as it does not take into account any of the aggregates above a sedimentation coefficient of 20 S. When Dual PI was expressed *in planta* it was not detectable in the feeding nematode (Urwin et al., 1998). The monomeric form

detected in both AUC and MS is smaller than GFP and more importantly mRFP, as such it would be expected to pass through the feeding tube. In light of our results it is likely that the Dual PI is aggregating and, as a consequence, either the aggregates are excluded by the feeding tube, or they are forming insoluble inclusions in the cytoplasm and are therefore inaccessible for uptake (Kiese et al., 2008, Carpenter et al., 1999).

3.5.3. Fluorescent proteins as a tool to test uptake

Although uptake of mRFP has been demonstrated convincingly (Valentine et al., 2007), and uptake of GFP has been demonstrated in some experiments (Goverse et al., 1998), we were unable to confirm either of these with fluorescence or confocal microscopy. This type of experiment is difficult with cyst nematodes in particular as the cuticle and internal lipid bodies have been shown to auto-fluoresce. Importantly, it was not possible to separate this auto-fluorescence from the actual spectrum of mRFP even with the use of spectral un-mixing. This highlights a limitation of confocal microscopy in uptake experiments, that may have had an impact on some of the conclusions of previously published works. Similarly, we were not able to achieve close to the levels of expression previously demonstrated *in planta* for iLOV (Chapman et al., 2008). This held true for all single iLOV lines tested and all iLOV fusion proteins, including the double iLOV-iLOV fusion.

3.5.4. Further resolution of cyst nematode feeding tube size exclusion

In this study we produced a range of different sized proteins to try to better resolve the size exclusion cut-off of the cyst nematode feeding tube. Even if they had been detectable in the nematode digestive system not all of these proteins could have their size predicted using RotaMol, as protein structures were not available. Computational prediction of protein structures is possible, and much work has been done in recent years to improve accuracy. The vast majority of available prediction methods are homology based, and are therefore improving with the steady increase of solved structures. Modelling the structure of a single iLOV protein is therefore trivial as the structure has been solved of a highly related protein. However in the case of fusion protein constructs, no information is available, and so protein structure prediction fails. Moreover, protein folding *in planta* is a complex process (as demonstrated by Dual PI), and one cannot simply assume that a fusion construct will have a structure similar to a combination of the two halves.

Determining protein structures is a difficult and time consuming task that relies on a number of assumptions, the most important of which is the ability for the protein in question to form crystals. It is not practical to experimentally determine the structure of all the fusion proteins expressed *in planta*. When designing novel compounds to target cyst nematodes, such as Dual PI, it is important that their size and or heterogeneity is assessed before being tested *in vivo*.

3.5.5. Comparing nematode feeding tube size exclusion to other biological membranes.

Exclusion across other biological membranes has been previously explored and, as would be expected, the basis for exclusion depends on the membrane in question. In plasmodesmata, hydrodynamic radius is quoted as the single determining factor of exclusion (Rutschow et al., 2011, Terry and Robards, 1987). This may not be the case for the cyst nematode feeding tubes. In terms of size, monomeric GFP is just 6% larger than mRFP and yet from some of the size exclusion experiments, is excluded by the feeding tube. A sharp cut-off would be expected if the pore size of the feeding tube was uniform. This may be the case for some nematode species, such as the root knot nematode *Meloidogyne incognita*. Electron microscope images of the feeding tube of *M. incognita* show a regular structure with discrete pores (Hussey and Mims, 1991). The conflicting data for the cyst nematodes suggests that pore size may follow a continuous probability distribution as would be expected from a non-uniform mesh, such as the glomerular epithelial filtration membrane (Edwards et al., 1999). Although this analogue has completely different evolutionary origins, it may share properties with the cyst nematode feeding tube. Indeed scanning electron micrographs of the cyst nematode feeding tube do not show a regular structure with discrete pores. The glomerular epithelial membrane is a meshwork of type IV collagens and a large number of experiments have been carried out to characterise the filtration properties of this membrane. Size exclusion of the glomerular epithelial membrane is typically defined as a sieving coefficient for a given structural unit, as opposed to a discrete limit, due to its heteroporous structure (Edwards et al., 1999). This is typical for the main manufactured form of ultrafiltration device, a depth filter. In depth filters, much of the rejection process occurs within the walls of the membrane, as opposed to at the surface in aptly named surface filters. It is therefore possible that although cyst and root knot nematodes, which evolved biotrophic parasitism independently, both create a feeding tube, each uses a different type of filtration system. The cyst nematode feeding tube is more analogous to a depth filter, whereas the root knot nematode feeding tube is more analogous to the regular structure of a surface filter.

Size alone is not sufficient to describe filtration across the glomerular epithelial membrane, transport of negatively charged molecules appears to be reduced relative to uncharged molecules (Guasch et al., 1993). Applying this logic to feeding tube size exclusion may explain some of the conflicting results. As previously suggested the presence of dimeric GFP will affect the outcome of size exclusion experiments. In addition the monomeric form itself may also pose problems. Monomeric GFP is only slightly larger than mRFP and the AUC peaks for GFP monomer and mRFP monomer lie directly on top of one another, suggesting they behave similarly in solution. However, GFP has a considerably larger external negative charge (Figure 3-18), and this may explain why analysis of the filtration of GFP across the feeding tube has been so difficult. The Debye-Hückel theory of charged solute interactions describes the difference in apparent size of charged molecules and pores, where a charged molecule of 29Å will behave in the same manner as an uncharged molecule of 37Å (Tencer et al., 1998). A relationship between size/shape and charge may therefore be needed to fully understand feeding tube size exclusion.

Technically, measuring sieving coefficients across the feeding tube membrane, if at all possible, will be a considerable challenge. Similarly, until a reliable method is developed to detect exclusion or passage through the feeding tube further work on resolving the limit will be difficult. It seems that a definitive size exclusion limit may never be found. However, based on the published literature for homogenous proteins (Valentine et al., 2007, Goverse et al., 1998), the size exclusion of the cyst nematode feeding tube is at least 1,864 Å².

Summary

- Created a novel method for measuring protein size *in silico*
- Validated the method against experimental measures of protein size
- Based on published works, the prediction of the maximum size of cyst nematode size exclusion is estimated at 1,864 Å²
- Unable to further resolve the size exclusion limit with different sized fluorescent proteins

4. Identification of candidate feeding structure genes

4.1. Introduction

As covered in detail in Chapter 1 and 3, electron microscopy studies have identified two key structures produced by endoparasitic nematodes that are directly associated with the plant-nematode interface; the feeding tube and the feeding plug (Holtmann et al., 2000, Endo, 1978, Rebois, 1980, Razak and Evans, 1976, Rebois et al., 1975). Feeding tubes are produced inside the feeding site and are thought to act as molecular sieves (Eves-van den Akker et al., 2014b). Feeding plugs appear as electron dense material intimately associated with the plant-nematode interface and are located in the apoplast between the anterior end of the nematode and the feeding site (Endo, 1978), although their function is unknown.

4.1.1. Effector identification strategies

Both feeding tubes and feeding plugs fall under the broad definition of effector given in Chapter 1; proteins (or other factors) secreted by a pathogen into its host in order to manipulate the host to the advantage of the pathogen. Effector identification strategies from other pathogens often rely on a conserved motif. For example the so called RXLR effectors of oomycete species contain a highly conserved 4 amino acid RXLR domain, involved in translocation into the host, downstream of a signal peptide (Whisson et al., 2007). Similarly, and also from oomycetes, the Crinkling and Necrosis (CRN) effectors always contain a conserved LXLFLAK domain at the N-terminal of the protein (Stam et al., 2013b, Stam et al., 2013a). Finally there are numerous prediction software packages available for the type secretion systems of bacteria. However, for nematode effectors no known motif, domain, or promoter element has been identified to date. Typical nematode effector identification strategies have centred around the analysis of ESTs from the oesophageal gland cells. This type of analysis has been very successful in identifying effectors from a range of plant parasitic nematodes (Torres, 2010, Siddique et al., 2014). As mixed reports exist for the origin of the feeding plug (Endo, 1978, Sobczak et al., 1999), and no experimental reports exist for the origin of the feeding tube (Grundler and Böckenhoff, 1997), this type of approach may be prone to false negatives in terms of identifying components of these structures. This is also evidenced by the fact that recent proteomics studies of root-knot nematode secretions have identified components of the secretome that may originate from structures other than the oesophageal gland cells (Bellafiore et al., 2008). More recently genomic analyses

have been used as a tool for identification of effectors based on assumptions about the characteristics of “effector genes”. This type of analysis identifies large numbers of genes, and is highly prone to false positives. This limitation is overcome by the subsequent localisation of candidate gene products to tissues with the ability to secrete proteins into the host (Chapter 1).

There have been very few studies of feeding plugs or feeding tubes since their identification in the 1970s [10,12], primarily due to the practical difficulties in working with sedentary endoparasites, which are small and embedded deep within the roots of their hosts at the biotrophic stage. However, the advent of next generation sequencing, in particular RNA sequencing, has provided new approaches to many questions in biology. Here we describe the use of RNAseq data coupled with evidence based assumptions on the characteristics of feeding tube and plug genes to identify putative, nematode-derived components of these structures.

4.2. Aims:

Make use of the recently assembled genome, and accompanying transcriptome sequences of the potato cyst nematode *G. pallida* to identify a range of candidate feeding structure genes for further study.

- Using assumptions about the characteristics of feeding structure genes, create a pipeline for identification of candidate genes.
- Clone and sequence candidate genes
- Characterise the gene family structure.

4.3. Materials and methods

4.3.1. Data mining for candidate feeding structure genes

4.3.1.1. Differential expression analysis

The *Globodera pallida* genome and life stage specific transcriptome project was completed in late 2013. Gene predictions along with replicated expression data for 8 life stages across the life cycle were made available in 2012. Normalised expression values for each gene in each life stage were compiled in an excel file. The average of the two replicates was taken to produce a single expression value for each gene for each life stage.

All predicted genes were tested against a series of assumptions made in relation to the expected characteristics of putative feeding structure genes. The first of these was that genes encoding components of these structures would show constitutively high expression during the feeding stages, and would not be expressed at the non-feeding stages. A simple algorithm was used to identify the genes most highly up-regulated during the feeding stages when compared to the non-feeding stages. The ratio of the average expression in all feeding stages to the average expression in all non-feeding stages was used to sort all the predicted genes. However, some expression values at non-feeding stages were less than 1 giving rise to disproportionately high values for ratio of expression. Therefore, the minimum value for expression in any stage was set to 1 in order to overcome this. In addition, genes with high expression values during some feeding stages but that were not expressed in others would still give a high average expression across all feeding stages, resulting in a high ratio assuming no expression during the non-feeding stages. Since such genes were unlikely to encode components of the feeding structures being sought (which should be constitutively present in all feeding stages) an arbitrary minimum value of >10 for expression in each of the feeding stages was set. All genes were then sorted by this ratio to give a rank order to the most likely candidates. The expression patterns of the top 2% were checked manually and if they matched the expected pattern they were taken for further analyses.

4.3.1.2. Predicted secreted protein pipeline of novel proteins

Signal peptides and transmembrane domains were predicted using two online tools from the Centre for Biological Sequence analysis (CBS). Putative signal peptides were predicted using SignalP v4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) and

transmembrane domains were predicted using TMHMM v2 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>).

4.3.2. DNA and protein alignment

Alignments of both DNA and protein were carried out using the command line tool MUSCLE v3.8.3.1 (Edgar, 2004). The “refine” command was used routinely to improve all alignments. Alignments were viewed in either BioEdit or Jalview (Waterhouse et al., 2009) Maximum likelihood phylogenetic trees were drawn in TOPALi V2 (Milne et al., 2009) with a minimum of 100 bootstraps.

4.3.3. 3'RACE

3' RACE was carried out using the 5'/3' RACE Kit (Roche) using the supplied oligo dT primer, a corresponding gene specific forward primer (Table 4-1) and RNA extracted from feeding females as per section 2.3.3. In the case of 444 genes, subfamily-specific forward primers were identified by aligning all cloned sequences and visualising in an alignment viewer (BioEdit). The RACE procedure was carried out according to the manufacturer's instructions, the resultant PCR products were analysed by typical gel electrophoresis and the relevant bands cloned and sequenced as described in section 2.4.

4.3.4. Single nematode DNA extractions

For the following three extraction methods 14 day post infection *G. pallida* females were flash frozen in liquid nitrogen in individual 1.5 ml microfuge tubes.

4.3.4.1. Whole Genome Amplification (WGA)

Two methods were used for whole genome amplification. In the first of these, 9 µl (the maximum permitted volume) of DNA extracted from single nematodes using the NucleoSpin DNA/RNA XS kit was added prior to the Lysis and Fragmentation step (step 3) of the GenomePlex Single Cell Whole Genome Amplification Kit (Sigma). In the second, the protocol suggested in the WGA kit was carried out following the manufacturer's instructions from start to finish.

4.3.4.2. NaOH digestion

Samples were suspended in 20 µl of 0.25 M NaOH, centrifuged to collect the contents, and frozen at -80 °C. The nematodes were incubated at 25 °C overnight, heated to 99 °C for 3 minutes, and cooled to room temperature. The contents were collected by centrifugation prior to the addition of 4 µl 1 M HCl, 10 µl 0.5 M Tris-HCl

(pH 8.0) and 5 μ l 2% Triton X-100. The contents were then vortexed to mix, collected by centrifugation, heated to 99 °C for 3 minutes, allowed to cool to room temperature and stored at -80 °C for future use. Between 0.25 and 1 μ l of each sample was used in a 25 μ l PCR reaction.

4.3.4.3. **Chaos nucleic acid extraction buffer**

Nematodes were suspended in 200 μ l "Chaos" buffer (Triant and Whitehead, 2009) (4.5 M guanidine thiocyanate, 2% N-lauryl sarcosine, 50 mM EDTA (pH 8.0), 0.1 M 2-mercaptoethanol, 0.2% antifoam-A), disrupted with a pipette tip, and lysed by vortexing. One volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the sample, which was vortexed and centrifuged at 10,000g for 5 minutes. The upper aqueous phase was mixed with 1 volume of 70% ethanol with the addition of 20 ng carrier RNA (NucleoSpin RNA XS). Total nucleic acid was extracted from the sample using a NucleoSpin RNA XS column, and the NucleoSpin RNA/DNA Buffer Set following manufacturer's instructions. Two microliters of purified DNA were used for a 50 μ l PCR reaction.

4.3.5. **Tandem repeat analysis**

Analysis of tandem repeat regions was carried out on genomic clones of 444 members only. For each subfamily, the central region of the gene (between the two introns) was translated so that frame would be consistent with a cDNA clone. The software XSTREAM (<http://jimcooperlab.mcdb.ucsb.edu/xstream/> 2013-8-8) was used to analyse the tandem repeats present for each subfamily.

4.3.6. ***In vitro* transcription**

Two micrograms of template plasmid DNA were linearised using 2 restriction enzymes for each construct in order to ensure complete linearisation. For 444 subfamily-clones Nde I and Spe I were used while Kpn I and Xho I were used for the GpCys control. Digests were carried out in the appropriate digestion buffer at 37 °C for 3 hours. 1/20th of each digest was electrophoresed on a 1.8% agarose gel to confirm that digestion was complete. Reactions were incubated in 0.5% SDS containing 200 μ g/ml proteinase K for 30 minutes at 50 °C. One volume of phenol:chloroform:isoamylalcohol (25:24:1) was then added, the sample was vortexed to mix and centrifuged for 10 minutes at 14,000g. The upper aqueous phase was removed to a fresh microfuge tube containing 2 volumes of EtOH, 1/20th volume 0.5 M EDTA and 1/10th volume of 3 M NaAc. The reaction was incubated at -20 °C overnight to precipitate the DNA. The following day the reaction was centrifuged at 11,000g for 10 minutes to pellet the DNA, the supernatant was

removed and the pellet allowed to air dry until clear. The pellet was resuspended in 10 µl of RNase-free water and the concentration measured using a NanoDrop spectrophotometer.

Four hundred nanograms of cleaned linear template DNA were used in the MEGAscript *in vitro* transcription kit (Ambion) following the manufacturer's instructions. The transcription reaction was stopped after 3 hours by addition of Turbo DNase. The reaction was then purified using the RNeasy Plant Mini Kit (Qiagen), following the manufacturer's instructions for "RNA cleanup", including the optional on-column DNA digestion. RNA was eluted in 30 µl of RNase free water.

Gene specific cDNA was synthesised from the purified RNA using SuperScript II reverse transcriptase (Invitrogen) or Maxima Reverse transcriptase (Thermoscientific) according to the manufacturer's instructions. In each case 220 ng of RNA was used, with the relevant gene specific primer. For SuperScript II, cDNA was synthesised at 42 °C. For Maxima reverse transcriptase the optional 65 °C incubation for 5 minutes was carried out, and cDNA was synthesised at 65 °C.

4.3.7. Denaturing RNA gel electrophoresis

Denaturing RNA gels were prepared by dissolving 1.8% agarose (w/v) in water, and allowing to cool to approximately 60 °C. An appropriate volume of 5 X running buffer was added (0.1 M MOPS pH 7.0, 40 mM NaAc ,5 mM EDTA pH 8.0) to give a final concentration of 1X. Finally 37% w/v formaldehyde solution (12.3 M) was added to obtain a final concentration of 2.2 M and GelRed (Biotium) at a ratio of 1:20,000. The solution was mixed by gentle agitation, poured into a gel tray and allowed to set in a fume hood for at least 2 hours. Each sample was prepared by mixing 2 µg of RNA with 5 X sample loading buffer (Bioline), and 10 µl of formamide to make a final volume < 20 µl and a final loading buffer concentration of 1X. Samples were heat denatured at 65 °C for 5 minutes followed by immediate chilling on ice prior to loading into the gel. The gel was electrophoresed in the dark at 200 volts for sufficient time to resolve the relevant bands (typically 2-3 hours)

Table 4-1 **Primer sequences, annealing temperatures and expected product sizes where applicable.**

Primer Name	Sequence 5' - 3'	TM (°C)	Expected product size (bp)
444_3'_RACE_Sub_A	GAGGTTATGACGAGCATCATC	54	N/A
444_3'_RACE_Sub_B	GAAAGGGCGGAGACAAAG	55	
444_3'_RACE_Sub_C	TGAGCATCGTCTCCGTGCTG	60	N/A
444_3'_UTR_Sub_A_R	AATGACATTACAATGGCAGAG	55	
444_3'_UTR_Sub_B_R	AGGCTACAAATTCACACAGC	55	N/A
444_3'_UTR_Sub_C_R	TGACACAATTACAATGGAAGG	56	
444_signalp1_F	ATG GTC GGC TAA CAA TTG A	55	N/A
444_signalp2_F	ATGGCCAACAACAATCTTAA	54	
444_no_signalp_F	TGGGGATGCAAGCCG	56	N/A
444_R	TTAATATTTGCATTCGCAAGC	59	
176_signalp_F	ATGCAATTATTCCTCTTTTTTGTG	58	139
176_R	TCACCAGCCTCCACCG	60	
3453_signalp_F	ATGATGACCCCTCTCCGCT	62	219
3452_R	TCACTTCTTTCCACCAGGGT	59	
1078_no_signalp_F	CAACTATTGCCGTGAGCTATCA	60	612
1078_R	TCAGTGCCACACTTCAGAG	60	
225_no_signalp_F	CTTAAACTCGACGAAAAAATC	54	243
225_R	TTAGTCATGGGAACACAATTC	55	

4.4. Results

4.4.1. Identification of genes that may encode components of the feeding structures

A computational approach was taken to identify an initial set of candidate feeding structure genes. This approach was based around the following three assumptions. Genes encoding components of both feeding structures would be specifically, and highly, up-regulated during all feeding stages of the life cycle, when compared to the non-feeding stages. Secondly, they would encode secreted proteins and therefore have a predicted signal peptide and no predicted transmembrane domain. Thirdly, they would lack conserved domains that had well described non-structural roles.

4.4.1.1. Expression analysis

Using the genome assembly, and accompanying RNAseq from various life stages mapped to the genome, expression of each gene at each life stage could be used to begin the process of identifying candidate genes. All genes in the genome were placed in rank order according to the ratio of average expression across feeding stages to average expression across non-feeding stages as described in section 4.2.2.1. The expression profile of the top 500 genes from this analysis was checked manually for consistently high expression across all stages (Figure 4-1), resulting in an initial set of 195 manually confirmed genes.

4.4.1.2. Signal peptide and transmembrane prediction

The 195 genes that passed the expression analysis filter were subjected to a typical “secreted protein prediction pipeline” that involved the computational prediction of both signal peptides and transmembrane domains. Signal peptides were predicted using SignalP 4.1 (Centre for Biological Sequence analysis – CBS), an example of a positive output is shown in Figure 4-2. Transmembrane domains were predicted on those genes that had predicted signal peptides, a typical negative output is shown in Figure 4-3. This analysis resulted in 29 unique candidate genes that contained putative signal peptides and that lacked putative transmembrane domains.

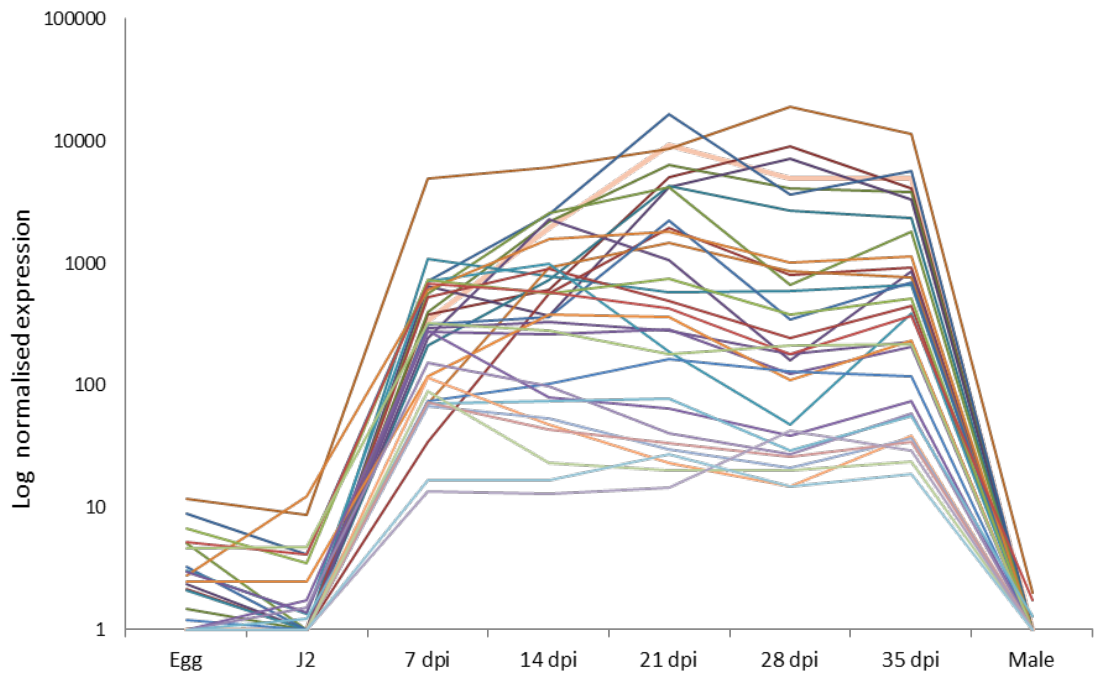


Figure 4-1 **Expression analysis of candidate feeding structure genes.** Expression profile of manually confirmed genes, consistently highly expressed across all feeding life stages, with secretion signals. The normalised average expression at each stage is plotted on a log scale, indicating up to 10,000 fold over expression during feeding stages compared to non-feeding stages.

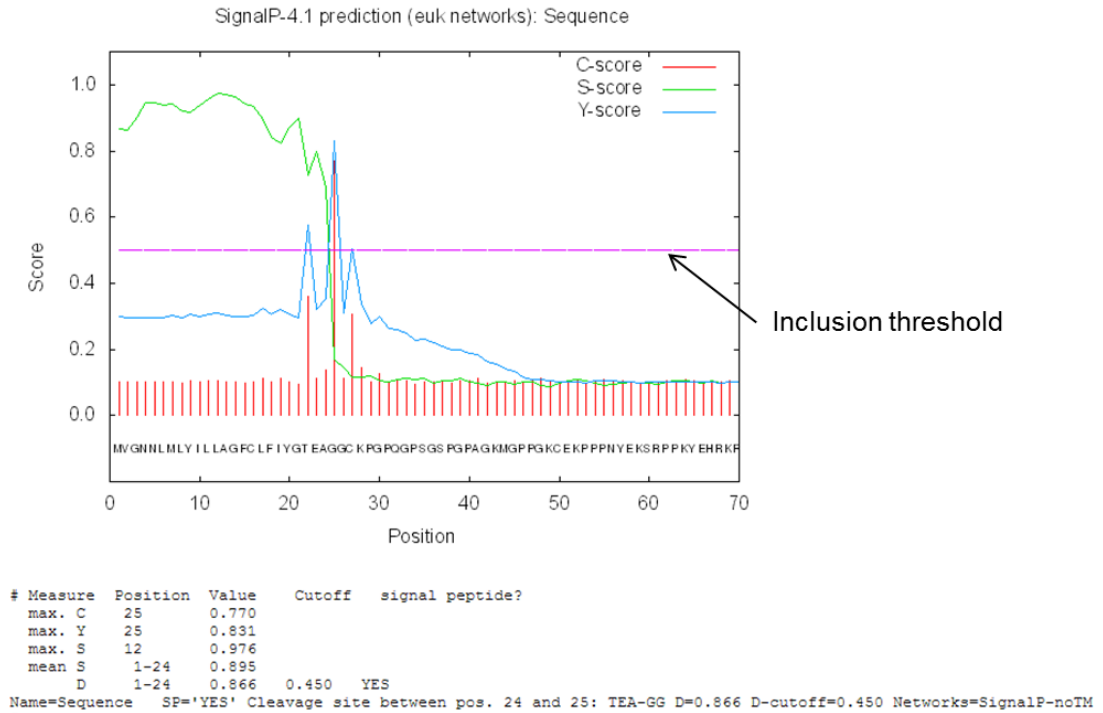


Figure 4-2 **Typical positive output from signal peptide prediction using SignalP v4.1.** For a 444-B subfamily gene the probability of the provided sequence including a signal peptide, and the point at which it is predicted to be cleaved (in this case between amino acid position 24 and 25) are indicated.

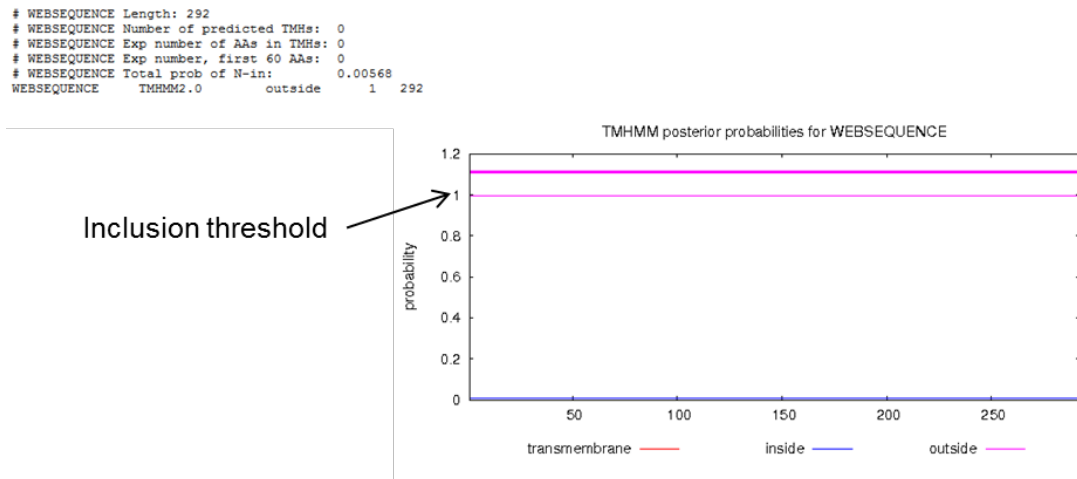


Figure 4-3 **Typical negative output from transmembrane domain prediction using TMHMM.** No transmembrane domain is detected across the length of the protein above the significance threshold (pink line).

4.4.1.3. Similarity search

The 29 candidates that passed the expression and secretion criteria were then compared to all published sequences in the NCBI non-redundant nucleotide collection (NR). Nine of the candidates had no significant similarity to previously described sequences (evalue ≥ 0.005) and were kept for further analysis. Twelve of the 29 sequences corresponded to a gene family of uncharacterised putative effectors previously identified (4D06) from the closely related cyst nematodes *Heterodera glycines* (Gao et al., 2003) and *Globodera rostochiensis* (unpublished). Five of the 29 had BLAST matches to glutathione synthetases and were therefore discarded as feeding structure candidates. Two of the 29 identified tBLASTx matches to *Caenorhabditis elegans* cuticle collagens with very high scores, and 1 of the 29 to a C-Lectin binding domain containing protein. These sequences were therefore discarded (Table 4-2). An overview of the entire candidate feeding structure identification pipeline is provided in Figure 4-4.

The following candidates were subject to further analysis, these comprised four genes and two gene families:

4.4.2. Single copy genes

Prior to further analysis, primers were designed to clone the full length coding sequence of selected genes from cDNA using primer3 software with default parameters. Cloning procedures were followed according to Section 2.4. Genes **225**, **3453** and **176** were cloned and, as expected, the cloned sequences matched those predicted from the genome assembly. These candidates were subject to further analysis described in detail in Chapter 5.

4.4.3. Candidate 1078 misprediction

The successfully cloned transcript of **1078** was compared to the gene coding region predicted from the genome sequence. Figure 4-5 shows the transmembrane results for predicted and cloned sequences. No significant transmembrane domain was predicted from the gene prediction, however cloning from cDNA revealed that the sequence did indeed contain a predicted transmembrane domain. Figure 4-6 shows an alignment of the predicted gene and the cloned gene, highlighting the region that was mispredicted and that corresponds to a transmembrane domain. Gene **1078** no longer matched the required criteria defined for candidate genes and was therefore discarded from further analyses.

Table 4-2 Summary of candidate feeding structure genes identified including expression at each life stage and BLAST hits.

<i>Gene/gene family name</i>	<i>Gene ID</i>	<i>Egg</i>	<i>J2</i>	<i>7 Days post infection</i>	<i>14 Days post infection</i>	<i>21 Days post infection</i>	<i>28 Days post infection</i>	<i>35 Days post infection</i>	<i>Male</i>	<i>BLAST</i>
444	GPLIN_000907700	0.4	0.0	648.5	373.5	4188.6	7144.4	3334.0	2.3	No Hits
444	GPLIN_001025300	0.2	0.0	382.6	602.7	4993.7	9047.1	4118.2	0.2	No Hits
3453	GPLIN_001526200	0.2	0.3	521.2	898.2	491.6	242.9	442.2	0.2	No Hits
176	GPLIN_000616800	0.0	4.7	325.1	277.1	181.5	209.9	216.8	4.6	No Hits
225	GPLIN_000713700	5.5	16.1	251.6	413.7	536.5	306.9	417.7	34.0	No Hits
444	GPLIN_001208400	6.4	8.5	33.7	89.9	27.6	65.1	60.1	1.1	No Hits
1078	GPLIN_001341400	0.7	2.0	15.8	10.7	8.7	12.8	11.2	5.2	No Hits
124	GPLIN_000488600	0.0	0.2	8.7	9.7	15.8	27.8	18.8	0.9	No Hits
210	GPLIN_000679600	0.0	13.5	16.3	39.1	14.0	15.9	21.0	1.2	No Hits
448	GPLIN_000481100	0.0	0.2	67.2	53.7	30.0	21.0	36.6	0.3	4D06-like effector (Gao et al., 2003)
448	GPLIN_000203300	0.4	2.4	62.3	33.3	46.5	15.7	31.9	1.6	4D06-like effector (Gao et al., 2003)
448	GPLIN_001255700	0.1	0.2	72.2	43.2	33.5	25.9	34.7	0.0	4D06-like effector (Gao et al., 2003)
448	GPLIN_000969900	0.0	1.1	69.1	57.1	25.5	16.3	40.2	0.1	4D06-like effector

										(Gao et al., 2003)
448	GPLIN_000970000	0.2	1.7	277.2	80.3	64.1	38.2	74.6	0.8	4D06-like effector (Gao et al., 2003)
448	GPLIN_000308900	0.0	0.0	11.5	10.4	10.7	13.6	11.5	0.0	4D06-like effector (Gao et al., 2003)
448	GPLIN_000950600	1.6	6.7	236.4	116.6	59.9	28.0	82.8	0.8	4D06-like effector (Gao et al., 2003)
448	GPLIN_000243700	0.8	1.8	85.4	51.5	26.9	25.3	40.1	0.4	4D06-like effector (Gao et al., 2003)
448	GPLIN_000243800	0.8	1.5	151.7	98.3	40.5	27.4	59.1	0.8	4D06-like effector (Gao et al., 2003)
448	GPLIN_001390400	0.0	0.2	41.6	47.6	17.8	18.0	29.3	0.5	4D06-like effector (Gao et al., 2003)
448	GPLIN_001390500	0.0	0.0	2.0	46.8	23.2	22.0	24.1	0.0	4D06-like effector (Gao et al., 2003)
448	GPLIN_000970100	0.0	14.3	186.5	58.5	39.9	46.4	76.2	0.0	4D06-like effector (Gao et al., 2003)
448	GPLIN_001606400	1.1	1.7	71.9	26.5	12.1	9.9	20.8	0.5	4D06-like effector (Gao et al., 2003)
448	GPLIN_000912100	3.9	2.6	121.9	48.3	21.3	11.9	36.6	0.9	4D06-like effector (Gao et al., 2003)
448	GPLIN_000860700	0.0	0.0	94.8	30.1	17.5	12.0	27.0	0.0	4D06-like effector (Gao et al., 2003)

GSH	GPLIN_000812300	0.6	0.1	103.1	53.1	22.5	12.4	32.6	1.2	glutathione synthetase
GSH	GPLIN_000167200	0.2	0.3	122.7	49.5	25.7	10.3	32.7	0.6	glutathione synthetase
GSH	GPLIN_000240700	1.0	0.5	149.9	61.5	27.9	13.5	36.4	1.5	glutathione synthetase
GSH	GPLIN_000241200	2.7	3.4	235.6	76.2	35.4	11.0	49.1	3.7	glutathione synthetase
GSH	GPLIN_000407200	0.0	0.1	206.5	53.7	40.9	22.4	49.4	0.2	glutathione synthetase

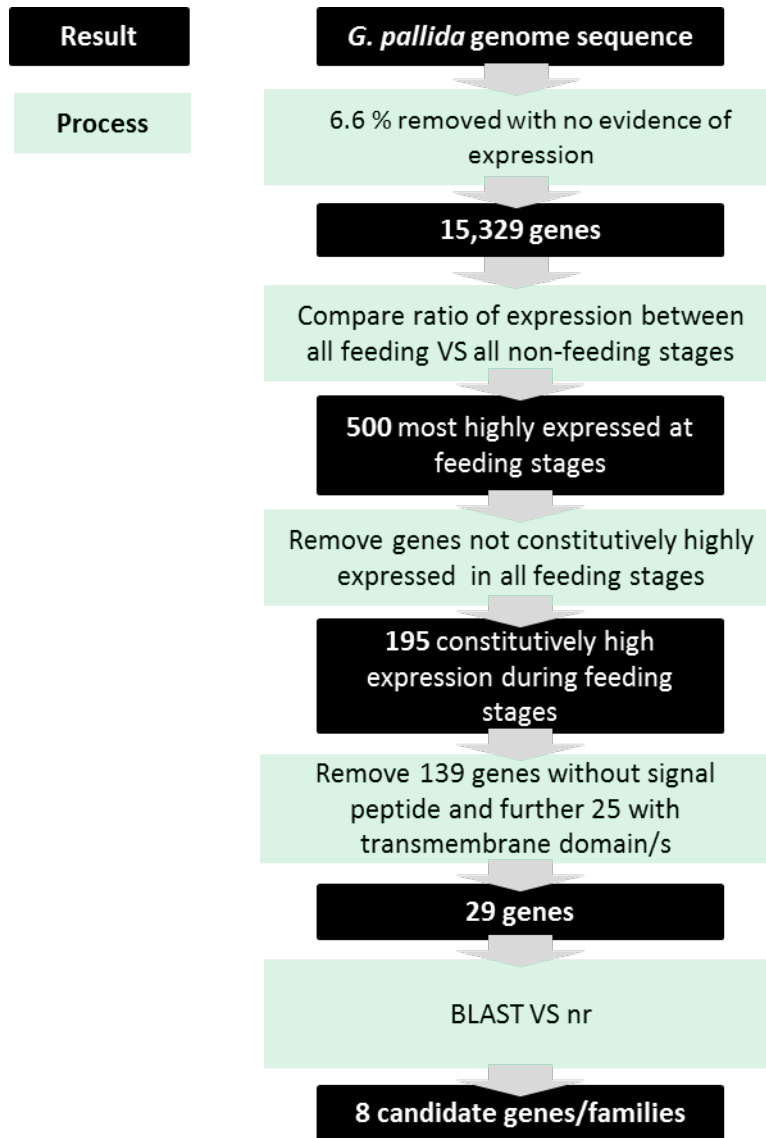
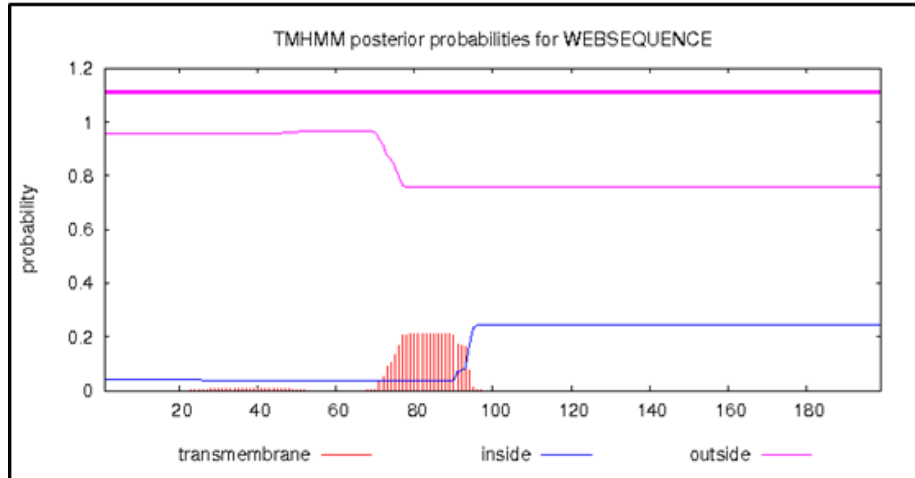


Figure 4-4 **Pipeline overview for identification of putative feeding structure components.** A summary of the process used for identifying candidate feeding structure genes from a complete genome. In brief, this involves differential expression analysis and manual confirmation of consistently high expression in all feeding stages, secreted protein prediction, and BLAST against the non-redundant nucleotide collection to exclude proteins with well-defined non-structural roles.

Predicted sequence



Cloned sequence

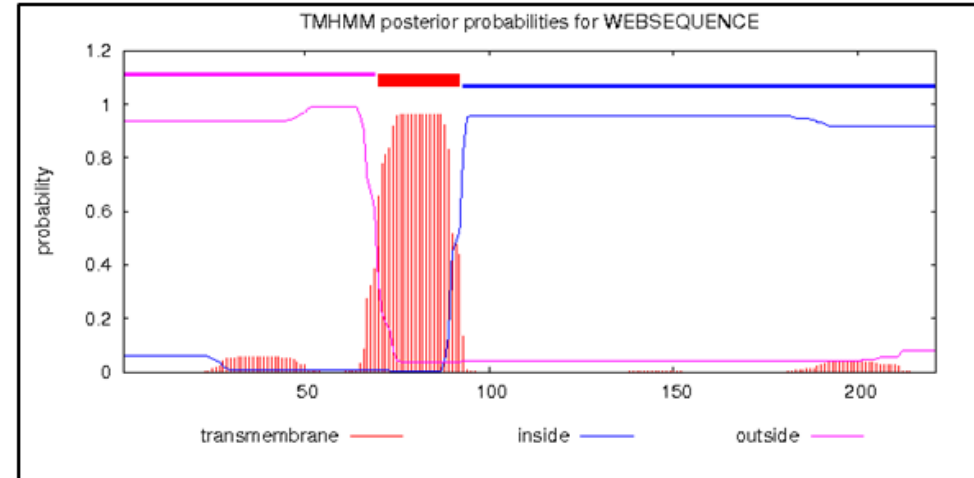


Figure 4-5 **Transmembrane predictions on predicted and cloned 1078 sequences.** No transmembrane domain is predicted above the inclusion threshold for the gene present in the genome sequence, whereas a domain is predicted from the cloned sequence. This was due to the gene prediction software missing an exon (below). Gene 1078 no longer matched the criteria used to choose candidate genes and was therefore discarded from further analyses.

```
CLUSTAL 2.1 multiple sequence alignment

Cloned_sequence      MSSITLLISIVLIVLISAGPFVQADNAGESHDLINRIKRNSYGYSGYGGSCSSCCSS 60
Predicted_Sequence  MSSITLLISIVLIVLISAGPFVQADNAGESHDLINRIKRNSYGYSGYGGSCSSCCSS 60
*****

Cloned_sequence      CGGGSSYQPTSYVQQAPAPP-TTTTTATSAAGLHAGAAVLLFLILQHLRQRWWQR 119
Predicted_Sequence  CGGGSSYQPTSYVQQAPAPPFFFFFFPVPVYTPVRAGVLLFLILQHLRQRWWQR 120
*****

Cloned_sequence      KRSVLQQQLERRWKLVLQLVEQFRQLGRFQRLRRLRLIHVGGQWRLWLRWRLLSVR 179
Predicted_Sequence  KRSVLQQQLERWKLVLQLVEQFRQLGRFQRLRRLRLIHVGGQWRLWLRWRLLSVR 180
*****
```

Figure 4-6 **Alignment of mispredicted and cloned 1078 protein sequences.** ClustalW alignment of the predicted and cloned 1078 protein sequences. The boxed region corresponds to an exon that was mispredicted by the automated gene annotation software and which encodes a putative transmembrane domain.

4.4.4. Candidate 448 gene family

Twelve of the 29 candidate genes were all similar to a gene family previously described (4D06), but as yet uncharacterised, from *Heterodera glycines* (Gao et al., 2003) and *Globodera rostochiensis*. These sequences were named “448” on the basis of the name provided to the *G. rostochiensis* sequences. The 12 448-like genes resulting from the candidate pipeline were used to identify further members of this gene family by BLAST. In total this identified 29 full length genes, that contain predicted signal peptides and no predicted transmembrane domains. The expression of all of these genes is summarised in Figure 4-7. The majority are expressed throughout infection, with a major peak at the early stages of infection (7 dpi) and a minor peak at 35 days post infection.

An alignment was made between all 29 putatively secreted 448 proteins and the eighteen previously identified in *G. rostochiensis* (Figure 4-8). The 448s in *G. pallida* can be broadly divided into three main groups (highlighted), one of which seems to group with the majority of *G. rostochiensis* sequences.

4.4.4.1. Candidate 448 highly conserved motif

All secreted 448 sequences, irrespective of the group, share a conserved motif as shown in Figure 4-9. Moreover, this consensus sequence, termed “PCCP”, has a conserved location within each protein. The PCCP motif is always present at the 5' end of the mature protein less than 10 amino acids after the predicted cleavage site of signal peptide.

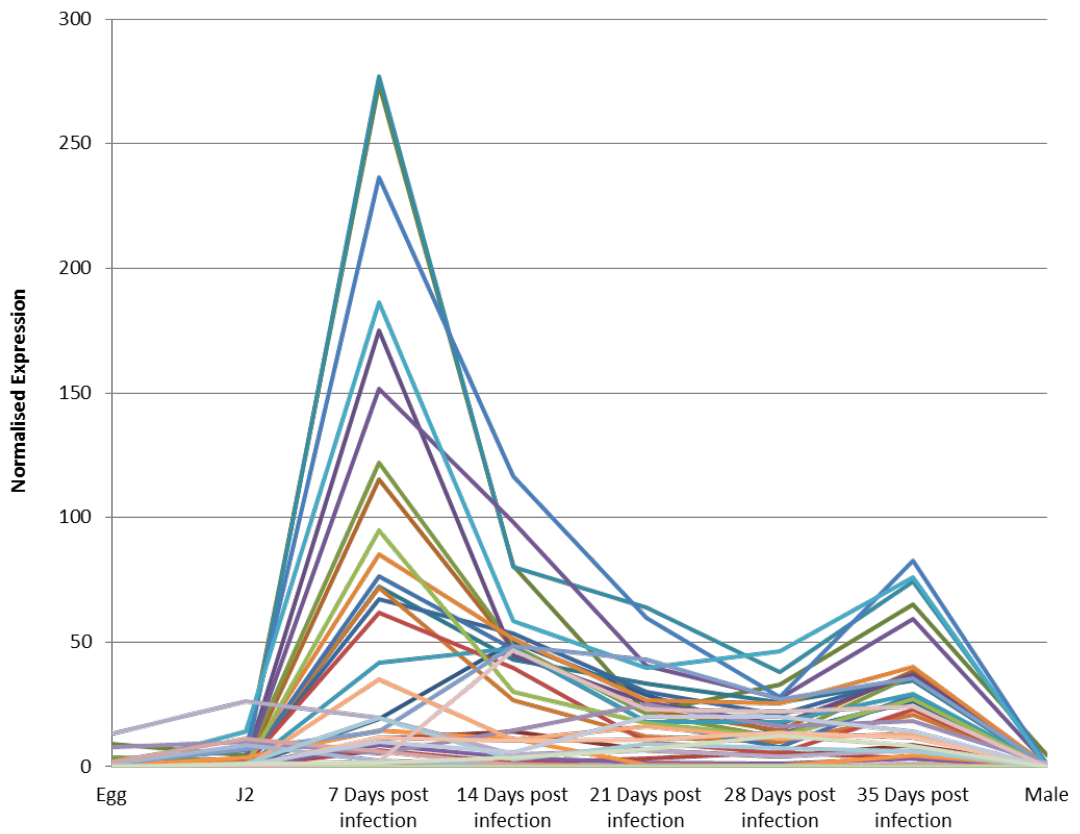


Figure 4-7 **Expression of all *G. pallida* 448 genes across the life cycle.** Summarised expressions of all 448 genes. The normalised average expression at each stage is plotted on the Y axis, indicating that most 448 genes have a peak of expression at 7 days post infection with a minor peak at 35 days post infection.

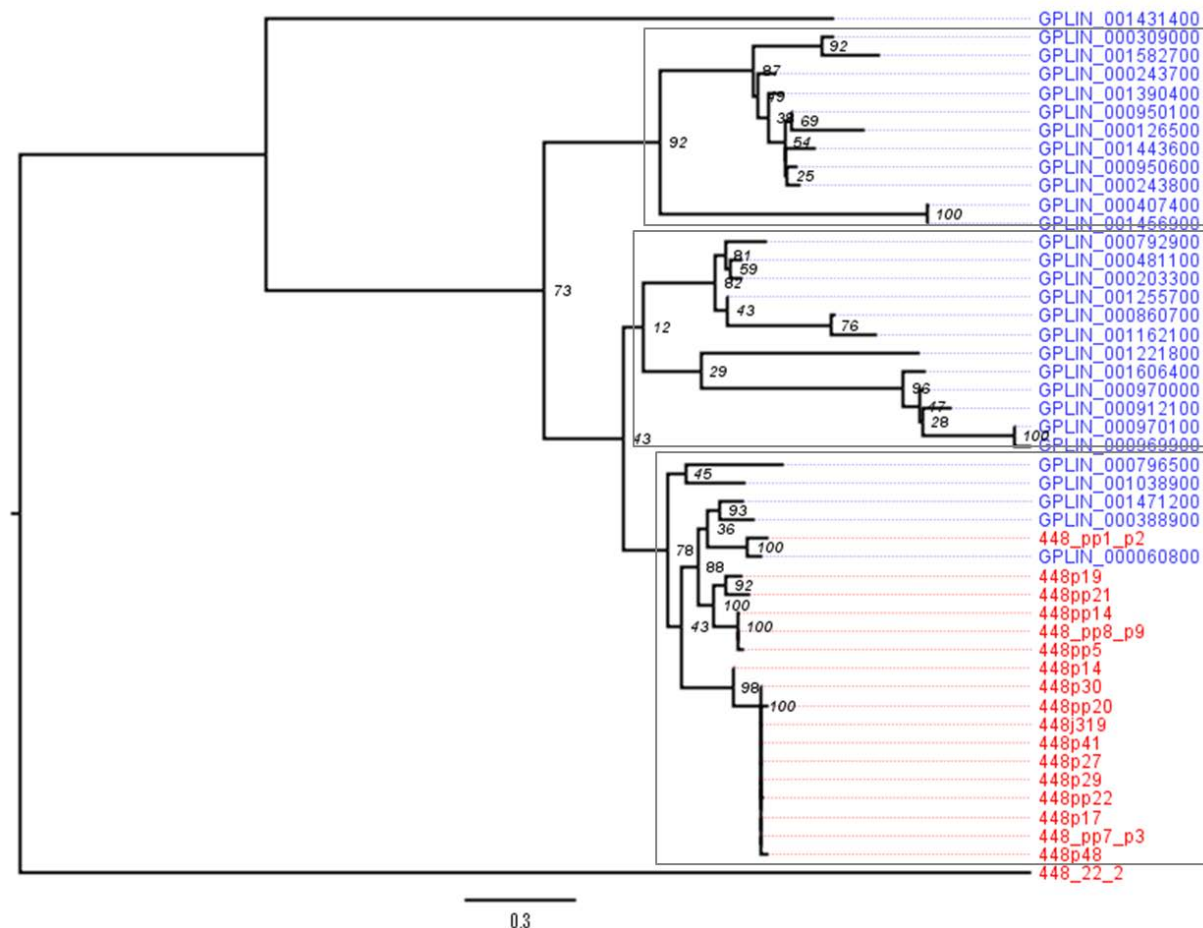


Figure 4-8 **Phylogeny of 448 genes of *G. pallida* and *G. rostochiensis*.** Maximum likelihood phylogeny using a protein alignment of the 29 putatively secreted *G. pallida* 448 sequences (blue) with the 12 available *G. rostochiensis* 448 sequences (red). Boot strap support values for 100 iterations are shown as node labels. The 448 genes can broadly be divided into three main groups, one of which contains most *G. rostochiensis* sequences.



Figure 4-9 **Motif common to all 448 sequences.** The program MEME was used to predict conserved motifs in the *G. pallida* 448 sequences. A highly conserved “PCCP” sequence was detected directly following the predicted signal peptide cleavage site.

4.4.5. Candidate 444 gene family

4.4.5.1. Cloning from cDNA

Based on the genomic sequence of GPLIN_001208400 (ftp://ftp.sanger.ac.uk/pub/pathogens/Globodera/pallida/Gene_Predictions/), primers were designed to amplify the coding region after the predicted cleavage site of the signal peptide to the stop codon (Table 4-1). This primer pair amplified a range of different sized products (Figure 4-10). Analysis of the sequences of these cloned products showed that they shared large stretches of conserved bases at the 5' and 3' ends of the genes, with more variable sequences in between these regions, and that they could be grouped into three subfamilies A, B and C. Each subfamily was then compared to the genome sequence of *G. pallida* by tBLASTn. Two complete genes were present, the previously identified GPLIN_001208400 and GPLIN_001025300, which were members of subfamily -A and -C respectively. There were also two gene fragments (either a partial sequence or a sequence containing poly-N regions), GPLIN_001135100 and the previously identified GPLIN_000907700, from subfamilies A and C respectively. Based on the genomic sequence of gene GPLIN_001208400, primers were designed to amplify this sequence from the start to the stop codon. These amplified a range of different sized products that could be placed into subfamilies A and C only.

No subfamily-B members were present in the assembled genome sequence. However, two 444-like sequences were present in the *de novo* transcriptome assembly of early sedentary stage nematodes (7 days post infection), both of which corresponded to subfamily -B. Primers designed to amplify from the start to the stop codon of these sequences amplified a range of products of similar sizes that all corresponded to subfamily -B. It was not possible to design PCR primers from coding regions that would specifically amplify full length subfamily-A or C.

4.4.5.2. 3'RACE

Using subfamily-specific primers identified from the 3' end of the coding sequence and oligodT primers, 3' untranslated regions (UTRs) were amplified from poly A tail amplified cDNA. Some apparently non-specific amplification occurred, however all clones that corresponded to subfamily-A members always had the same 3' UTR. This was also the case for subfamilies B and C, where all 3' UTRs identified were identical within, but different between, subfamilies (Figure 4-11). Thus, it was possible to design subfamily-specific primers using a combination of the signal peptide and 3' UTR sequences. Figure 4-12 shows the gene group structure,

available primer combinations, and their resultant PCR products on DNA extracted from a pool of nematodes.

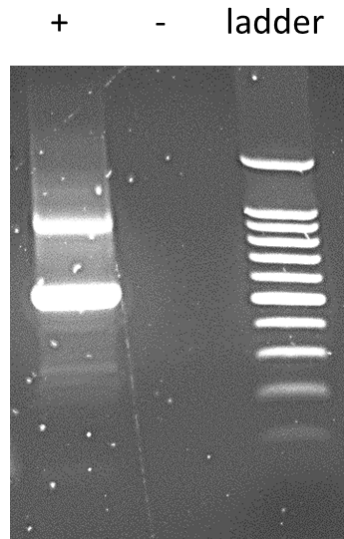


Figure 4-10 **First 444 PCR amplification of cDNA.** PCR using the 444 truncated primers on cDNA extracted from 5 week females. Multiple amplification products were amplified from the single primer pair. Amplification products were cloned, sequenced and could be readily distinguished into three subfamilies.

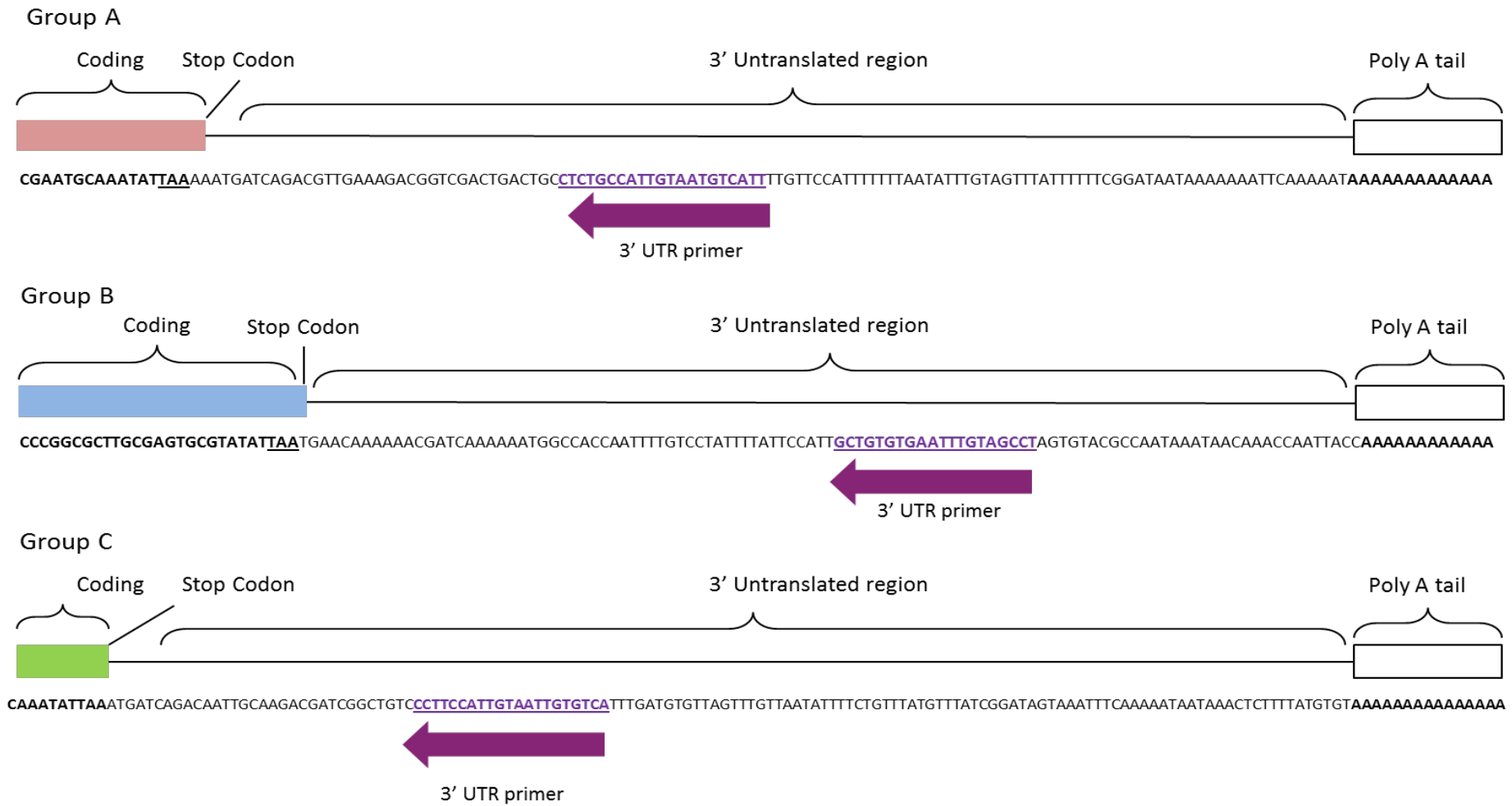


Figure 4-11 **3' Untranslated regions of 444 groups A, B and C identified by 3' RACE.** 3' RACE identified a unique 3' UTR for each subfamily that was conserved within groups but different between groups leading to the identification of subfamily specific 3' UTR primers for subsequent PCR analysis.

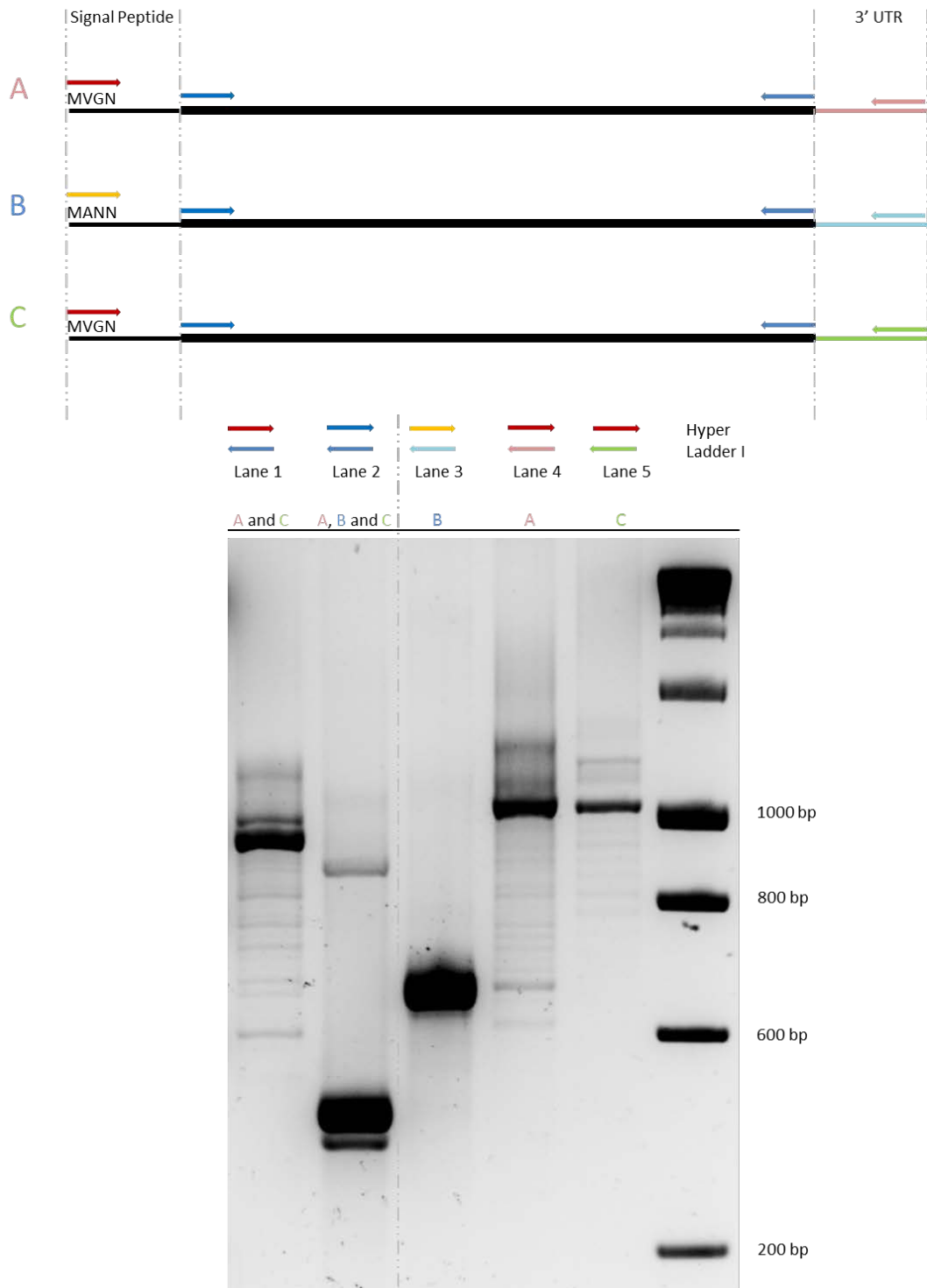


Figure 4-12 **Schematic representation of 444 family gene structure, available primer pairs, and resultant PCR products on genomic DNA.** Generic primers (blue) are able to amplify all groups. Groups are distinguished by PCR using a combination of 3' UTR primers (coloured by subfamily) and one of the two types of signal peptide primers (red or yellow). Multiple PCR products are detected for each primer pair.

4.4.5.3. **Single nematode analysis**

When cloning candidate 444 members it was noted that variation could be seen within subfamilies using the same primer pair (also shown by the multiple bands obtained from a single primer pair Figure 4-12). The template used in the PCR reaction was DNA extracted from a pool of nematodes. It was hypothesised that the variation seen in cloned DNA sequences may be explained by one of two scenarios. Either each nematode contained all of the cloned sequences, or each nematode contained a subset, and the large number of sequences identified was as a result of pooling hundreds of different nematodes. To clarify this, the PCR and cloning was repeated on an individual worm basis for each subfamily.

- Whole Genome Amplification

Whole genome amplification (WGA) was used on individual worms in order to generate sufficient template for a PCR reaction. The whole genome amplification for two different nematodes appeared to be successful (Figure 4-13i). However, although actin sequence could be amplified (Figure 4-13ii) the generic 444 primers (Figure 4-13ii), and the group specific primers (Figure 4-13iii) yielded negative or inconsistent results.

- NaOH digestion

An alternative method for extracting DNA from individual nematodes was attempted. This involved simple NaOH digestion, and subsequent PCR of lysate from single individuals. From these samples, the 18S ribosomal gene was successfully amplified, but no combination of 444 primers resulted in PCR amplification (Figure 4-14).

- Chaos DNA extraction

Finally "Chaos" extraction was attempted on 6 individual feeding stage nematodes. Figure 4-15 shows the products amplified using each of the three group specific primer sets on the same 6 nematodes, along with Elongation Factor 1 alpha control (EF1 α). Subfamily-A shows considerable differences between individuals. In the case of subfamily-C it appears that there is not just variation between genes but also either presence or absence. This PCR pattern for subfamily-C was repeatable. No such variation is seen using the control gene EF1 α . Amplification products from each subfamily were sequenced to confirm bands did indeed correspond to 444 genes. No two nematodes tested had the same complement of 444 genes.

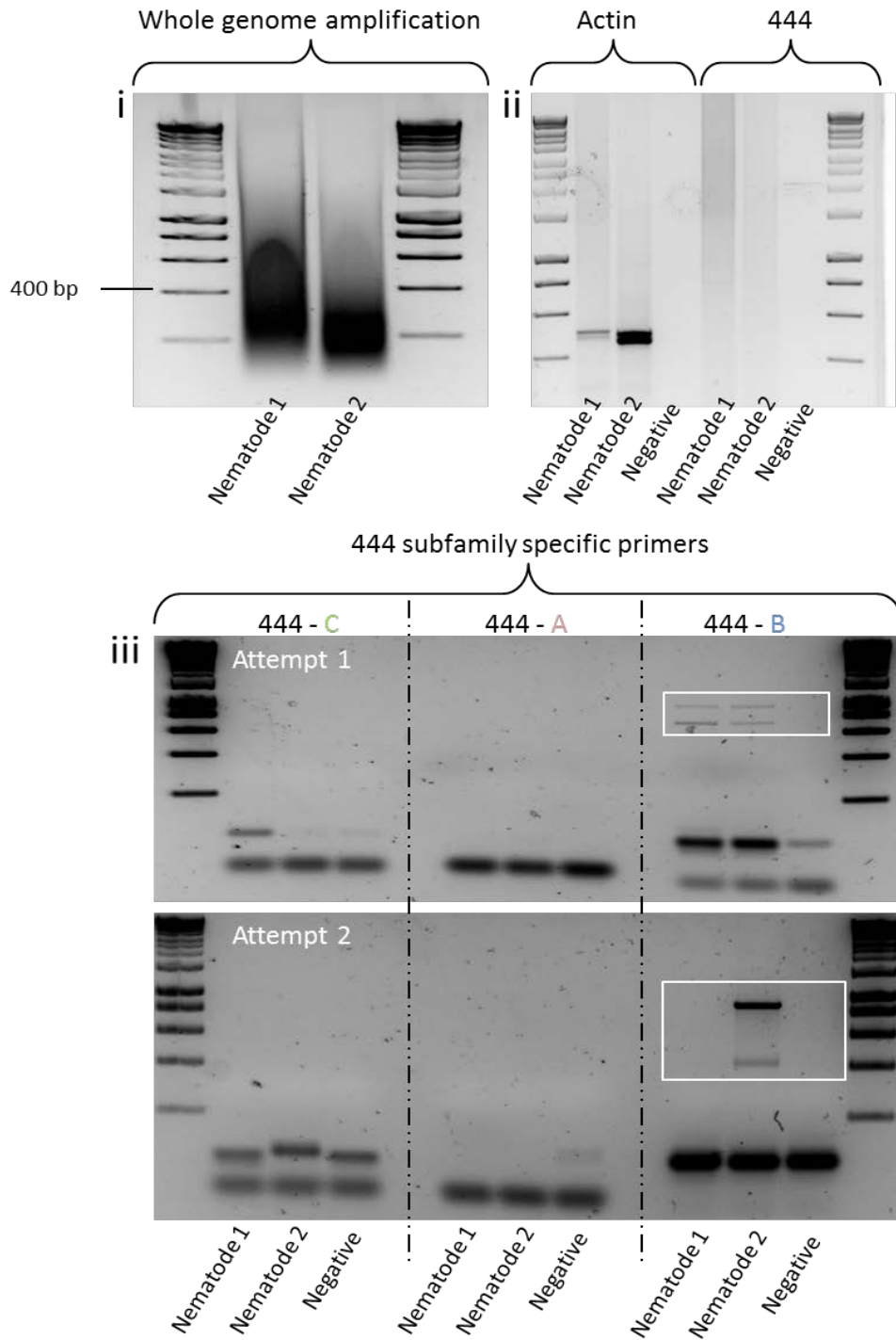


Figure 4-13 **Whole genome amplification of single nematodes.** i) shows the resultant amplified DNA sample when electrophoresed on an agarose gel, with the an expected average size of fragments to be 400bp. ii) PCR on both amplified samples with actin and 444 generic primers. No products were detected using the 444 generic primers. iii) for each attempt, 444 gene specific primers on both amplified samples, amplification is either absent or inconsistent between replicates.

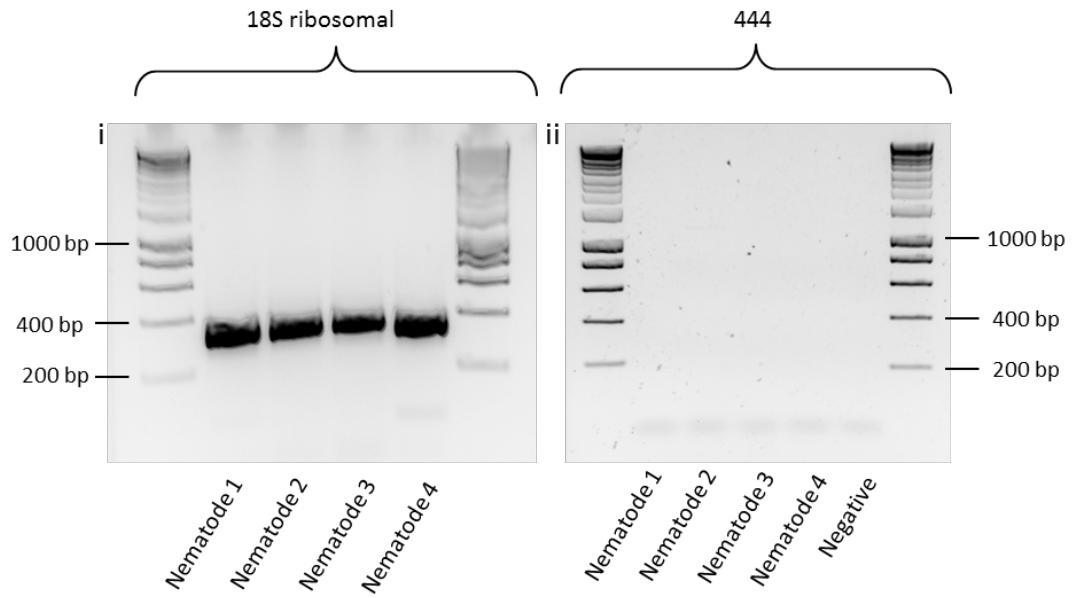


Figure 4-14 **PCR on single nematode NaOH DNA extractions.** Four individual nematodes were digested with NaOH and the lysate was analysed by PCR using both 18S ribosomal control primers and 444 generic primers. Actin primers were able to amplify a sequence of the expected size. No amplification could be detected using the 444 generic primers.

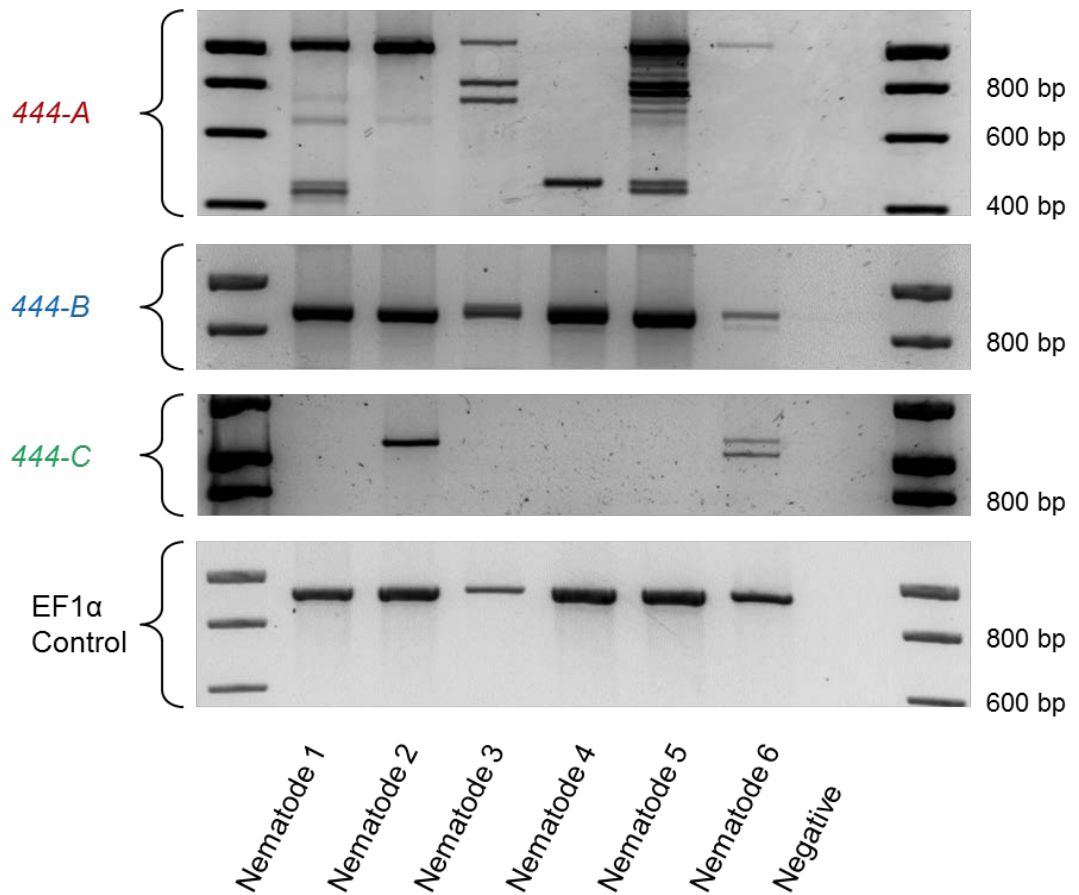


Figure 4-15 **Single nematode Chaos DNA extraction.** PCR using 444 gene specific primers and EF1 α control primers on Chaos DNA extraction from 6 individual nematodes. Considerable variation is seen between individuals in all 444 groups but not in the EF1 α control.

4.4.5.4. **Apparent non-canonical splicing**

Additional variation in the tandem repeat regions was noticed when cloning 444 sequences from cDNA compared to cloning from genomic DNA. Although the known list of genomic DNA sequences is not exhaustive, not all of these additional variants could be aligned with known splice sites when comparing to cloned genomic DNA sequences. Gene splicing typically follows a well conserved CAG:GTAAGT motif at the 5' splice site, where the splice cut will form before the underlined GT and is joined after a corresponding 3' AG. Sequences more similar to this exon/intron boundary are more faithfully spliced than more diverged sequences. In less than 1% of cases a GC can replace the GT with no discernable change to splicing efficiency (Burset et al., 2000). A combination of GT:AG or GC:AG splice sites could be used to explain some, but not all, of the additional variation observed in cDNA samples compared to gDNA samples. Recently it has been shown that during the cDNA synthesis process *in vitro*, RT enzymes are capable of "skipping" sections of RNA, in a homology dependent manner, to produce a truncated cDNA that resembles a typical gene splicing event (Cocquet et al., 2006, Houseley and Tollervy, 2010). An experiment was carried out to determine if this was occurring with 444 sequences. pGEM T-Easy plasmid containing a single 444-A gene cloned from cDNA was linearised, and RNA was transcribed *in vitro*. This single RNA species, as confirmed on denaturing gel electrophoresis (Figure 4-16iv), was used in a cDNA synthesis reaction. When gene specific PCR was carried out on this cDNA, multiple products were amplified (Figure 4-16i). No such banding pattern could be amplified by PCR on the vector used for RNA synthesis (Figure 4-16i). Taken together, the results suggest that these alternate transcripts were generated *in vitro*, by the cDNA synthesis process, and are temperature and PCR independent. This was also true for subfamily-C, and not true for a control gene (GpCys), treated in exactly the same manner. Clones of each were sequenced to ensure this was not a result of mispriming. Using this information, all cDNA clones were removed from the phylogenetic analysis, as it is impossible to determine if they are real, or non-canonical apparent splice events.

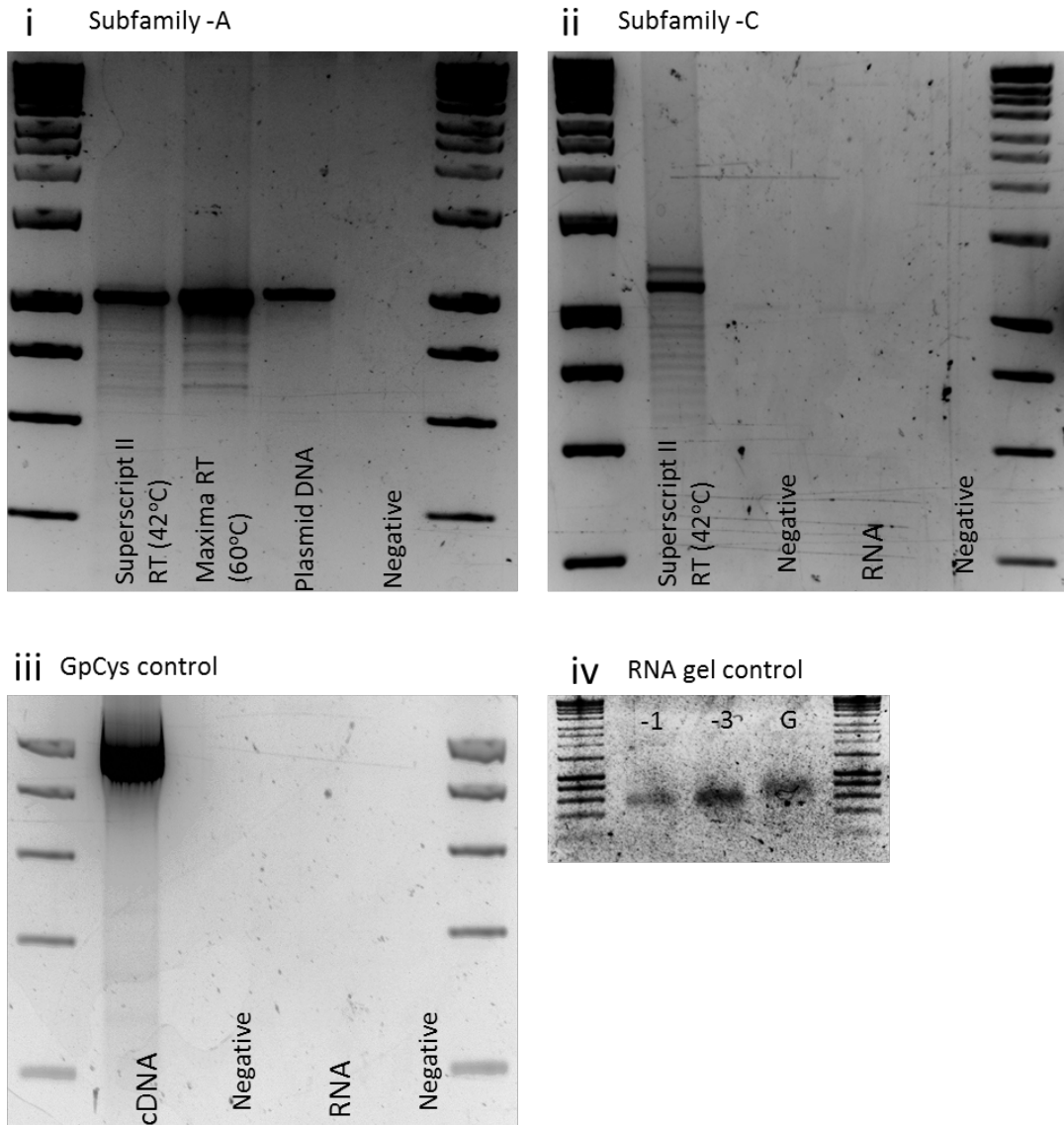


Figure 4-16 **Non-canonical apparent splicing of 444 genes.** A single RNA species was synthesised *in vitro*. cDNA was reverse transcribed from this RNA followed by PCR analysis. **i)** For subfamily-A, multiple amplification products can be seen resulting from PCR on cDNA produced with either SuperScript II or Maxima RT. No such banding pattern is observed for PCR on a single gene from plasmid DNA. **ii)** Similar additional banding patterns can be seen for subfamily-C. **iii)** No such band pattern can be seen for a non-tandem repeat containing control GpCys. **iv)** Denaturing RNA gel confirms a single RNA species in each case. Additional amplification products are therefore introduced as a result of the reverse transcriptase enzymes.

4.4.5.5. **444 gene family tree and transcript structure**

After removal of all cDNA sequences for this analysis, all cloned genomic sequences were aligned using MUSCLE v3.8.3.1, and a maximum likelihood phylogeny generated using 100 bootstrap iterations in TOPALi v2.5. Figure 4-17 show that subfamilies A, B and C are grouped into three distinct clades. Figure 4-18 shows an alignment of all genomic 444 sequences within subfamilies, and highlights the variable regions. Using a combination of genomic DNA, cloned sequences, and cDNA cloned sequences, a schematic of the transcript structure was created (Figure 4-19). All genes share a >95% conserved region at both the N and C terminus of the transcript. Within each of these regions there is one faithfully spliced intron with GT:AG splice sites. The intronic sequence within these regions, and the 3' UTRs, are conserved within subfamilies, and are different between subfamilies. No instances have been found, in over 200 clones, of introns either being absent in gDNA clones or being present in cDNA clones. For subfamilies A and C the middle region of the genes contains a region with a variable number of tandem repeats, unique to each group. Group B does not have this region, instead its second intron is slightly longer.

4.4.5.6. **Tandem repeat analysis**

In all cases, the entire tandem repeat region consists of a single open reading frame. For subfamily-A and -C only, large variation is observed in the number, sequence and order of tandem repeats in the deduced amino acid sequences corresponding to different genomic sequences. Within this region, subfamily-A genes contain four motifs, two of which are present as tandem repeats with complex organisations (Figure 4-20). The most common motif (1.1) consists of 6 amino acids, the first two are variable followed by a highly conserved RGGG. This glycine rich motif is present on average 12 times per gene, although this varies greatly. The second motif (1.2) consists of 5 amino acids, with a variable first position followed by conserved DRGD. This motif is present approximately 4 times per gene. The other two motifs (1.3 and 1.4) are usually present no more than once, often at the start or end of the tandem repeat domain. Including all variable regions of all motifs in subfamily-A, 20 unique amino acid sequences are encoded by 43 unique nucleotide sequences.

The subfamily-B tandem repeat region always contains four proline rich tandem repeats, comprised of three different motifs (2.1, 2.2 and 2.3). Motif 2.1 consists of 7 amino acids in the sequence EKPPPKY. Motif 2.2 is identical, except for the inclusion of an additional proline in the proline rich repeat (EKPPPPKY) and has no

variation in sequence. Motif 2.3 consists of 9 amino acids, also has no sequence variation and is usually present only once. If motif 2.1 is present in positions 1 and 3, the position 3 motifs are more similar at the nucleotide level to other 2.1 motifs in position 3, than they are to 2.1 motifs in position 1, even if they are identical at the amino acid level. The four subfamily-B tandem repeat variants are encoded by 10 unique nucleotide sequences. Interestingly, both subfamily-A and -C also contain the first and last of this type of tandem repeat.

Subfamily-C tandem repeat regions contain two lysine and glutamic acid rich motifs (3.1 and 3.2) the first of which (3.1) occurs in tandem repeats on average 15 times, although this varies greatly. Motif 3.1 consists of 11 amino acids with highly conserved amino acids in position 2, 4, 5, 7, 10 and 11. Motif 3.2 consists of 17 amino acids and is always present as a single copy in the final position of the tandem repeat domain of subfamily-C. Including all variable regions of all motifs in subfamily-C, 27 unique amino acid sequences are encoded by 42 unique nucleotide sequences.

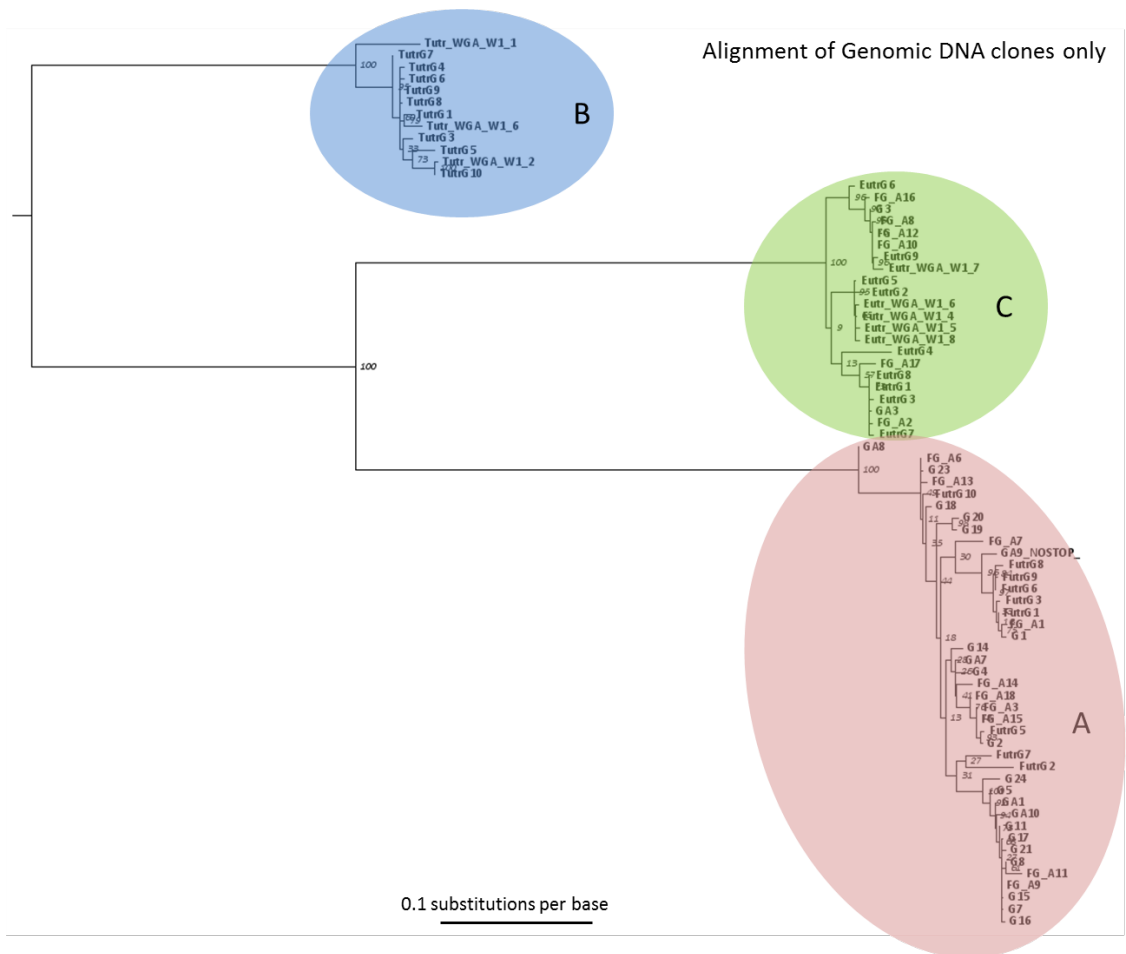


Figure 4-17 **Phylogenetic tree of genomic 444 DNA sequences.** Maximum likelihood phylogeny based on a protein alignment of all 444 sequences cloned from genomic DNA. Node labels represent boot strap support values for 100 iterations.

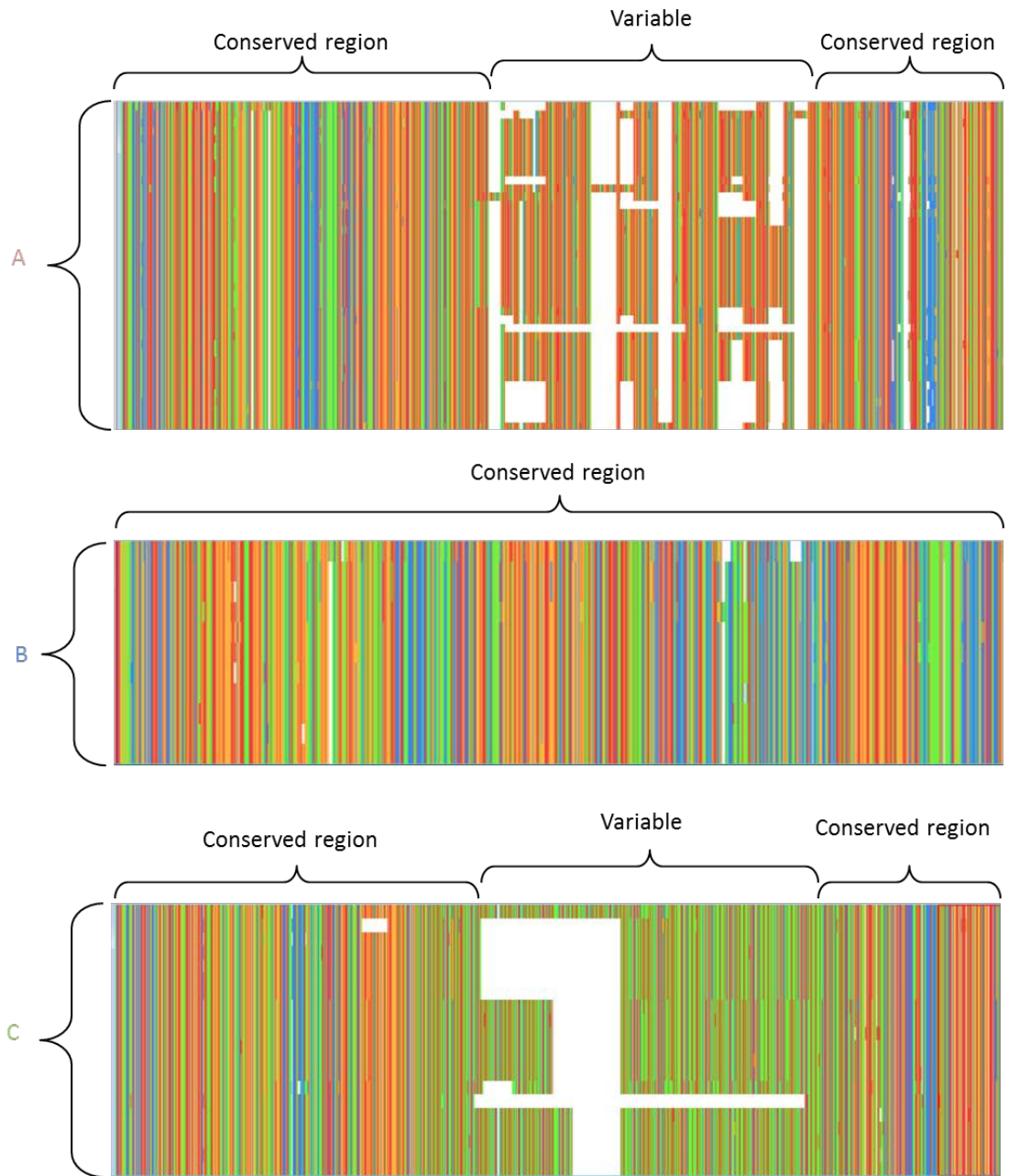


Figure 4-18 **Within subfamily deduced amino acid sequence from 444 genomic sequences.** For subfamily A and C the variable region in the middle is flanked by two conserved regions. There is no such variation in subfamily B genes.

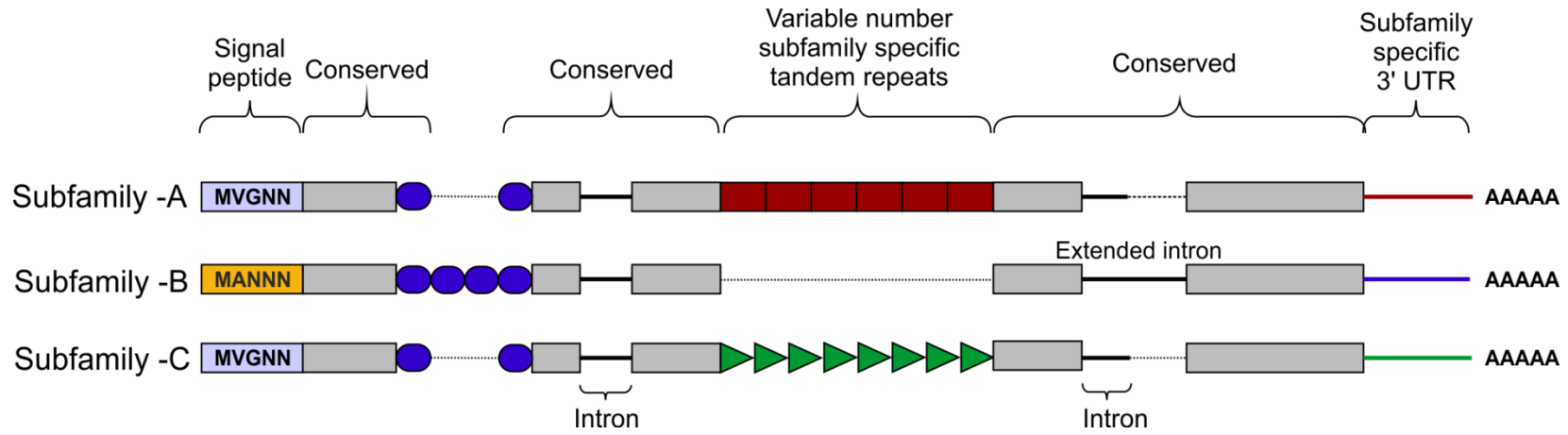


Figure 4-19 **Schematic representation of 444 gene family transcript structure.** All 444 genes identified to date encode a secretion signal followed by two highly conserved regions flanking a region of variable number tandem repeats, and a subfamily specific 3' UTR. Conserved regions of *Gp-hyp* genes share > 95 % identity between subfamilies. Subfamily-2 has a slightly different signal peptide (highlighted). 3' UTRs are identical within subfamilies, irrespective of number of tandem repeats. Members of all subfamilies can be amplified using the F-ALL and R-ALL primer pair, subfamilies can be distinguished by PCR using a combination of signal peptide and 3' UTR primers. Solid lines represent introns, while dashed lines represent alignments.










Subfamily-A	Number	Position	Consensus
1.1	Average 12	N/A	
1.2	Average 4	N/A	
1.3	Usually 1	1 or 2	
1.4	Usually 1	Usually N-1	
Subfamily-B			
2.1	Usually 2	1 and 3	
2.2	Usually 1	2	
2.3	Usually 1	4	
Subfamily-C			
3.1	Average 15	N/A	
3.2	Always 1	Last	

Figure 4-20 **Tandem repeat analysis of 444 genes.** Deduced amino acid sequences from the tandem repeat regions of genomic DNA sequences are shown. For each subfamily, the motifs present in the tandem repeat domain are shown along with their average copy number, position and consensus sequence. For 444-A and -C no pattern has emerged to the positions/organisations of the highest frequency motifs. Motif 1.1 has the greatest variation in tandem repeat number, although motif 3.1 has less variation in the number of tandem repeats it has the greatest variation in the motif consensus sequence.

4.4.5.7. **444 domain structure and modular organisation**

Approximately 80 unique genomic DNA 444 sequences were identified. The variation is not only due to the variable number of tandem repeats. In addition, within the highly conserved regions, several non-synonymous sequences were identified. Figure 4-21 shows a schematic representation of these domains, and the various combinations and modularity identified to date. Each combination of domains can be referred to as a “type”. Types are not subfamily-specific. Despite using the subfamily-specific 3' UTR primers, variable numbers of tandem repeats, and different types can still be identified. Moreover, sequences encoding identical tandem repeat regions can have different amino acid sequences at every domain locus, and similarly identical types at every domain locus can have different tandem repeat regions.

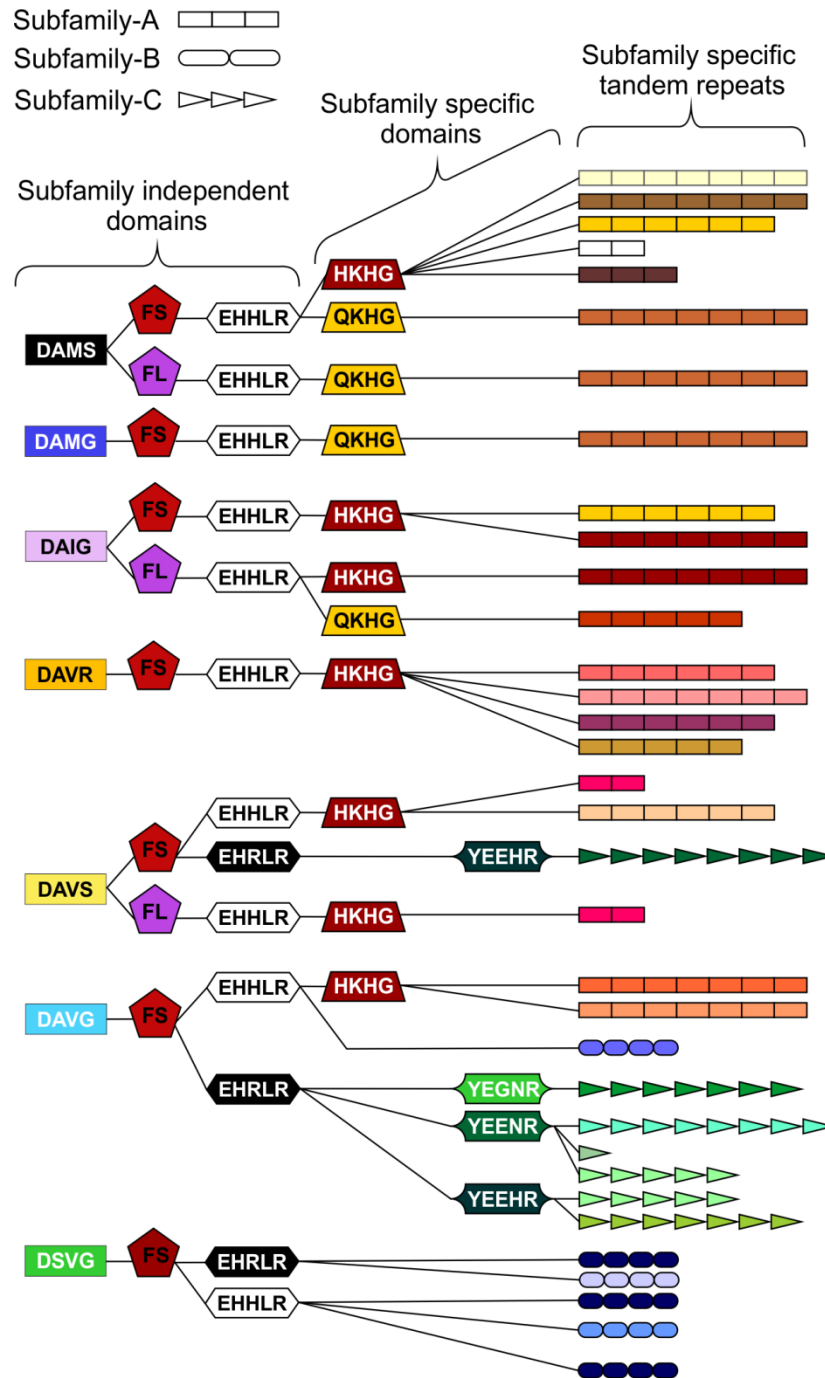


Figure 4-21 **Schematic representation of 444 domain organisations.** Within the conserved regions, non-synonymous SNPs result in various domains. Combinations of domains give rise to “types”. Certain types are subfamily specific, whereas others are not. For each column only, the same colour indicates the same sequence. By organising the sequences using types as indicators (not in the order they appear in the genes), 444 genes do not group by subfamily. The three different subfamily specific tandem repeats are indicated in the final column as different shapes, where the size roughly correlates to number of repeats. Sequences encoding similar tandem repeat regions can have different domains at every locus, and similarly sequences encoding identical domains at every locus can have different tandem repeat regions.

4.5. Discussion

Using a pipeline to identify genes significantly up-regulated during the biotrophic stages of infection we identified a range of putative feeding tube/plug candidate genes for further study (Chapter 5).

4.5.1. Quality of genome assembly

Although only a draft *G. pallida* genome sequence is currently available (Cotton et al., 2014), this is adequate for large scale identification of candidate genes. Combined with the life stage specific transcript data, the genome represents a tremendous resource for studying plant-nematode interactions. The draft genome contains 85% of the Core-Eukaryotic Genes (CEGs) and is thus predicted to contain 85% of all *G. pallida* genes, totalling 16,419 gene models. Despite a manual annotation step being used to train the gene prediction software, there will inevitably be gene predictions that need to be confirmed by cloning. With the exception of 1078 and 444, primers designed from predicted genes produced amplicons from cDNA that matched the predicted sequences. In the case of the 1078 sequence misprediction of an intron meant that the cloned sequence from cDNA contained a putative transmembrane domain not present in the genome sequence (Figure 4-5 and -6), and was therefore discarded from further analysis. In addition, the 444 gene family is heavily underrepresented in the assembled genome as discussed in more detail below.

4.5.2. Convergence of identification strategies

Two main approaches have been used to identify genes that may be involved in the interaction between nematode and plant. Both of these centre around the oesophageal gland cells described in Chapter 1. Proteins produced by these gland cells are secreted into the plant and are hypothesised to carry out a range of functions such as induction and maintenance of the feeding site and suppression of host defences (Haegeman et al., 2012, Hewezi and Baum, 2013). The first, and to date arguably the most effective approach, has been to identify all genes expressed in the gland cells by micro aspiration of cell contents, RNA extraction and subsequent RNAseq (Hewezi and Baum, 2013). In addition, cDNA-AFLPs have been used to identify candidate effectors from *G. rostochiensis* (Qin et al., 2000) in a laboratory-based approach that mirrors the *in silico* approach adopted here. An inclusive approach was taken to identify all genes that matched a set of reasonable assumptions based on the expected characteristics of feeding structure genes. The pipeline described here, and the sequencing of gland cell libraries, converge on the identification of the 448/4D06 gene family. This convergence gives some

confidence in the approach, however, it is surprising that more examples of previously described effectors were not identified. In particular, the inclusive approach taken here was expected to include numerous false positives (i.e. secreted genes up-regulated at feeding stages that are not effectors). The sequences identified in the analysis presented here were subsequently analysed by *in situ* hybridisation in order to determine if they are expressed in nematode tissues that would allow them to be secreted into the plant (Chapter 5).

4.5.3. Non-canonical splicing.

Reverse Transcriptase (RT) enzymes synthesise complementary DNA (cDNA) from an RNA template. The ability of RT enzymes to faithfully convert RNA to cDNA is the keystone of many molecular biology applications. Recently it has been shown that during the cDNA synthesis process *in vitro*, RT enzymes are capable of “skipping” sections of RNA, in a homology dependent manner, to produce a truncated cDNA that resembles a typical gene splicing event (Cocquet et al., 2006, Houseley and Tollervey, 2010). RT enzymes used for cDNA synthesis are derived from retroviruses, either Moloney Murine Leukemia Virus (MMLV) or Avian Myeloblastosis Virus (AMV). This template switching ability is a remnant of their activity in viruses as successful retroviral replication requires at least two template switching events between areas of high homology to one another (Gilboa et al., 1979).

More sequence variation was seen when cloning 444 genes from cDNA than from gDNA and this was initially assumed to be due to alternative or aberrant splicing. Gene splicing typically follows a well conserved CAG:GTAAAGT motif at the 5' splice site, where the splice cut will form before the underlined GT and is joined after a corresponding 3' AG. Sequences more similar to this exon/intron boundary are more faithfully spliced than more diverged sequences. These are known as GT:AG splice sites. In less than 1% of cases a GC can replace the GT with no discernable change to splicing efficiency (Burset et al., 2000). A combination of GT:AG or GC:AG splice sites could be used to explain some, but not all, of the additional variation observed in cDNA samples compared to gDNA samples.

In the case of the 444 genes we have demonstrated that the RT enzyme is capable of introducing “non-canonical” splice events, that are both PCR and temperature independent (Figure 4-16). The “non-canonical” splice events observed are always present in the tandem repeat region in the middle of the gene/transcript. As discussed, RT can skip sections by switching templates from one section of gene to another where multiple regions with high similarity are present. The large number of

almost identical tandem repeats present in 444 genes introduces several possible permutations of skipping events, that result in the observed additional variation.

These factors made it impossible to definitively identify genuine alternative splicing events. Therefore, where possible, all subsequent experiments describing the variation in 444 genes were carried out at the DNA level (Figure 4-17). These results have profound implications for all experiments that use RT enzymes, in particular the rapidly advancing field of RNAseq.

4.5.4. The complex 444 multi-gene family

Despite the presence of just three complete 444 sequences in the assembled genome and transcriptome sequences of *G. pallida* (one corresponding to each subfamily), the complexity of the 444 gene family was identified by conventional PCR and sequencing of individual clones. In addition, each time more clones were sequenced more unique sequences were identified suggesting that this list of ~80 unique sequences is far from exhaustive and that the full complexity of the gene family is yet to be catalogued. The absence of the full gene family in the *G. pallida* genome assembly, and the presence of two fragmented genes in poly-N regions, highlights a limitation of sequencing and assembly of short reads, generated from sequencing a population of non-genetically identical individuals. All 444 sequences, irrespective of subfamily, share stretches of 410 and 94 nucleotides with >90% identity at the 5' and 3' ends respectively which may underlie the difficulty in assembly.

All unique genomic 444 sequences identified can be readily assigned to one of three subfamilies primarily, although not only, based on the amino acid sequence of the tandem repeat region. Numbers of subfamily-specific tandem repeats range greatly. The 444 tandem repeats are short, some of which contain highly conserved regions directly preceded by variable di-residues. Tandem repeats of subfamilies -1, -2, and -3 contain conserved glycine, proline and lysine residues. Glycine residues often create flexible linkers between domains. These data may suggest that variable residues interspersed by highly conserved linker regions may play a role in ligand binding or multimerisation. In addition, several non-synonymous variations were identified in 444 sequence, within the highly conserved regions, that do not group by subfamily. Various combinations of these domains are described here as "types". The amino acid changes of 444 domains are usually physiochemically similar, and may be structurally superficial. The "type" structure may therefore reflect evolutionary origins and rearrangements rather than function, whereas the different subfamily tandem repeats, and the different number of

tandem repeats within subfamilies, may reflect different functions. It has been noted however, that subtle amino acid changes that should conserve physiochemical properties can have an impact on effector function (Whisson et al., 2007).

4.5.5. Copy number variation of 444 sequences between individuals of the same population

The 444 gene family shows unparalleled genomic diversity between individuals: no two nematodes tested had the same complement of 444-A or 444-C sequences present. Nematodes differed in the length, number, and even presence/absence of gene subfamilies. Due to the nature of the draft genome assembly of *G. pallida* we are unable to confirm that sequences are paralogues, although the fact that individual nematodes differ in the size, and particularly number, of sequences within subfamilies suggests this is the case. The number and variability of 444 genes suggests that this gene family is under strong selection pressure. Gene expansions of cytochrome P450 genes have been described in *Anopheles* species (Scott et al., 1994), where estimates are of approximately 30 to 40 genes (Ranson et al., 2002). This is a good example as CYP450 genes, and in particular their copy number, have been linked to the extremely high selection pressure provided by resistance to DDT (Wondji et al., 2009, Djouaka et al., 2008, Ranson et al., 2002). It is possible that the expansion of the 444 gene family in *G. pallida* may reflect a similarly high selection pressure from the host. Secreted components are the pathogen factors that are recognised by both pattern recognition receptors and resistance gene products. Diversity in 444 sequences may reflect the need to evade recognition in order to avoid detection.

As discussed above the genome sequence of *G. pallida* was challenging to assemble (Cotton et al., 2014). This may also be due to the inherent genetic variation that we have demonstrated between individuals of the same population, even infecting the same plant. Genetic diversity in plant parasitic nematodes has never been studied on such a scale, the most relevant works being between (as opposed to within) populations (Alenda et al., 2013, Plantard et al., 2008, Thiery et al., 1997, Blok et al., 1998). Interestingly it has been suggested that genetic variability observed at the scale of a field or even of a region is already observed at the scale of a single plant within a field (Plantard et al., 2008).

4.5.6. Inter-specific differences

In the case of the 444 genes we have found that a UK population of *G. pallida* has considerably more variation than *G. rostochiensis*. Due to the highly conserved

regions of 444 genes, and the genetic similarity of *G. pallida* and *G. rostochiensis*, it is expected that the difference observed is not an artefact of primer affinity. It has been suggested that the UK populations of *G. pallida* are from a broader genetic introduction than *G. rostochiensis*. If the diversity seen between individuals of *G. pallida* for the 444 genes is true on a wider genomic scale, this may explain the difficulty in identifying a broad spectrum resistance against *G. pallida* compared to *G. rostochiensis* (Barone et al., 1990). Recent technological advances in sequencing the transcriptome/genome of single nematodes may identify wider differences between individuals (Sasagawa et al., 2013). In addition, previously published works identified 448 genes from gland cell libraries of other cyst nematode species. Identifying the candidate genes, characterised here, in other related plant parasitic nematodes may inform function.

Summary

- Starting from entire genome and transcriptome sequence, Five candidate genes have been identified that match the expected characteristics of a feeding structure gene and require further characterisation
- The five candidates include two large gene families, one of which (444) is novel
- 444 genes show high sequence diversity within a population at the level of the individual
- Non-canonical splicing of 444 genes makes identifying bonafide cDNA clones challenging.

5. Characterisation of candidate feeding structure genes

5.1. Introduction

Various large scale identification strategies have been successful in identifying effectors as discussed in Chapter 4. However, all methods of identification, including those involving direct gland cell library sequencing, require further characterisation of identified genes to support function.

5.1.1. Validating effector identification

It is generally assumed that genes specifically expressed at certain life stages have a function that is important at those stages. The effector discovery pipeline described in Chapter 4 is based on this assumption. However, the sequences identified by this pipeline may have numerous different functions in the biotrophic stages of the nematodes and only a subset of these sequences will represent genuine effectors. It is therefore essential to confirm the spatial as well as temporal expression patterns of candidate effectors. Genes encoding genuine effectors will be expressed in a tissue from which proteins can be secreted into the plant. To identify the spatial expression pattern of genes of interest, a technique known as *in situ* hybridisation is carried out (de Boer et al., 1998). This involves the homology-dependent hybridisation of a labelled DNA probe to endogenous mRNA in fixed tissue samples. Histochemical detection of the DNA probe allows visualisation of the specific tissue in which the gene is expressed (de Boer et al., 1998).

5.1.2. Structures with the capacity to secrete proteins into the host

As introduced in Chapter 1, cyst nematodes contain a number of tissues with the capacity to secrete effectors into the host. The most obvious of these are the dorsal and subventral gland cells, many studies have identified proteins secreted from these cells into plant tissues via the stylet. However, other nematode tissues may also represent sources of effectors. The amphids are the primary sense organs and their function was considered to be confined to the migratory stages of sedentary endoparasitic nematodes. However, structural changes occur in the amphids during the transition from migratory juvenile to sedentary feeding stages (Jones et al., 1994, Perry, 1996), suggesting distinct roles at each stage. In addition, two previous studies showed that the feeding plugs of cyst nematodes are continuous with the amphid openings of sedentary females (Endo, 1978, Sobczak et al., 1999),

one of which concluded that the feeding plug originates from the amphidial canal (Endo, 1978).

A gene encoding a protein secreted from the amphids of *Meloidogyne* species has been identified (*map-1*) which is a member of a family of genes, with a function in virulence, restricted to root knot nematodes. These proteins contain numerous tandem repeats, the number and arrangement of which correlates with nematode (a)virulence (Castagnone-Sereno et al., 2009). Although this was initially hypothesised to be an expansin-like protein (Tomalova et al., 2012) and thought to be secreted from the amphids (Semblat et al., 2001), recent studies suggest it, or a similar gene, may encode a plant peptide hormone mimic and that the gene may be expressed in the dorsal gland cell (Rutter et al., 2014).

The nematode surface is also a potential source of proteins that come into direct contact with the host. A gene encoding a secreted glutathione peroxidase is expressed in the hypodermis of the cyst nematode *Globodera rostochiensis* (Jones et al., 2004). The nematode protein is functional, and is able to break down a range of hydroperoxides *in vitro* (Jones et al., 2004). It is therefore possible that this glutathione peroxidase is involved in counteracting host reactive oxygen species produced as a defence response during infection. Similarly, a peroxiredoxin that metabolises hydrogen peroxide has been identified on the surface of *G. rostochiensis* (Robertson et al., 2000). Finally, a Cellulose Binding Module2 (CBM2)-bearing protein accumulates near the vagina of gravid female *M. incognita* 55 days post infection *in planta* (Vieira et al., 2011) and may originate from the rectal glands. It is hypothesised that this protein may alter host cell walls to facilitate egg laying (Vieira et al., 2011). Nematodes therefore have a variety of tissues that can produce proteins with a potential role in host-parasite interactions.

5.1.3. Functional characterisation of effectors

Having confirmed putative-effector expression in a tissue with the capacity to secrete proteins, numerous functional studies can be carried out to elucidate effector function.

5.1.3.1. Localisation

Expression in secretory tissues is important to determine validity of putative effectors. In addition, the *in planta* localisation of the encoded proteins can inform function. Methods to localise nematode effectors *in planta* include direct detection of secreted proteins *in situ* in fixed samples using corresponding antibodies, or

heterologous over-expression of effectors in translational fusions with fluorescent proteins. These methods have revealed both the site of delivery, and the potential site of function of a range of effectors. Such studies can also be used to help interpret the outputs of yeast-two-hybrid screens.

The first nematode effector detected in the cytoplasm of the feeding site was a CLE peptide. Subsequently, numerous studies have characterised nematode parasitism proteins apparently localised to the nucleus of plant cells (Jones et al., 2009, Elling et al., 2007, Jaouannet et al., 2012). In the case of the *M. incognita* effector EFF, direct delivery to the cytoplasm of the giant cell from the gland cells was observed by immunolocalisation. Interestingly, this antibody targeted to the EFF gene co-localised in plant cells with stained nuclei (Jaouannet et al., 2012). Mi-EFF1 does encode a canonical Nuclear Localisation Signal (NLS) that is therefore functional *in planta*. However, presence or absence of canonical NLSs does not always correlate with nuclear localisation (Jones et al., 2009). Interestingly, a ubiquitin protein with a short C-terminal extension appears to be cleaved *in planta*, where the ubiquitin domain is localised in the cytoplasm and C-terminal extension alone is transported to the nucleus (Tytgat et al., 2004). Nuclear localisation may suggest a functional role in transcription regulation that is directly relevant to either feeding site developmental re-programming or suppression of host defences.

The studies described above relate to effectors that are introduced into the cytoplasm of a host cell. However, it has become apparent that the apoplast is also an important recipient compartment for nematode effectors (Vieira et al., 2011). Nematodes have evolved several different mechanisms to introduce proteins into the apoplast, from numerous different secretory tissues. The plant CLE peptide mimics, described in Chapter 1, are secreted into the feeding site from the dorsal gland cell, before being transported to the apoplast by the host cell's own machinery in order to fulfil their function. This export is controlled by a short non-canonical secretion signal N-terminal to the plant-peptide hormone domain mimic and is unique to nematode CLE peptides (Wang et al., 2010). Other effectors, such as the MAP1 proteins, are secreted from the amphids directly into the apoplast by parasitic stage *M. incognita* (Vieira et al., 2011). In addition, several other effectors are secreted into the apoplast from the subventral gland cells. These include an aspartyl protease-like protein (Mi-ASP2), as well as a range of plant cell wall degrading and modifying enzymes secreted during migration (Vieira et al., 2011, Haegeman et al., 2011).

5.1.3.2. *In planta* interactions

The functional role of an effector can be determined by identifying the host proteins that it targets. This is often carried out by Yeast-Two-Hybrid (Y2H) analysis against a host cDNA library. Despite the inherent lack of biological context (with interactions tested in yeast rather than plants), and its limitation to pairwise interactions, Y2H analyses have significantly advanced knowledge of nematode effector function. Various binding partners have been identified for nematode effectors, as discussed in Chapter 1, some of which have putative roles in virulence. For example, a secreted SPRY domain-containing protein (SPRYSEC) from *G. rostochiensis* interacts with an intracellular Coiled Coil (CC) -Nucleotide Binding (NB) – Leucine Rich Repeat (LRR) receptor (CC-NB-LRR). Interestingly, despite the interaction with the LRR domain, this interaction does not result in a resistance response, suggesting that the SPRYSEC is not recognised by the CC-NB-LRR but rather perturbs its endogenous function (Rehman et al., 2009, Postma et al., 2012).

5.1.3.3. Loss of function knockdown

Despite numerous attempts, transformation of cyst nematodes has not yet been achieved. However, gene knockout studies can be performed using RNA interference (RNAi). RNAi is a conserved mechanism in eukaryotes in which double stranded RNA (dsRNA) causes post-transcriptional gene silencing in a sequence-dependent manner. dsRNA corresponding to the sequence of interest, when taken up by the nematode, is used by endogenous RNAi machinery resulting in gene knockdown. RNAi has been used to pheno-copy gene knockout studies since its description in *Caenorhabditis elegans* (Fire et al., 1998). For *C. elegans* dsRNA is most commonly delivered by feeding nematodes on bacteria expressing an inverted repeat or “hair-pin” construct. However, as the majority of plant-parasitic nematodes are obligate biotrophs, and therefore only feed from living plant tissue, dsRNA has to be introduced using a different strategy (Lilley et al., 2007, Lilley et al., 2012). dsRNA targeting a gene of interest can be synthesised *in vitro* and resuspended in a buffer containing octopamine. This neurotransmitter stimulates feeding behaviour and thus facilitates uptake. Eggs or J2 nematodes are soaked in the dsRNA solution for between 4 hours to 3 days (Rosso et al., 2009). While this process has been useful for studies on some plant parasitic nematodes it is clear that there is marked variability in terms of the susceptibility of nematode species to RNAi (Dalzell et al., 2011, Maule et al., 2011). Alternatively the gene of interest can be cloned in an inverted repeat and transformed into plants. dsRNA is then produced by the plant transcription machinery and can be delivered to nematodes while they feed (Huang et al., 2006, Ibrahim et al., 2011). This is often referred to as Host

induced Gene Silencing (HIGS). Scoring phenotypes is difficult for an obligate parasite but phenotypes range from reduced hatching efficiency (Fanelli et al., 2005), changes in J2 motility (Kimber et al., 2007), changes to sex ratio (Urwin et al., 2002) and female fecundity (Bakhetia et al., 2005).

5.2. Aims

To further characterise candidate feeding structure genes of *G. pallida* identified in Chapter 4

- Identify those candidates with expression in a tissue capable of secretion into the host
- Determine RNAi knockdown phenotype of candidates
- Localise the proteins encoded by candidate genes in the host.

5.3. Materials and methods

5.3.1. *In situ* hybridisation

A two-step process is required to generate *in situ* hybridisation probes. Initially a short double stranded DNA (dsDNA) template is amplified by PCR from cDNA or a cDNA clone of a gene of interest. Secondly, this dsDNA is used as a template in an asymmetric (single primer) PCR to incorporate digoxigenin labelled dUTP (derived from a DIG-DNA labelling mix, Roche) into single stranded DNA probes. In general, oligonucleotide primers were designed to amplify regions of between 130 – 250 base pairs from the corresponding cDNA clones (Summarised in Table 5-1). In the case of the 444 genes, *in situ* hybridisation probes were designed to target a region of 134 conserved nucleotides at the 3' end of the translated region of all 444 transcripts. In the case of the 448 gene family, a single gene was used to design an *in situ* probe (GPLIN_000950100) and the nucleotide identity to other 448 members was calculated using BLAST. PCR reactions to generate template fragments were carried out as described in Section 2.4.1.

Asymmetric PCR was carried out with the same primers that were used to amplify the template. Asymmetric PCR using either the reverse or forward primers only was carried out by incubating at 94 °C for 2 minutes followed by 35 cycles of 94 °C for 15 seconds, 55 °C annealing for 30 seconds, and 72 °C extension for 90 seconds. Incorporation of DIG-labelled dUTP was confirmed by an apparent increase in size on agarose gel electrophoresis compared to template dsDNA. In all experiments, the DNA probe generated from the reverse primer was complementary to the endogenous mRNA of interest and was therefore the positive probe. The DNA probe generated from the forward primer was identical in sequence to the endogenous mRNA, would therefore not be able to bind, and was used as a negative control. *In situ* hybridisation was carried out as described by de Boer *et al.* (de Boer *et al.*, 1998) with the following alterations. Cleaned potato roots heavily infected with 7-14 dpi feeding female *G. pallida* were lightly macerated using a bench top blender, and soaked in 10% formaldehyde for 3 days at room temperature. Fixed nematodes were collected by additional blending and subsequent sucrose gradient centrifugation (40% w/v). Feeding females were collected between 200 and 150 µm mesh sieves. The protocol was continued as described from the cutting stage (de Boer *et al.*, 1998).

Table 5-1 *In situ* hybridisation primers, annealing temperatures and expected amplicon size.

Primer Name	Sequence 5' - 3'	T _M (°C)	Expected product size (bp)
444_insitu_F	ACGGAGGTTATGACGAG	50	134
444_insitu_R	ATATTTGCATTCGCAAGC	53	
448_insitu_F	GTCAACTTCACGAACTCGGTGG	62	230
448_insitu_R	CCTGTCGCTTTTGC GGCCAAAT	69	
3453_insitu_F	CTTTTTCGCCATTGTCCTTCTGAT	54	200
3453_insitu_R	ACTTCTTTCCACCAGGGTAGT	52	
225_insitu_F	TGCGCTCTTCCGTTTTGATC	56	203
225_insitu_R	ACACTGTCCACACGCATTAC	56	
176_insitu_F	ATGCAATTATTCCTCTTTTTTG TG	58	139
176_insitu_R	TCACCAGCCTCCACCG	60	

5.3.2. *In planta* RNA interference

The effects of knocking out expression of candidate effectors were examined using a HIGS approach. Where possible inverted repeat constructs were designed so that the full length cDNA clones were included in the repeated sequence. In the case of the 444 gene family, both a full length inverted repeat of a single subfamily-A member, and an inverted repeat using only the conserved 3' end of the cDNA were created. For the 444, 176 and 3453 sequences, forward and reverse primers designed with restriction enzyme sites XhoI and KpnI respectively at the 5' end were used to amplify sequences in the sense orientation. A second set of forward and reverse primers with the restriction enzyme sites XbaI and HindIII added were used to amplify sequences in the anti-sense orientation. Correct amplification of both fragments for each gene was confirmed by sub cloning into pGEM-T Easy and subsequent Sanger sequencing. Sequences with no errors were cloned into the vector pHannibal (Wesley et al., 2001) under the control of a CaMV 35S promoter and OCS terminator using the relevant restriction enzymes and a typical restriction enzyme cloning procedure as described in section 2.4. The entire construct from promoter to terminator was cloned from pHannibal into the plant binary vector pART27 (Gleave, 1992) using SacI and SpeI. As a control, a full length GFP inverted repeat was created as described above using the oligonucleotide primers detailed in Table 5-2. In the case of the 448 gene family, a single construct was designed containing inverted repeats targeting three of the main groups of 448 genes based on a DNA alignment. Sense orientation fragments were amplified using oligonucleotide primers introducing XhoI - BamHI, BamHI - EcoRI and EcoRI - KpnI on the 5' and 3' ends of fragments 1, 2 and 3 respectively. This allowed all three fragments to be cloned in a tandem array in pHannibal. Antisense orientation fragments were amplified using oligonucleotide primers introducing XbaI - BamHI, BamHI - EcoRI and EcoRI - HindIII on the 5' and 3' ends of fragments 1, 2 and 3 respectively and cloned in a similar manner. The two tandem arrays were then cloned into pHannibal as above using the XhoI - KpnI and XbaI - HindIII pairs in sense and antisense orientation respectively. The entire cassette including promoter and terminator was cloned into pART27 as described above. All pART27 IR constructs were transformed into *Agrobacterium rhizogenes* strain R1000 as described in section 2.4.12.

A single colony of *A. rhizogenes* strain R1000 containing the relevant pART27 IR construct, or no-construct control, was incubated in 5 ml liquid Luria Bertani medium (containing 50 µg/ml rifampicin and 100 µg/ml kanamycin) overnight at 28 °C. Potato hairy root transformation was carried out by incubating 1 cm squares of *Solanum tuberosum* (cv 'Desirée') leaf material in 9 cm petri dishes containing

liquid MS20 medium (4.3 g/l Murashige and Skoog (with vitamins), 20 g/l sucrose, pH 5.3 - 5.6) containing 100 µl of *Agrobacterium* culture, for 3 days at room temperature. Leaf squares were dried on filter paper and placed on MS20 agar plates (2.4 g/l agar), containing 50 µg/ml kanamycin and 400 µg/ml cefotaxime. Roots originating from different locations on each leaf square were considered individual transformation events, and were removed and cultured on MS20 agar plates containing 50 µg/ml kanamycin. Expression of the inverted repeat constructs was confirmed by RNA extraction (RNeasy Plant Mini Kit, Qiagen), cDNA synthesis (SuperScript II Reverse Transcriptase, Invitrogen), and PCR using the relevant transgene-specific primers.

Table 5-2 **RNAi construct primers, expected annealing temperatures and product sizes.**

Primer Name	Sequence 5' - 3'	TM (°C)	Expected product size (bp)
444_RNAi_IR_F_xho	CTCGAG ATGGTCGGCAACAATTTG	56	793
444_RNAi_IR_R_kpn	GGTACC TTAATATTTGCATTCGCAAGC	56	
444_RNAi_IR_F_xba	TCTAGA ATGGTCGGCAACAATTTG	56	793
444_RNAi_sh_F_xba	TCTAGA ACGGAGGTTATGACGAG	50	
444_RNAi_sh_F_xho	CTCGAG ACGGAGGTTATGACGAG	50	134
444_RNAi_IR_R_hind	AAGCTT TTAATATTTGCATTCGCAAGC	56	
176_RNAi_IR_F_xho	CTCGAG ATGCAATTATTCCTCTTTTTTGTG	58	348
176_RNAi_IR_R_kpn	GGTACC TCACCAGCCTCCACCG	60	
176_RNAi_IR_F_xba	TCTAGA ATGCAATTATTCCTCTTTTTTGTG	58	348
176_RNAi_IR_R_hind	AAGCTT TCACCAGCCTCCACCG	60	
GFP_RNAi_IR_F_xho	CTCGAG ATGAGTAAAGGAGAAGAAGACTTTTC	53	717
GFP_RNAi_IR_R_kpn	GGTACC CTATTTGTATAGTTCATCCATGCC	55	
GFP_RNAi_IR_F_xba	GGTACC CTATTTGTATAGTTCATCCATGCC	53	717
GFP_RNAi_IR_R_hind	AAGCTT CTATTTGTATAGTTCATCCATGCC	55	
3453_RNAi_IR_F_xho	CTCGAG ATGATGACCCCTCTCCGCT	62	219
3453_RNAi_IR_R_kpn	GGTACC TCACTTCTTTCCACCAGGGT	59	
3453_RNAi_IR_F_xba	TCTAGA ATGATGACCCCTCTCCGCT	62	219
3453_RNAi_IR_R_hind	AAGCTT TCACTTCTTTCCACCAGGGT	59	
448_RNAi_frag1_F_Xho	CTCGAG ATCCATGTTGTCCTGGCAGT	60	206
448_RNAi_frag1_F_xba	TCTAGA ATCCATGTTGTCCTGGCAGT	60	
448_RNAi_frag1_R_Bam	GGATCC GAGTTGTGCGTCGATTTTCGT	61	
448_RNAi_frag2_F_Bam	GGATCC GTGTTGCCGGCAGTCAG	62	134
448_RNAi_frag2_R_Eco	GAATTC AGCTCAGTGCACCTTTTGACG	59	
448_RNAi_frag3_F_Eco	GAATTC AAGTGGCCGCACTGATGT	60	135
448_RNAi_frag3_R_Kpn	GGTACC AGCTGTTTCATTGCCTCCAA	59	
448_RNAi_frag3_R_Hind	AAGCTT AGCTGTTTCATTGCCTCCAA	59	

5.3.3. Acid fuchsin staining and nematode infection quantification

Hatched J2 of *G. pallida* were sterilised for 20 minutes in an appropriate volume of hexadecyltrimethylammonium bromide (CTAB, 0.5 mg/ml – Sigma) containing 0.1 % v/v chlorhexidine digluconate (Sigma) and 0.01% v/v Tween-20, followed by three washes in sterile tap water. J2s were suspended at a concentration of approximately 1 nematode per μl and 30 μl of suspension were pipetted onto each infection point. Three infection points were used per hairy root plate and 10 plates were used per line, with three independent lines for each inverted repeat, one line for the GFP control, and one line for the empty vector control. Two weeks after infection roots were stained by soaking in 1% sodium hypochlorite for 5 minutes, washing 3 times with tap water for 1.5 minutes, followed by boiling in 1 x acid fuchsin stain (0.035% acid fuchsin (w/v), 2.5% glacial acetic acid (v/v)) for 2 minutes. Stained roots were cleaned in acidified glycerol (1 drop glacial acetic acid per 100 ml glycerol) before total nematode numbers per root system were counted.

5.3.4. Yeast-two-hybrid screens

The 176 gene, and a representative from each 444 subfamily, were cloned into two yeast transformation vectors for pairwise interactions analysis. Coding sequences of interest were amplified after the predicted signal peptide cleavage point with the addition of 5' NdeI and 3' BamHI restriction enzyme sites. For 176 the forward and reverse oligonucleotide primers CATATGCTTATTTCCGCCTTATTTTTGG and GGATCCTCACCAGCCTCCACCG were used respectively. For 444 genes the same primer pair was used for all three subfamily members and sequences of interested were amplified from existing cloned and sequenced 444 genes in pGEM-T Easy vector. The forward and reverse oligonucleotide primers CATATG GGCTGCAAGCCAGGACC and GGATCCTTAATATTTGCATTCAACAAGCC were used respectively. Amplified products were sub-cloned into the pGEM T Easy vector as described in section 2.4. Clones were sequenced to confirm correct amplification. Each insert was cloned into both yeast transformation vectors pGBKT7 and pGADT7AD using the introduced restriction sites as described in section 2.4. Yeast competent cells were produced and transformed according to the Two-Hybrid Kit (ClonTech) manufacturer's instructions. Pairwise interactions were carried out by resuspending a single colony containing the relevant plasmid of interest in 100 μl of sterile H_2O . One microliter of each colony suspension to be mated was pipetted onto a YPDA plate to give rise to every pairwise interaction (10 g/l Bacto yeast extract, 20 g/l Bacto peptone, 20 g/l Glucose monohydrate, 40 mg/l Adenine hemisulfate and 20 g/l Bacto Agar). Colonies were allowed to grow for 20

hours at 30 °C on mating medium before being stamped using a velvet replica onto low stringency selection medium (SD – Leu (ClonTech)). After 5 days colonies that developed were replicated onto high stringency media. Interaction phenotypes on selection media were scored.

5.3.5. Antigenicity prediction

Antigenic regions of proteins of interest were predicted using Antigenicity Plot to identify those regions most amenable to raise antibodies against (Hopp and Woods, 1981). Both 448 and 444 are present as large gene families allowing the antigenicity plot to be viewed in the context of full gene family protein alignments. The 444 proteins were similar enough to identify a single antigenic peptide sequence (VVRVARGEYENKCPAGPAGDVGPPGPPGPSG) present at a high level of identity in all subfamilies. Subfamily-specific protein alignments were used to generate a consensus sequence for each subfamily. The first 20 amino acids of this peptide match with 100% identity to a consensus 444-A protein, the first 29 amino acids match with 100% identity to a consensus 444-B protein and the last 13 amino acids match with 100% identity to a consensus 444-C protein.

The 448 gene family encodes a more diverse set of proteins; as a result designing a single peptide which was present in all 448 genes was impossible. The antigenicity plot was viewed in the context of a whole gene family protein alignment again; in this case two peptides were identified each specific to a different group of 448 proteins in regions of high antigenicity (IVNFTNSVDTDDKQTLC and CKNANDVETAVKSALS). All peptides were synthesised and antibodies were subsequently produced by Yorkshire Bioscience Ltd (UK).

5.3.6. Embedding, sectioning and immunochemistry

Pieces of potato root, 14 days post infection with J2 of *G. pallida*, were fixed in 4% paraformaldehyde in PEM buffer (50 mM PIPES, 10 mM EGTA, 10 mM MgSO₄ pH 6.9) for 3 days at 4 °C. Fixed samples were dehydrated by incubating at 4 °C in the following ethanol series: 10% ethanol for 30 minutes, 20% ethanol for 30 minutes, 30% ethanol for 30 minutes, 50% ethanol for 30 minutes, 70% ethanol for 60 minutes, 90% ethanol for 60 minutes and 100% ethanol for 60 minutes. Following dehydration samples were embedded in LR White resin (Agar Scientific) using the following resin series (diluted in 100% ethanol): 10% for 30 minutes, 20% for 30 minutes, 30% for 30 minutes, 50% for 30 minutes, 70% for 60 minutes, 90% for 60 minutes, 100% for 60 minutes, 100% overnight, 100% for 8 hours and finally 100%

overnight. Each section of root was transferred to a single gelatine capsule containing 100% resin, and the resin was allowed to set at 37 °C for five days.

0.5 – 2 µm serial sections were produced from resin embedded tissues using an Ultracut Ultra-microtome (Reichert-Jung). Sections were collected onto multi-well slides treated with Vectabond (Vectabond Laboratories). Sections were blocked with 5% milk powder diluted in 1 x PBS for 30 minutes. Block solution was replaced with primary antibody diluted 1 in 5, or pre-immune serum at the same dilution, in 0.5% milk-PBS and incubated at room temperature for 2 hours. Wells were washed three times with PBS and incubated in FITC-conjugated anti-rabbit secondary antibody diluted 1 in 100 in PBS for 90 minutes at room temperature in the dark. Wells were washed three times in PBS prior to incubation in 0.4 x calcofluor white (Sigma, 1g/L) diluted in PBS for 5 minutes at room temperature in the dark. Wells were washed three times in PBS followed by addition of a single drop of antifade solution (Citifluor, Agar Scientific) and covering with a coverslip. Sections were imaged on a Leica Leitz DMRB binocular fluorescent microscope (Leica microsystems, Milton Keynes, UK) using Ultraviolet light for Calcofluor and the GFP filter for FITC.

5.4. Results

Several candidate feeding structure genes were identified in Chapter 4 from genome sequence of *G. pallida*. This chapter focuses on characterisation of these candidates in more detail.

5.4.1. Candidate 225

Candidate 225 was a single copy gene identified as being highly expressed during the feeding stages. A 200 bp region of candidate 225 was cloned from cDNA and used as a template for asymmetric PCR. Asymmetric PCR, using either the forward or the reverse primer only, was used to incorporate DIG-labelled dNTPs into a single stranded DNA probe. Single stranded probes were electrophoresed alongside the template to confirm DIG incorporation by apparent increase in molecular weight (Figure 5-1). In all experiments, the DNA probe generated from the reverse primer was complementary to the endogenous mRNA of interest and was therefore the positive, the non-complementary probe was used as a negative control.

For candidate 225, *in situ* hybridisation using the reverse probe showed constitutive dark staining across the majority of the body of the nematode (Figure 5-2). No such dark staining was observed with the forward (negative) probe. Although the digestive systems appeared slightly darker than the rest of the nematode body in negative controls, this was clearly distinct from the dark purple staining obtained with the reverse (positive) probe. From the staining patterns it was concluded that candidate 225 is not expressed in a tissue with the capacity to secrete proteins into the plant tissue it no longer conformed to the selection criteria and was not pursued further.

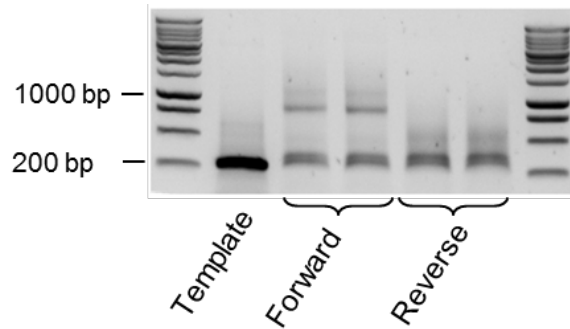


Figure 5-1 **Preparation of probes for candidate 225 *in situ* hybridisation.** Single stranded DNA probes, generated by either forward or reverse primers each in technical replicate, containing DIG-labelled dUTP are electrophoresed alongside template DNA. The incorporation of DIG-dUTP is confirmed by an apparent increase in molecular weight when compared to the template. Single stranded DNA electrophoresed on a non-denaturing gel often takes multiple conformations which migrate as separate bands.

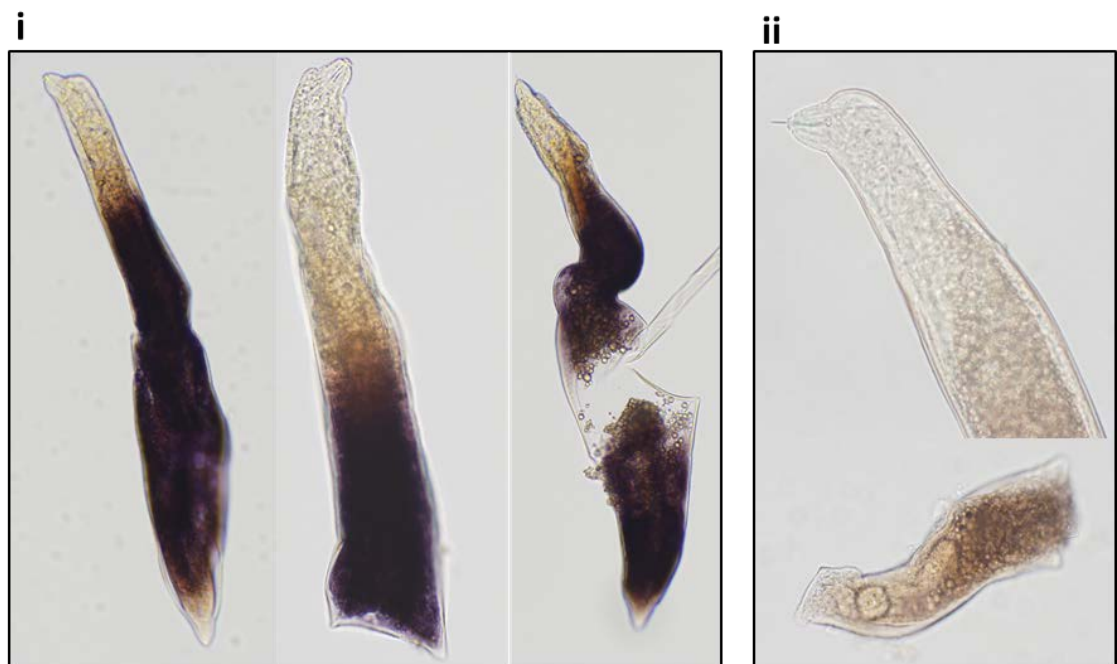


Figure 5-2 **Candidate 225 *in situ* hybridisation.** **i)** Dark staining using the reverse probe (positive) can be seen across the majority of the nematode body, in particular in the digestive system. **ii)** No such staining can be seen using the forward probe (negative).

5.4.2. Candidate 3453

A single 3453 gene was present in the final version of the *G. pallida* genome assembly. Constitutively high expression was observed in all feeding stages.

5.4.2.1. Candidate 3453 *in situ* hybridisation

In situ hybridisation probes were generated in a similar manner to that previously described, and electrophoresed alongside template to confirm the incorporation of DIG-labelled dNTPs (Figure 5-3). *In situ* hybridisation using the reverse probe showed dark staining in a paired structure posterior to the metacarpal bulb (Figure 5-4). Comparison with a schematic diagram showed that this staining pattern is consistent with the position of either the nerve ring, or the two most posterior amphidial sensory neurons associated. No similar staining pattern was observed with the negative control (Figure 5-4ii).

5.4.2.2. Candidate 3453 *in planta* RNA interference

The full length 3453 coding sequence was cloned in an inverted repeat (IR) arrangement using the pHannibal/pART plasmid system (as described in section 5.3.2). A “wild type” hairy root line (no transgene control) and a GFP inverted repeat construct were generated and tested alongside each candidate. Stable transformations of potato hairy roots were carried out with each construct. Figure 5-5 shows a typical result following transformation; transgenic hairy roots can be seen emerging from calli of *Solanum tuberosum* leaf material incubated on kanamycin selection media. Due to the proximity of calli to one another, all roots originating from the same region (red circle – Figure 5-5) were considered to be from the same transformation event. A single length of root was taken from each circle onto a separate plate to increase biomass prior to infection trials (Figure 5-5ii).

Several lines were generated for each construct. Expression of the IR construct was confirmed, and relatively high expressing lines were identified by semi-quantitative RT-PCR (Figure 5-6). For each line, relative intensity of the gene specific PCR amplicon was compared to that of the gene encoding elongation factor 1 α . The four most highly expressing candidate 3453-IR lines, the highest expressing GFP-IR line, and a growth phenotype matched wild type line were used in an infection trial. The same GFP-IR and wild type lines were used as controls for all future *in planta* RNAi experiments. For each line, 10 replicates were infected with nematodes as described in section 3.4.2.7. Although a difference in number of nematodes per root system was seen between all transgenic lines and the wild type, no difference was seen between the 3453 inverted repeat lines and the GFP

inverted repeat lines (Figure 5-7). As it was unclear whether candidate 3453 is expressed in a tissue with the capacity to secrete proteins into the plant, and has no observable RNAi phenotype compared to the control, it was not explored further.

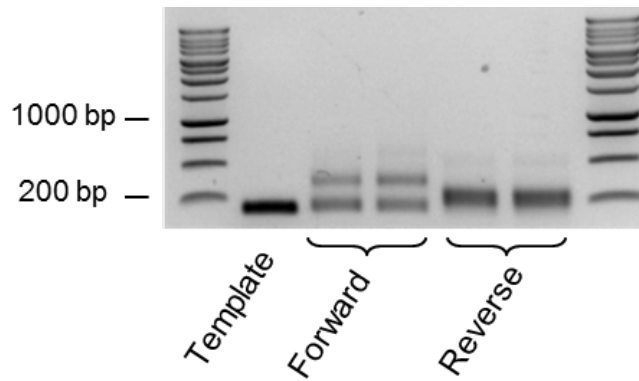


Figure 5-3 **Candidate 3453 *in situ* hybridisation probe preparation.** DIG-labelled single stranded DNA probes electrophoresed alongside template DNA. The incorporation of DIG-dUTP is confirmed by an apparent increase in molecular weight when compared to the template.

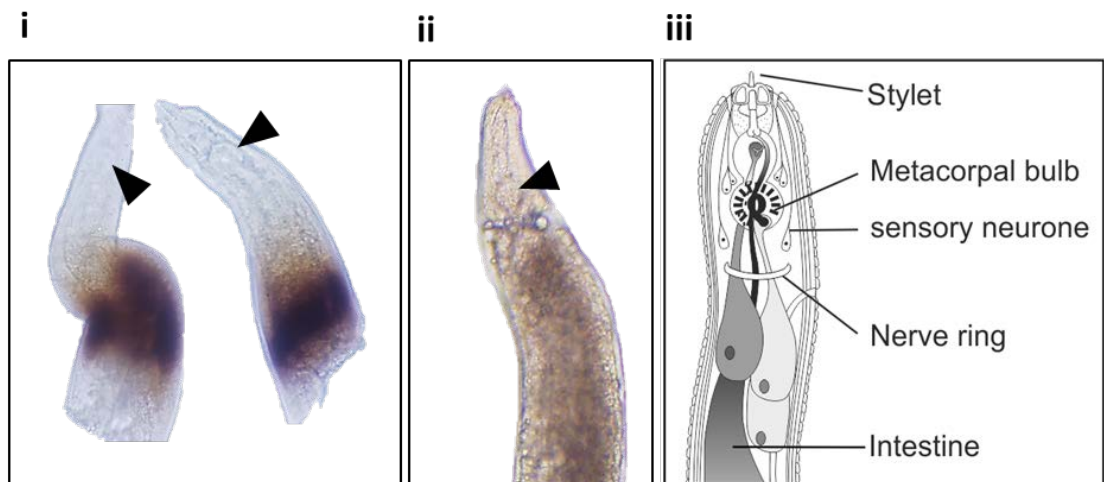


Figure 5-4 **Candidate 3453 *in situ* hybridisation.** **i)** *In situ* hybridisation using the reverse probe results in dark staining of a paired structure posterior to the metacorporeal bulb. **ii)** No staining is observed with the negative control probe. **iii)** Comparison with a schematic diagram does not indicate staining in a tissue with the capacity to secrete proteins.

i



ii



Figure 5-5 **Typical transgenic hairy root transformation in tissue culture.** **i)** Hairy roots emerging from calli on leaf material. Roots originating from each circle are considered to be from individual transformation events. **ii)** Individual lines transferred to separate plates to increase biomass prior to infection.

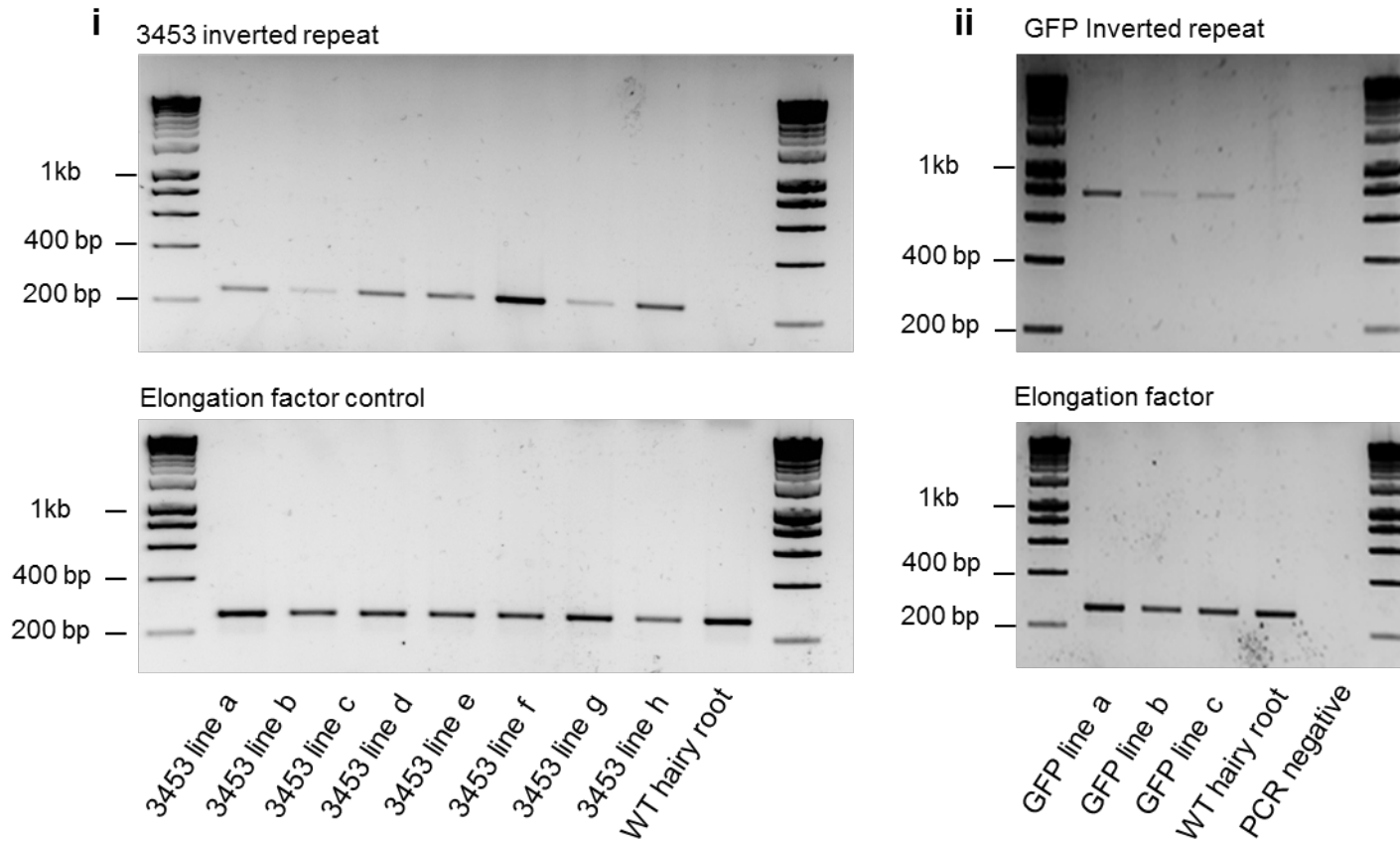


Figure 5-6 **Semi-quantitative RT-PCR of candidate 3453 and GFP inverted repeat RNAi construct expression in hairy roots.** Comparison of the relative intensity of the gene specific PCR products with those of the Elongation factor control highlight 3453-IR lines C, D, E and H and GFP-IR line A as being high expressing.

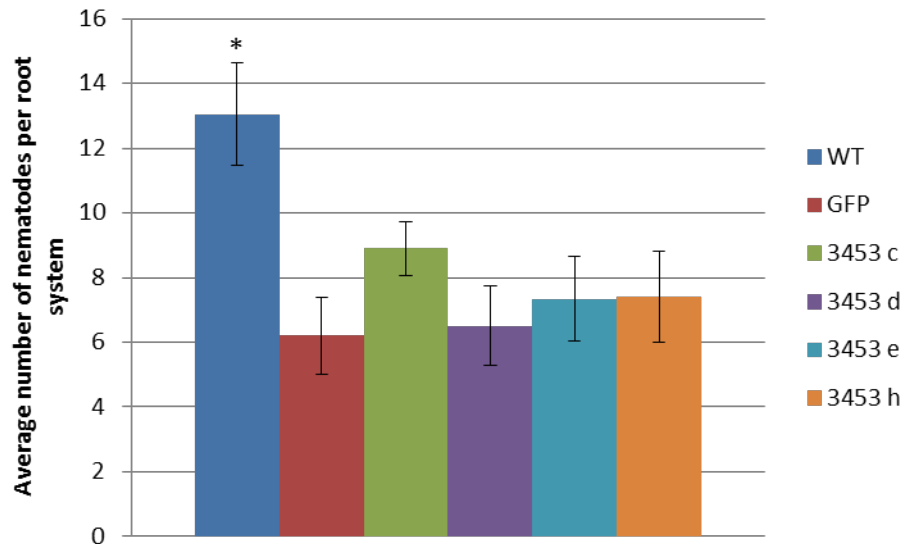


Figure 5-7 **Candidate 3453 RNAi hairy root infections.** The average number of nematodes present at 14 days post infection on transgenic hairy roots. A comparison was made between the Wild Type control (WT), the GFP control, and 4 independent 3453 inverted repeat lines. Although a statistically significant difference ($p < 0.05$) can be seen between the wild type hairy root line and the others, no difference could be seen between the GFP control and any of the four 3453 hairy root lines tested.

5.4.3. Candidate 176

A single 176 gene was identified in the *G. pallida* genome assembly. Very high expression was observed across all feeding stages, increasing in the later feeding stages.

5.4.3.1. Candidate 176 *in situ* hybridisation

Due to the short length of the 176 transcript an *in situ* hybridisation probe was designed to the full length of the transcript. Figure 5-8 shows the apparent increase in size of the DIG-labelled single stranded DNA probe compared to the template, using both the forward (negative) and reverse (positive) primers. The reverse probe bound strongly and specifically to a paired structure anterior to the pump chamber. A comparison with the schematic diagram suggests that this corresponds to the amphid sheath cells, as these are the only paired structures of this size in this region of the nematode (Figure 5-9). No staining was observed with the forward probe.

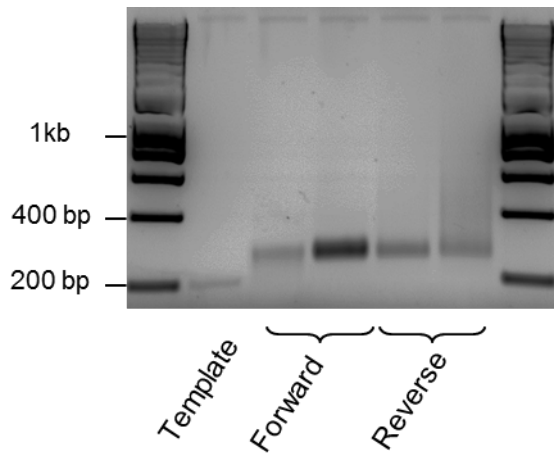


Figure 5-8 **Candidate 176 *in situ* hybridisation probe.** Asymmetric PCR using forward and reverse primers for 176 *in situ* probe preparation. Incorporation of DIG-labelled dUTP can be visualised as an apparent increase in size when electrophoresed alongside the template.

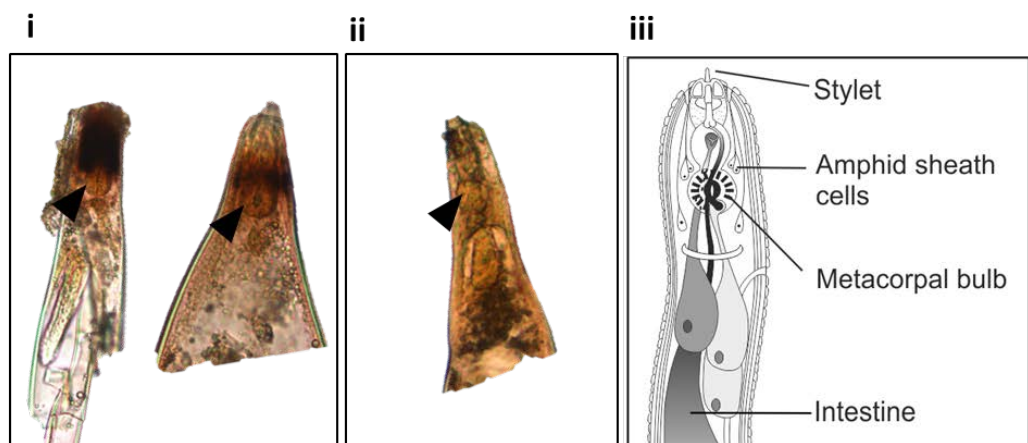


Figure 5-9 **Candidate 176 *in situ* hybridisation.** **i)** Dark staining can be observed in a paired structure anterior to the median bulb (arrow). **ii)** No such staining pattern can be observed in the negative control. **iii)** Comparison to a schematic representation indicates this tissue corresponds to the amphid sheath cells.

5.4.4. Candidate 444 gene family

As discussed in Chapter 4, 444 sequences are part of a large gene family, of unknown absolute size, highly up regulated during the feeding stages. Individual nematodes differ in the number, size and in some cases presence or absence of certain 444 subfamilies. As a result of this variability, where possible, experiments were designed to target as many 444 genes in as many of the three subfamilies as possible through the conserved regions.

5.4.4.1. Candidate 444 *in situ* hybridisation

Figure 5-10 describes the design of the *in situ* probe which targets a 3' region that is conserved between all 444 subfamilies. This probe shares a minimum of 91% identity (average >97%) across all cloned 444 sequences (n = 195). It is thus expected that this probe will bind to every member of every subfamily irrespective of the individual 444 complement of each nematode. Figure 5-11 shows that probe was synthesised and that DIG was incorporated into the single stranded probe. The antisense probes derived from 444 showed specific binding to a paired structure, anterior to the metacorpal bulb. No such staining is observed using the negative control (forward) (Figure 5-12). As seen for candidate 176, this staining pattern is consistent in position with the amphid sheath cells (Figure 5-12iii).

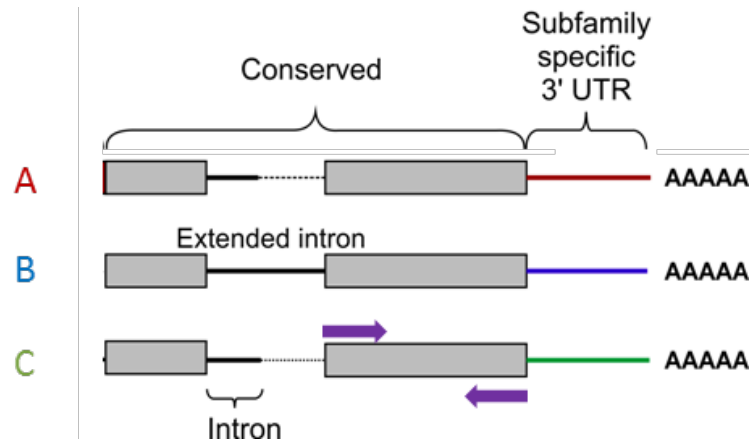


Figure 5-10 **Generic 444 *in situ* hybridisation probe design.** Due to the large number of highly similar genes, a single probe was designed to cover the 3' conserved region of the transcripts (purple arrows). This probe shares a minimum of 91 % identity across all cloned 444 genes (97 % average).

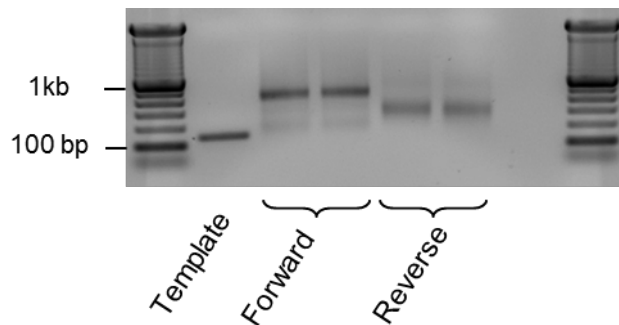


Figure 5-11 **Candidate 444 *in situ* hybridisation probe PCR.** Asymmetric PCR on double stranded template using forward or reverse primers. An increase in apparent size when electrophoresed alongside template (134 bp) indicates incorporation of DIG-labelled dUTP.

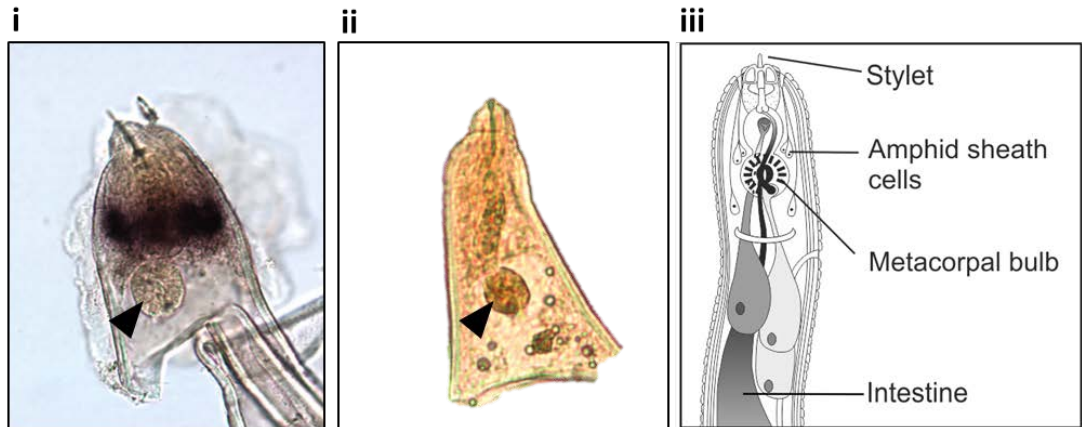


Figure 5-12 5-13 **Candidate 444 *in situ* hybridisation.** **i)** *In situ* hybridisation using the reverse probe highlights dark staining of a paired structure anterior to the median bulb (arrow) in 14 days post infection females. **ii)** No such staining pattern was seen with the negative control forward probe. **iii)** Comparison to the schematic diagram indicates the staining is consistent with the amphid sheath cells.

5.4.4.2. Yeast-two-hybrid analysis of candidates 444 and 176

Both 444 and 176 genes appear to be expressed in the same tissue; the amphid sheath cells. In addition, they are both expressed at the same time in the nematode life cycle and both encode tandem repeat proteins. Potential interactions between and within 444 subfamilies or with 176 protein were therefore investigated. The 176 gene and a representative of each 444 subfamily were cloned into both bait and prey vectors for Yeast-two-Hybrid (Y2H) analysis. Every pairwise interaction between all baits and preys was carried out. An interaction between sequence 176 as bait and prey was seen as indicated by yeast growth on selection medium (Figure 5-13). When re-streaked on high stringency medium the interaction was confirmed by the growth of yeast colonies and activation of the LacZ gene giving rise to blue colonies. No interactions were seen with any of the 444 subfamily proteins, either with themselves or with the 176 protein, on the low or high stringency selection media. Repeating this experiment on three separate occasions produced the same pattern of interactions.

5.4.4.3. Candidate 444 and 176 bacterial expression and mass spectrometry

In order to confirm the interactions, or lack thereof, from the yeast-two-hybrid experiment a second method to test protein interactions (mass spectrometry) was used. The protein sequences used for Y2H were expressed in, and purified from, bacteria (as described in Section 2.5). The 176 protein was insoluble when expressed in bacteria and was therefore not amenable to mass spectrometry. 444-B was the only protein that expressed well and was soluble. It was therefore tested for interaction with itself, however no such interaction could be seen. For 444-B this confirmed the negative result seen in Y2H.

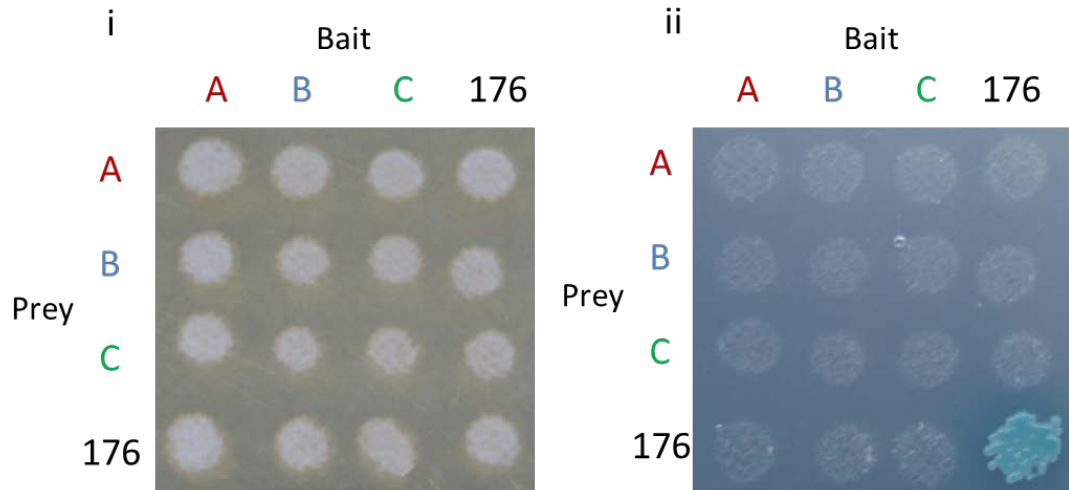


Figure 5-14 **444 and 176 Yeast-two-hybrid analysis.** Every pairwise interaction between a member of each 444 subfamily (A, B and C) and 176 was carried out. **i)** Growth of yeast on mating medium indicates roughly equal quantities of yeast for each interaction. **ii)** An interaction, as indicated by yeast growth on selection medium, can be seen between 176 protein and its self. No interaction is present between any of the 444 subfamily proteins.

5.4.4.4. **444 – Antigen design and immunolocalisation**

To determine if the 444 proteins are secreted from the amphids, one representative member from each 444 subfamily was used to define antigenic regions of the mature peptides common to all subfamilies (Figure 5-14). This region was used to design a synthetic peptide of 29 amino acids in length, with a minimum of 13 amino acids with 100% identity to each of the 444 subfamilies (Figure 5-15). The specificity of the antibody produced was tested by western blotting (Figure 5-15ii). A sample of purified 444-B protein was electrophoresed alongside a sample of 176 protein. 176 protein was used as a control as it is a protein that also contains tandem repeats but is not a 444. The 444 antibody only detected the 444-B protein and not the 176 protein. The 444 antibody was able to detect several proteins in a total nematode extraction, but did not detect any proteins in the same quantity of plant root protein (Figure 5-16). Pre-immune serum did not bind to purified 444-B protein and did not bind to the same range of proteins in nematode extracts (Figure 5-16). The 444 antibody bound to a protein present in the intercellular spaces at the plant-nematode interface between the anterior end of the nematode and the feeding site (Figure 5-17).

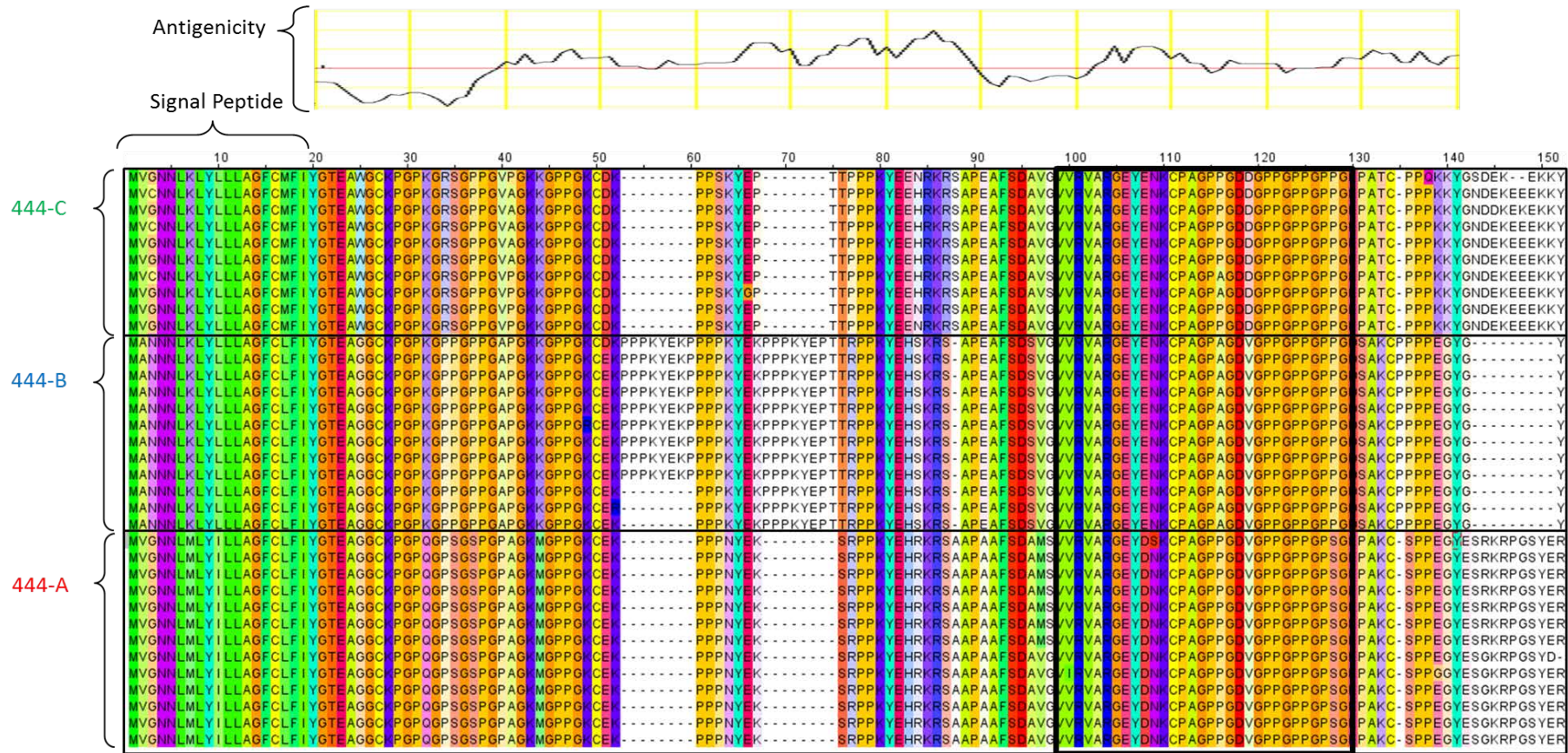


Figure 5-15 444 antigenic regions for peptide design. An alignment of the N-terminal region of proteins from all 444 subfamilies. Antigenic regions of the proteins are compared with conserved sections of the alignment to identify a suitable location for generating antibodies. The regions varied slightly between subfamilies, although at least 13 amino acids of the synthetic peptide (boxed) match with 100% identity to each of the three subfamilies.

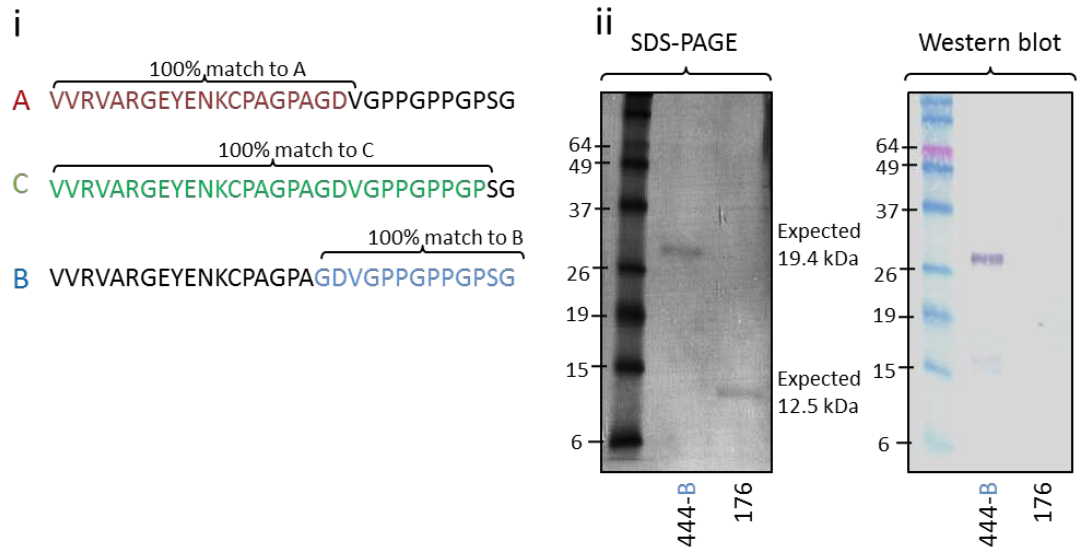


Figure 5-16 **444 Immunolocalisation - Peptide design and antiserum quality control**. A single peptide was designed that contained a minimum of 13 amino acids identical to a conserved region between the three subfamilies (i). The peptide was used raise a polyclonal antibody able to detect all 444 proteins. Specific binding of the antisera is seen to a 444 subfamily B protein expressed in bacteria, when compared to the control nematode gene 176 (ii).

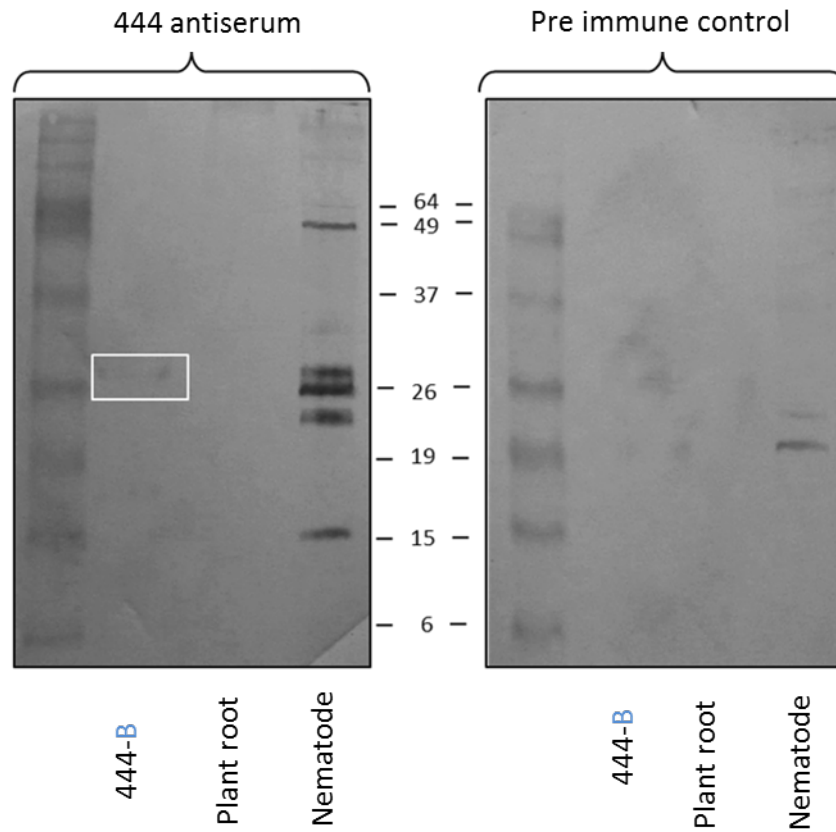


Figure 5-17 **444 antiserum specificity control**. Equal quantities of whole nematode protein extract, plant root protein extract are electrophoresed alongside very low amount of 444-B purified protein. 444 antibody specifically detects 444-B proteins expressed and purified from bacteria and a range of proteins from a total nematode extract that are not detected using the same concentrations of pre-immune control. Neither 444 antisera or pre-immune control can detect plant proteins at these concentrations.

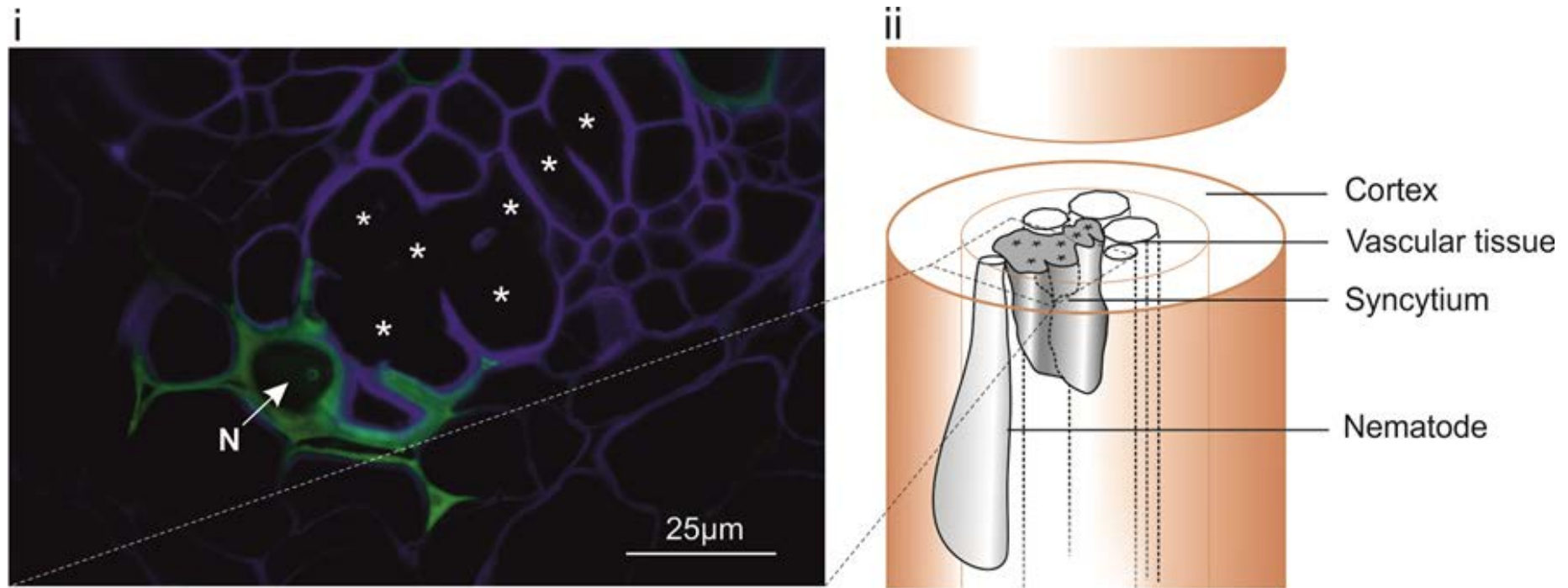


Figure 5-18 **Immunolocalisation of 444 proteins in infected root tissue.** **i)** 2 µm section of an infected potato root showing 444 protein (green) detected between cell walls (blue) in the apoplast between the feeding site (*) and the anterior end of the nematode (N) 14 days post infection. Cell walls are stained with calcofluor white (blue). **ii)** Comparison with a schematic highlights the nematode in the context of the syncytium and the root tissue.

5.4.4.5. RNA interference of candidate 444

A construct was designed to express a hairpin transcript derived from a single member of subfamily A as double stranded RNA *in planta* using either the full length transcript (F) or the 3' conserved region (S). The ability of each construct to knockout all members of all subfamilies was assessed computationally. A consensus sequence for each subfamily was generated by aligning all cDNA clones within groups in Sequencher. Every possible 21 nucleotide section across the length of the full length (F) or 3' conserved (S) inverted repeat was compared on the basis of 100% identity to the consensus sequence for each group. The full length construct, as expected, showed considerably better coverage for all subfamilies. Coverage across the group A consensus sequence was almost ubiquitous, with 400 fragments matching across the length of the sequence. 26 individual 21 nucleotide fragments matched across most of the length of the group B consensus sequence. 10 fragments matched to the group C consensus sequence, including 4 at the 3' end. The full length inverted repeat construct would therefore be expected to down regulate the 444 transcript/s of any nematode feeding on the transgenic plants, irrespective of the complement of 444 genes in each nematode. This IR construct was transformed into *A. rhizogenes* (R1000) for hairy root transformation. Successfully transformed hairy root lines were tested for expression of the IR construct by sqRT-PCR (Figure 5-18). Expressing lines, whose growth phenotypes matched most closely those of the WT and GFP IR lines, were used for an infection trial (Figure 5-19). 444-A-IR lines showed a significant reduction (of approximately 50-60%) in total nematode infection when compared to wild type hairy roots or GFP ($p < 0.05$).

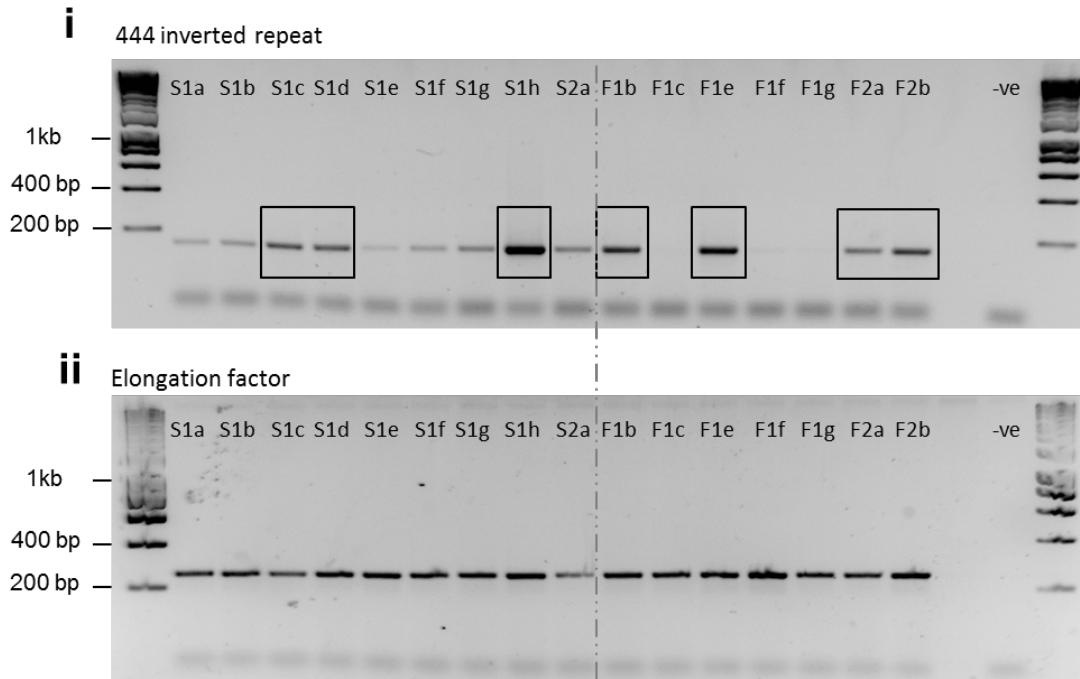


Figure 5-19 **Semi-quantitative RT-PCR of 444-A-IR hairy roots.** Relative quantity of PCR product amplified using gene specific primers after 25 cycles (i) compared to elongation factor primer after 23 cycles (ii) was used as an indication of transgene expression level. High expressing 444 IR lines are indicated by boxes. Individual transformation events (a-h) are indicated as either short 3' inverted repeat (S) or full length inverted repeat (F), of either the first (1) or second (2) transformation occasions.

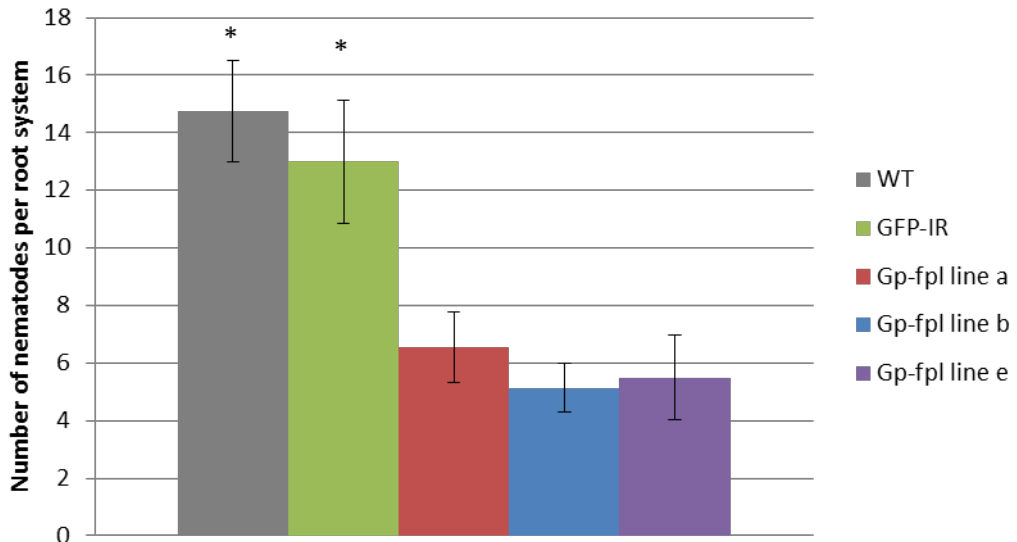


Figure 5-20 **In planta RNAi of 444 genes.** Three lines of transgenic hairy roots expressing the 444-A inverted repeat were tested against wild type hairy roots, and transgenic hairy roots expressed a GFP inverted repeat. All three 444-A lines showed a significant reduction ($p < 0.05$) in the total number of successfully established *G. pallida* nematodes at 14 dpi when compared to the GFP IR line.

5.4.5. Candidate 448 gene family

Twenty eight 448 gene family members were identified that encoded proteins with predicted secretion signals. These were broadly categorised into four major groups based on an alignment of the protein sequences. All secreted 448 proteins had similar expression profiles, being highly up regulated at 7 and 14 days post infection, with reduced expression in later feeding stages.

5.4.5.1. Candidate 448 *in situ* hybridisation

The 448 DNA sequences from *G. pallida* can be divided into several groups. The assumption was made that because all 448 sequences are very similar, within and between groups, and all have the same temporal expression profile, they would all have the same spatial expression profile. A generic *in situ* probe was designed to target one of the groups, without the aim of being specific to that group. The probe designed was based on the cloned sequence of GPLIN_000950100. This probe shared an average of 77% nucleotide identity with other members of its group and an average of 46% identity with all other 448 sequences. Figure 5-20 shows the characteristic apparent increase in molecular weight as expected for DIG incorporation. Figure 5-21 shows a typical hybridisation pattern highlighting expression in the dorsal gland cell. It is unclear which of the 448 family members the probe is binding to. However, based on the sequence identity of the probe to other 448 genes, it is likely that it binds to the mRNA from more than one 448 gene. This may also account for the extremely strong staining.

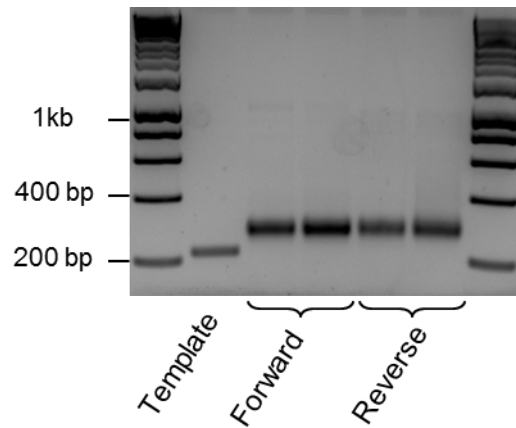


Figure 5-21 **Candidate 448 *in situ* hybridisation probe.** Single stranded DIG-labelled DNA probe for the forward and reverse primers electrophoresed alongside template double stranded DNA. The increase in apparent molecular weight of the probes is as a result of the incorporation of DIG labelled dUTP.



Figure 5-22 **Candidate 448 *in situ* hybridisation.** **i)** *In situ* hybridisation shows dark specific staining in a single large cell posterior to the median bulb (arrow) of feeding 14 days post infection female *G. pallida*. **ii)** No staining is observed using the opposite sense probe. **iii)** Comparison with the schematic suggests this staining corresponds to the position of the dorsal gland cell.

5.4.5.2. **Candidate 448 RNA interference**

Due to the large number, and possible redundancy, of 448 genes, a construct was designed to knock out as many sequences as possible as it is impractical to make 28 individual inverted repeat constructs or a single construct containing 28 different inverted repeats. The 448 sequences can be subdivided into several groups, with three 'major' groups (Figure 5-22). An inverted repeat construct for each of the major groups was designed, and concatenated to form a triple inverted repeat construct that will generate 21 nucleotide silencing fragments binding to most of the known 448 sequences. The entire triple inverted repeat construct was compared by BLAST to the non-redundant nucleotide database and the *G. pallida* assembled genome sequence. The only sequences identified were from the *G. pallida* genome and corresponded to 448 genes. The construct was generated and transformed into potato hairy roots. However there was insufficient time to conduct infection trials on these lines.

5.4.5.3. **Candidate 448 antibody design**

One representative member from two of the major groups containing 448 genes previously cloned was used to define antigenic regions of the mature peptide (Figure 5-23). This, coupled with a protein alignment identified antigenic regions specific to each. Synthetic peptides of 16 and 17 amino acids in length were synthesised for the group containing 448-1 and 448-4 respectively. The group 448-4 peptide shares 100% similarity with all members of its group across at least 15 amino acids whereas the group 448-1 peptide share 100% similarity with one member of its group, and 1 amino acid mismatch with the rest of its group, either a T to N substitution in position 8 or an E to K substitution in position 9 (Figure 5-23). There was insufficient time to fully test or characterise the 448 protein localisation in planta.

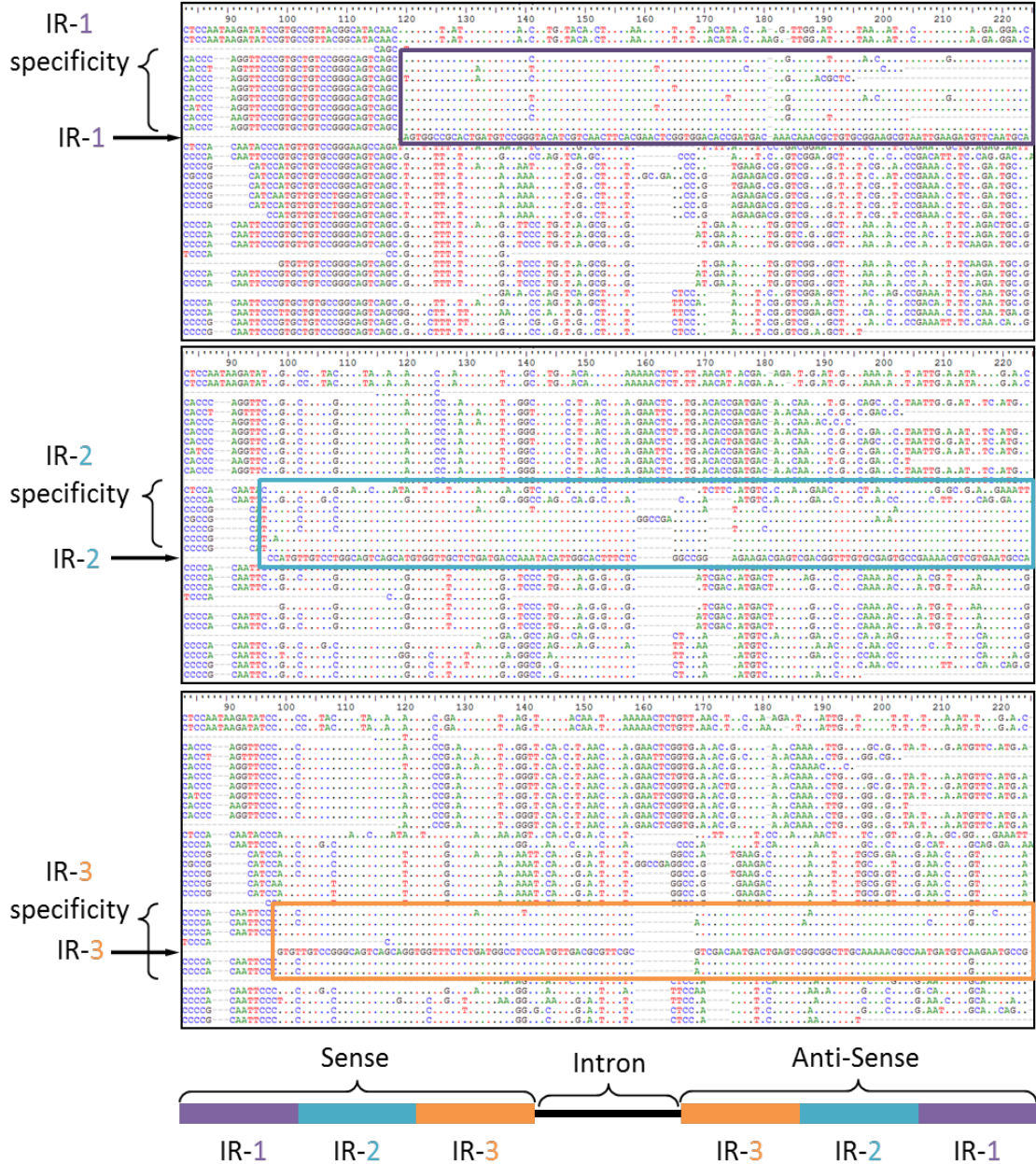


Figure 5-23 **448 triple inverted repeat construct**. DNA alignment of 448 transcripts encoding secreted proteins. For each panel conserved nucleotides in 448 transcripts compared to the Inverted Repeat (IR) are indicated by a point. Three separate sections of DNA that broadly correspond to each of the major groups in a DNA alignment of 448 transcripts were concatenated to produce a single construct. The single construct should have at least one 21 nt fragment overlapping most of the sequences present.

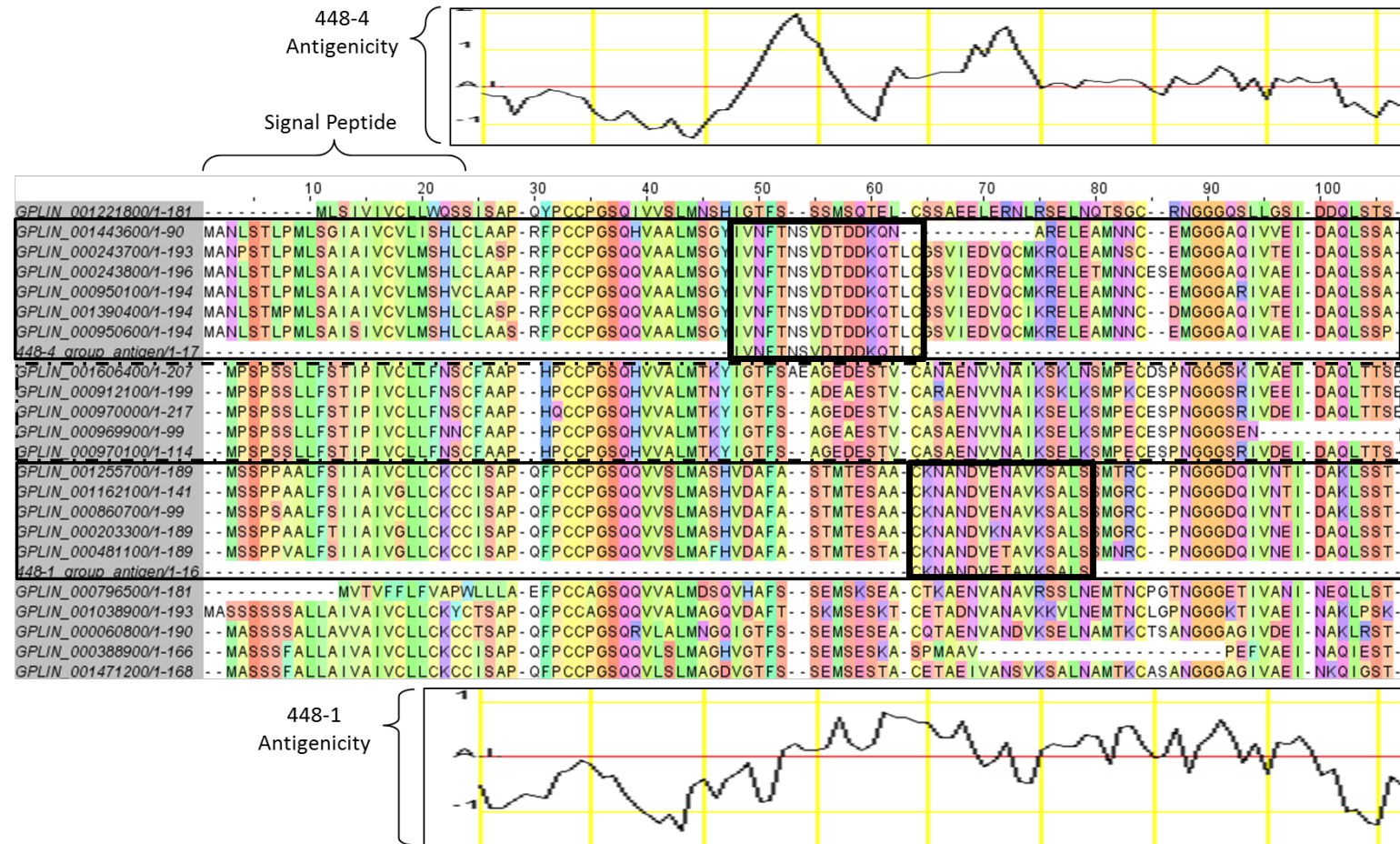


Figure 5-24 **448 antigen design**. Antigenic regions of two groups of 448 genes previously described (448-1 and 448-4) were predicted and compared to a protein alignment of 448s from all groups. Antigenic regions specific to group 448-1 and 448-2 were used to synthesise synthetic peptides.

5.5. Discussion

5.5.1. Effector identification strategies

The pipeline described in Chapter 4 aimed to identify candidate feeding structure genes. These fall under the broad definition of effectors “proteins (or other factors) secreted by a pathogen into its host in order to manipulate the host to the advantage of the pathogen”. Sequences similar to the 448 genes have been previously identified in gland cell library sequencing studies (Gao et al., 2003) representing a convergence of identification strategies and providing confidence that the bioinformatic pipeline described here can be used to identify effectors. It is perhaps surprising that more cases of convergence in identification strategies were not found. As discussed in Chapter 4, the approach taken is prone to false positives (i.e. the identification of genes that are not effectors). In the case of 225 this was indeed true: *in situ* hybridisation localised gene expression to the entire body including the digestive system of the nematode. Digestive expression might be expected as the characteristics of a hypothetical feeding tube/plug gene are similar to those of a digestive enzyme. Both require expression during the feeding stages and in order for both to be functional they require secretion. It is possible that the extremely strict expression profiling criteria used in this study prevented more known effectors from coming through the pipeline. The identification strategy used here also allows effectors originating from tissues other than the gland cells to be identified and therefore offers a complementary approach.

5.5.2. The role of amphids during the feeding stages of *Globodera pallida*

As previously suggested (Perry, 1996, Jones et al., 1994), the data presented here indicate a changing role for the amphids in *G. pallida* throughout the life cycle. Members of two putative effector families, 444 and 176, were expressed in large secretory cells associated with the amphids (amphid sheath cells). The amphid sheath cells produce material that is present in the amphid canal (Endo, 1978). Although two previous studies suggested that the feeding plug originates from the opening of the amphid canal, 444 proteins are detected extending further into the apoplasm than would be expected of feeding plugs (Endo, 1978). Moreover it is challenging to rationalise the high sequence variability and individual variation of 444 gene complement in the context of a feeding plug. As discussed in Chapter 4 the variable tandem repeat structure is complex and contains highly conserved residues interspersed with variable single- or di-residues. It was hypothesised that the tandem repeats may be involved in homo- or hetero-complex formation.

However, no evidence for 444 protein-protein interactions was found in Y2H or mass-spectrometry. It is unclear what the role of the different subfamilies, or indeed the variable number tandem repeats, is in the context of plant parasitic nematodes. It is possible that they may interact with a plant protein or glycan situated in the apoplasm. Screening either a plant-derived Y2H library or an array of apoplastic glycans may identify binding partners of 444 proteins (Liang et al., 2008). Once identified, the interaction affinity with different subfamilies or numbers/organisations of tandem repeats may elucidate their biological function. Other apoplastic effectors of plant-parasitic nematodes include a calreticulin effector (Mi-CRT) of *M. incognita* (Jaouannet et al., 2013). This protein is secreted from the subventral gland cells into the apoplasm and functions to suppress host defences by preventing calcium influx. Future work will elucidate if 444 genes have a similar function.

The 176 gene is also expressed in the amphid sheath cells. Although no *in planta* localisation studies have been carried out, 176 did self-associate in Y2H, indicating that it may form complexes *in planta*. It was not possible to confirm the Y2H interaction in mass-spectrometry because the protein was insoluble when expressed in bacteria. It is reasonable to speculate that the components of a feeding plug, which has a dense fibrous appearance in electron micrographs, would indeed be insoluble. Further evidence is required to determine if 176 genes do encode components of the feeding plug.

The data here for the 444 and 176 gene families, in addition to those of other effector families described in Section 5.1.3.1, suggest that the amphids play a major role in delivering proteins to the apoplasm during successful interactions.

5.5.3. RNA interference

Gene knockdown studies for plant-parasitic nematodes using RNAi either rely on dsRNA soaking or *in planta* delivery (Lilley et al., 2007, Lilley et al., 2012). RNA soaking was not attempted in this study as one of the criteria for identifying these genes was that expression is restricted to the feeding stages. RNAi requires some transcript to be present before transcript degradation can maintain the knockdown effect, therefore exposing J2s (that do not express these genes) to dsRNA would most likely be ineffective. Instead, a hairpin construct was expressed in potato hairy roots. It has been reported that dsRNA expression in hairy roots is not ubiquitous and that phenotypes of the root cultures can vary, which may affect nematode infection (Wubben et al., 2009, Triplett et al., 2008). Phenotypes of no vector control, GFP IR control, and candidate feeding structure gene IR hairy root lines were therefore matched as well as possible before infecting with nematodes. Due to

the patchy nature of hairy root transgene expression (Triplett et al., 2008, Wubben et al., 2009), it is possible that successfully established nematodes had induced feeding sites from cells not expressing the dsRNA. It is therefore difficult to confirm the knockdown of expression from the nematodes that successfully establish. Similarly due to the amplification step and transient nature of RNAi in nematodes (Dalzell et al., 2011), highly specific knockout studies of individual sequences within large gene families (448 or 444) are not possible.

Despite the inherent variability of the hairy root system (Wubben et al., 2009), knockdown of the 444 genes resulted in a significant reduction in nematode numbers compared to the GFP control. Although the dsRNA construct used has the potential to knock down all members of all subfamilies characterised to date, the list of cloned 444 sequences we have characterised is unlikely to be complete. Future work will focus on generating transgenic potato plants expressing the inverted repeat constructs of all three 444 gene subfamilies, potentially combating the diversity of 444 genes. RNAi has been proposed as a method to develop novel synthetic disease resistance in crop species (Heinemann et al., 2013). More work will be needed to determine if a triple 444-IR construct will provide sufficient levels of resistance to constitute a reliable solution. Additionally, work will be carried out to identify the stage at which infection is unsuccessful. Interestingly knockdown of the 3453 gene family did appear to result in a reduction in nematodes compared to the wild type control. However, the GFP-IR control showed similar levels of infection to the 3453-IR lines. This again highlights the inherent variability of transgenic hairy roots for host-derived RNAi.

RNAi signal trafficking between host and parasite has been demonstrated on numerous occasions. Examples include chewing insects that lyse cells and feed directly from host cell cytoplasm (Baum et al., 2007). The stylets of Clade 12 plant-parasitic nematodes do not cross the plasma membrane but rather cause local invaginations. However, these plant-parasitic nematodes ingest host cell cytoplasm as shown by uptake of fluorescent proteins and dextrans (Valentine et al., 2007, Bockenhoff and Grundler, 1994). *In planta* dsRNA expression does result in nematode gene silencing (Huang et al., 2006) and the dsRNA is presumably introduced to the nematode by the same means. Indeed this appears to be common to other host-pathogen interactions from a range of kingdoms. Obligate biotrophic fungi, such as *Bulmeria graminis*, form specialised interactions with their hosts (Nowara et al., 2010) in which finger-like appendages also form local invaginations of host cell plasma membrane. Similar to sedentary plant-parasitic nematodes, fungal haustoria do not rupture the plasma membrane. Despite this,

host induced gene silencing has still been demonstrated (Nowara et al., 2010, Pliego et al., 2013) and the signal must cross the haustorial membrane. In addition, plant-parasitic plants also form structurally similar haustoria with similar RNAi signal trafficking (Westwood et al., 2009, Tomilov et al., 2008). It has been suggested that the crossing of the plasma membrane may be facilitated by plasmodesmata or intense vesicle trafficking at the haustorial membrane (Nowara et al., 2010). It has been shown that plasmodesmata are active in syncytia between plant cells (Hofmann and Grundle, 2006) although no studies have directly analysed the plant-nematode interface for plasmodesmata or increased vesicle trafficking. Interestingly, recent work (Weiberg et al., 2013) has demonstrated that parasites use small RNA molecules as effectors, demonstrating that mechanisms for transfer of RNA between host and parasite exist.

Summary

- The spatial expression patterns of 5 genes/gene families have been characterised. Three of the five candidates are expressed in tissues with the capacity to secrete proteins into the host, reinforcing their function as putative-effectors.
- The 444 proteins are secreted into the host, accumulate in the apoplasm, and are required for parasitism.
- 176 genes are expressed in secretory cells and the proteins form self-associated complexes. 176 monomers, or complexes, are insoluble when expressed in bacteria.

6. Analysis of the transcriptomes of cyst nematodes and related species

6.1. Introduction

As discussed in Chapter 1, feeding tubes and feeding plugs have been identified in a range of biotrophic plant parasitic nematode species. Having identified (Chapter 4) and characterised (Chapter 5) putative feeding plug and feeding tube components from *G. pallida*, it was decided to investigate the presence of these candidates in other related nematode species for which feeding structures have been described, in the hope that this would provide further evidence for their proposed function(s). An analysis of the biology of related species and their feeding structures allows an informed decision to be made on which species to compare.

6.1.1. Related plant parasitic nematodes

6.1.1.1. The root knot nematode *Meloidogyne incognita*

As introduced in Chapter 1, root-knot nematodes are sedentary biotrophic pathogens in Clade 12 of the phylum Nematoda. Although they induce feeding sites known as giant cells, these are morphologically distinct from the cyst nematode syncytia.

6.1.1.2. The reniform nematode *Rotylenchulus reniformis*

The reniform nematodes, so named due to their kidney shaped body, are comprised of 10 valid species, for which *Rotylenchulus reniformis* is the most economically important and consequently the most well studied (Robinson et al., 1998). *R. reniformis* has a very broad host range of > 300 plant species, many of which are common weeds (Quénéhervé et al., 2006, Agudelo et al., 2005, Robinson et al., 1998). J2 hatch from eggs deposited in the soil where they moult, without feeding, into J3, J4 and immature females or males at a roughly 50:50 ratio (Bird, 1983). Females penetrate through the cortex destructively and induce feeding site formation in the stele. The feeding site of *R. reniformis* is similar morphologically to that of the cyst nematodes and is so termed a syncytium (Agudelo et al., 2005, Rebois et al., 1975). Although less information is available regarding the reniform nematode syncytium, it is assumed it shares similar metabolic characteristics as that of the cyst nematodes due to its morphological similarity and its phylogenetic proximity (van Megen et al., 2009).

6.1.1.3. The cyst nematode *Globodera rostochiensis*

The life cycle, host range, and biology of the cyst nematode *Globodera rostochiensis* is very similar to that of *G. pallida* and is described in detail in Chapter 1.

6.1.2. Feeding structures in related plant parasitic nematodes

6.1.2.1. Feeding tubes

Feeding tubes have been identified from a wide range of nematode species in several different clades of plant parasitic nematodes, suggesting an essential role in plant parasitism. Within the Clade 12 plant parasites (van Megen et al., 2009), which includes the sedentary endoparasitic nematodes, feeding tubes have been observed in many species that induce feeding structures including *Meloidogyne* spp, *Rotylenchulus reniformis*, and *Globodera* spp (Rahman Razak and Evans, 1976, Rebois, 1980). Individual species within each of these three genera have very similar feeding strategies and structures (Paal et al., 2004, Wyss and Grundler, 1992, Wyss, 1992). Feeding tubes are not produced by related nematodes that do not induce feeding structures (e.g. migratory endoparasites) suggesting that they represent an adaptation to biotrophic interactions.

The differences in the morphology and composition of the feeding tubes of root knot, cyst and reniform nematodes are summarised in Figure 6-1. Feeding tubes of *M. incognita* have a regular crystalline structure with a diameter ranging from 820-990 nm, consisting of an electron dense wall approximately 190-290 nm thick and an electron transparent lumen 340-510 nm in diameter. The feeding tubes are up to 73 µm in length (Hussey and Mims, 1991), and are often associated with whorls of membrane systems of unknown function. Cyst nematode feeding tubes, such as those of *G. pallida* and *G. rostochiensis*, are of a similar length but their walls are composed of an electron dense mesh with no regular structure observable under a transmission electron microscope. They contain an electron transparent lumen of diameter approximately 600nm with a considerably thinner wall of 80-180 nm and are not associated with membrane systems. Reniform nematode feeding tubes are morphologically similar to those of the cyst nematodes (Figure 6-1) inasmuch that they lack the regular crystalline structure of the root knot nematode feeding tube (Rebois, 1980). Reniform nematodes produce a spiral feeding tube with a maximum diameter of 0.6 – 1.2 µm that decreases as the distance from the stylet aperture increases (Rebois, 1980). *R. reniformis* feeding tubes are approximately 118 µm in length (Rebois et al., 1975), longer than those produced by root knot nematodes

and cyst nematodes (Hussey and Mims, 1991). Membrane systems, similar to those associated with root knot nematode feeding tubes, have been described for *R. reniformis* (Rebois, 1980).

6.1.2.2. Feeding plugs

Feeding plugs have been reported in a range of cyst nematodes (Sobczak et al., 1999, Holtmann et al., 2000, Endo, 1978) and appear to be specifically associated with syncytia, as they have never been described as being in association with root-knot nematode giant cells (Hussey and Mims, 1991). A similar structure (variously referred to as a feeding plug or feeding peg) has also been described in *R. reniformis* (Razak and Evans, 1976, Rebois, 1980). However, the *R. reniformis* feeding plug appears to be morphologically distinct from the cyst nematode feeding plug.

6.1.2.3. Feeding structures in different species

The morphological differences between the root knot nematode feeding tube and that produced by cyst and reniform nematodes probably reflects independent origins of sedentary endoparasitism in these groups. It is therefore likely that the genes that encode components of these structures are different. Similarly, feeding plugs have never been described for root knot nematodes but a similar structure has been described for the reniform nematodes. We would therefore not expect to find any of the *G. pallida* candidate feeding structure genes identified in Chapter 4 in the root knot nematodes. However, we expect them to be present in other cyst nematodes, such as *G. rostochiensis* (Figure 6-1), and they may be present in the reniform nematode *R. reniformis* (Figure 6-1).

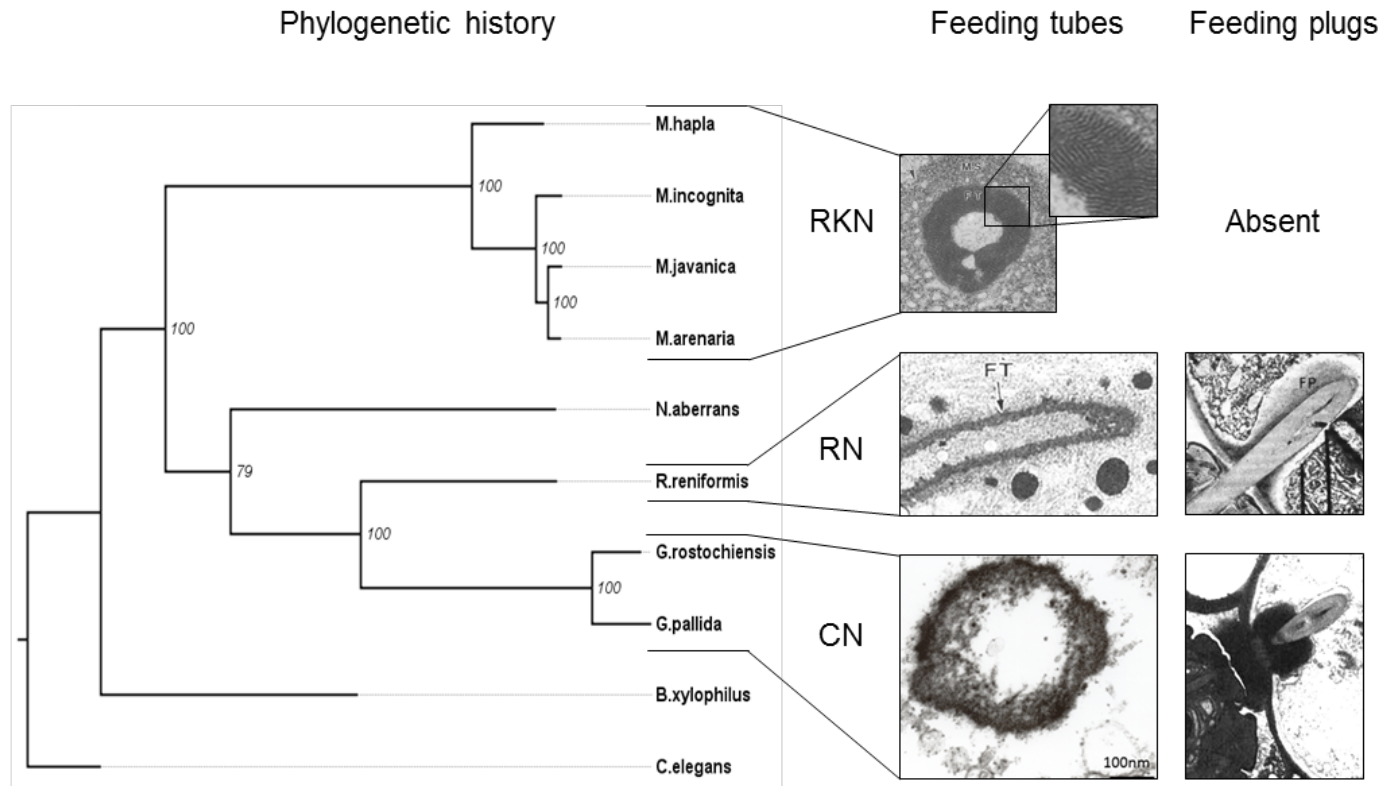


Figure 6-1 **Phylogenetic distribution and morphology of feeding tubes and feeding plugs.** Local phylogenetic history of related plant parasitic nematodes adapted from Eves-van den Akker *et. al.*, (in preparation). Feeding tube images adapted from Rebois (1980) (Rebois, 1980) and P. E. Urwin (per. comm.). Feeding plug images adapted from Endo (1978) and Rebois (1980) (Endo, 1978, Rebois, 1980).

6.1.3. DNA Sequencing and next generation technology

6.1.3.1. History of DNA sequencing

In 1977 Frederick Sanger developed the first generation of sequencing technology based on chain termination. This procedure was subsequently named Sanger sequencing (Sanger et al., 1977), and was the first widely accepted sequencing technology. Variations of this technology are still used to the present day for many low throughput sequencing applications. Sanger sequencing can generate long reads (up to 900 bases) with very high accuracy (99.999% (Liu et al., 2012)). This was the primary technology used for the human genome project (Collins et al., 2003). This project cost \$ 4.7 billion USD (adjusted for inflation) and these high costs fuelled research into sequencing technology, leading to the development of the Next Generation Sequencing (NGS) technologies (Liu et al., 2012), that are orders of magnitude faster, cheaper and more efficient.

6.1.3.2. Next Generation Sequencing (NGS)

There are three main NGS sequencing platforms, each of which has its own advantages and disadvantages, but all of which have vastly greater throughput than Sanger sequencing.

Roche 454 sequencing was the first commercially available technology. Modern 454 machines are capable of producing approximately 1 million 700 bp single end reads in 24 hours. The main disadvantages of this technology are its low accuracy across polyN tracts and its relatively high cost. 454 sequencing is by far the most expensive of the NGS technologies. costing > \$ 12 per million bases. Although this is vastly less expensive than Sanger sequencing the other two NGS technologies are 100 – 200 times less expensive. SOLiDv4, produces short reads (< 100bp) at comparable accuracies to Roche 454, but has considerably higher read depth. Illumina sequencing is by far the most widely used NGS technology, due to its very low cost (\$ 0.07 per million bases for the HiSeq) and incredibly high output (3 billion 100 bp paired end reads in 3 – 10 days). It has slightly lower accuracy than the other technologies (98%) but this is compensated for by considerable increases in read depth. Comparisons between the three platforms by re-sequencing well characterised organisms have shown very little difference in overall performance (Suzuki et al., 2011). Next generation sequencing has revolutionised almost every field of life sciences, including nematology. Genome sequencing, previously only attainable for a few carefully selected model organisms, is now accessible for almost all organisms. For example, sequencing of individual strains of economically

important eukaryotic pathogens can be carried out rapidly in response to emerging threats (Chowdappa et al., 2013).

6.1.3.3. RNAseq

One of the increasingly common and valuable applications of NGS technology is RNA sequencing (RNAseq). A total RNA sample can be converted to first strand cDNA, and sequenced on any of the NGS platforms. Depth of coverage is sufficiently great that simultaneous detection and quantification of all transcripts in a given RNA sample is feasible. This RNA sequence can be mapped to an assembled genome sequence or assembled *de novo*. This is an incredibly powerful tool that can generate transcriptome data sets, for organisms that do not have well characterised genome sequences, at very low costs.

As part of the *G. pallida* genome sequencing project, a draft *G. rostochiensis* genome sequence was produced. In addition, RNA was sequenced from two biological replicates of second stage juveniles of *G. rostochiensis*. In order to take advantage of this pre-existing information, *G. rostochiensis* was chosen as one of the species to compare to *G. pallida*. *R. reniformis* was chosen as the second species to compare, due to its close phylogenetic proximity and similarity in feeding structure morphology. *M. incognita* was not chosen due to its independent origin of sedentary parasitism, lack of feeding plug, and morphologically distinct feeding tube. RNA seq is an ideal tool for generating reference transcriptome sequences for intra and inter-specific comparisons.

6.2. Aims

The aim of this part of the work was to use Illumina sequencing to generate transcriptome data sets for two species of plant parasitic nematode related to *G. pallida*; *G. rostochiensis* and *R. reniformis*.

- For each species, sequence a parasitic and a juvenile life stage, with biological replicates.
- Carry out differential expression analysis and comparative transcriptomics for analysis within and between species.
- Confirm the presence or absence of the candidate feeding structure genes identified from *G. pallida* in these two related species.

6.3. Materials and methods

6.3.1. Biological material

For *G. rostochiensis*, the same batch of cysts that was used for the genome and transcriptome sequencing carried out prior to this project was used to generate biological material. This allowed direct comparison between the two datasets. A highly inbred line of *R. reniformis*, originally created by Dr. Forrest Robinson, was obtained from Dr Martin Wubben, and used for all sequencing of this species.

6.3.2. Nematode collection and RNA extraction

6.3.2.1. *G. rostochiensis* second stage juveniles

G. rostochiensis cysts, stored dry at 4 °C, were hatched in potato root diffusate to release the second stage juvenile (J2) nematodes, as described in section 2.3.2. J2s were collected into 15 ml polystyrene tubes (Starstedt, Leicester) and allowed to stand for 30 seconds. The supernatant was removed to a clean 15 ml polystyrene tube to separate J2s from any large, heavy debris. Small, lighter debris was removed by allowing the nematode suspension to settle for 30 minutes and discarding the supernatant.

6.3.2.2. *G. rostochiensis* sedentary female feeding females

Nematodes were extracted 14 days post infection (dpi) using the protocol described in section 2.3.3 and flash frozen in liquid nitrogen.

6.3.2.3. *R. reniformis* J2

R. reniformis J2 were hatched from eggs collected from infected plants. Root tissue was removed from the soil and cleaned. Clean root tissue was cut into ~ 2 cm sections and homogenised in a bench top blender for 10 seconds. Homogenised tissue was passed over 150 µm, 63 µm and 25 µm sieves. Eggs were collected from the 25 µm sieve and cleaned by sucrose centrifugation (Section 2.3.3). Eggs were allowed to hatch at room temperature in water over a 25 µm mesh. J2 nematodes were collected daily from under the 25 µm mesh, nematodes collected in the first 2-3 days were discarded. Collected J2s were cleaned, as described above, and flash frozen in liquid nitrogen.

6.3.2.4. *R. reniformis* feeding stage

Feeding stage females of *R. reniformis* were extracted using a similar method to that for cyst nematodes described in section 2.3.3 except that cotton was used as

the host, and nematodes were collected from 5-10 week old plants. Using plants of this age allowed multiple rounds of reproduction to occur and therefore populations were un-synchronised. Only the relevant, sedentary female equivalent, stages were collected manually under the microscope using a pipette.

6.3.2.5. RNA extraction and quality control

RNA extractions were carried out using RNeasy Spin columns (Qiagen) using the slightly modified protocol as described in Section 2.4.7. 4 µl of RNA was removed from each sample for quality control, the remainder was stored at -80 °C. Two methods were used to assess the quality of extracted RNA. Firstly, 2 µl of each sample was analysed using a Nanodrop spectrophotometer (NanoDrop products, Delaware, USA), and the absorbance at 230, 260 and 280 nm was recorded. Samples where the ratio of absorbance at 260/280 was > 2 and the ratio of absorbance at 260/230 was between 2 and 2.2 were also processed on a Bioanalyser (Agilent Technologies, California, USA). Samples with an RNA Integrity Number (RIN) greater than 8 were considered sufficiently high quality for library construction and sequencing.

6.3.3. Library construction, sequencing and quality control

Library construction and sequencing was carried out using the service provided by The Genome Analysis Centre (TGAC, Norwich, UK). In brief, libraries were constructed using the PerkinElmer Sciclone using the TruSeq RNA protocol (Illumina). mRNA was separated from 1 µg of total RNA by poly-A pull down using biotin beads. mRNA was fragmented, cDNA synthesised, and overhanging ends repaired to create blunt ended DNA. 3' A over hangs were added using *Taq* DNA polymerase, and ligated to corresponding 3' T overhangs present on adapter sequences. Un-ligated adapters were removed using size selection XP beads (Beckman Coulter).

Nine libraries for three conditions (*G. rostochiensis* sedentary female, *R. reniformis* J2 and *R. reniformis* sedentary female), each in biological triplicate, were prepared with different barcoded adapters. Library quantities were normalised and pooled to a concentration of 10 nM each. The pooled sample was diluted to a final concentration of 10 pM, spiked with 1% PhiX control v3 and sequenced on a single lane of Illumina HiSeq using 100 cycles for each paired end read.

Raw DNA reads were supplied with a quality score for each base ranging from 10 to 50, representing the confidence with which each base was called. Each library

was visualised in FastQC to assess the per base GC content, average quality score per base, and the presence of any over-represented sequences. Over-represented sequences for all libraries were pooled into a single text file for later filtering.

Reads were trimmed based on quality score using the program Trimmomatic. Two approaches were used for quality trimming, depending on the application of the reads. A high stringency approach was used for trimming reads prior to assembly, and a less stringent approach was used for trimming reads prior to mapping. Leading or trailing bases were removed if they had a quality score of less than 28 or 22 for assembly and mapping respectively. A sliding window of 10 bases was passed over each read, if the average quality score was less than 28 or 22 for assembly or mapping respectively the entire window was removed. In both approaches after trimming a minimum read length of 50 nucleotides was specified. All over represented sequences corresponding to Illumina adapters were removed by comparison to the over represented sequences file prepared previously.

Trimmed RNA reads were then loaded back into FastQC to verify Trimmomatic had successfully cleaned the reads. Trimmed reads for assembly were pooled by concatenating all files for each pair in the same order.

6.3.4. Assembly

6.3.4.1. *In silico* normalisation

An *in silico* normalisation step was applied to the reads prior to assembly using the `normalize_by_kmer_coverage.pl` script supplied with Trinity. This step is designed to reduce the computational power and time required by removing any reads that are represented more than once. Approximately 1GB of RAM was used per million reads, and 10 CPUs were requested from the server. Following normalisation the number of remaining reads was counted; typically this should result in a reduction of 90% of the reads but only a loss of 0.6% of the information.

6.3.4.2. Trinity assembly

Trinity (version r2013-02-25) was used to assemble transcriptomes *de novo* using the *in silico* normalised reads. All parameters were left at default, and 200GB RAM was requested from the cluster. Following assembly, the `TrinityStats.pl` script was used to count the number of transcripts, components, and the N50 value for each species.

6.3.4.3. CLC bio assembly

CLC version 4.0 was used to assemble *de novo* transcriptomes using both the *in silico* normalised reads and the non-normalised reads for comparison. All parameters were set to default.

6.3.5. Mapping

6.3.5.1. Mapping to genome - TopHat2

TopHat2 was used to map cleaned reads back to the assembled genome for *G. rostochiensis*. Maximum intron size was estimated at 11,000 nucleotides based on the published information available for the closely related cyst nematode *G. pallida*. Library insert size was calculated to 400 from average insert size plus two times the average read length.

6.3.5.2. Mapping to transcriptome - Bowtie2

Bowtie2 was used to map cleaned reads back to assembled transcriptomes using default parameters. The Trinity wrapper script `run_RSEM_align_n_estimate.pl` was used (discussed in more detail below). For mapping raw reads to fasta files of interest from different species, a less restrictive approach was taken by increasing the allowed mismatches using the following parameters: `-N 1, -L 15, -i S,1,0.50, -D 20, -R 3`

6.3.6. Differential expression

The Trinity wrapper script `run_RSEM_align_n_estimate.pl` automatically runs both Bowtie2 (mapping) and RSEM (read counts). For each library the cleaned mapping reads were mapped back to the relevant transcriptome using default parameters. This generates one file per library and these were then merged to give a gene counts matrix for all data libraries using `merge_RSEM_frag_counts_single_table.pl`. Differential expression was calculated using the perl script `run_DE_analysis.pl` and specifying the method as edgeR. The data were then normalised using `run_TMM_normalisation_write_FPKM_matrix.pl`. Finally, differential expression was analysed using the script `analyse_diff_expr.pl`. For *R. reniformis* differentially expressed transcripts were defined as a minimum of 4 fold change between conditions with a minimum p-value of 0.001. For *G. rostochiensis* only two replicates were available for each condition (section 6.4.4.2) and so to ensure confidence in the results the minimum p-value was decreased by two orders of magnitude to 0.00001. The significantly differentially expressed transcripts were clustered based on their expression using the `define_clusters_by_cutting_tree.pl`

script and setting the percentage height to 40. This produced 4 clusters for each species, an up-regulated and a highly up-regulated for each life stage. The above pipeline was carried out for Trinity transcripts and Trinity components but only Trinity components were used in downstream analyses.

6.3.7. Gene predictions

6.3.7.1. Open reading frame prediction - Transdecoder

Individual libraries for *G. rostochiensis* were mapped to the available genome as described above. Cufflinks was used on each library to generate a set of predicted transcripts. These were then merged using Cuffmerge to create a genome wide data set. Finally transdecoder was used on the Cuffmerge output to generate a best scoring open reading frames (ORFs) file.

6.3.7.2. CEGMA

CEGMA was used to predict highly conserved Core Eukaryotic Genes (CEGs) present in both transcriptomes and the available genome. These predictions were used as a measure of quality for both genome and transcriptome assemblies. Default parameters were used for all CEGMA predictions.

6.3.8. Bioinformatic comparisons

6.3.8.1. BLAT

For *G. rostochiensis* the assembled transcriptome was compared to the assembled genome using BLAT, a BLAST-Like program with considerable speed increased for large files (Kent, 2002). The default setting of a minimum of 90% identity was used for all BLAT comparisons.

6.3.8.2. BLAST

Typical BLAST analyses were carried out using BLAST 2.2.29+ (Boratyn et al., 2013) either locally on custom databases or against the non-redundant server at <http://BLAST.ncbi.nlm.nih.gov/Blast.cgi> using the relevant parameters for each job.

Reciprocal best BLAST hits

Reciprocal Best BLAST Hits (RBBH) were generated by creating BLAST databases for the two relevant data sets using the make BLASTdb command. Each data set was compared by BLAST to the opposing database specifying the extended tabular output option. The output was filtered to leave only those BLAST hits with a minimum identity of 70% and a minimum coverage of 50% for both query and

subject. Best BLAST hits were defined by Bit score not E-value as the latter varies as a function of query/subject length.

Blast2GO

Blast2GO (Conesa et al., 2005) was used to annotate the functional categories up-regulated in each cluster. For each species, BLASTx against the non-redundant protein database was carried out on the up-regulated transcripts from each cluster, specifying xml output. The results were loaded into Blast2GO to assign Gene Ontology (GO) terms to each BLAST hit.

Bacterial/plant contamination removal

The entire transcriptome of *G. rostochiensis* was compared to the non-redundant nucleotide database using BLASTn. Transcripts were removed if the best BLAST hit had a nucleotide sequence identity greater than 70% across more than 70% of the length of the query to either a non-eukaryote or a plant, and no other nematode or animal sequences were identified.

6.3.8.3. Secreted protein pipeline

A pipeline for extracting secreted proteins from a data set was made on Galaxy. The first step in the pipeline was to predict signal peptides using SignalP 4.0 (Petersen et al., 2011), and discard any proteins without signal peptides. The remaining proteins were then tested for presence of transmembrane domains using TMHMM v2.0, and were discarded if they contained > 0 transmembrane domains.

6.4. Results

6.4.1. RNA extractions and quality assessment

Two methods were used to assess the quality of RNA extracted from nematode samples. Firstly Nanodrop was used to estimate the quantity and presence of any contaminants. Figure 6-2i shows a typical output from the Nanodrop. Samples with absorbance ratios of 260/280 nm greater than 2 and 260/230 nm between 2 and 2.2 were also analysed using the Bioanalyser (with the exception of sample S5). Figure 6-2ii shows a typical output from the Bioanalyser. Samples with an RNA Integrity Number (RIN) greater than 8 were considered sufficiently high quality for library preparation and RNA sequencing. Nanodrop and Bioanalyser results for all samples are summarised in Table 6-1.

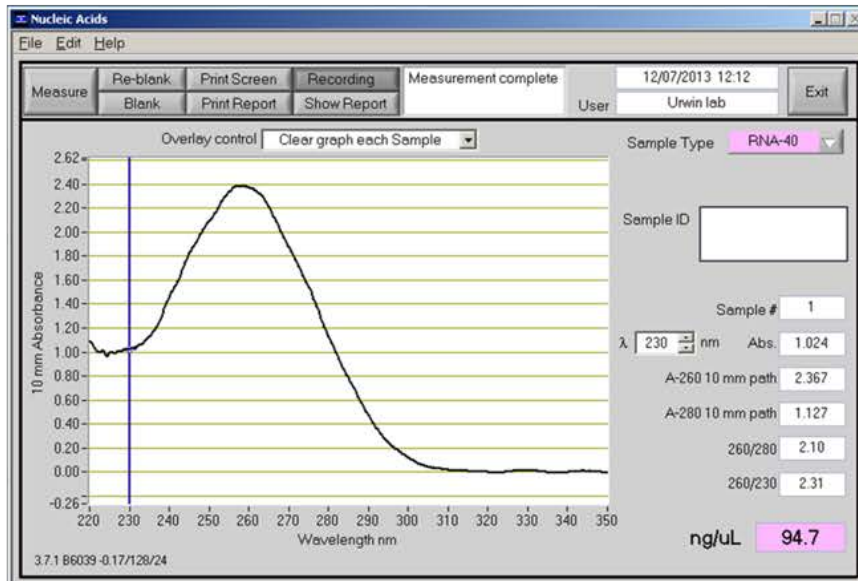
6.4.2. RNAseq reads and quality control

41 million paired end reads, from two replicates of *G. rostochiensis* J2s, were generated prior to this work, as part of the *G. pallida* genome sequencing project. An additional 9 libraries containing a total of 198 million paired end reads were generated in this project. Three replicates were generated for each of *G. rostochiensis* feeding stage, *R. reniformis* feeding stage and *R. reniformis* second stage juveniles. In total this represents approximately 240 million paired end reads containing 24.2 billion nucleotides (Table 6-2). For each library the reads were visualised in FastQC and the GC content, overrepresented sequences, and base quality were assessed. Figure 6-3 shows a typical FastQC output for GC content. A “shoulder” on the right hand side of the distribution for all *G. rostochiensis* feeding samples probably represents contamination from a bacterial/fungal species with a distinct GC content (Figure 6-3). Following visualisation, reads for each library were trimmed based on the quality score of each base. Figure 6-4 shows the read length distribution before trimming (all reads 100 bp), and after trimming prior to either assembly (more stringent), or mapping (less stringent). For *G. rostochiensis*, approximately 100 million reads for assembly and 104 million reads for mapping remained after cleaning. For *R. reniformis* approximately 84 million reads for assembly and 90 million reads for mapping remained after cleaning. In total the cleaning removed almost one quarter of all reads, and reduced the average length of each read by one third. The less stringent approach for the mapping reads resulted in approximately 7% more reads remaining after cleaning.

Table 6-3 Summary of quality control data for RNA samples.

Sample name	260/280	260/230	RIN	conc. Nanodrop (ng/ μ l)	conc. Bioanalyser (ng/ μ l)	Quantity (μ g)
JJ_S1_G.rostochiensis	2.15	2.46	9.7	317	367	9.51
JJ_S2_G.rostochiensis	1.98	2.1	9.2	52	45	2.184
JJ_S3_G.rostochiensis	2.17	2.39	9.6	106	173	4.452
JJ_S4_R.reniformis	2.1	2.31	9.2	94	163	3.948
JJ_S5_R.reniformis	2.02	1.7	9.2	82	92	3.444
JJ_S6_R.reniformis	2.14	2.04	9	110	136	4.62
JJ_S7_R.reniformis	2.09	2.29	9.5	178	242	7.476
JJ_S8_R.reniformis	1.98	2.31	9.7	70	85	2.94
JJ_S9_R.reniformis	2.04	2.19	9.7	87	118	3.654

i



ii

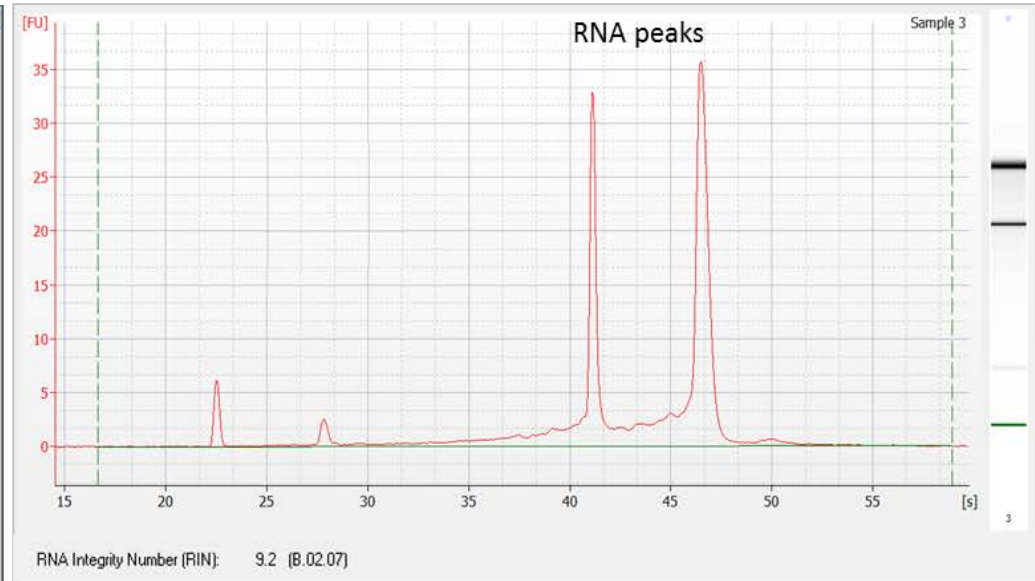


Figure 6-2 **RNA extraction quality control.** i) Typical output from Nanodrop. Absorbance ratios of 260/280 > 2 and 260/230 between 2 and 2.2 are considered free of contaminants. ii) Typical output from bioanalyser. RNA integrity number (RIN) greater than 8 is generally considered high quality.

Table 6-4 Summary of raw and trimmed reads.

Species	file name	sample alias	life stage	Pairs of reads	Pairs of reads after Trimmomatic for assembly	Total reads per species - assembly	Pairs of reads after Trimmomatic for mapping (%increase from assembly)
<i>G. rostochiensis</i>	ERR202487	J2 ros_s-1	J2	28,593,901	23,505,755	97,207,945	24,764,630 (5%)
<i>G. rostochiensis</i>	ERR202479	J2 ros_s0	J2	13,394,038	11,747,598		11,894,422 (1.2%)
<i>G. rostochiensis</i>	723_LIB5154_LDI39 95_CGATGT_L008	JJ_S1_G.rostochiensis	sedentary female	27,930,440	20,442,434		22,026,179 (7.7%)
<i>G. rostochiensis</i>	723_LIB5155_LDI39 96_TGACCA_L008	JJ_S2_G.rostochiensis	sedentary female	14,536,453	10,728,498		11,529,099 (7.4%)
<i>G. rostochiensis</i>	723_LIB5156_LDI39 97_ACAGTG_L008	JJ_S3_G.rostochiensis	sedentary female	41,207,045	30,783,660		33,036,159 (7.3%)
<i>R. reniformis</i>	723_LIB5157_LDI39 98_GCCAAT_L008	JJ_S4_R.reniformis	J2	21,494,288	16,084,603	84,795,396	17,240,653 (7.2%)
<i>R. reniformis</i>	723_LIB5158_LDI39 99_CAGATC_L008	JJ_S5_R.reniformis	J2	21,596,215	16,133,600		17,302,156 (7.2%)
<i>R. reniformis</i>	723_LIB5159_LDI40 00_CTTGTA_L008	JJ_S6_R.reniformis	J2	15,552,692	11,470,292		12,319,441 (7.4%)
<i>R. reniformis</i>	723_LIB5160_LDI40 01_ATGTCA_L008	JJ_S7_R.reniformis	sedentary female	12,480,345	9,092,531		9,764,782 (7.3%)
<i>R. reniformis</i>	723_LIB4989_LDI40 02_CCGTCC_L008	JJ_S8_R.reniformis	sedentary female	23,362,097	17,285,378		18,575,173 (7.5%)
<i>R. reniformis</i>	723_LIB5161_LDI40 03_GTGAAA_L008	JJ_S9_R.reniformis	sedentary female	19,998,683	14,728,992		15,800,532 (7.2%)

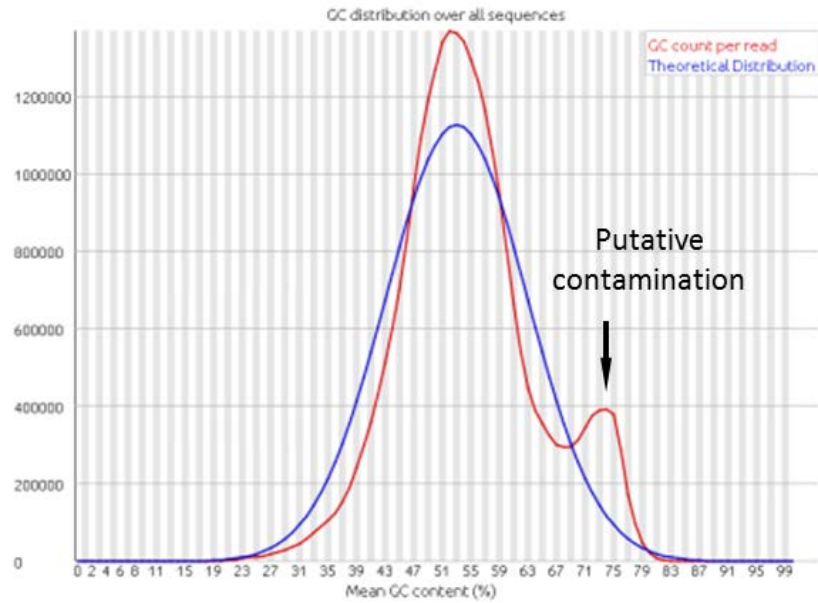


Figure 6-3 **FastQC report of GC content of *G. rostochiensis* sedentary female reads identifies putative contamination.** One of the sedentary female replicate samples for *G. rostochiensis* visualised in FastQC. This analysis of mean GC content highlighted a potential source of contamination, represented by a peak of higher GC content to the right of the main peak. If this were a *bona fide* feature of the transcriptome of *G. rostochiensis* it should be present in all samples.

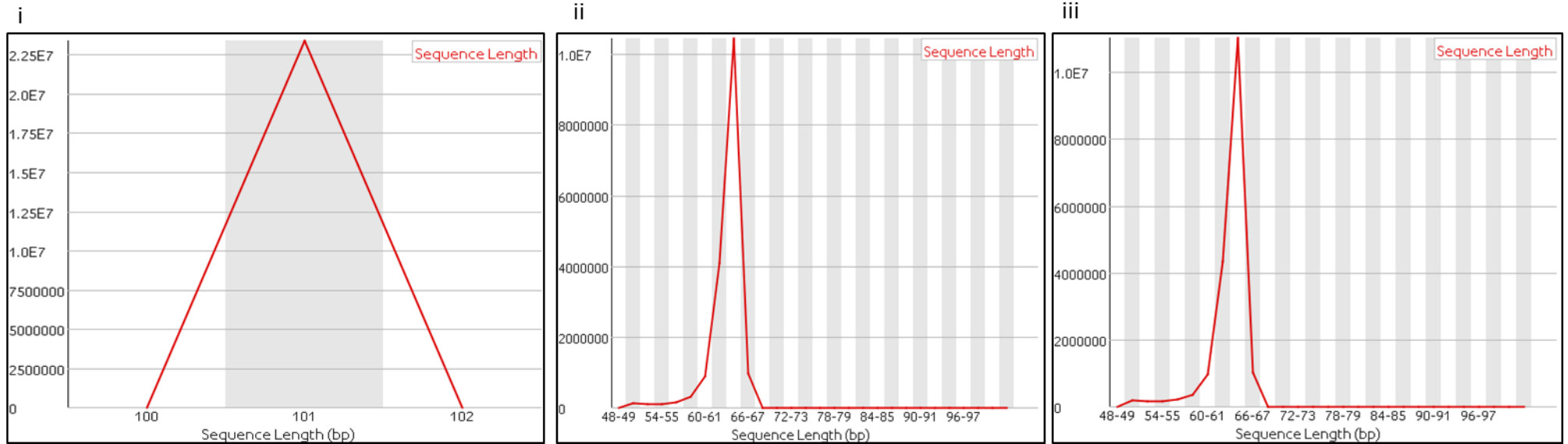


Figure 6-4 **Length distribution of reads following quality score trimming.** Panels show sequence length distribution from FastQC of RNAseq reads for the same library before trimming low quality reads (i), after trimming for assembly (ii), and after trimming for mapping (iii). Approximately 25% of reads were discarded between steps i and ii, and approximately 7% recovered between ii and iii.

6.4.3. Assembled transcriptomes and quality assessment

The process of removing duplicate reads and *in silico* normalisation was carried out prior to assembly. For both organisms two methods were used to assemble *de novo* transcriptomes and only the best was used for downstream analysis. For each species, all reads for all life stages were pooled prior to assembly.

6.4.3.1. R. reniformis - Trinity

Assembly using Trinity generated 103,957 transcripts containing 27,749 components with an N50 value of 1961 (components are used loosely as a proxy for genes). Core Eukaryotic Genes (CEGs) were predicted using CEGMA to assess the quality of the assembled transcriptome. Of the 458 CEGs expected to be present, 86% were present as full length transcripts and 96% were present as partial matches (partial defined as < 70% full length). We therefore make the assumption that the transcriptome of *R. reniformis* is largely complete, and the majority of what is present is largely full length.

6.4.3.2. R. reniformis – CLC

Assembly using CLC was carried out on both the *in silico* normalised reads and the non-normalised reads. The most complete CLC assembly was generated using the normalised reads. This assembly contained 42,641 transcripts with an N50 value of 898. Approximately 73% of CEGs were present although less than 50% of CEGs were present as full length transcripts.

6.4.3.3. G. rostochiensis – Trinity

Assembly using Trinity produced 113,928 transcripts containing 22,346 components or “genes”, with an N50 of 1617. The quality of the transcriptome for *G. rostochiensis* was evaluated in three ways. Firstly CEGMA identified 81% full length CEGs and 94% partial. For *G. rostochiensis* a high quality genome sequence was available, and therefore the transcriptome was compared against the genome assembly. The genome sequences contained more complete CEGs (93%) although there were 2 CEGs present in the transcriptome that were absent from the genome. This suggests the transcriptome has the ability to improve the quality of the genome assembly by 0.4%. Secondly the whole *G. rostochiensis* transcriptome was compared to the genome using BLAT. 94,047 of the 113,928 transcripts (82%) mapped with at least 90% id and 90% coverage. We therefore make the assumption that between 81% and 82% of all genes in the *G. rostochiensis* genome are present as full length transcripts in the *G. rostochiensis* transcriptome.

As a final confirmation, genes were predicted from the *G. rostochiensis* genome, and the number of genes was compared to the Trinity proxy for genes (components). There was remarkable similarity between the two methods, Trinity predicted 22,346 components from the transcriptome, and Transdecoder predicted 22,368 genes from the genome sequence.

6.4.3.4. *G. rostochiensis* – CLC

Assembly using CLC was carried out on both normalised and non-normalised reads. Like *R. reniformis*, the most complete *G. rostochiensis* CLC assembly was obtained from normalised reads. This contained 35,716 transcripts with an N50 value of 858. 70% of CEGs were present as partial transcripts and only 36% were present as full length genes.

Figure 6-5 shows a comparison between the assembly methods. For both species the Trinity assemblies produced more transcripts with a higher average length and higher N50 values. Trinity assemblies contained more CEGs, and more complete CEGs as summarised in Table 6-5. The lower N50, coupled with the CEG complete:partial ratio suggests CLC assemblies are less complete and transcripts present are more fragmented. For both species, only Trinity assemblies were used for all further downstream analysis.

6.4.3.5. Removal of contamination

In the *G. rostochiensis* sedentary female samples only, a subset of sequences had a higher GC content than the mean (Figure 6-3). These were hypothesised to originate from contamination. Therefore, all assembled sequences in the Trinity transcriptome of *G. rostochiensis* were removed if they had a best BLAST hit against either a non-Eukaryote or any species in the plant kingdom. This removed 139 transcripts from the *G. rostochiensis* transcriptome. The average GC content of the removed contaminants was less than expected based on Figure 6-3 (49.9%) compared to the rest of the *G. rostochiensis* transcriptome (47.1%) (Figure 6-6). This suggests that these reads putatively originating from contamination are largely underrepresented in the transcriptome assembly. Trinity has a default minimum transcript length of 200 base pairs, therefore suggesting that the potential contamination may have assembled poorly into <200 bp fragments and so was not present in the assembled transcriptome. This is to be expected as contamination is usually poorly represented.

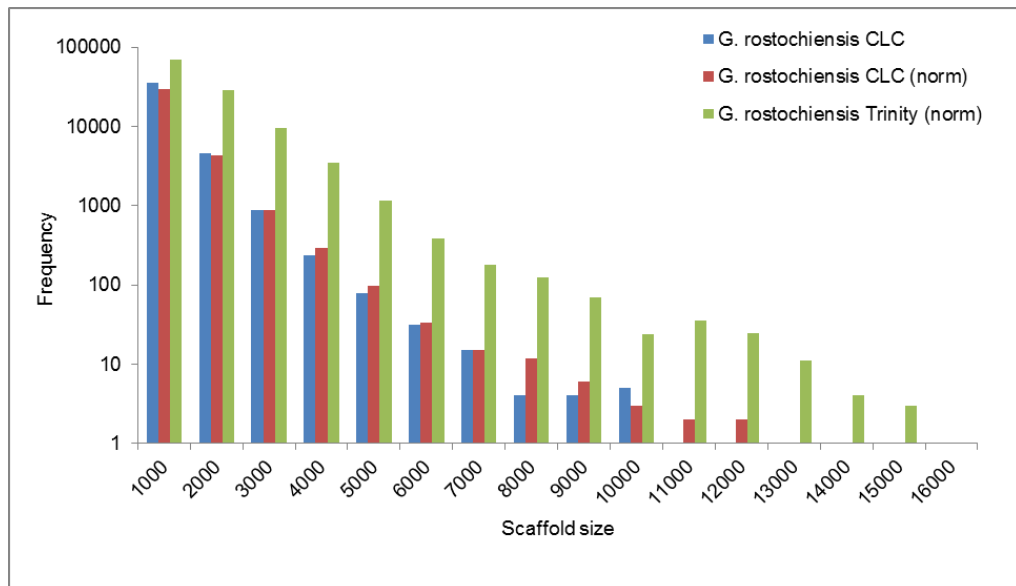
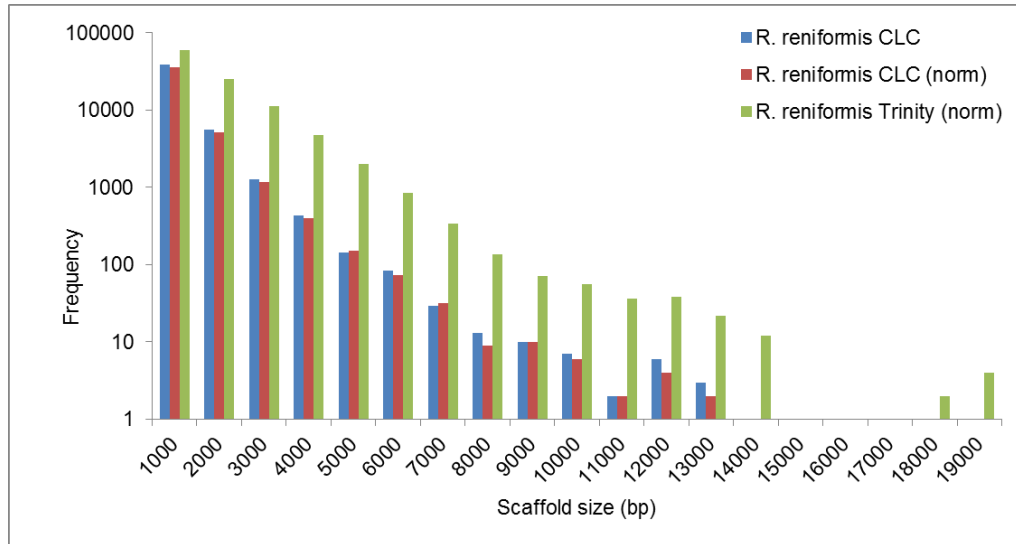


Figure 6-5 **Comparison of contig size between assembly methods.** Frequency distribution for size of transcripts produced by various assembly programs. For both species Trinity produced more transcripts, of a longer average length and a higher N50 compared to CLC. Surprisingly, CLC on *in silico* normalised reads (i.e. less overall information) produced better quality transcriptomes than on the full data set.

Table 6-5 Transcriptome assembly quality assessment.

Species	Assembly method	Full length CEGs	Partial CEGs
<i>G. rostochiensis</i>	CLC	24%	59%
<i>G. rostochiensis</i>	CLC on normalised reads	36%	70%
<i>G. rostochiensis</i>	Trinity on normalised reads	81%	94%
<i>R. reniformis</i>	CLC	43%	70%
<i>R. reniformis</i>	CLC on normalised reads	50%	73%
<i>R. reniformis</i>	Trinity on normalised reads	86%	96%

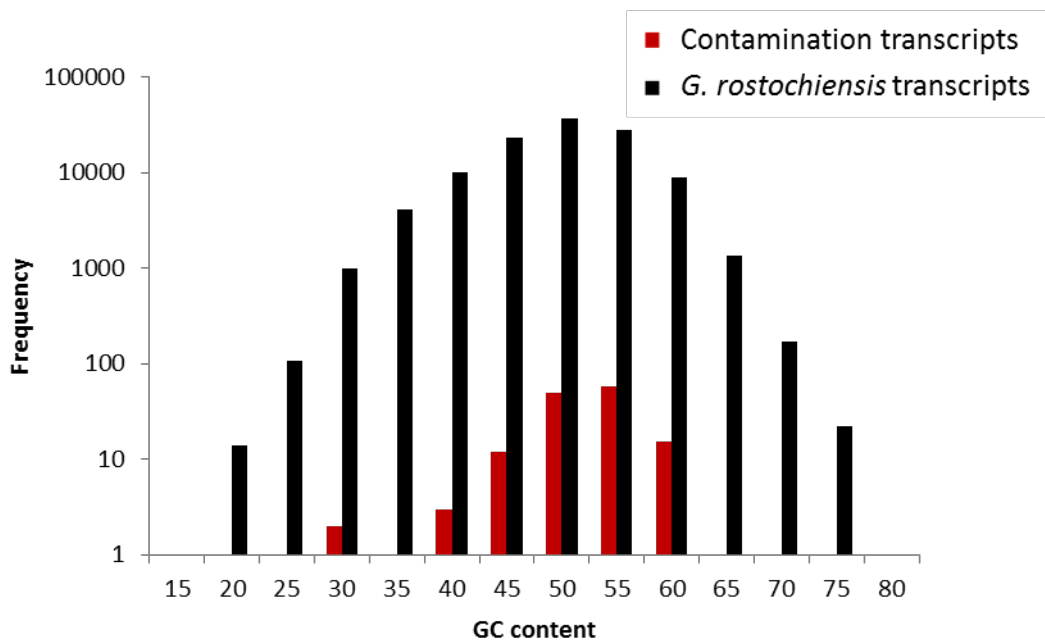


Figure 6-6 Analysis of GC content of transcripts removed from *G. rostochiensis* assembly. In total 133 of the 113,928 transcripts were removed. The removed transcripts had an average GC content of 49.9 %.

6.4.4. Gene predictions and differential expression

6.4.4.1. *R. reniformis* – within species gene cluster comparison

As Trinity assembles multiple transcripts for each gene, a non-redundant set of transcripts was made using CD-HIT-EST. This was then used for downstream analysis such as BLAST and functional Gene Ontology (GO) analysis. From an original set of 103,957 transcripts 63,493 were non-redundant at 90% nucleotide similarity.

RNA reads from each library were aligned to the assembled transcriptome. The RSEM package supplied with Trinity was used to estimate the number of reads mapping to each transcript. EdgeR was used to determine differentially expressed transcripts between J2 and sedentary female using a p value cut off of 0.00001. Differentially expressed transcripts were then identified from the non-redundant set for downstream analysis. Significantly differentially expressed transcripts could be assigned to 4 clusters based on their expression profile. This resulted in two clusters for each life stage. The first cluster, containing the most highly up-regulated sequences, contained 133 and 229 transcripts for J2 and sedentary female respectively. The second cluster, containing less highly up-regulated sequences, contained 1766 and 1440 transcripts for J2 and sedentary female respectively. In total 1899 transcripts were significantly up-regulated at J2 and 1669 transcripts were significantly up-regulated at sedentary female (Figure 6-7). This corresponded to 1254 and 1274 non-redundant transcripts for J2 and sedentary female respectively. These non-redundant sequences were used for downstream GO term predictions.

Figure 6-8 shows the GO terms assigned to various cellular components for each of the gene clusters. Remarkable similarity can be seen in the process represented in each category, especially considering these genes have already been separated by differential expression. Comparing all genes in each expression cluster identified very few significantly over represented sequences. For J2 the most over represented genes were involved in muscle contraction and locomotion located at the synapse, and for the sedentary female the most over represented genes were involved in oogenesis and gamete generation and macro molecular complexes. Although no difference is seen in the number of secreted proteins, these GO graphs were generated after significant differential expression analysis and so the genes encoding secreted proteins at J2 are not the same as at sedentary female, indicating a different suite of secreted proteins is required for each stage.

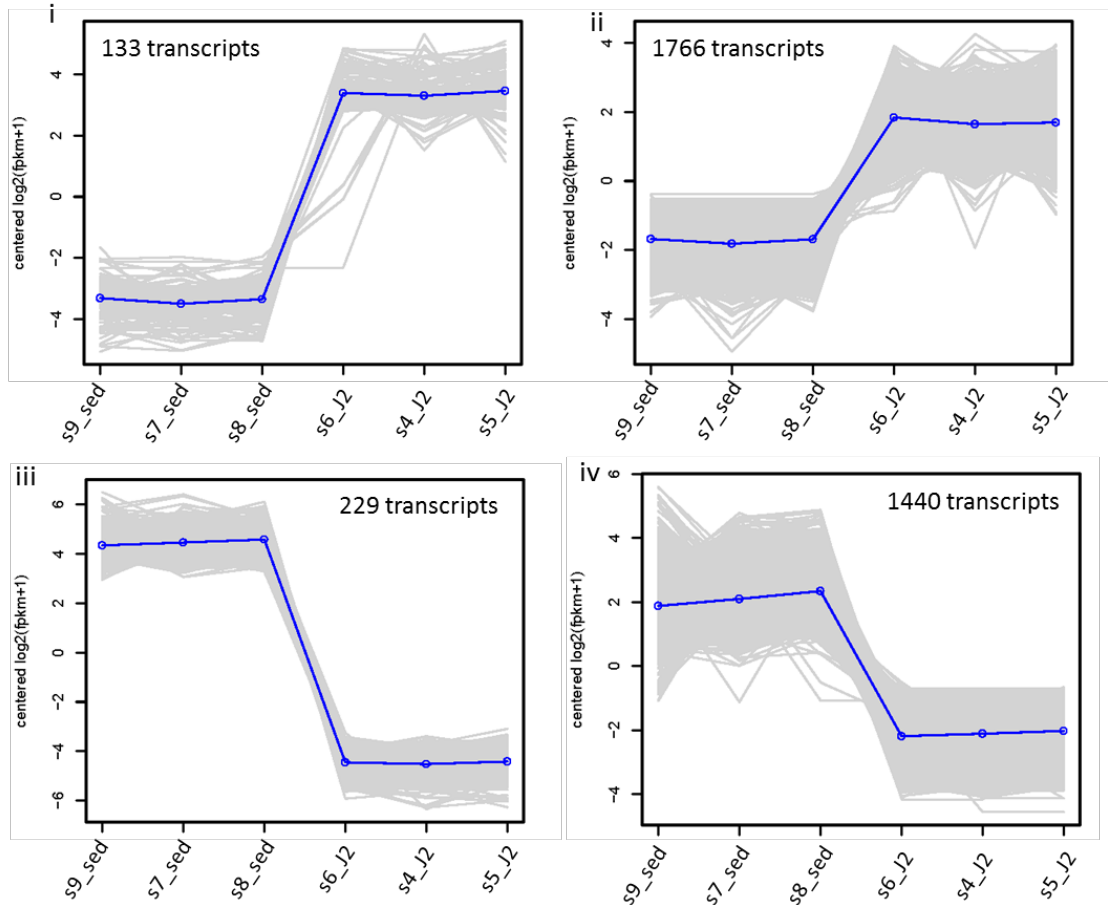


Figure 6-7 *R. reniformis* gene expression clusters. Four clusters were generated that describe differentially expressed transcripts between conditions. These can be subdivided into ‘life stage specific’ clusters describing genes up-regulated at J2 and genes up-regulated at sedentary female. The first for each life stage was the most highly up-regulated and contained 113 and 229 transcripts for J2 (i) and sedentary female (iii) respectively. The second for each life stage contained the less highly up-regulated and consisted of 1766 and 1440 for J2 (ii) and sedentary female (iv) respectively.

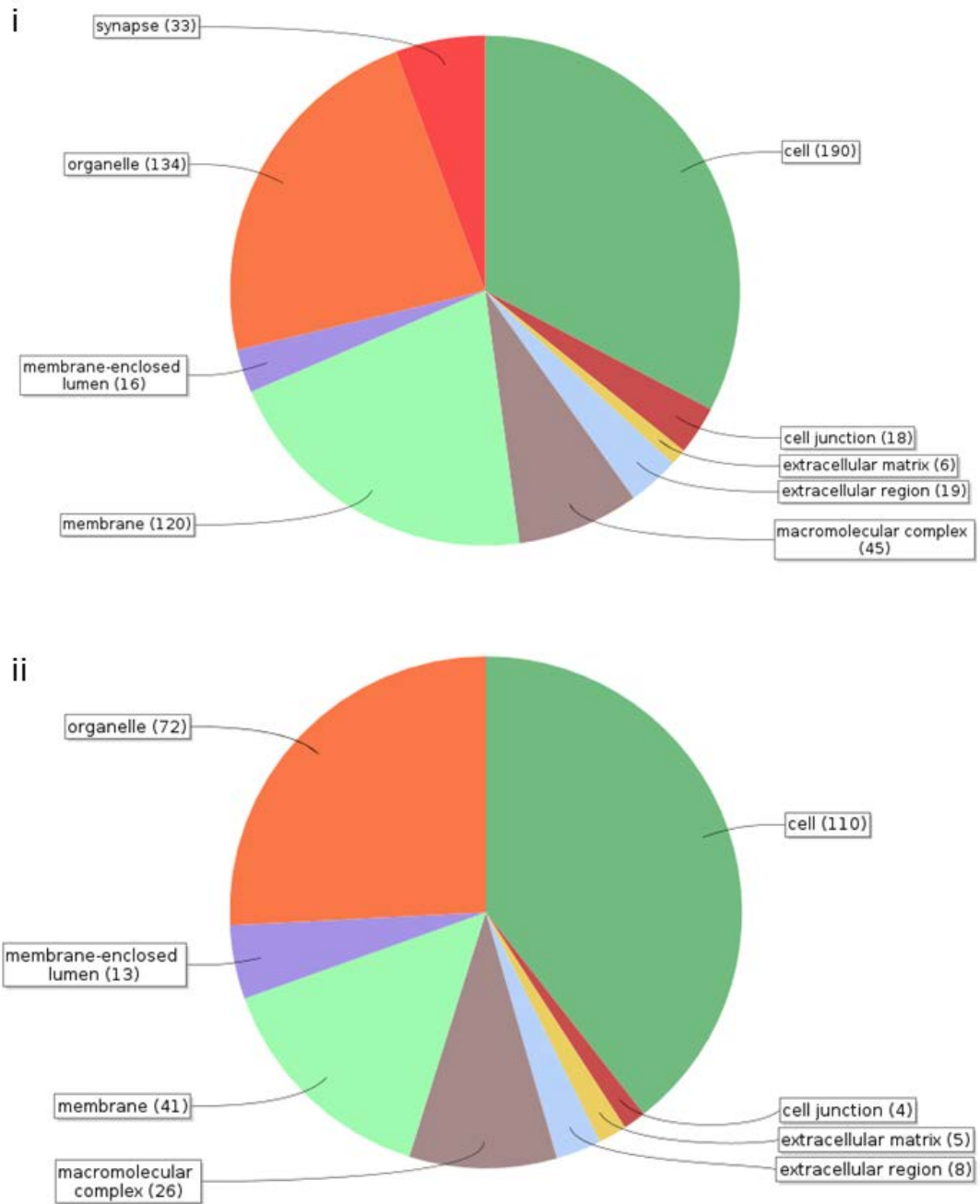


Figure 6-8 **Blast2GO cellular component analysis of differentially expressed gene clusters for *R. reniformis***. Over represented cellular components associated with the J2 stage (i) are frequently involved in muscle contraction and locomotion. Although no difference in the number of secreted proteins (extracellular matrix) between J2 and sedentary female (ii) was seen, this analysis was carried out after differential expression, indicating a different suite of secreted proteins is important for each stage.

6.4.4.2. *G. rostochiensis* – within species gene cluster comparison

The Trinity assembly of *G. rostochiensis* generated over 113,928 transcripts. These were grouped into non-redundant sequenced using CD-HIT. This reduced the number of transcripts to 59,839 non-redundant at 90% nucleotide similarity.

RNA reads from each library were mapped to the transcriptome using Bowtie2 and counts of mapped reads were made using RSEM. The data were initially checked and it appeared that one of the three samples from sedentary female nematodes was in very poor agreement with the other two. It was assumed this would drastically affect the validity of the downstream analysis and so this sample was removed from the mapping stage. This resulted in the use of just two samples for J2 and sedentary female. Although this will result in less statistical power, the use of a stringent p value cut-off allows confidence in the differentially expressed genes.

Normalised read counts were then used for differential expression analysis using EdgeR. Using a minimum p value cut-off of 0.00001 (two orders of magnitude lower than the default) and a minimum of 4-fold change, resulted in a total of 2079 differentially expressed transcripts. These were clustered based on expression profile into 4 clusters, two for each life stage. The first cluster, containing most highly up-regulated at each life stage, contained 177 and 85 transcripts for J2 and sedentary female respectively. The second cluster, containing less highly up-regulated sequences, contained 875 and 942 transcripts for J2 and sedentary female respectively. This corresponded to 699 and 637 non-redundant transcripts for J2 and sedentary female respectively (Figure 6-9).

For functional GO term analysis the full set of non-redundant up-regulated sequences was used. Figure 6-10 shows the GO terms assigned to various cellular components. Most notable significant differences between the two clusters are the over representation of locomotion genes located at the synapse in the J2 cluster and the over representation of secreted proteins in the sedentary female cluster.

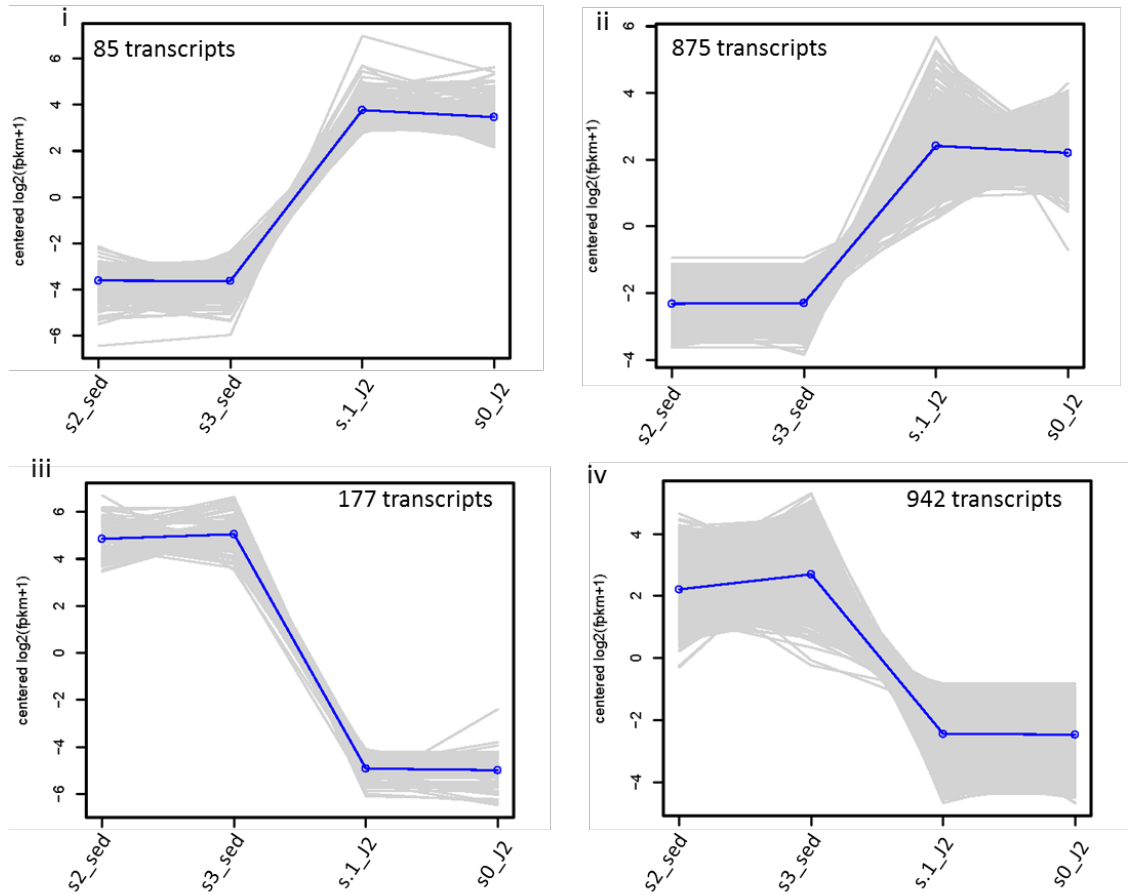


Figure 6-9 ***G. rostochiensis* gene expression clusters.** Genes are grouped into four clusters, two for each life stage. The first cluster, containing the most highly up-regulated sequences, contained 85 and 177 transcripts for J2 (**i**) and sedentary female (**iii**) respectively. The second cluster, containing less highly up-regulated sequences, contained 875 and 942 transcripts for J2 (**ii**) and sedentary female (**iv**) respectively.

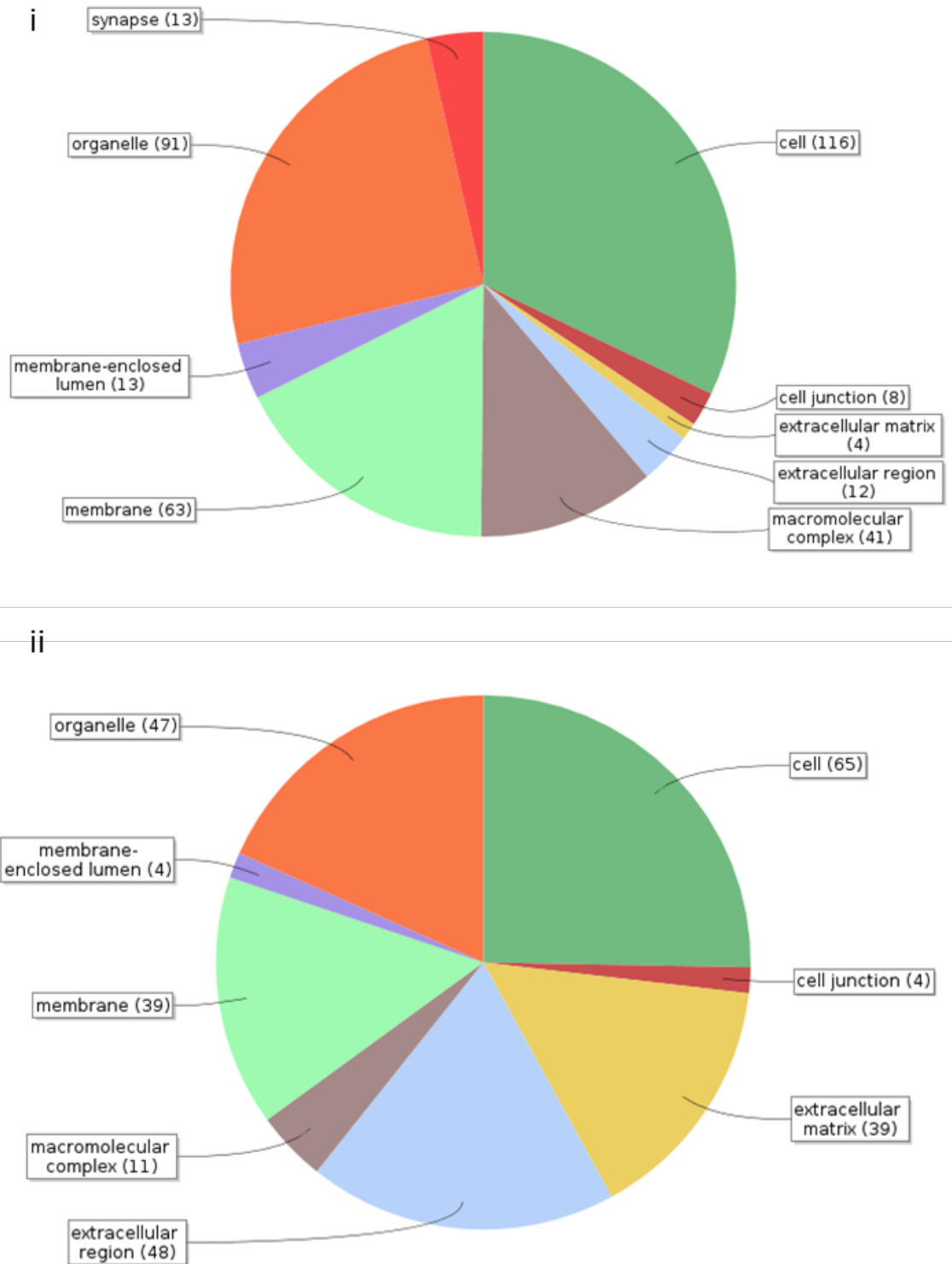


Figure 6-10 **Blast2GO cellular component analysis of differentially expressed clusters for *G. rostochiensis*.** A significant enrichment of genes involved in locomotion can be seen for the J2 gene clusters (**i**) by enrichment of the cellular components associated with the synapse. The sedentary female cluster (**ii**) has a significant increase in the number of genes associated with the extracellular matrix.

6.4.5. Comparing gene expression clusters between species

6.4.5.1. Putative homologs between species

A pair wise comparison was made between genes up-regulated in the same stages from the two species using Reciprocal BLAST Hit (RBH) to identify putative homologs. Of the 1,254 and 699 sequences up-regulated in *R. reniformis* and *G. rostochiensis* respectively, 164 were identified as potential homologs. However, for the 637 and 1274 sequences up-regulated at sedentary female, for *G. rostochiensis* and *R. reniformis* respectively, just 57 were identified as putative homologs. As the initial starting numbers of genes are almost identical this suggests that either the juvenile stages of the two species are more similar to one another than the feeding stages or that the developmental stages are not perfectly matched between species.

6.4.5.2. Functional categories in gene clusters between species

Comparing the Blast2GO results from *R. reniformis* and *G. rostochiensis* revealed several interesting differences. For both species, as expected, sequences associated with locomotion/neuromuscular function were over represented at the J2 stage, which is the motile stage in both organisms. For *G. rostochiensis* the sedentary female stage had a significant up regulation of cuticle collagens. Cuticle collagens were also present in *R. reniformis* at the sedentary female sample, but they were not over represented as several collagens were present at J2. Similarly, genes involved in reproduction were significantly over represented in *G. rostochiensis* feeding stages. Conversely in *R. reniformis* some these functions were represented at the J2 stage.

6.4.5.3. Functional categories in secreted proteins at sedentary female stage between species

A comparison was also made between the two species only using the secreted proteins up-regulated in the sedentary female sample. 26 GO IDs are common in the sedentary female secreted proteins of the two species (Figure 6-11i). The functional categories that are represented by these common GO IDs are shown in Figure 6-12. "Protein binding" is abundant in the common secreted proteins, as is "oxidoreductase activity". Unsurprisingly the most common prediction for localisation of these secreted proteins is to the extracellular matrix (Figure 6-12).

A comparison of the GO descriptions, as shown in Figure 6-11ii, shows that there is a single common description between the two species that corresponds to

glutathione synthetase (represented in the previous analysis as oxidoreductase), highlighting a potentially important and conserved function in the plant-parasite interaction. Several glutathione synthetase genes were identified from the candidate feeding tube pipeline in *G. pallida* (as described in Chapter 4). All glutathione synthetase genes were extracted from the *R. reniformis* and *G. rostochiensis* transcriptomes. A summary of their expression at sedentary female on a per component basis can be seen in Figures 6-13 and 6-14 for *G. rostochiensis* and *R. reniformis* respectively. All normalised expression values <1 were set to 1 in order to give conservative estimates on fold change from non-feeding to feeding. Expression values are expressed as a ratio of sedentary female:J2, therefore bars above 1 correspond to an up regulation in feeding stages and bars below 1 correspond to an up regulation at non-feeding stages. Bars in red highlight those component names for which at least one isoform is predicted to be secreted. For *G. rostochiensis* 9 of the 15 components have at least one isoform predicted to be secreted, for *R. reniformis* 20 of the 40 components contain at least one isoform that is predicted to be secreted.

Moreover, a superoxide dismutase and a glutathione peroxidase were present in the *R. reniformis* sedentary female cluster (also involved in oxidoreductase activity). These sequences were used as BLAST queries against *G. pallida* and *G. rostochiensis* to identify homologs. Several sequences could be identified for each species, some of which encoded secreted proteins expressed during the feeding stages.

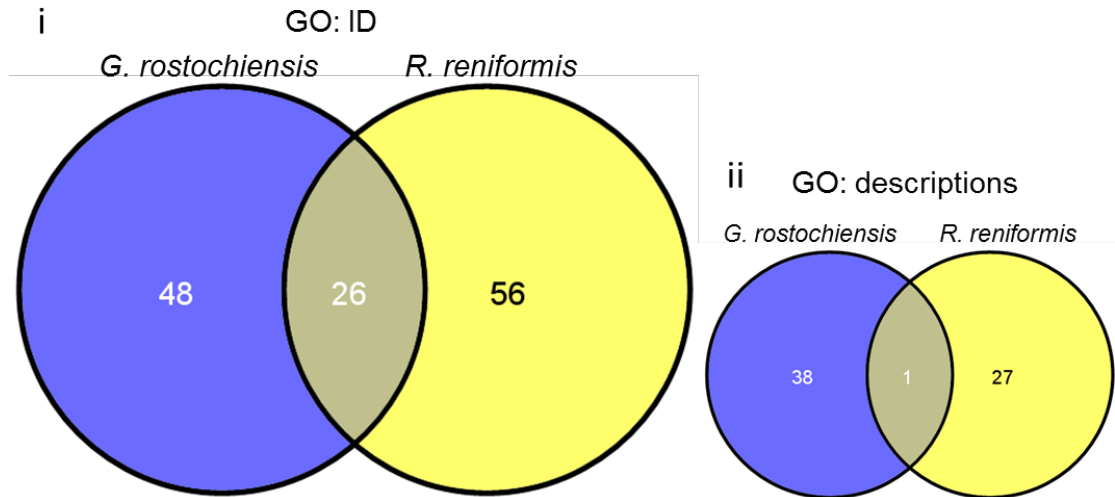


Figure 6-11 **Gene ontology terms for secreted proteins at sedentary female for *G. rostochiensis* and *R. reniformis*.** i) 26 Gene Ontology (Go) IDs are common between the two sets of secreted proteins. ii) If just the GO descriptions are used for comparison, only one is common between the two sets of secreted proteins. This description corresponds to a glutathione synthetase.

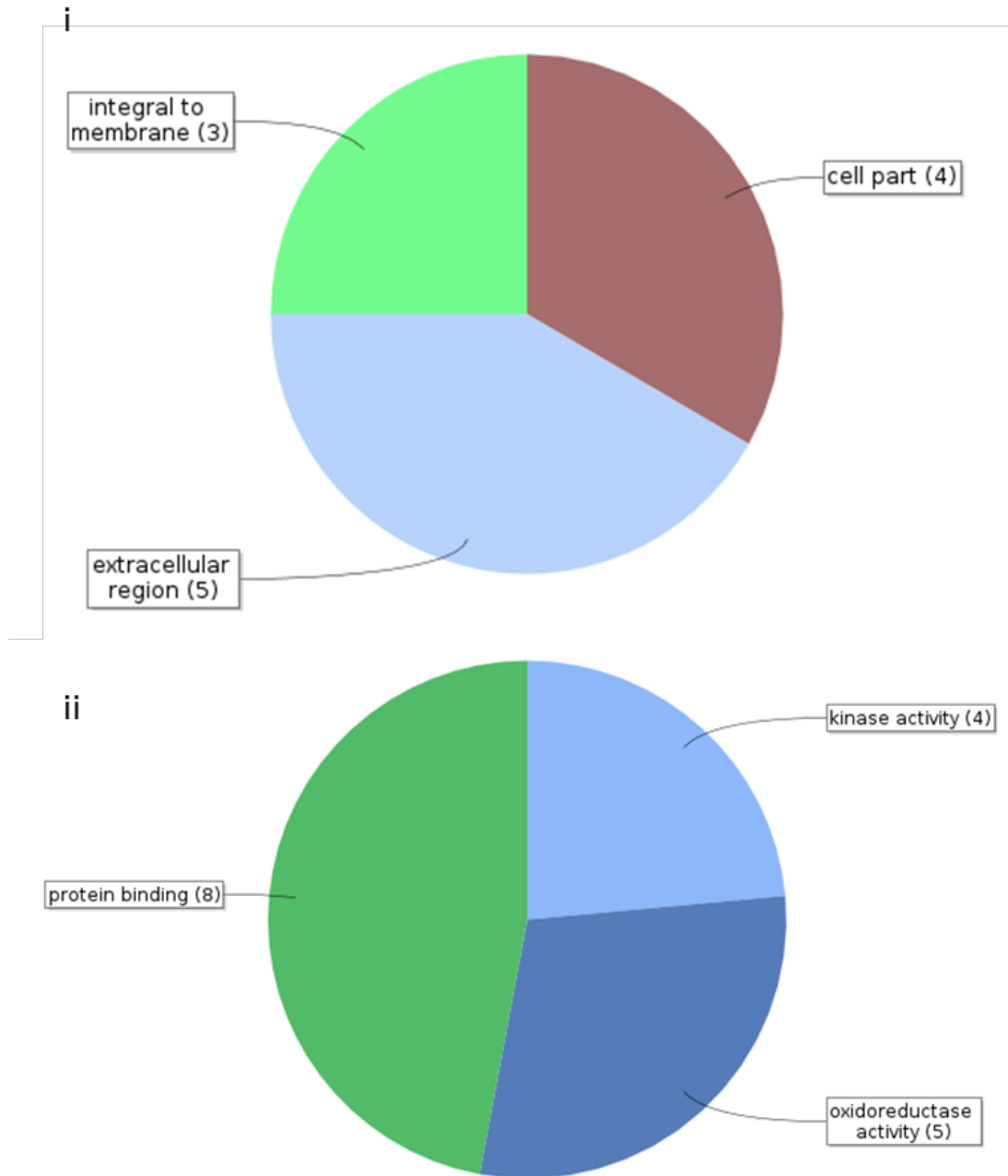


Figure 6-12 **Blast2GO comparison between secreted proteins up-regulated at sedentary female for *R. reniformis* and *G. rostochiensis*.** **i)** The cellular component associated with the common secreted proteins, unsurprisingly the most common is the extracellular matrix. **ii)** The molecular function associated with the common secreted proteins, protein binding and oxidoreductase activity make up the majority of functions.

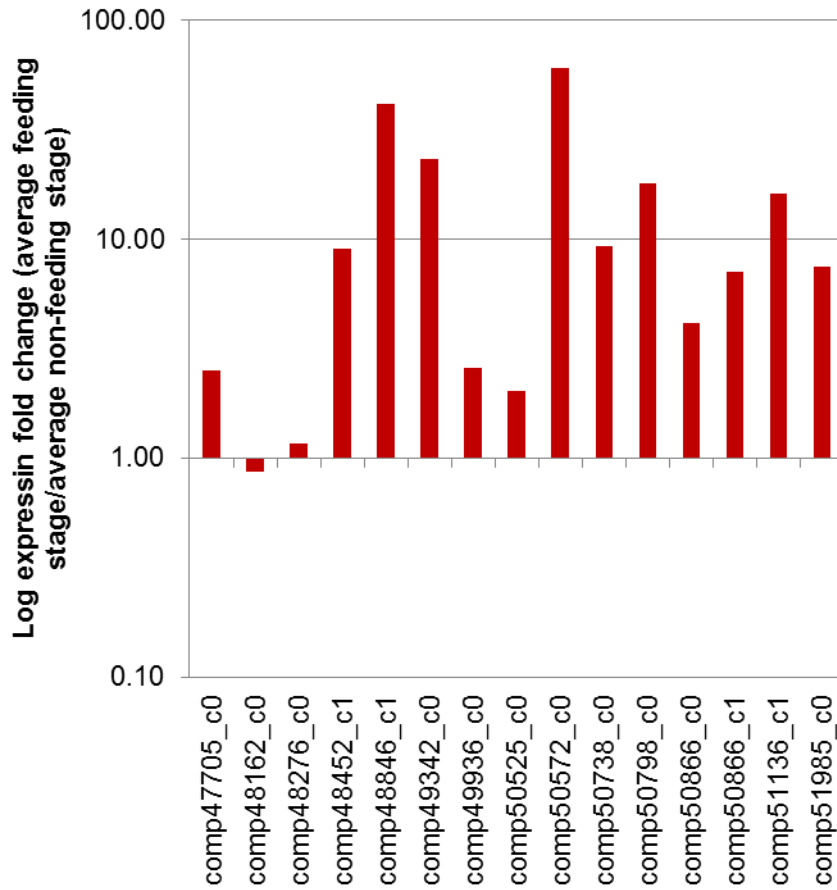


Figure 6-13 **Expression of glutathione synthetase genes in *G. rostochiensis*.** Expression is displayed as fold change between average normalised non-feeding stage compared to feeding stage on a per gene basis. All genes have at least one isoform predicted to be secreted.

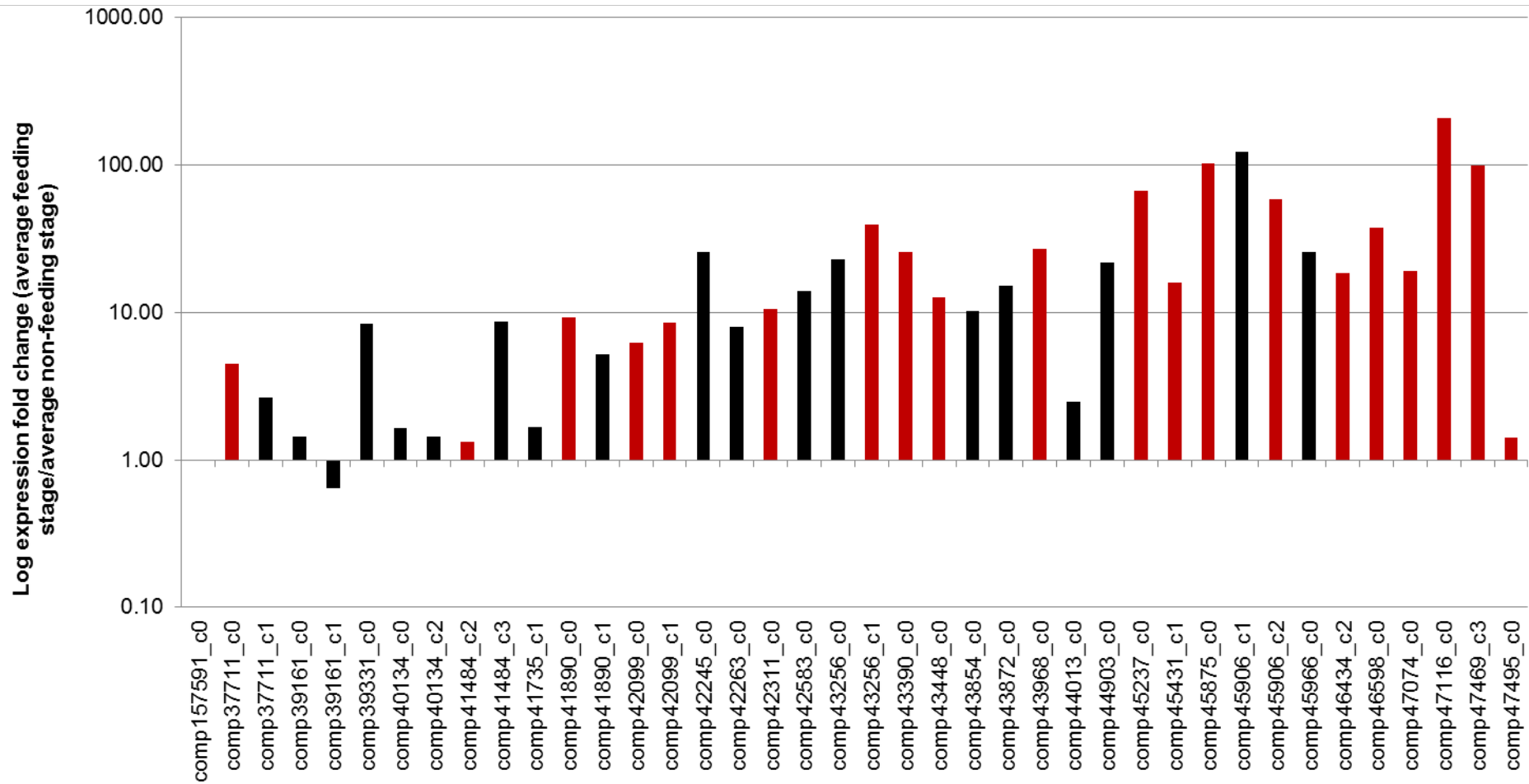


Figure 6-14 **Expression of glutathione synthetase genes in *R. reniformis*.** Expression is displayed as fold change between average normalised non-feeding stage compared to feeding stage on a per gene basis. Red bars indicate genes that have at least one isoform predicted to be secreted.

6.4.6. Whole transcriptome comparison between *R. reniformis* and *G. rostochiensis*

6.4.6.1. BLAT

The whole transcriptomes of *R. reniformis* and *G. rostochiensis* were compared to the *G. rostochiensis* genome using BLAT. For *G. rostochiensis* 94047 of the 113928 transcripts (82%) mapped to the *G. rostochiensis* genome with > 90% coverage and > 90% identity. For *R. reniformis* 213 transcripts could be aligned to the *G. rostochiensis* genome with > 90% coverage and > 90% identity, and 2363 with 70% identity and 70% coverage.

6.4.6.2. Reciprocal Best BLAST Hit (RBBH) analysis

The transcriptomes of *R. reniformis* and *G. rostochiensis* were also compared by reciprocal best BLAST hit at the nucleotide level. This identified 2831 putative homologs between the two transcriptomes. The percentage identity for the putative homologs ranged from 70% to 100% with an average value of 76.4% (Figure 6-15). This analysis revealed a group of 43 highly conserved homologues that are between 99 and 100% identical. For *R. reniformis* one of these highly conserved genes was present in the J2 expression cluster, and seven were in the sedentary female expression cluster. For *G. rostochiensis*, none of the highly conserved genes could be identified from the J2 expression cluster, and 19 could be identified from the sedentary female expression cluster. However, these two lists were mutually exclusive. None of the highly conserved genes in any of the *R. reniformis* clusters corresponded to any of the highly conserved genes in any of the *G. rostochiensis* clusters. This suggests that the most highly conserved genes in each species are not expressed at the same time in the nematode life cycle (Figure 6-16). Further, that the genes involved in the feeding stages for *G. rostochiensis* are present in *R. reniformis* at 99 – 100% identity but are not expressed at the feeding stages, and vice versa, which may reflect differential expression on different host species.

Therefore a comparison was made to another cyst nematode, feeding on the same host as *G. rostochiensis* (*G. pallida*). The majority of the 19 highly conserved genes in the *G. rostochiensis* feeding cluster are present with very high similarity in *G. pallida*. Of those that are present, all but one are expressed exclusively during the feeding stages (Figure 6-16).

Of the 19 highly conserved genes present in the *G. rostochiensis* feeding stage cluster, 16 have no assigned GO terms. The remaining three correspond to a

nematode cuticle collagen, a hydroxyacyl-coenzyme and a C-type lectin domain containing protein. The first two are present as GO terms in the *R. reniformis* feeding stage secreted cluster but are not part of the subset of highly conserved genes. Of the seven highly conserved genes present in the *R. reniformis* sedentary female cluster one represents a collagen-like protein and the other a C-lectin domain containing protein. These are not the direct orthologues of the similar sequences in the *G. rostochiensis* sedentary female cluster. None of the *R. reniformis*-specific feeding stages genes has corresponding homologous sequences in the *G. pallida* genome assembly, despite them being present at 99/100% identity in *G. rostochiensis*.

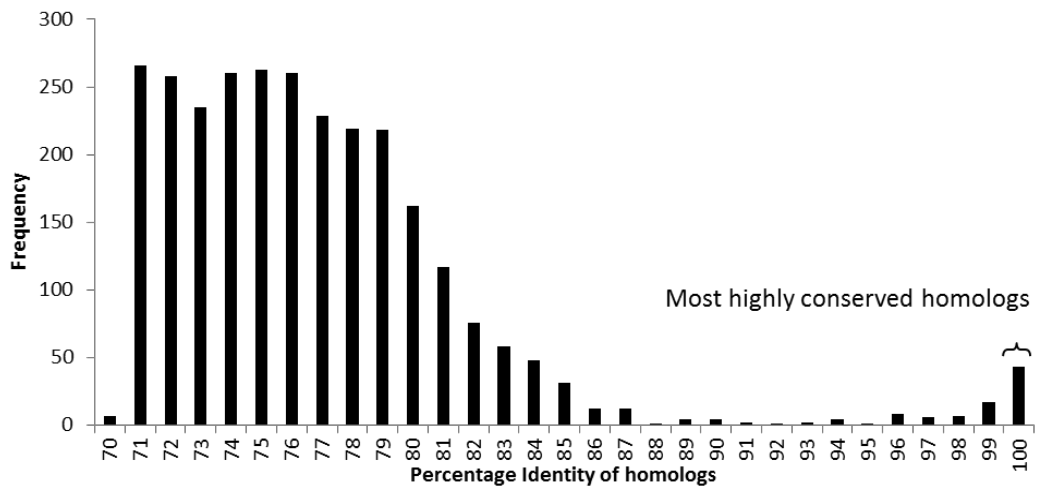


Figure 6-15 Frequency distribution of percentage identity of putative homologs between *R. reniformis* and *G. rostochiensis*. The frequency distribution clearly highlights a two tailed distribution with a sub set of 43 very highly conserved homologs.

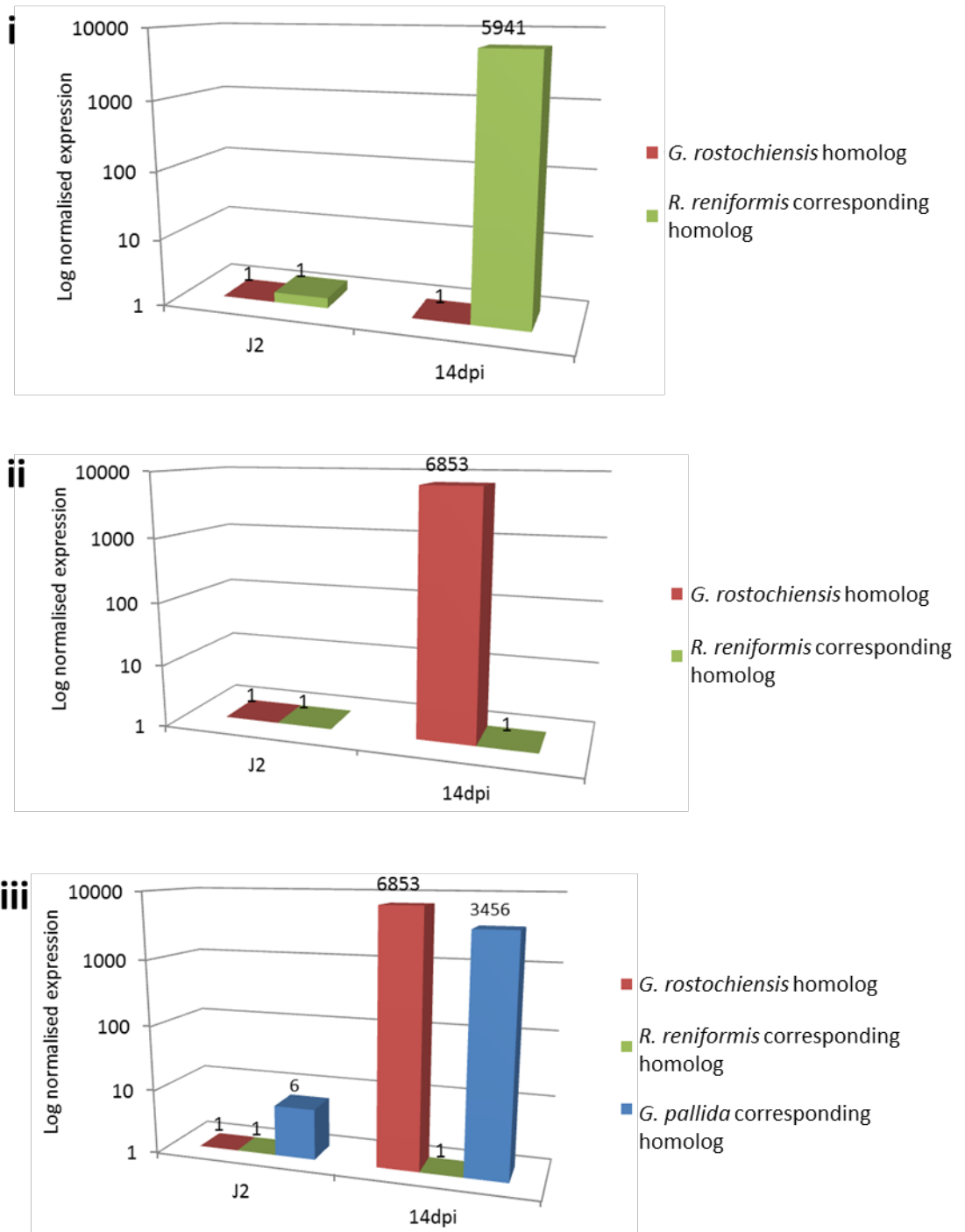


Figure 6-16 **Comparison of expression between species of two example highly conserved homologs.** **i)** *R. reniformis* and *G. rostochiensis* contain a very similar transcript (99/100% identity), yet only *R. reniformis* expresses it. **ii)** Conversely for a completely unrelated transcript, both *R. reniformis* and *G. rostochiensis* contain very similar sequences (99/100% identity), yet only *G. rostochiensis* expresses it. **iii)** In this case the corresponding *G. pallida* homolog is also expressed at sedentary female.

6.4.7. Presence of *G. pallida* candidate feeding structure genes in the *R. reniformis* and *G. rostochiensis* transcriptomes

A series of candidate feeding structure genes were identified (Chapter 4) and characterised (Chapter 5) from *G. pallida*. In brief, these candidates include the 444 gene family, the 448 gene family, 3453, and 176. The presence of these genes in the newly assembled *G. rostochiensis* and *R. reniformis* transcriptomes was evaluated using BLAST. A good quality draft genome assembly of *G. rostochiensis* was available, as well as a very poor quality draft genome assembly for *H. glycines*. Where possible both genome assemblies were used for these analyses but the nature of the *H. glycines* assembly made it very difficult to draw firm conclusions about the presence or absence of genes. For continuity in all phylogenetic trees *G. pallida* sequences are shown in blue, *G. rostochiensis* sequences are shown in red, *R. reniformis* sequences are shown in green and *H. glycines* sequences are shown in black.

6.4.7.1. Gene 176

A single complete 176 gene was identified from the *G. pallida* genome assembly. A second highly similar partial copy was present slightly 3' to the first, although it seemed to be missing a start codon. The single complete 176 gene was used to BLAST against the newly assembled *R. reniformis* and *G. rostochiensis* transcriptomes, and the available *G. rostochiensis* and *H. glycines* genomes. For all other cyst nematodes tested we were able to identify two different 176 genes, named A and B. However, 176 was not present in the *R. reniformis* transcriptome. The raw RNA seq reads for *R. reniformis* were mapped back to all 176 sequences from all cyst nematodes with very permissive settings but no matches were found. We therefore conclude that 176 appears to be cyst nematode specific. Figure 6-17 shows a phylogenetic tree derived from an alignment of all genomic and cDNA/transcriptomic sequences for 176. Despite the inclusion of genomic and cDNA sequences in the same phylogenetic tree the sequences always group by species. Expression of the 176 genes from the *G. rostochiensis* transcriptome and the *G. pallida* genome is summarised in Figure 6-18. Expression is displayed as fold change on a per gene basis, where bars above 1 indicate up regulation at sedentary female and bars below 1 indicate up regulation at J2. In both species of PCN gene 176 was highly up-regulated in feeding stage nematodes.

6.4.7.2. Gene 3453

A single 3453 gene was present in the latest version of the *G. pallida* genome assembly. It was clear that the assembly software had difficulty with genomic regions containing this class of sequences. In previous versions of the *G. pallida* genome assembly there were either two similar sequences present or it was absent. 3453-like sequences were identifiable in both *R. reniformis* and *G. rostochiensis* transcriptome assemblies. Interestingly the Trinity assemblies for both *R. reniformis* and *G. rostochiensis* generated multiple components. In an alignment of all 3453-like sequences they did not group by species, but instead by component. It appears there is a second copy that is absent from the current *G. pallida* genome or transcriptome assemblies (Figure 6-19).

Figure 6-20 summarises the expression of the 3453 genes across species on a per gene basis. Although 3453 is present in *R. reniformis*, and is very similar to the *G. pallida* and *G. rostochiensis* sequences, it is slightly up-regulated in juveniles in *R. reniformis* whereas the PCN sequences are highly up-regulated at feeding stages.

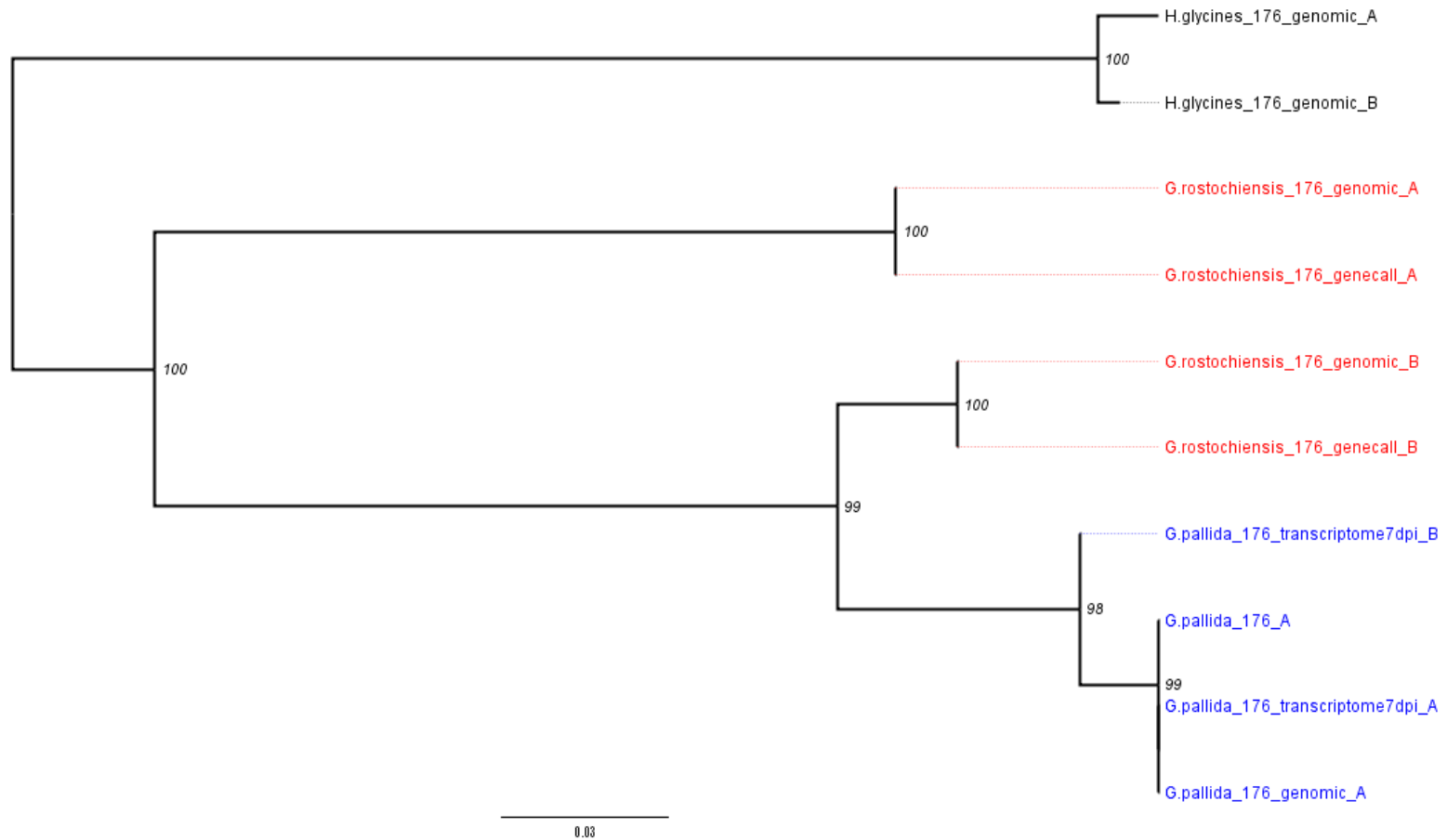


Figure 6-17 **Phylogenetic tree of 176 genes in cyst nematodes.** Phylogenetic tree was based on DNA sequence and generated using a maximum likelihood model with 100 bootstraps (shown as node labels). 176 genes group by species as opposed to isoform. 176-like sequences are not present in the *R. reniformis* transcriptome assembly. Scale bar indicates expected substitutions per base.

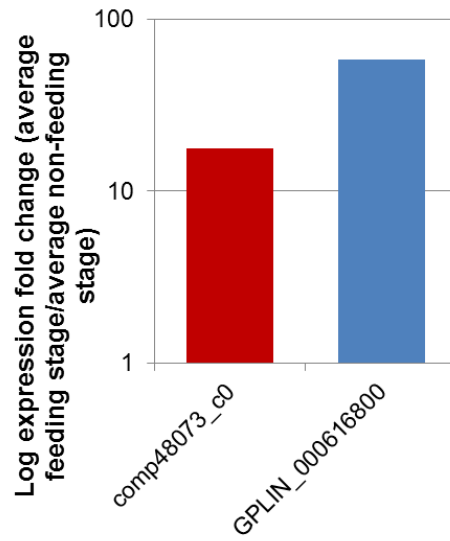


Figure 6-18 **Comparison of expression of 176-like sequences in *G. rostochiensis* and *G. pallida*.** Gene expression is displayed as fold-change between feeding and non-feeding stages. 176 genes are 17-fold up-regulated at feeding compared to non-feeding stages in *G. rostochiensis* (red) and 58-fold in *G. pallida* (blue).

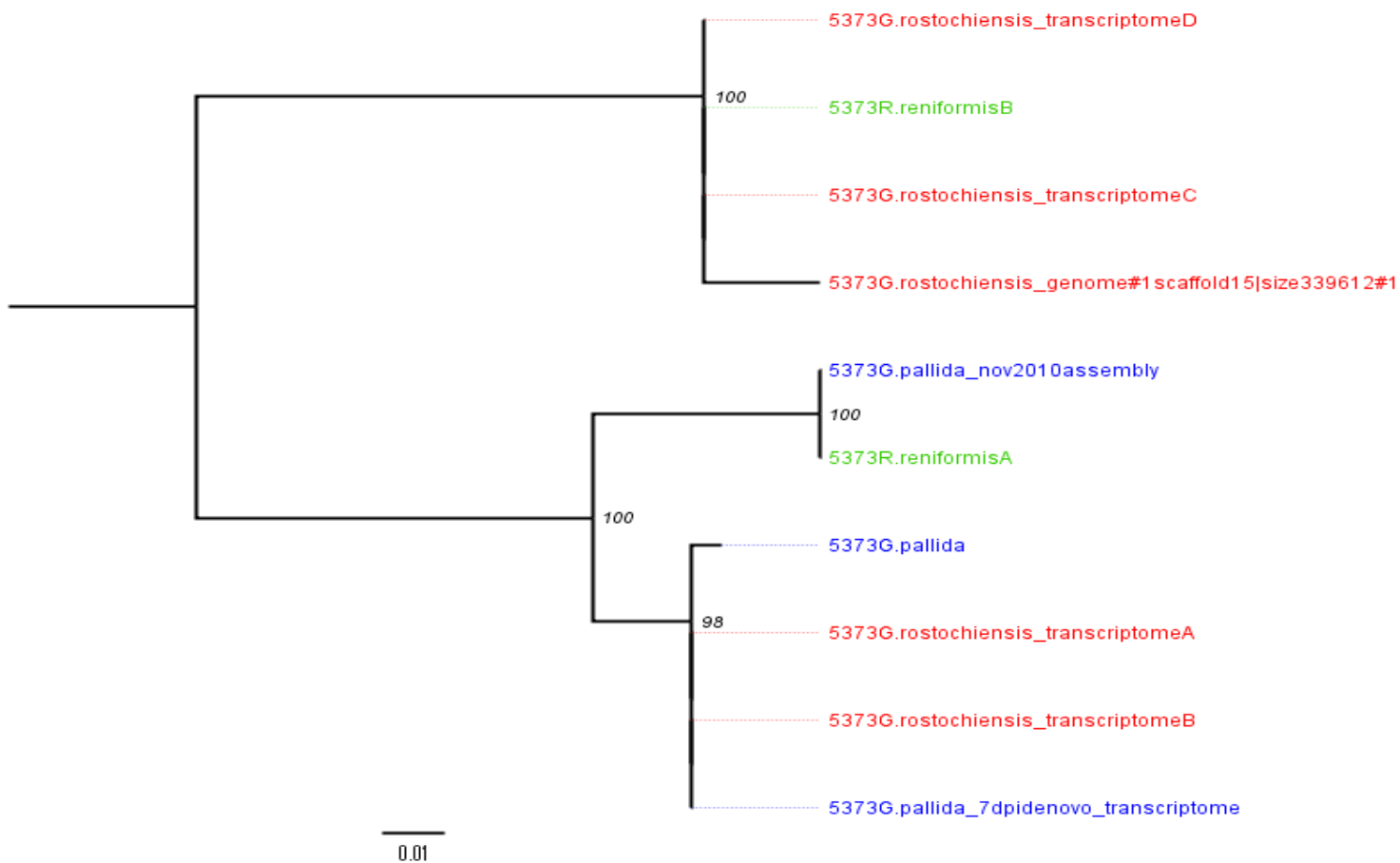


Figure 6-19 **Phylogenetic tree of 3453 genes in cyst and reniform nematodes.** Phylogenetic tree was based on DNA sequence and generated using a maximum likelihood model with 100 bootstraps (shown as node labels). 3453 genes group by sequence subgroup as opposed to species. Scale bar indicates expected substitutions per base.

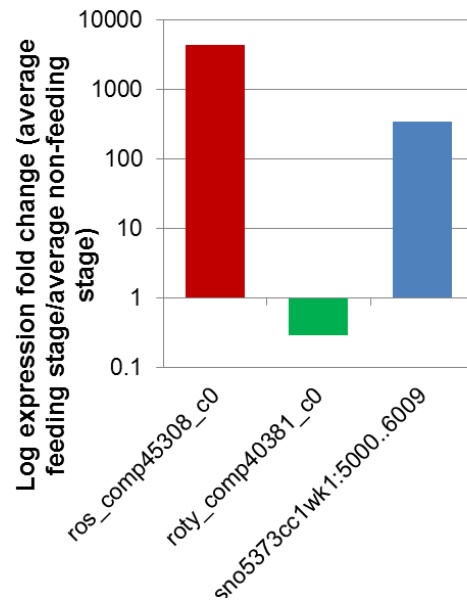


Figure 6-20 **Comparison of expression of 3453-like genes in *G. rostochiensis*, *R. reniformis* and *G. pallida*.** Gene expression is displayed as fold change for feeding to non-feeding stages. 3453 transcripts are 3.5-fold up-regulated at J2 for *R. reniformis* (green), but 4300 and 360 fold up-regulated at sedentary female for *G. rostochiensis* (red) and *G. pallida* (blue).

6.4.7.3. 444 gene family

As described in detail in Chapters 4 and 5, 444 genes are part of a complex multi-gene family. For BLAST analysis a single representative from each of the three subfamilies A, B and C was used. As described in section 5.3.5.7 several 444-like sequences were identified from *G. rostochiensis* genomic DNA by PCR. It was therefore expected to be present in the transcriptome. We were able to identify members of subfamilies A, B and C in the *G. rostochiensis* transcriptome. However, they were all grouped under the same component, as different isoforms, despite them locating to the individual groups of the phylogenetic tree (Figure 6-21). Only subfamilies A- and B- like sequences were present in the *R. reniformis* transcriptome (Figure 6-21).

The 444 family appears to have very high conservation at the protein level, even outside the cyst nematodes. A protein alignment of an *R. reniformis* subfamily B sequence and the corresponding *G. pallida* and *G. rostochiensis* sequences shows over 93% conservation across the full length of the protein (Figure 6-22). Sequence diversity is much higher in the introns between *G. pallida* and *G. rostochiensis* (no intron sequences are available for *R. reniformis*).

Expression of the 444-like sequences present in the various genomes and transcriptomes is summarised in Figure 6-23. Although we have identified over 80 unique genomic sequences from *G. pallida* only three are present across all genome assembly versions. Despite the remarkable conservation of the *R. reniformis* 444 genes these sequences are not up-regulated at sedentary female in the manner observed for the similar sequences in both *G. pallida* and *G. rostochiensis* (Figure 6-23).

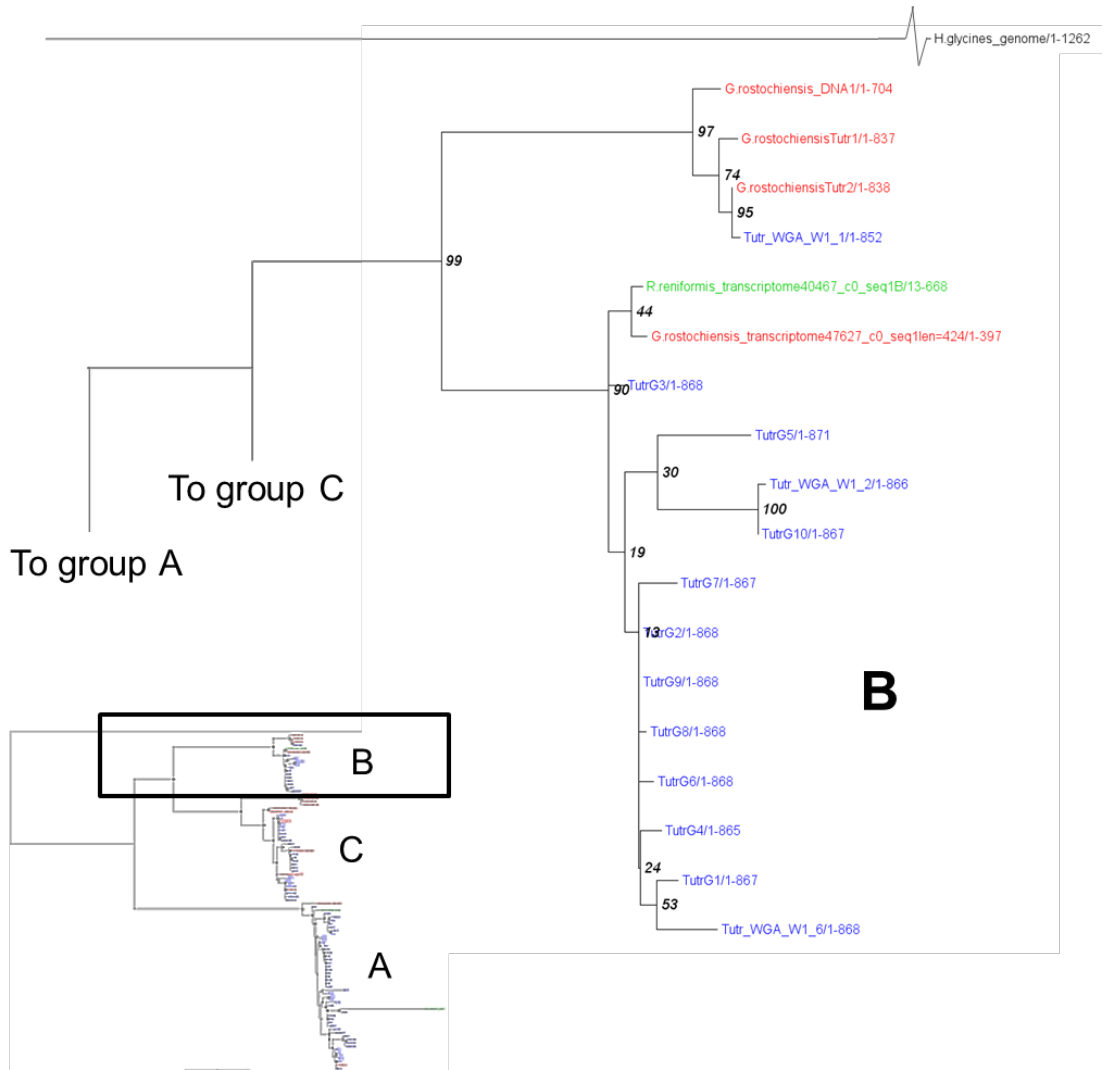
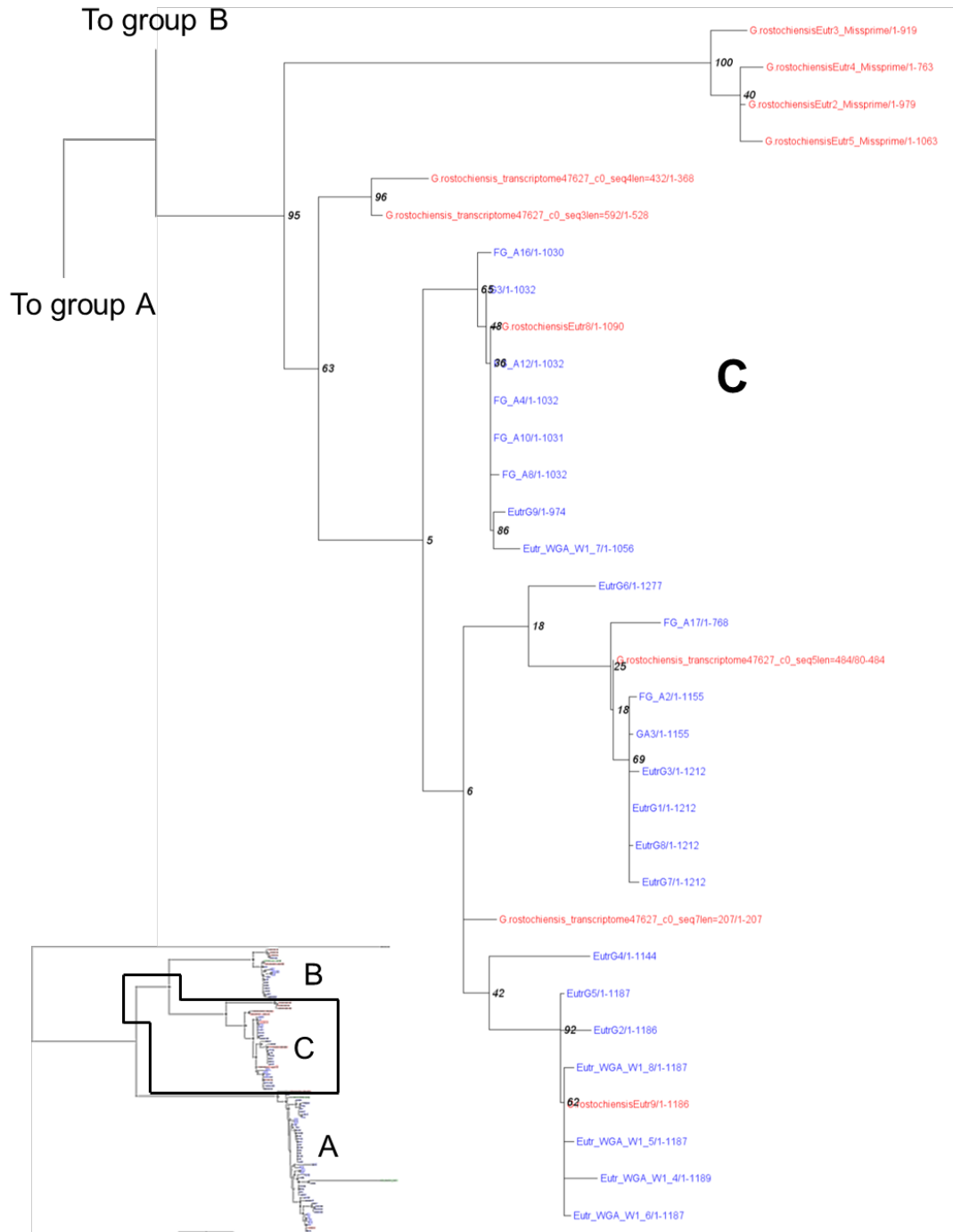


Figure 6-21 **Phylogenetic analysis of all 444 sequences from *G. pallida*, *G. rostochiensis*, *R. reniformis* and *H. glycines*.** Maximum likelihood tree of DNA sequences with 100 boot straps. Members of all subfamilies are present for *G. rostochiensis*, and member of just subfamilies A and B are present for *R. reniformis*.



To group B and C



To rest of group A

To rest of group A, and groups B and C



```

R.reniformis_transcriptome4046  MANNLKLKLYLLLAGFCLFIYGTTEAGGCKPGPKGPPGPPGAPGKKGPPGKC 50
G.pallida_ButrG3                MANNLKLKLYLLLAGFCLFIYGTTEAGGCKPGPKGPPGPPGAPGKKGPPGKC 50
G.rostochiensis_Butr2          MANNLKLKLYLLLAGFCLFIYGTTEAGGCKPGPKGPPGPPGAPGKKGPPGKC 50
*****:*****

R.reniformis_transcriptome4046  EKPPPKYEKPPP-----KYEPTRPPKYEHSKRSAPEAFSDSVGVV 91
G.pallida_ButrG3                EKPPPKYEKPPPKYEKPPP-KYEPTTRPPKYEHSKRSAPEAFSDSVGVV 99
G.rostochiensis_Butr2          EKPPPKYEKPPPKYEKPPPKYEKPPP-PPKYEHSKRSAPEAFSDAVGVV 99
*****          ***  .  *****

R.reniformis_transcriptome4046  RVARGEYENKCPAGPAGDVGPPGPPGPPGDSAKCPPPEEGYGYGHEGGGK 141
G.pallida_ButrG3                RVARGEYENKCPAGPAGDVGPPGPPGPPGDSAKCPPPEEGYGYGHEGGGK 149
G.rostochiensis_Butr2          RVARGEYENKCPAGPAGDAGPPGPPGPPGDSAKCPPPEEGYGYGHEGGGK 149
*****:*****

R.reniformis_transcriptome4046  GGDKEEKYEHRLRAVRGSGYEEVGGPPGPPGAGPPGADGPPGACECKY 189
G.pallida_ButrG3                GGDKEEKYEHHLRAVRGSGYEEVGGPPGPPGAGPPGADGPPGACECKY 197
G.rostochiensis_Butr2          GGDKEEKYEHHLRAVRGSGYEEVGGPPGPPGAGPPGADGPPGACECLY 197
*****:*****

```

Figure 6-22 **Protein alignment of 444-B like sequences from *G. pallida*, *G. rostochiensis* and *R. reniformis*.** Considerable conservation can be seen across the full length of the protein between the three species.

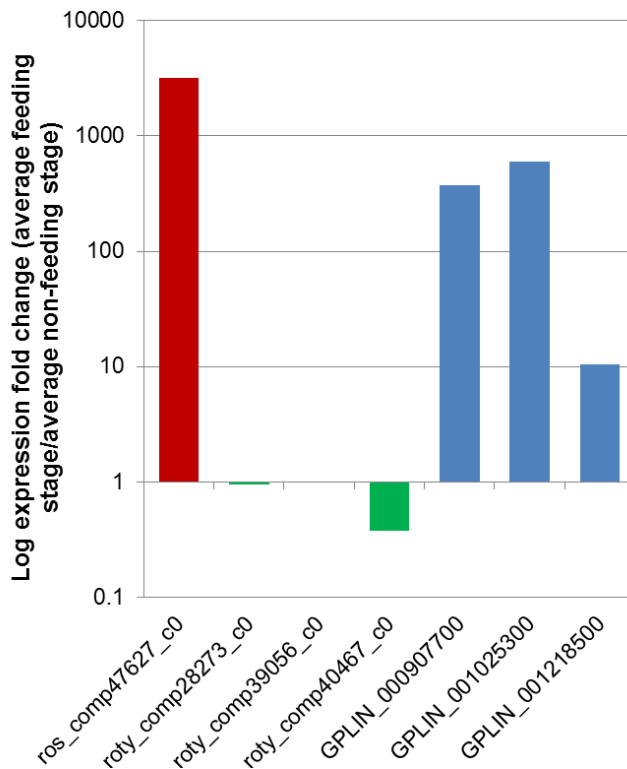


Figure 6-23 **Comparison of expression of 444 genes in *G. rostochiensis*, *R. reniformis* and *G. pallida*.** Expression fold change between feeding and non-feeding stages is shown on a per gene basis. Despite over 93 % amino acid conservation, *R. reniformis* appears to have a different expression pattern for 444-like sequences compared to *G. pallida* and *G. rostochiensis*.

6.4.7.4. **Candidate 444 diversity in *G. rostochiensis***

Due to the nature of 444 genes, and the unusual non-canonical apparent splicing observed, the variability of 444 genes present in *G. rostochiensis* was assessed by PCR. The subfamily specific *G. pallida* 444 primers were used to amplify related sequences from *G. rostochiensis* genomic DNA. (Figure 6-24). Amplification was possible from *G. rostochiensis* using these primers, although considerably fewer amplicons were identified compared to PCR on *G. pallida* DNA. Cloning and subsequent sequencing of the *G. rostochiensis* PCR products confirmed they were 444-like sequences. Although it is challenging to directly compare similarity for most subfamilies, subfamily B was considerably longer than a typical subfamily B member from *G. pallida*. The subsequent sequencing identified this difference in length is primarily contained within the intron sequences.

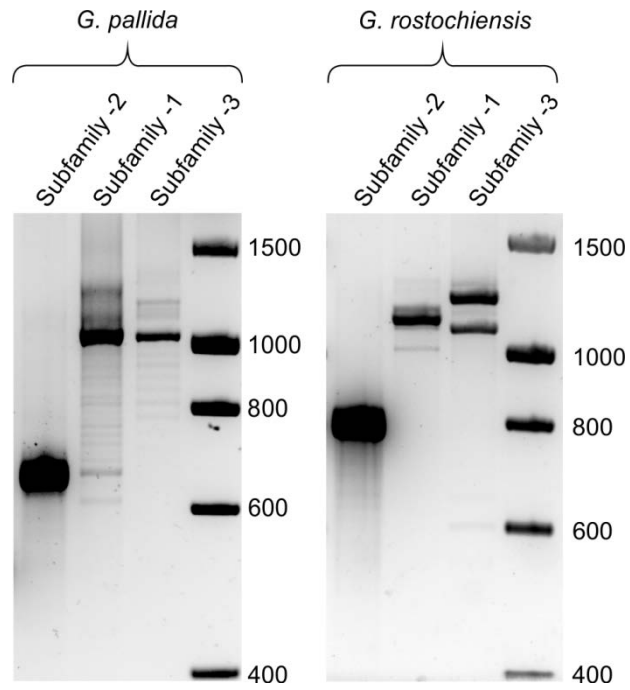


Figure 6-24 **Comparison of candidate 444 genes in *G. pallida* and *G. rostochiensis*.** Gel electrophoresis of PCR products generated by using 444 primers, designed to amplify all subfamilies. Considerably more amplicons are obtained from *G. pallida* compared to *G. rostochiensis*.

6.4.7.5. 448 gene family

In *G. pallida* the 448 gene family is complex. The genome sequence contained at least 28 full length genes that encoded putative secreted proteins. The full list of *G. pallida* 448 proteins was used as a BLAST query against the *R. reniformis* and *G. rostochiensis* transcriptomes. Members of the 448 family were identifiable in both species. Figure 6-25 shows a phylogenetic tree derived from an alignment of the extracted protein sequences. These sequences clearly group into two distinct clades (i and ii). Clade i contains mostly *R. reniformis* and *G. rostochiensis* sequences and Clade ii contains mostly *G. pallida* sequences.

As discussed in Chapter 4, the 448 gene family in *G. pallida* contained a very highly conserved PCCP domain present in almost all of the sequences. The presence of this motif was analysed in all *G. rostochiensis* and *R. reniformis* sequences. No *R. reniformis* sequences contained the PCCP motif although the *R. reniformis* sequences in Clade ii did contain a slightly different motif of L/V CC K in the same position in the sequence. The *R. reniformis* sequences in Clade i did not. All *G. rostochiensis* sequences in the Clade ii do contain either a PCCP motif or a variation that includes a double cysteine. Analysis of the expression patterns of the 448 genes in *R. reniformis* and both PCN species showed in general 448-like sequences are up-regulated during the sedentary stages of the life cycle when compared to the non-sedentary (Figure 6-26).

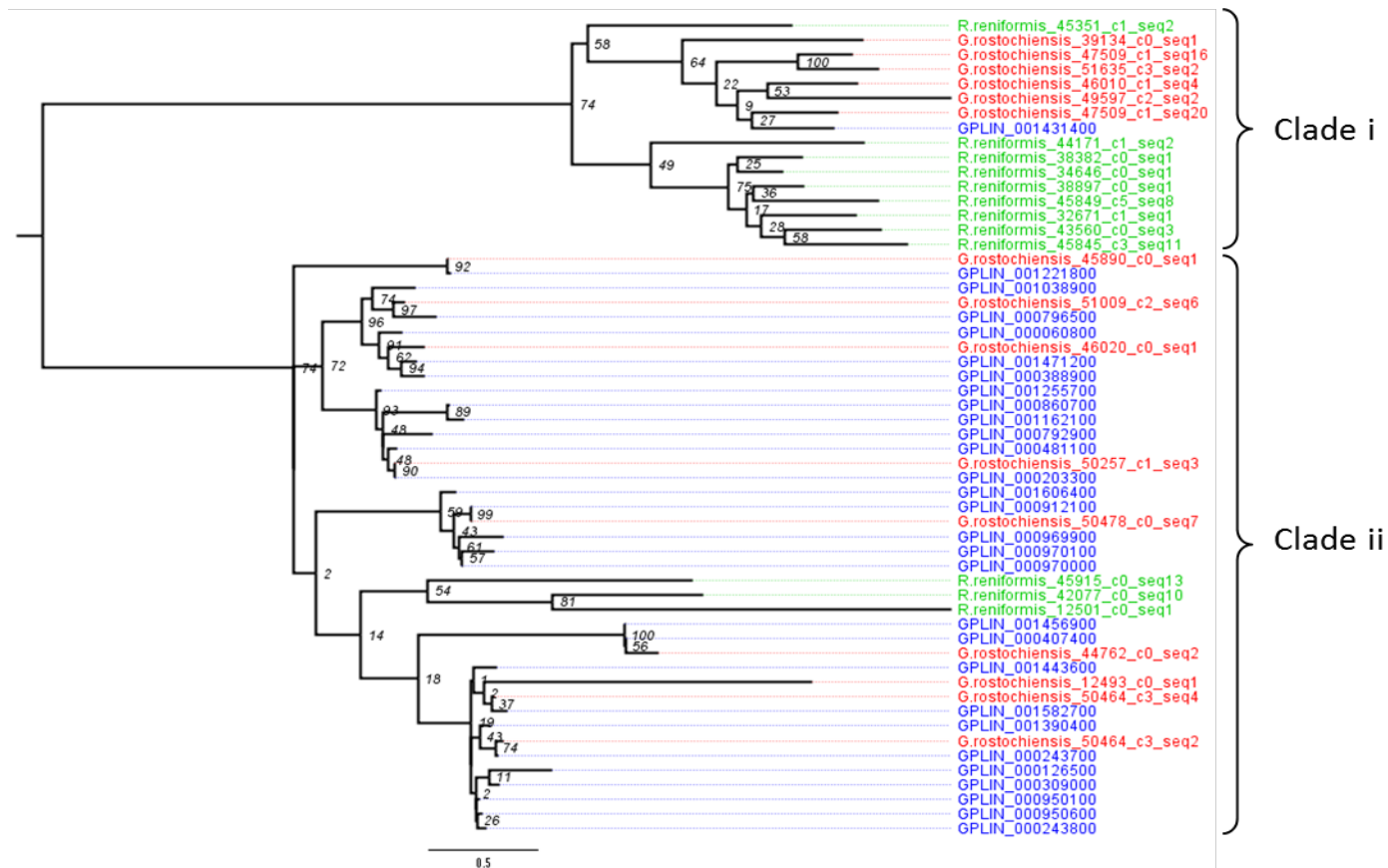


Figure 6-25 **Phylogenetic tree of 448 proteins in *G. pallida*, *G. rostochiensis* and *R. reniformis*.** The phylogenetic tree can be broadly split into two clades. Clade i contains mostly *R. reniformis* and *G. rostochiensis* sequences and typically does not contain the PCCP motif (although the *G. pallida* sequence does). All other full length sequences from any species in the rest of the tree contain the PCCP motif or similar XCCX. Scale bar indicates expected substitutions per base.

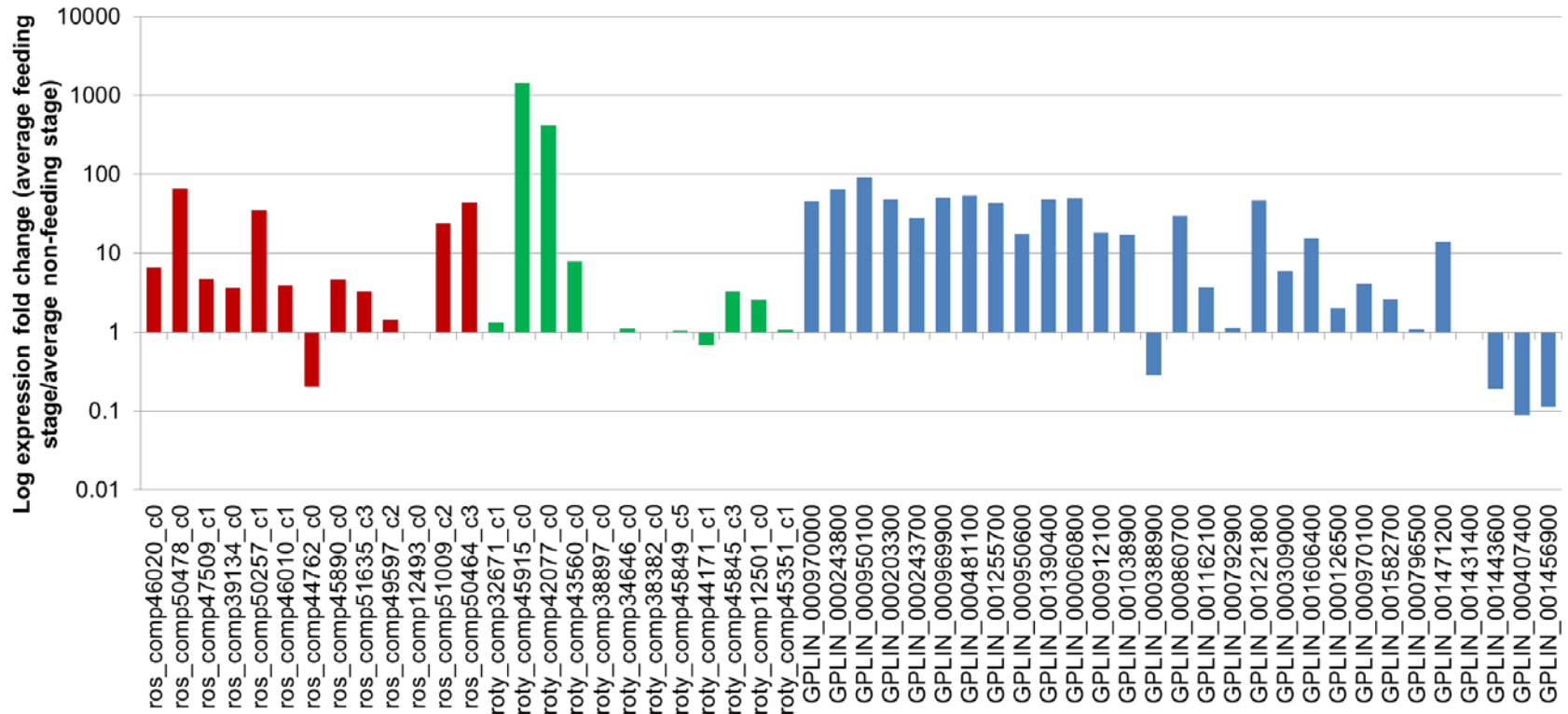


Figure 6-26 **Comparison of 448 gene expression in *G. rostochiensis*, *R. reniformis* and *G. pallida*.** Expression is displayed as fold change for feeding to non-feeding stages on a per gene basis. *G. rostochiensis* (red), *R. reniformis* (green) and *G. pallida* (blue) all have roughly the same expression patterns for 448s, the majority of which are up-regulated at sedentary female compared to J2.

6.4.8. Putative feeding tube components

The pipeline used to predict putative feeding tube components from *G. pallida* (Chapter 4) was used on the non-redundant significantly up-regulated clusters from *G. rostochiensis* and *R. reniformis*. Not all of the non-redundant sequences contained full length transcripts required for secreted protein prediction. Transdecoder was used to predict the best scoring Open Reading Frames (ORFs) from the non-redundant sequences up-regulated at sedentary female, and only those that were complete at the 5' end (*i.e.* contained a potential start codon) were used for secreted protein prediction. This set included sequences that may have been incomplete at the 3' end. This resulted in 438 complete transcripts and 119 3' partials for *R. reniformis* and 301 complete transcripts and 117 3' partials for *G. rostochiensis*.

6.4.8.1. *Rotylenchulus reniformis*

238 of this initial set of 557 sequences had predicted signal peptides and 157 of these had no predicted transmembrane domains. This corresponded to 181 isoforms as identified by Trinity. These 181 isoforms were then compared to the non-redundant protein database. Of the 181 secreted isoforms, 85 identified no significantly similar sequences in the non-redundant protein database ($p < 0.0001$), or the previously identified candidates from *G. pallida* identified in Chapter 4, by BLASTp. This corresponded to 77 new candidate genes for further analysis.

6.4.8.2. *Globodera rostochiensis*

179 of the initial set of sequences had predicted signal peptides, and 103 of these had no predicted transmembrane domains. This corresponded to 151 isoforms as defined by Trinity. Of the 151 secreted isoforms 22 identified no significantly similar sequences in the non-redundant protein database ($p < 0.0001$), or the previously identified candidates from *G. pallida* identified in Chapter 4, by BLASTp. This corresponded to 19 unique genes for further analysis.

6.5. Discussion

6.5.1. Transcriptome assembly and differential expression

The advent of next generation sequencing, in particular RNA sequencing, has provided new approaches that can be used to address many questions in biology (Wang et al., 2009). The ability to simultaneously detect and quantify all transcribed genes in a sample at any given time is enormously powerful and has revolutionised many aspects of biological sciences.

Next Generation Sequencing (NGS) and RNAseq are very young disciplines. Several methods have been described to generate, assemble, and characterise RNAseq data with no single approach currently used as the agreed standard in the field. To further complicate matters, there are many different software packages available for each step in the bioinformatics pipeline (Zerbino and Birney, 2008, Grabherr et al., 2011). For this study Illumina sequencing was chosen. Illumina sequencing produces very high sequence depths of 100 base-pair paired-end reads. As a result, Illumina platforms produce very high redundancy (Rodland, 2013), and the data were therefore *in silico* normalised prior to assembly. Two assembly methods were tested; Trinity and CLC bio. Interestingly, although both CLC assemblies were less complete than the Trinity assembly (Table 6-5), the more complete of the two CLC assemblies was made using the normalised read set. This read set has approximately 90% fewer reads than the non-*in silico* normalised read set, and contains slightly less information. This suggests that the noise present with high sequencing depths is detrimental to the assembly process for de Bruijn graph assemblers. This was true for both species sequenced (Table 6-5).

6.5.1.1. Limitations of Trinity assembly

Due to the lack of context when performing *de novo* transcriptome assemblies of short reads, two broad assumptions have to be made. Two very similar genes can be wrongly assumed to be isoforms of the same gene and, conversely, two very different isoforms can be wrongly assumed to be two separate genes. For this reason Trinity does not assign the term “gene” to transcripts, it does however group transcripts into “components” which is a loose proxy for genes using the two above assumptions, with each component potentially having multiple “isoforms”. This grouping of different genes under the same component is clearly the case for the *G. rostochiensis* 444-like sequences. Similarly, due to sequencing errors, and non-canonical splicing events discussed in Chapter 4, some “isoforms” may simply be artefacts. As a result, small sequence differences between isoforms were removed

at 90% nucleotide identity, and all remaining isoforms were treated independently and the term “gene” was used tentatively.

Assembly of short (100 bp) reads can also produce chimeric transcripts that are either the product of two separate genes (multi-chimeras) or incorrect assembly of the same gene (self-chimeras). Of available assembly methods, Trinity produces one of the lowest chimera and redundancy rates (Yang and Smith, 2013). When we compare the number of “components” produced by the *G. rostochiensis* transcriptome assembly, and the number of genes predicted from the *G. rostochiensis* genome assembly, the numbers are remarkably similar, giving some confidence in the term “component” being used as a proxy for “genes”. Although no genome assembly is available for *R. reniformis*, the transcriptome assembly produced roughly the same number of “components” as for *G. rostochiensis*, which might be expected for a closely related organism.

6.5.1.2. Normalisation and differential expression

Although it was initially suggested that RNAseq experiments do not require sophisticated normalisation procedures, it is in practice extremely important as raw reads counts are often not comparable between samples (Wang et al., 2009). Multiple biological replicates were sequenced for each life stage of each organism. Differential expression requires the direct comparison between samples, both within and between life stages. However, each sample will sequence differently producing different numbers of total reads. Normalisation is therefore a critical part of the process to ensure that gene counts can be compared between biological replicates (Soneson and Delorenzi, 2013, Dillies et al., 2013, Rapaport et al., 2013). Again there are several methods for normalisation. RPKM (Reads Per Kilobase per Million reads) was until recently the method of choice as it accounts for gene length, which is correlated with number of reads. This has on a number of occasions been shown to be inaccurate (Oshlack and Wakefield, 2009, Dillies et al., 2013, Wagner et al., 2012). TMM normalisation (Trimmed Mean of M-values) is currently considered to be one of the best performing normalisation methods available (Dillies et al., 2013) and is included with the EdgeR package used here.

Unsurprisingly, several methods exist for analysing differential expression of normalised read counts, some of which have been created specifically for RNAseq experiments (Anders and Huber, 2010, Robinson et al., 2010, Trapnell et al., 2013) and others that have been adapted from microarray experiments (Hardcastle and Kelly, 2010). Interestingly, those adapted from microarray experiments perform comparably to those specifically designed for RNAseq (Rapaport et al., 2013).

Several reviews of available software packages and methodology have not identified a clearly superior package, but have identified limitations of particular packages. For example, Cuffdiff produces many false positives when tested on technical replicates, in particular for low expressed transcripts where noise is more of a factor (Rapaport et al., 2013). Unsurprisingly, number of replicates and sequencing depth are both important factors in differential expression analysis. However, It has been noted that for most methods (including EdgeR used here) over 90% of highly expressed differentially expressed genes are reliably detected with as few as 2 replicates and just 5% of total reads (Rapaport et al., 2013), giving confidence in the *G. rostochiensis* data.

6.5.2. Contamination removal and Horizontal gene transfer (HGT)

Contigs originating from putative contamination were removed by selecting all contigs that only had significant BLAST hits to anything plant or non-metazoan. Genes acquired by horizontal gene transfer (HGT) have been discussed for plant parasitic nematodes on numerous occasions (reviewed in (Haegeman et al., 2011) (Danchin et al., 2010)). In both the *G. rostochiensis* and *R. reniformis* transcriptomes several putative HGT genes were identified from the J2 gene expression cluster. Important to note is that the protocol used for removing contigs that are likely to be derived from contamination may also remove HGT events that should not be removed. It is however likely that previously described HGT genes are now more closely related to other nematodes than the bacteria or fungi they originated from. More work needs to be carried out to determine if these are functional proteins secreted by the nematode.

6.5.3. Between species functional annotation comparison

By assigning functional categories to genes expressed at different life stages, we are able to make links between gene sequence and biology. As described in the introduction, the life cycles of the cyst nematodes and the reniform nematodes are different. *G. rostochiensis* enters the root as a juvenile, initiates its feeding site, where it undergoes moults and develops into either male or female depending on host cues (Sobczak and Golinowski, 2011). In contrast, *R. reniformis* undergoes its moults in the soil without feeding, and enters the root as an adult female (Robinson et al., 1998). For both species to carry out these moults re-modelling of the collagen based cuticle is required. For *G. rostochiensis* gene expression of the cuticle collagens was over represented in the sedentary female sample whereas in *R. reniformis* this was not the case as some are expressed during the J2 stage. A

similar pattern was also observed with genes involved in reproduction and sex determination, which may reflect the differing biology of these two species. Unsurprisingly, both species had over representation of genes involved in locomotion in the Juvenile stages compared to the sedentary stages. However, the only common Gene Ontology (GO) term between the two feeding stage samples was oxidoreductase activity (discussed in detail below).

Some of these differences may reflect the considerably faster life cycle of *R. reniformis* compared to *G. rostochiensis*. *R. reniformis* can complete its life cycle from egg to egg in 9 days, making matching of the two species life cycles difficult.

6.5.4. Oxidoreductase activity

When comparing the functions assigned to the up-regulated genes at the sedentary stage between species, a single common term was identified corresponding to oxidoreductase. When explored in more detail it appeared that there were several genes present in both species that carry out this function; glutathione synthetase, glutathione peroxidase, and superoxide dismutase. The majority of these genes were present in the sedentary gene expression clusters, suggesting a role during parasitism. Indeed, as mentioned in Chapter 4, several glutathione synthetases were identified as being significantly up-regulated in the *G. pallida* feeding stages. This implies an underlying function shared both within and outside the cyst nematodes.

Glutathione peroxidase expression has been demonstrated in the hypodermis of *G. rostochiensis* where it presumably acts to detoxify plant derived reactive oxygen species, a typical plant defence response (Haegeman et al., 2012). Glutathione synthetases have been linked to preservation of the host-symbiont interaction in the endosymbiont *Rhizobium tropici* (Muglia et al., 2008). The plant triggers senescence of the root nodule after a number of days, which is proposed to be controlled, at least in part, by an oxidative burst. This burst is significantly reduced by the symbiont, as a result of secreted glutathione. This highlights a potentially interesting insight into symbiotic evolution but also provides a potential role for the nematode derived glutathione synthetases in the host, as opposed to simply detoxifying plant defence compounds.

In general, nematodes such as *C. elegans* contain only a single glutathione synthetase gene. The plant-parasitic nematode species described here contain between 10-20 times more than this. In addition, the vast majority of these nematode derived glutathione synthetase genes identified contain a secretion

signal. No information is available about the spatial expression of the nematode derived glutathione synthetase genes. However, if they were simply present to detoxify reactive oxygen species consumed by the nematode, the enzyme itself would not be secreted to the gut, but rather the product; glutathione. It is therefore possible that glutathione synthetase genes are secreted into, and function within, the host. The effect of redox potential on gene expression has been demonstrated in a range of organisms including animals (Kunsch and Medford, 1999) and plants (Luis Garcia-Gimenez et al., 2013). Although the details are not fully understood, eukaryotic cells require glutathione to survive and cancer cells have extremely high levels of this compounds while stem cells require low levels to retain pluripotency (Luis Garcia-Gimenez et al., 2013). These data suggest that glutathione plays a key role in global gene expression changes. Similarly, tracking glutathione levels in the host during infection could provide interesting insights into their potential role in feeding site formation/maintenance. Finally, a set of transcription factors differentially expressed in response to different glutathione concentrations have been identified in root meristems (Luis Garcia-Gimenez et al., 2013). It would be interesting to examine expression of these transcription factors in the nematode feeding sites.

6.5.5. *G. pallida* candidate feeding structure genes in *G. rostochiensis* and *R. reniformis*.

The rationale for generating the transcriptome data described in this Chapter was to identify the “candidate feeding structure genes” from *G. pallida* in other related plant parasitic nematodes. All candidate genes characterised in Chapter 5 were present in *G. rostochiensis*, lending weight to the idea that these proteins have a conserved role. Surprisingly, despite the fact that *R. reniformis* is not a cyst nematode, it contained members of the 448 gene family, the 448 gene family and the 3453 gene family, all with remarkable sequence conservation.

6.5.5.1. 448 gene family

The 448 gene family in *G. pallida* has a highly conserved domain at the N terminal end of the mature protein, as discussed in Chapter 4. It is possible that the two cysteines present in this conserved domain may play an important structural role. A subgroup of 448-like sequences was identified, consisting mainly of sequences from *G. rostochiensis* and *R. reniformis*, that lack the PCCP motif. These may have a different structure/role. Similar to *G. pallida*, the majority of 448s in *G. rostochiensis* and *R. reniformis* are up-regulated during the sedentary stage of the life cycle, again suggesting a conserved function. However *in situ* hybridisation

needs to be carried out to confirm that the spatial expression patterns of 448-like sequences in *R. reniformis* and *G. rostochiensis* are consistent with the dorsal gland staining pattern observed with *G. pallida*. 448s were originally identified from an *H. glycines* gland cell library (Gao et al., 2003), where gland cell expression has been confirmed, suggesting a highly conserved role in parasitism.

6.5.5.2. 176 genes

No 176-like sequences were identified in the *R. reniformis* transcriptome by BLASTp. Due to the incompleteness of the *R. reniformis* assembly that was generated from only two life stages, we cannot be confident that the 176 gene family is absent in this nematode. However, the raw reads for *R. reniformis* were mapped back to all cloned 176 sequences from all species, with very permissive settings, and not a single read successfully mapped. This suggests that 176 genes are at most, cyst nematode specific.

6.5.5.3. 444 gene family

Our analysis has shown that 444 genes are present in all cyst nematode species sampled and are also present in the closely related reniform nematode *R. reniformis*. Given the phylogenetic proximity of the reniform nematodes to the cyst nematodes (van Megen et al., 2009), and the similarities in the feeding sites that they induce (Jones et al., 2013), it could be speculated that 444 genes share a similar function in both groups. However, of the three transcripts corresponding to 444 genes in *R. reniformis*, two have very low expression and the third is slightly up-regulated at the juvenile stage. This is in direct contrast to the >1,000 fold up-regulation during feeding stages of the cyst nematodes.

Despite the presence of just three 444 sequences in the current *G. pallida* genome assembly, over 80 unique sequences were identified by conventional cloning (Chapter 4). Similarly for *G. rostochiensis*, several 444 sequences were identified by PCR (section 5.3.5.7), while just three are present in the assembled transcriptome (all grouped under a single component). Since no conventional PCR experiments have been undertaken using *R. reniformis* it is not possible to draw any conclusions about the number of 444 genes present from the transcriptome assembly alone. The absence of the full gene family in the *G. pallida* genome assembly and *G. rostochiensis* transcriptome assembly again highlights a limitation of sequencing and assembly of short reads. All 444 sequences, irrespective of subfamily, share stretches of 410 and 94 nucleotides with >90% identity at the 5'

and 3' ends respectively which may underlie the difficulty in assembling 444 sequences in both species included here.

6.5.6. Putative orthologs between *R. reniformis* and *G. rostochiensis*

Whole transcriptome Reciprocal Best BLAST Hit (RBBH) analysis identified a number of putative orthologs between *G. rostochiensis* and *R. reniformis*. When the similarity distribution of these putative orthologs was examined, it was clear there was a subset of very highly conserved genes present in the two species (Figure 6-15). 43 genes were identified as being between 99 and 100% similar between the two species at nucleotide level. 19 of these 43 were present in the *G. rostochiensis* feeding stage cluster, the vast majority of which had no assigned gene ontology terms.

6.5.6.1. Proposed functions of super-conserved orthologs

It is difficult to assign functions to genes without downstream analysis, particularly when no gene ontology terms are associated with the gene of interest. Of the highly conserved orthologs expressed during the parasitic stages of *G. rostochiensis*, three have characterised domains, the first of which corresponds to a C-type lectin. C-type lectins have many different functions, and are broadly characterised for their role in protein-protein interactions. It is also suggested that they play roles in both animal (Loukas et al., 1999) and, more recently, plant-parasitic (Ganji et al., 2014) nematode infections although due to the variable nature of C-type lectins these roles are likely to be quite different. Further analysis of their role in parasitism, potentially using yeast-two-hybrid, will be required to elucidate their function.

The second of the annotated highly conserved orthologues expressed at feeding stages is a collagen. This is presumably involved in cuticle re-modelling as the feeding females grow. The final annotated super-conserved ortholog is a hydroxyacyl-coenzyme. These play a role in fatty acid elongation/metabolism. The role of these proteins in plant parasitic nematodes is not clear, although they may be important for processing of food taken up by the nematodes. The remaining genes that are highly conserved between the two plant parasitic nematode groups and that are up-regulated during infection are novel proteins. Blast2GO is an inadequate tool for characterising these genes, as very few (three) of the 19 highly conserved orthologs have assigned GO terms. The fact that most of these genes are "novel" is a reflection of how little is known about the biology of plant parasitic nematodes.

Interestingly, the 19 super-conserved orthologs that are present in the *G. rostochiensis* feeding stage cluster are not present in either the feeding or non-feeding stage cluster of *R. reniformis*. In the case of these orthologs *R. reniformis* has almost exactly the same genes present, and yet does not express them at the same time. The host for *R. reniformis* for this experiment was cotton (cv. Δpine90), and the host for *G. rostochiensis* for this experiment was potato (cv. Desiree). It is possible that the differences in expression of the gene family members reflect these differences in the host. The majority of these 19 highly conserved orthologs, were also identifiable in *G. pallida* and are also up-regulated during the feeding stages. The host for *G. pallida* in this experiment was also potato (cv. Desiree). In each case, the two nematodes feeding on potato (*G. pallida* and *G. rostochiensis*) express the same orthologues during feeding stages and the nematode on cotton (*R. reniformis*) does not. As discussed the host range of *R. reniformis* is broad (Robinson et al., 1998) covering > 300 species, including potato (Linford and Yap, 1940). More work will be required to determine if putative host specific genes are up-regulated during feeding stages of *R. reniformis* on potato. Precedent has been demonstrated for highly specific pathogen gene expression with tissue specific effectors of *Ustilago maydis* (Skibbe et al., 2010) and, more recently, the host specific effectors of *Zymoseptoria tritici* (Kellner et al., 2014).

Summary

- De novo transcriptomes of two species of plant-parasitic nematode (*G. rostochiensis* and *R. reniformis*) have been sequenced and assembled
- The quality and completeness of the assemblies has been assessed
- Whole transcriptome comparisons between and within species have been performed
- The candidate feeding structure genes from Chapter 4 have been identified in these new resources to determine the evolutionary history of the gene families.

7. General discussion

There were two main aims to the work described; to characterise the molecular sieving capabilities of the feeding tube and to use the newly available genomic resources for *G. pallida* to identify genes encoding components of the feeding tube and/or feeding plug. These structures play key roles in the plant-nematode interaction and are potentially viable targets for novel control strategies.

7.1. Feeding tube size exclusion

The capacity of feeding tubes to selectively exclude macromolecules based on size is well supported and has been reported on numerous occasions (Urwin et al., 1997b, McCarter et al., 2010, Valentine et al., 2007, Goverse et al., 1998, Bockenhoff and Grundler, 1994). Here, the size exclusion limit was defined using a pragmatic measurement of protein size *in silico*. Although it is clear that feeding tubes behave as molecular sieves, their biological function remains uncertain. Feeding tubes are present in both cyst and root-knot nematodes. Sedentary endoparasitism is hypothesised to have evolved independently in these two groups (van Megen et al., 2009). Feeding tubes are not reported in the most closely related migratory endoparasitic species and are also therefore likely to have evolved independently. This may explain the differences in feeding tube morphology and size exclusion limit between cyst and root-knot nematodes (Figure 1-3). The absence of feeding tubes from migratory endoparasitic species also implies that feeding tubes have a biological function related to the sustained biotrophic interactions of root-knot and cyst nematodes with their hosts. It has been hypothesised that the feeding tube prevents uptake of large macromolecules/organelles which could cause total or partial blockage of the stylet (Razak and Evans, 1976). However, this is unlikely to be the case as migratory endoparasitic nematodes within the same phylogenetic clade feed with similar stylets but without feeding tubes, and ingest host cell cytoplasm with no apparent difficulty.

Selective uptake of certain proteins could have obvious benefits for nematodes, particularly if nematicidal molecules could be specifically excluded by a feeding tube. However, although some large proteins (such as the *Bacillus thuringiensis* crystal protein) are toxic to nematodes and show differential uptake by cyst and root-knot nematodes (Li et al., 2007) it is not generally the case that larger

molecules exclusively possess nematicidal activity. It is therefore unlikely that exclusion of certain macromolecules is the primary or sole function of feeding tubes.

Cyst and root-knot nematodes that have induced a feeding site do not have the ability to move and are unable to induce further feeding sites. It is therefore necessary for the nematode to maintain the integrity and efficacy of the feeding site throughout its feeding stages. Feeding sites, and in particular syncytia, are under high pressure. Syncytial cells appear to be at a higher pressure (5000 – 7000 hPa more) than surrounding cells (Grundler and Böckenhoff, 1997, Böckenhoff and Grundler, 1994). Insertion of a stylet alone into the feeding site across the plasma membrane would be likely to compromise the syncytium. Feeding tubes may therefore act as a way for the nematode to withdraw host cell assimilate without disrupting the integrity of the feeding site.

Hydraulic resistance of a filter is determined by the properties of the filter and the viscosity of the fluid moving across it (Jensen et al., 2014). For a given pressure and viscosity, fluid will travel more slowly through several small pores (filter), than through a single large pore of the same total area (Jensen et al., 2014). The biological relevance of this principle has already been demonstrated in plants. Phloem sieve plates, and their hydraulic resistance, have been shown to limit flow rate of sap after external attack from animals (Jensen et al., 2012). In the context of plant-parasitic nematodes, withdrawal of host cell assimilate from a pressurised feeding site through a feeding tube, rather than through the stylet alone, will reduce flow rate and may prevent the feeding site from collapsing.

Root-knot nematode feeding tubes allow uptake of larger macromolecules than cyst nematode feeding tubes presumably as a result of their different pore size, as suggested by electron microscopy studies (Figure 1-3, Page 12). Minor modifications in pore size of a filter can have a large impact on hydraulic resistance (Jensen et al., 2014). This suggests that the flow rate of solutes through feeding tubes of root-knot nematodes will be less restricted than across cyst nematode feeding tubes. This may be explained by the differences in feeding sites of root-knot and cyst nematodes. Root-knot nematodes feed from several individual giant cells whereas cyst nematodes feed from a single large syncytium. No studies exist that describe the turgor pressure of giant cells. However, the proposed hydraulic resistance hypothesis, coupled with the less restrictive feeding tubes of root-knot nematodes, suggest that giant cells are likely to be under considerably less turgor

pressure than syncytia. In addition, it has been noted that giant cell walls are approximately twice as thick as those of syncytia in *Arabidopsis* and so are perhaps less flexible/elastic. Moreover, the composition of syncytial cells walls suggests that they are particularly flexible (Davies et al., 2012). Cell wall elasticity may therefore explain the differences in apparent pressure.

As described in Chapter 6, some reniform nematodes also induce and feed from a single large syncytium (Agudelo et al., 2005, Rebois et al., 1975). The feeding tubes of these reniform nematodes are morphologically similar to those of the cyst nematodes (Rahman Razak and Evans, 1976, Rebois, 1980), further supporting the proposed hydraulic resistance hypothesis. Sieving characteristics and size exclusion limits have not yet been defined for reniform nematodes, although the hypothesis proposed above predicts that these characteristics will be similar to those of the cyst nematodes. Interestingly, despite the apparent phylogenetic proximity of the false-root knot nematode *Nacobbus aberrans* to the root knot nematodes (van Megen et al., 2009, Eves-van den Akker et al., 2014a), it too induces and feeds from a single large syncytium. Although feeding tubes have not been described for *N. aberrans* (Jones and Payne, 1977), it could be predicted that it too would require feeding tubes similar to those of the cyst and reniform nematodes. In addition, a feeding tube has been described for *Rotylenchulus macrodoratus*, at the stylet orifice inside the uninucleate giant cell it produces (Cohn and Mordechai, 1977). Although ultrastructural studies are reported, there are no images of the feeding tube. It would be interesting to determine whether the structure and functional sieving capabilities of the *R. macrodoratus* feeding tubes reflect its phylogenetic position (i.e. similar to *R. reniformis*/cyst nematodes) or its feeding site (i.e. similar to giant cells of root-knot nematodes). No intracellular feeding tubes have been described for migratory endoparasites within Clade 12 (*Pratylenchus* and *Radopholus* species.), or any plant parasites outside of Clade 12. Indeed, the hydraulic resistance function of feeding tubes would not be required by these species and would, most probably, be detrimental, as they ingest entire host cell contents.

The only other feeding tubes described are those of the Clade 1 plant-parasite *T. similis*, although these appear to be functionally distinct from the Clade 12 feeding tubes (Section 1.4.7). They comprise an open-ended tube that forms around the onchiostyle during insertion through host cell walls. During feeding, entire cell contents are withdrawn down this feeding tube through the opening in the distal end. Therefore, the feeding tubes of *T. similis* do not act as molecular sieves.

7.2. Feeding tube and feeding plug candidate genes

Several feeding tube and feeding plug candidates were identified from the genome sequence of the cyst nematode *G. pallida* based on predictions about the characteristics of a feeding structure gene. Subsequent analysis of the molecular characteristics and phylogenetic distribution was used to elucidate function. Although numerous candidate genes have been identified, it is still unclear whether or not a feeding tube or feeding plug gene has been identified. *In vivo* observations on the timing of feeding tube formation and the release of secretory granules suggest that feeding tube components originate from the pharyngeal gland cells, although no direct evidence in support of this has been provided (Grundler and Böckenhoff, 1997). One of the candidates described here, the 448 gene family, is expressed in the dorsal gland cell during infection and the phylogenetic distribution of 448-like genes correlates well with feeding tube morphology (i.e. present in cyst/reniform nematodes and absent in root-knot nematodes). However, numerous genes that do not encode components of the feeding tube could also satisfy these criteria; significantly more work will need to be carried out to determine function. Future work in this area could include immunolocalisation of the 448 proteins in the host using the antibodies that have already been produced.

Two candidates (444 and 176) were identified as being expressed in the amphid sheath cells; secretion of the former from the amphids into the apoplasm was confirmed by immunolocalisation. Although it has been suggested that cyst nematode feeding plugs may originate from the amphid canal (Endo, 1978) the 444 protein extends further in the apoplasm than indicated by electron micrographs for feeding plugs (Endo, 1978). In addition, the phylogenetic distribution of the 444 genes does not correlate well with the distribution of feeding plugs. The feeding pegs/plugs of *R. reniformis* are difficult to distinguish from plant material and may not be nematode derived. However, 444 genes were identified in the transcriptome of *R. reniformis* that are extremely similar to the 444 genes in cyst nematodes (Chapter 6). In addition, the genetic variability of 444 genes between individuals of the same population is difficult to reconcile with the function of a feeding plug. Alternatively, the data presented do not exclude a role for candidate 176 in forming the feeding plug. The 176 genes are also expressed in amphid sheath cells and their distribution mirrors the distribution of feeding plugs; they were not identified in the *R. reniformis* transcriptome but were present in every cyst nematode genetic resource tested. In addition, the 176 proteins are insoluble when expressed in bacteria. Due to the circumstantial nature of these data, more work will be required

to determine the exact location of 176 proteins *in planta* in order to identify their function.

7.3. Success of the identification strategy

The pipeline described was successful in identifying both novel and previously described, putative and verified, effectors with apparently important functions in parasitism. However, no direct evidence has been presented for identification of a feeding structure gene. The first stage of the pipeline identified those genes most highly up-regulated during the feeding stages compared to non-feeding stages. Only the top 500 genes were analysed (equating to < 5%). However, this may have presented an over simplification. Firstly, effector-secreting tissues of plant parasitic nematodes do not constitute a large proportion of total body mass. The size of tissues and number of cells within may blur the lines between expression values. Expression of genes expressed in small tissues may artifactually appear to be lower than those expressed across the whole body. The effector-secreting-tissue-RNA present in a total nematode RNA extraction will roughly correlate with the size of the tissue in relation to the total size of the body. This phenomenon should be reduced by comparing expression fold change rather than expression value, however, the method used to reduce false positives may exacerbate the problem. If expression in the control stage (J2) is zero, and the test stage (14 dpi) is > zero, the expression fold change will be infinite (even if 14 dpi expression = 1 or 10,000). If the expression of a given gene in a given stage was < 1, it was therefore set to 1 to reduce the skew of high expression fold changes from non-integers < 1. This approach removes infinite expression fold changes, but also reduces sensitivity of fold change for genes with low expression, or apparently low expression due to small relative tissue size. In the attempt to reduce false positives, false negatives may have been introduced, and this may explain the lack of more convergence between the method here and direct gland cell sequencing, and perhaps the failure to identify feeding structure genes. In addition, this method does not account for protein turnover, re-localisation, or transcript storage (Hammond et al., 2000). The second criterion was the presence of a signal peptide to target the protein for secretion although similarly this ignores non-classical secretion mechanisms (Bendtsen et al., 2004).

Alternative methods to identify the feeding tube may include direct isolation and protein sequencing. Feeding tubes are apparently insoluble and accumulate within the feeding site up to 14 days post production (Hussey and Mims, 1991). Laser capture of syncytial cells and subsequent protein purification may amass enough

feeding tube material to sequence using mass spectrometry. Importantly, if the feeding tube is assumed to be insoluble 90% of all proteins can be removed immediately by only collecting the insoluble protein fraction. Such purification of intact complexes has been achieved for the *Shigella* type III secretion needle complex (Tamano et al., 2002). The type III secretion needle complex is a protein structure used to deliver bacterial effectors across plasma membranes. Feeding tubes apparently form prior to or concurrently with effector 'injection'. It may be speculated that feeding tubes facilitate effector delivery rather than cytoplasmic withdrawal. Indeed it is not known how nematode effectors are actually translocated into the host cells. It is often quoted in the literature that effectors are 'injected' into host cells (Jaouannet et al., 2012, Quentin et al., 2013), yet the stylet orifice does not appear to physically cross the plasma membrane (Rebois, 1980). Therefore, an alternative import mechanism will be required, whether via/through the feeding tube or otherwise. It could be speculated that this offers an explanation for the often small molecular weight of cyst nematode effectors. Feeding plugs, however, may not be amenable to this type of purification strategy due to their amorphous structure.

There is a general consensus that the feeding plug is nematode derived. As discussed in Chapter 6, this may not be the case for the callose-like feeding peg of reniform nematodes. Feeding plugs form in the host apoplasm, and as demonstrated, there are numerous mechanisms that are used by nematodes to deliver proteins to this compartment. However, the feeding tube forms inside the host cell, and despite circumstantial evidence for nematode origin (timing and parasite-specificities), it may be partially of plant origin but stimulated by nematode apoplastic proteins. The hypothetical protein that stimulates this formation may not be encapsulated by the criteria used here to define a putative feeding structure gene.

No direct evidence has been found that unequivocally confirms that any of the candidates described here are feeding plug or feeding tube genes. However, the 176 and 448 gene families are considered putative effectors, and the 444 gene family can be considered as verified effectors. Although the identification pipeline was designed in such a way to maximise the possibility of feeding structure gene identification, it was noted from the outset that other effector genes would also be identified using this protocol. The spatial expression and *in planta* localisation of the 444 gene family has contributed to our understanding on the delivery mechanisms and cellular compartments utilised by plant-parasitic nematodes in parasitizing plants.

7.3.1. Tandem repeat proteins

It is notable that the 176, 444 and 3453 gene/gene families all encode tandem repeat proteins. In general, the term tandem repeat protein refers to proteins that contain internal, linear, tandem repeats of two iterations or more. There is no defined maximum limit for the number of amino acids within such proteins, however large scale analyses have demonstrated they can range from di-residues to over 100 (Marcotte et al., 1999). This definition cannot be used to wholly describe the complex rearrangements of the 444 gene family tandem repeats; nevertheless, for the purposes of this discussion they will be considered together with 176 and 3453. It has been suggested that error-prone repeat expansions facilitate rapid evolution when compared to non-repeat proteins (Andrade et al., 2001, Marcotte et al., 1999). Repeat proteins are three times more abundant in eukaryotes than prokaryotes (Marcotte et al., 1999) and this may compensate for the long generation time of eukaryotes compared to prokaryotes, and provide an alternative means of rapid evolution.

Although repeat proteins do not have a single common function they are often involved in recognition or ligand binding. The largest class of plant resistance proteins, termed NLRs, is characterised by a central nucleotide-binding domain and a C-terminal Leucine-Rich Repeat (LRR) domain. The characteristic 'horseshoe' structure of the LRR domain (Figure 7-1) has been determined by comparison to the solved structure of a leucine rich repeat-containing porcine ribonuclease inhibitor (Kobe and Deisenhofer, 1993). The LRR domain is often responsible for ligand binding, that results in a conformational change in other parts of the protein which subsequently trigger downstream signaling (Slootweg et al., 2013).

The first and basal-layer of plant defences is triggered by the recognition of conserved molecular patterns by pattern-recognition receptors (PRRs) which also contain LRR domains. These patterns are either Pathogen-Associated Molecular Patterns (PAMPs – highly conserved pathogen molecules), such as the N-terminus of bacterial elongation factor Tu (elf18 (Kunze et al., 2004)), or Damage-Associated Molecular Patterns (DAMPs), which are endogenous structural components released by partial cell wall hydrolysis (Sanabria et al., 2008). Oligogalacturonides are a major class of DAMP and numerous studies have characterised their ability to induce defence genes (reviewed in (De Lorenzo et al., 2011)). Recognition of PAMPs by the corresponding receptors in plants leads to PAMP-triggered immunity (PTI). Similarly, recognition of DAMPs leads to DAMP-triggered immunity (DTI). Recognition of PAMPs or DAMPs is often encoded by the extracellular LRR domain of recognition receptors. Hyper-variable solvent-exposed residues in the LRR

domain of some NLR proteins are responsible for binding specificity. Interestingly, two membrane-anchored extracellular tomato disease resistance proteins Cf-9 and Cf-4 are 91% identical yet each recognizes sequence-unrelated fungal avirulence proteins (Wulff et al., 2009). This highlights the diversity in binding specificity with relatively high receptor sequence homology within tandem repeat domains.

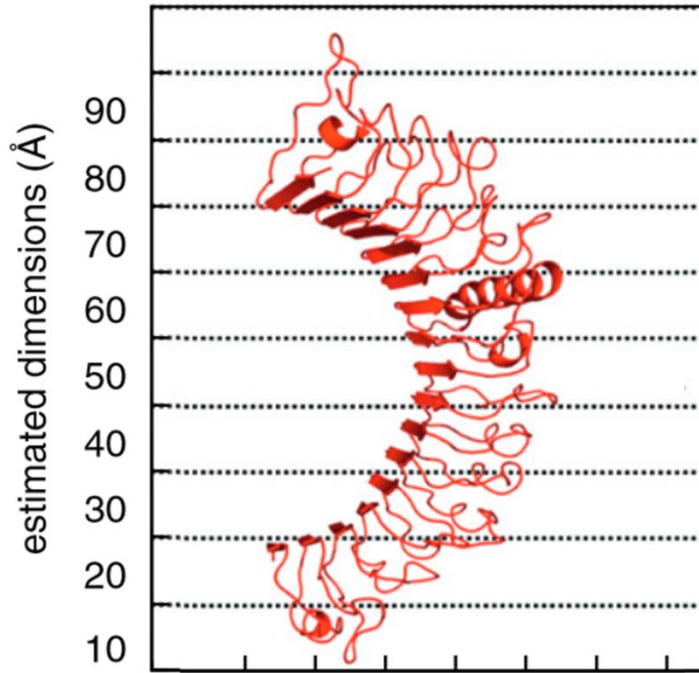


Figure 7-1 **Predicted structure of the leucine rich repeat domain of plant resistance proteins.** The hypothetical structure of a leucine rich repeat (LRR) domain of a plant resistance protein (Takken and Goverse, 2012). Structure based on a leucine rich repeat-containing porcine ribonuclease inhibitor (Kobe and Deisenhofer, 1993). This is one example of a tandem repeat protein with ligand binding capacity.

Other noteworthy variable tandem repeat proteins include the Transcription Activator-Like (TAL) effectors of *Xanthomonas spp* (Doyle et al., 2013). TAL effectors are highly adaptable phyto-bacterial virulence factors that contain a 34 amino acid tandem repeat present in 17.5 iterations. DNA-binding base pair specificity is conferred by variable di-residues in positions 12 and 13 within each tandem repeat. 444 proteins contain numerous shorter tandem repeats that are considerably more variable in both sequence and organisation. Tandem motif 1.1 does contain highly conserved regions directly preceded by variable di-residues. However, 444 proteins are localised in the apoplasm and so DNA binding is unlikely, although other ligand binding may be possible. It is unclear what the role of the different 444 subfamilies, variable number/organisation of tandem repeats, or domains is in the context of ligand binding. 444 tandem repeats of subfamilies -A, -B, and -C contain conserved glycine, proline and lysine residues. Glycine residues often create flexible linkers between domains. Taken together, these data may suggest that variable residues interspersed by highly conserved linker regions play a role in ligand binding in the apoplasm.

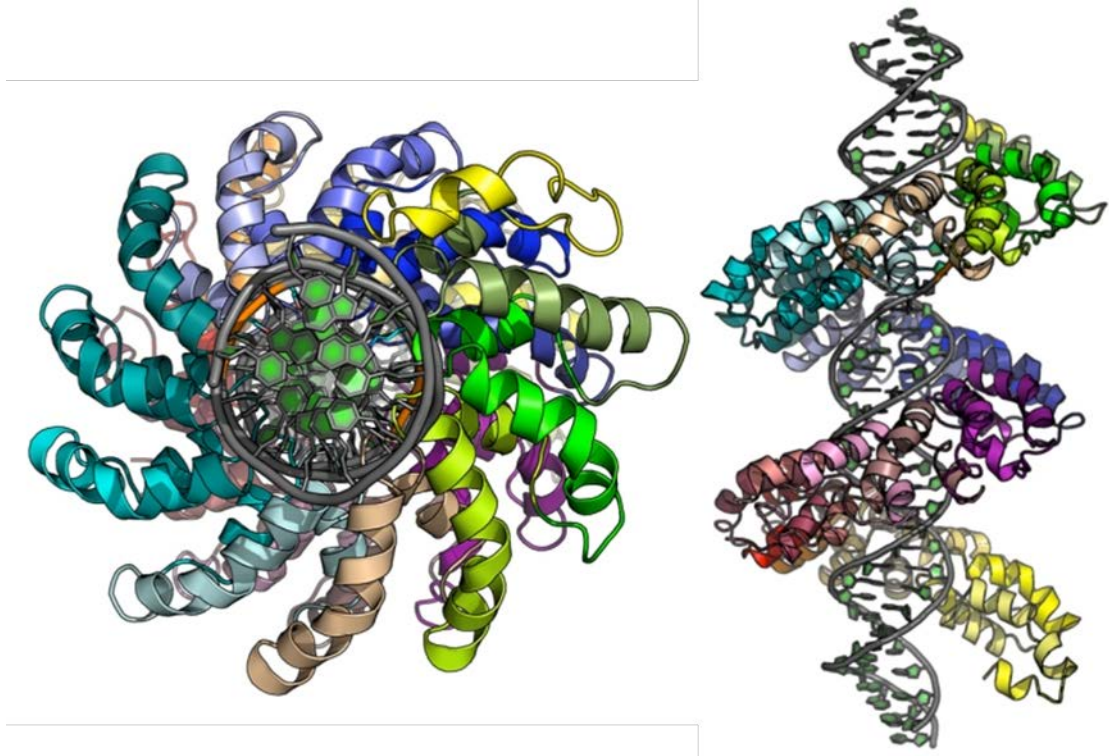


Figure 7-2 **Transcription activator like effectors DNA-bound structure.** The regular tandem repeats of the Transcription Activator Like (TAL) effectors of *Xanthomonas* bacteria form a super-helix around the double stranded DNA (DeFrancesco, 2012).

Other apoplastic effectors of plant-parasitic nematodes have been identified with distinct roles in parasitism, including suppression of host defence responses. A calreticulin (Mi-CRT) is secreted into the apoplasm from the subventral gland cells of *M. incognita*. Over expression of Mi-CRT in *A. thaliana* reduced expression of endogenous defence marker genes and defence responses, including callose deposition, induced by the bacterial PTI elicitor elf18 (Jaouannet et al., 2013). In addition, plants over-expressing Mi-CRT were more susceptible to *M. incognita* and the oomycete root pathogen *Phytophthora parasitica* (Jaouannet et al., 2013). It is hypothesised that Mi-CRT prevents calcium influx by direct binding in order to suppress immunity. Calcium influx is an early and essential step of the PTI signalling pathway. Presumably elf18 was used as at present no direct PTI-inducing elicitors have been identified from nematodes.

It has been known for some time that plant-parasitic nematode genomes encode a range of cell wall degrading enzymes. However, recent direct detection of several nematode-derived cell wall degrading enzymes in the apoplasm has been described (Vieira et al., 2011). This degradation of endogenous cell walls, in addition to the cell wall rearrangements that take place in syncytial development (Davies et al., 2012), suggest that there may be large scale induction of DAMP-triggered immunity (DTI). DTI also requires a similar calcium influx to PTI. In the absence of any identified nematode PTI elicitor it is possible that for the root-knot nematodes Mi-CRT is also able to suppress DTI. Only relatively recently has cell wall degrading enzyme-induced DTI suppression been demonstrated in bacteria (Sinha et al., 2013). Mi-CRT is taxonomically restricted to the root-knot nematodes and is absent from the cyst nematodes.

These data, coupled with the possibility that the 444 tandem repeat domain has a role in ligand binding, may suggest a similar defence suppression function for the 444 proteins. It is possible that the 444 tandem repeat domain variation provides glycan binding specificity. 444 proteins present in the apoplasm may sequester DAMPs and out-compete plant pattern recognition receptors. Variable number tandem repeats may encode specificity for variable chain oligogalacturonides. More work will be required to determine glycan/protein binding capacity of 444 proteins by either glycan array or yeast-two-hybrid analysis (Laurent et al., 2008). Presumably if this is the case, Mi-CRT *in-planta* over expression should rescue 444 RNAi-dependent reduced susceptibility.

Precedent has been demonstrated for an effector able to sequester elicitors of the PTI response. The LysM effectors of fungal pathogens contain numerous LysM

domains in tandem repeats (Bolton et al., 2008). These are secreted into the apoplast where they bind parasite-chitin with micromolar affinity (de Jonge et al., 2010). The strong binding of LysM effectors to chitin makes it unavailable to the plant PRRs, therefore evading PTI. No effectors to date have been demonstrated to suppress DTI by sequestering DAMPs.

7.3.2. Tandem repeated genes

In addition to tandem repeat proteins, entire genes in tandem arrays needs to be considered. The 176 gene was present in the *G. pallida* genome <10 kilbases downstream of an almost identical copy. This identical copy appeared to be non-functional as it lacked a signal peptide or obvious start codon. Interestingly two 176 genes were present in the transcriptome assembly for the closely related *G. rostochiensis*, perhaps suggesting that the duplication event occurred before the divergence of these two species. Although it remains to be confirmed, the 444 gene family may also be present in tandem repeats. The *G. pallida* genome assembly reflects less than 1/10th of the known 444 sequence variation within the 'Lindley' population. 444 sequences are collapsed into just two individual genes. This highlights a limitation of 76 base-pair illumina reads for assembly of repetitive genes, and consequently the genomic organisation of the 444 genes remains unclear.

Multiple tandem duplications of a single effector, or clusters of effectors, in the genome sequence of pathogens are known as virulence or pathogenicity islands (Hacker et al., 1997). Pathogenicity islands are well described for bacteria and can be present on mobile genetic elements such as transposons or plasmids (Hacker et al., 1997). However, in eukaryotic pathogens, such as oomycetes, virulence genes often occur in 'plastic' gene-sparse regions of the genome. Selecting gene-sparse genome segments unique to *P. infestans* compared to two related pathogenic species will identify approximately 3% of all genes but 62% of known *P. infestans* effectors (Raffaele et al., 2010). This suggests that frequent occurrence of genomic rearrangements has contributed to effector diversity in *Phytophthora* species. For 444 genes, individual nematodes vary in the number, size, and type of effector subfamilies. It is unclear what the underlying genetic mechanism/s is/are that account for this variation in copy-number spanning an order of magnitude. Two alternative hypotheses can be presented, 1) The variation observed in 444 genes results from sexual recombination between several different original populations of *G. pallida* or, 2) the diversity is regularly introduced from within a population through frequent genomic rearrangements and recombination events.

If the latter is true, this would suggest that each generation has a different 444 gene complement to the previous generation and that these would presumably be selected for in each feeding female. The variation in 444 effector complement was catalogued from successful feeding females due to the constraints of carrying our PCR from small samples. All successful individuals contained at least one member of at least one subfamily; some subfamilies are apparently non-essential. Determining the 444 complement of non-successful individuals would be problematic but may begin to elucidate the genetic origin of the diversity within the 444 gene family.

State-of-the-art PacBio technology is capable of producing read lengths over 8000 base pairs; >6 times the length of the longest 444 effector genomic sequence. *De novo* assembly of such long read data may generate the necessary genetic context to determine 444 effector genomic arrangements. Comparing 'perfect' genome assemblies originating from two independent lines of *G. pallida* unique in their 444 gene complement will determine whether local tandem duplications produce arrays of 444 effectors, or if regular partial-genomic rearrangements contribute to the variation.

8. References

- 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., HAMAMOUCHE, N., HEWEZI, T., JAILLON, O., JUBIN, C., LEONETTI, P., MAGLIANO, M., MAIER, T. R., MARKOV, G. V., MCVEIGH, P., PESOLE, G., POULAIN, J., ROBINSON-RECHAVI, M., SALLET, E., SEGURENS, B., STEINBACH, D., TYTGAT, T., UGARTE, E., VAN GHELDER, C., VERONICO, P., BAUM, T. J., BLAXTER, M., BLEVEZACHEO, T., DAVIS, E. L., EWBANK, J. J., FAVERY, B., GRENIER, E., HENRISSAT, B., JONES, J. T., LAUDET, V., MAULE, A. G., QUESNEVILLE, H., ROSSO, M.-N., SCHIEX, T., SMANT, G., WEISSENBAACH, J. & WINCKER, P. 2008. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, 26, 909-915.
- AGUDELO, P., ROBBINS, R. T., STEWART, J. M., BELL, A. & ROBINSON, A. F. 2005. Histological observations of *Rotylenchulus reniformis* on *Gossypium longicalyx* and interspecific cotton hybrids. *Journal of Nematology*, 37, 444-447.
- AIST, J. R. 1976. Papillae and related wound plugs of plant-cells. *Annual Review of Phytopathology*, 14, 145-163.
- ALEND, C., GALLOT-LEGRAND, A., FOUVILLE, D. & GRENIER, E. 2013. Sequence polymorphism of nematode effectors highlights molecular differences among the subspecies of the tobacco cyst nematode complex. *Physiological and Molecular Plant Pathology*, 84, 107-114.
- ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biology*, 11.
- ANDRADE, M. A., PEREZ-IRATXETA, C. & PONTING, C. P. 2001. Protein repeats: structures, functions, and evolution. *Journal of Structural Biology*, 134, 117-131.
- ATKINSON, H. J., URWIN, P. E., HANSEN, E. & MCPHERSON, M. J. 1995. Designs for engineered resistance to root-parasitic nematodes *Trends in Biotechnology*, 13, 369-374.
- BAKHETIA, M., CHARLTON, W., ATKINSON, H. J. & MCPHERSON, M. J. 2005. RNA interference of dual oxidase in the plant nematode *Meloidogyne incognita*. *Molecular Plant-Microbe Interactions*, 18, 1099-1106.
- BARONE, A., RITTER, E., SCHACHTSCHABEL, U., DEBENER, T., SALAMINI, F. & GEBHARDT, C. 1990. Localization by restriction-fragment-length-polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Molecular & General Genetics*, 224, 177-182.
- BAUM, J. A., BOGAERT, T., CLINTON, W., HECK, G. R., FELDMANN, P., ILAGAN, O., JOHNSON, S., PLAETINCK, G., MUNYIKWA, T., PLEAU, M., VAUGHN, T. & ROBERTS, J. 2007. Control of coleopteran insect pests through RNA interference. *Nature Biotechnology*, 25, 1322-6.

- BAUTERS, L., HAEGEMAN, A., KYNDT, T. & GHEYSEN, G. 2014. Analysis of the transcriptome of *Hirschmanniella oryzae* to explore potential survival strategies and host–nematode interactions. *Molecular Plant Pathology*, 15, 352-363.
- BELLAFIORE, 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, e1000192.
- BENDTSEN, J. D., JENSEN, L. J., BLOM, N., VON HEIJNE, G. & BRUNAK, S. 2004. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design and Selection*, 17, 349-356.
- BERG, R., FESTER, T. & TAYLOR, C. 2009. Development of the root-knot nematode feeding cell. In: BERG, R. & TAYLOR, C. (eds.) *Cell Biology of Plant Nematode Parasitism*. Springer Berlin / Heidelberg.
- BIRD, A. F. 1983. Growth and molting in nematodes - changes in the dimensions and morphology of *Rotylenchulus reniformis* from start to finish of molting. *International Journal for Parasitology*, 13, 201-206.
- BLAXTER, M. 2011. Nematodes: the worm and its relatives. *PLoS biology*, 9, e1001050.
- BLEVE-ZACHEO, T., RUBINO, L., MELILLO, M. T. & RUSSO, M. 1997. The 33 K protein encoded by cymbidium ringspot tomosvirus localizes to modified peroxisomes of infected cells and of uninfected transgenic plants. *Journal of Plant Pathology*, 79, 197-202.
- BLOK, V. C., MALLOCH, G., HARROWER, B., PHILLIPS, M. S. & VRAIN, T. C. 1998. Intraspecific variation in ribosomal DNA in populations of the potato cyst nematode *Globodera pallida*. *Journal of Nematology*, 30, 262-274.
- BOCKENHOFF, A. & GRUNDLER, F. M. W. 1994. Studies on the nutrient-uptake by the beet cyst-nematode *Heterodera schachtii* by in-situ microinjection of fluorescent-probes into the feeding structures in *Arabidopsis thaliana*. *Parasitology*, 109, 249-&.
- BOLTON, M. D., VAN ESSE, H. P., VOSSSEN, J. H., DE JONGE, R., STERGIOPOULOS, I., STULEMEIJER, I. J., VAN DEN BERG, G., BORRÁS-HIDALGO, O., DEKKER, H. L. & DE KOSTER, C. G. 2008. The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. *Molecular Microbiology*, 69, 119-136.
- BORATYN, G. M., CAMACHO, C., COOPER, P. S., COULOURIS, G., FONG, A., MA, N., MADDEN, T. L., MATTEN, W. T., MCGINNIS, S. D. & MEREZHUK, Y. 2013. BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, 41, W29-W33.
- BURSET, M., SELEDTSOV, I. & SOLOVYEV, V. 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research*, 28, 4364-4375.
- CARPENTER, J. F., KENDRICK, B. S., CHANG, B. S., MANNING, M. C. & RANDOLPH, T. W. 1999. Inhibition of stress-induced aggregation of protein therapeutics. *Amyloid, Prions, and Other Protein Aggregates*, 309, 236-255.
- CASTAGNONE-SERENO, P., SEMBLAT, J.-P. & CASTAGNONE, C. 2009. Modular architecture and evolution of the *map-1* gene family in the root-knot nematode *Meloidogyne incognita*. *Molecular Genetics and Genomics*, 282, 547-554.

- CHAPMAN, S., FAULKNER, C., KAISERLI, E., GARCIA-MATA, C., SAVENKOV, E. I., ROBERTS, A. G., OPARKA, K. J. & CHRISTIE, J. M. 2008. The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 20038-20043.
- CHOWDAPPA, P., KUMAR, N. B., MADHURA, S., KUMAR, M. S., MYERS, K. L., FRY, W. E., SQUIRES, J. N. & COOKE, D. E. 2013. Emergence of 13_A2 blue lineage of *Phytophthora infestans* was responsible for severe outbreaks of late blight on tomato in south-west India. *Journal of Phytopathology*, 161, 49-58.
- CLEMMER, D. E. & JARROLD, M. F. 1997. Ion mobility measurements and their applications to clusters and biomolecules. *Journal of Mass Spectrometry*, 32, 577-592.
- COCQUET, J., CHONG, A., ZHANG, G. & VEITIA, R. A. 2006. Reverse transcriptase template switching and false alternative transcripts. *Genomics*, 88, 127-131.
- COHN, E. & MORDECHAI, M. 1977. Uninucleate giant cell induced in soybean by the nematode *Rotylenchulus macrodoratus*. *Phytoparasitica*, 5, 85-93.
- COLLINS, F. S., MORGAN, M. & PATRINOS, A. 2003. The Human Genome Project: lessons from large-scale biology. *Science*, 300, 286-290.
- CONESA, A., GÖTZ, S., GARCÍA-GÓMEZ, J. M., TEROL, J., TALÓN, M. & ROBLES, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21, 3674-3676.
- 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. 2014. The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology*, 15, R43.
- CREER, S., FONSECA, V., PORAZINSKA, D., GIBLIN-DAVIS, R., SUNG, W., POWER, D., PACKER, M., CARVALHO, G., BLAXTER, M. & LAMBSHEAD, P. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, 19, 4-20.
- DALZELL, J. J., MCVEIGH, P., WARNOCK, N. D., MITREVA, M., BIRD, D. M., ABAD, P., FLEMING, C. C., DAY, T. A., MOUSLEY, A. & MARKS, N. J. 2011. RNAi effector diversity in nematodes. *PLoS neglected tropical diseases*, 5, e1176.
- DANCHIN, E. G. J., ROSSO, M.-N., VIEIRA, P., DE ALMEIDA-ENGLER, J., COUTINHO, P. M., HENRISSAT, B. & ABAD, P. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 17651-17656.
- DAVIES, L. J., LILLEY, C. J., KNOX, J. P. & URWIN, P. E. 2012. Syncytia formed by adult female *Heterodera schachtii* in *Arabidopsis thaliana* roots have a distinct cell wall molecular architecture. *New Phytologist*, 196, 238-246.
- DAVIS, E. L., HUSSEY, R. S. & BAUM, T. J. 2004. Getting to the roots of parasitism by nematodes. *Trends in Parasitology*, 20, 134-141.
- DE ALMEIDA ENGLER, J., DE VLEESSCHAUWER, V., BURSSSENS, S., CELENZA, J. L., INZÉ, D., VAN MONTAGU, M., ENGLER, G. & GHEYSEN, G. 1999. Molecular markers and cell cycle inhibitors show the importance of

cell cycle progression in nematode-induced galls and syncytia. *The Plant Cell Online*, 11, 793-807.

- DE ALMEIDA ENGLER, J. & GHEYSEN, G. 2013. Nematode-induced endoreduplication in plant host cells: why and how? *Molecular Plant-Microbe Interactions*, 26, 17-24.
- DE BOER, J. M., YAN, Y., SMANT, G., DAVIS, E. L. & BAUM, T. J. 1998. *In situ* hybridization to messenger RNA in *Heterodera glycines*. *J Nematol*, 30, 309-312.
- DE JONGE, R., VAN ESSE, H. P., KOMBRINK, A., SHINYA, T., DESAKI, Y., BOURS, R., VAN DER KROL, S., SHIBUYA, N., JOOSTEN, M. H. & THOMMA, B. P. 2010. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science*, 329, 953-955.
- DE LORENZO, G., BRUTUS, A., SAVATIN, D. V., SICILIA, F. & CERVONE, F. 2011. Engineering plant resistance by constructing chimeric receptors that recognize damage-associated molecular patterns (DAMPs). *FEBS Letters*, 585, 1521-1528.
- DEFRANCESCO, L. 2012. TAL effector-DNA structure. *Nat Biotech*, 30, 158-158.
- DEVINE, K., BYRNE, J., MAHER, N. & JONES, P. 1996. Resolution of natural hatching factors for golden potato cyst nematode, *Globodera rostochiensis*. *Annals of Applied Biology*, 129, 323-334.
- DILLIES, M.-A., RAU, A., AUBERT, J., HENNEQUET-ANTIER, C., JEANMOUGIN, M., SERVANT, N., KEIME, C., MAROT, G., CASTEL, D., ESTELLE, J., GUERNEC, G., JAGLA, B., JOUINEAU, L., LALOE, D., LE GALL, C., SCHAEFFER, B., LE CROM, S., GUEDJ, M., JAFFREZIC, F. & FRENCH STATOMIQUE, C. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, 14, 671-83.
- DJOUAKA, R. F., BAKARE, A. A., COULIBALY, O. N., AKOGBETO, M. C., RANSON, H., HEMINGWAY, J. & STRODE, C. 2008. Expression of the cytochrome P450s, CYP6P3 and CYP6M2 are significantly elevated in multiple pyrethroid resistant populations of *Anopheles gambiae* s.s. from Southern Benin and Nigeria. *Bmc Genomics*, 9, 538.
- DOYLE, E. L., STODDARD, B. L., VOYTAS, D. F. & BOGDANOVA, A. J. 2013. TAL effectors: highly adaptable phyto-bacterial virulence factors and readily engineered DNA-targeting proteins. *Trends in Cell Biology*, 23, 390-398.
- EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792-1797.
- EDWARDS, A., DANIELS, B. S. & DEEN, W. M. 1999. Ultrastructural model for size selectivity in glomerular filtration. *American Journal of Physiology-Renal Physiology*, 276, F892-F902.
- ELLING, A. A., DAVIS, E. L., HUSSEY, R. S. & BAUM, T. J. 2007. Active uptake of cyst nematode parasitism proteins into the plant cell nucleus. *International Journal for Parasitology*, 37, 1269-1279.
- ENDO, B. Y. 1978. Feeding plug formation in soybean roots infected with the soybean cyst nematode. *Phytopathology*, 68, 1022-1031.
- EVANS, K., FRANCO, J. & DE SCURRAH, M. M. 1975. Distribution of species of potato cyst-nematodes in South America. *Nematologica*, 21, 365-369.
- EVES-VAN DEN AKKER, S., LILLEY, C., DANCHIN, E., RANCUREL, C., COCK, P., URWIN, P. & JONES, J. 2014a. The transcriptome of *Nacobbus*

- aberrans* reveals insights into the evolution of sedentary endoparasitism in plant-parasitic nematodes. *Genome Biology and Evolution*, evu171.
- EVES-VAN DEN AKKER, S., LILLEY, C. J., AULT, J. R., ASHCROFT, A. E., JONES, J. T. & URWIN, P. E. 2014b. The feeding tube of cyst nematodes: Characterisation of protein exclusion. *Plos One*, 9, e87289.
- FANELLI, E., DI VITO, M., JONES, J. T. & DE GIORGI, C. 2005. Analysis of chitin synthase function in a plant parasitic nematode, *Meloidogyne artiellia*, using RNAi. *Gene*, 349, 87-95.
- FIELDS, G. B., ALONSO, D. O. V., STIGTER, D. & DILL, K. A. 1992. Theory for the aggregation of proteins and copolymers. *Journal of Physical Chemistry*, 96, 3974-3981.
- FIRE, A., XU, S., MONTGOMERY, M. K., KOSTAS, S. A., DRIVER, S. E. & MELLO, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391, 806-11.
- GANJI, S., JENKINS, J. N. & WUBBEN, M. J. 2014. Molecular characterization of the reniform nematode C-type lectin gene family reveals a likely role in mitigating environmental stresses during plant parasitism. *Gene*, 537, 269-278.
- GAO, B. L., ALLEN, R., MAIER, T., DAVIS, E. L., BAUM, T. J. & HUSSEY, R. S. 2003. The parasitome of the phytonematode *Heterodera glycines*. *Molecular Plant-Microbe Interactions*, 16, 720-726.
- GHEYSEN, G. & FENOLL, C. 2002. Gene expression in nematode feeding sites. *Annual Review of Phytopathology*, 40, 191-219.
- GILBOA, E., MITRA, S. W., GOFF, S. & BALTIMORE, D. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell*, 18, 93-100.
- GILES, K., PRINGLE, S. D., WORTHINGTON, K. R., LITTLE, D., WILDGOOSE, J. L. & BATEMAN, R. H. 2004. Applications of a travelling wave-based radio-frequency only stacked ring ion guide. *Rapid Communications in Mass Spectrometry*, 18, 2401-2414.
- GLEAVE, A. P. 1992. A versatile binary vector system with a T-DNA organizational-structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Molecular Biology*, 20, 1203-1207.
- GOVERSE, A., BIESHEUVEL, J., WIJERS, G. J., GOMMERS, F. J., BAKKER, J., SCHOTS, A. & HELDER, J. 1998. *In planta* monitoring of the activity of two constitutive promoters, CaMV 35S and TR2', in developing feeding cells induced by *Globodera rostochiensis* using green fluorescent protein in combination with confocal laser scanning microscopy. *Physiological and Molecular Plant Pathology*, 52, 275-284.
- GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCALI, E., HACOEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N. & REGEV, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29, 644-U130.
- GRAY, J. E., PICTON, S., GIOVANNONI, J. J. & GRIERSON, D. 1994. The use of transgenic and naturally-occurring mutants to understand and manipulate toato fruit ripening. *Plant Cell and Environment*, 17, 557-571.

- GRENIER, E., FOURNET, S., PETIT, E. & ANTHOINE, G. 2010. A cyst nematode 'species factory' called the Andes. *Nematology*, 12, 163-169.
- GRUNDLER, F. M. & BÖCKENHOFF, A. 1997. Physiology of nematode feeding and feeding sites. In: FENOLL, C., GRUNDLER, F. M. & OHL, S. A. (eds.) *Cellular and molecular aspects of plant-nematode interactions*. Springer.
- GRUNDLER, F. M. W., SOBCZAK, M. & GOLINOWSKI, W. 1998. Formation of wall openings in root cells of *Arabidopsis thaliana* following infection by the plant-parasitic nematode *Heterodera schachtii*. *European Journal of Plant Pathology*, 104, 545-551.
- GUASCH, A., DEEN, W. M. & MYERS, B. D. 1993. Charge selectivity of the glomerular-filtration barrier in healthy and nephrotic humans. *Journal of Clinical Investigation*, 92, 2274-2282.
- HACKER, J., BLUM-OEHLER, G., MÜHLDORFER, I. & TSCHÄPE, H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Molecular Microbiology*, 23, 1089-1097.
- HAEGEMAN, A., JONES, J. T. & DANCHIN, E. G. J. 2011. Horizontal Gene Transfer in Nematodes: A Catalyst for Plant Parasitism? *Molecular Plant-Microbe Interactions*, 24, 879-887.
- HAEGEMAN, A., MANTELIN, S., JONES, J. T. & GHEYSEN, G. 2012. Functional roles of effectors of plant-parasitic nematodes. *Gene*, 492, 19-31.
- HAMAMOUCHE, N., LI, C., HEWEZI, T., BAUM, T. J., MITCHUM, M. G., HUSSEY, R. S., VODKIN, L. O. & DAVIS, E. L. 2012. The interaction of the novel 30C02 cyst nematode effector protein with a plant β -1, 3-endoglucanase may suppress host defence to promote parasitism. *Journal of Experimental Botany*, 63, 3683-3695.
- HAMMOND, S. M., BERNSTEIN, E., BEACH, D. & HANNON, G. J. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*, 404, 293-296.
- HARDCASTLE, T. & KELLY, K. 2010. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, 11, 422.
- HEINEMANN, J. A., AGAPITO-TENFEN, S. Z. & CARMAN, J. A. 2013. A comparative evaluation of the regulation of GM crops or products containing dsRNA and suggested improvements to risk assessments. *Environment International*, 55, 43-55.
- HEWEZI, T. & BAUM, T. J. 2013. Manipulation of plant cells by cyst and root-knot nematode effectors. *Molecular Plant-Microbe Interactions*, 26, 9-16.
- HOFMANN, J. & GRUNDLER, F. M. W. 2006. Females and males of root-parasitic cyst nematodes induce different symplasmic connections between their syncytial feeding cells and the phloem in *Arabidopsis thaliana*. *Plant Physiology and Biochemistry*, 44, 430-433.
- HOGAN, C. J. J., RUOTOLO, B. T., ROBINSON, C. V. & DE LA MORA, J. F. 2011. Tandem differential mobility analysis-mass spectrometry reveals partial gas-phase collapse of the GroEL complex. *Journal of Physical Chemistry B*, 115, 3614-3621.
- HOLTMANN, B., KLEINE, M. & GRUNDLER, F. M. W. 2000. Ultrastructure and anatomy of nematode-induced syncytia in roots of susceptible and resistant sugar beet. *Protoplasma*, 211, 39-50.

- HOOLAHAN, A. H., BLOK, V. C., GIBSON, T. & DOWTON, M. 2012. A comparison of three molecular markers for the identification of populations of *Globodera pallida*. *Journal of Nematology*, 44, 7-17.
- HOPP, T. P. & WOODS, K. R. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences*, 78, 3824-3828.
- HOPPER, J. T. S. & OLDHAM, N. J. 2009. Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: The effect of ligand binding on conformational stability. *Journal of the American Society for Mass Spectrometry*, 20, 1851-1858.
- HOUSELEY, J. & TOLLERVEY, D. 2010. Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. *Plos One*, 5, e12271.
- HUANG, G., ALLEN, R., DAVIS, E. L., BAUM, T. J. & HUSSEY, R. S. 2006. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proceedings of the National Academy of Sciences*, 103, 14302-14306.
- HUSSEY, R. S. 1989. Disease-inducing secretion of plant-parasitic nematodes. *Annual Review of Phytopathology*, 27, 123-141.
- HUSSEY, R. S. & MIMS, C. W. 1991. Ultrastructure of feeding tubes formed in giant-cells induced in plants by the root-knot nematode *Meloidogyne incognita*. *Protoplasma*, 162, 99-107.
- IBRAHIM, H. M., ALKHAROUF, N. W., MEYER, S. L., ALY, M. A., GAMAL EL-DIN, A. E. K. Y., HUSSEIN, E. H. & MATTHEWS, B. F. 2011. Post-transcriptional gene silencing of root-knot nematode in transformed soybean roots. *Experimental Parasitology*, 127, 90-99.
- JAOUANNET, M., MAGLIANO, M., ARGUEL, M., GOURGUES, M., EVANGELISTI, E., ABAD, P. & ROSSO, M. 2013. The root-knot nematode calreticulin Mi-CRT is a key effector in plant defense suppression. *Molecular Plant-Microbe Interactions*, 26, 97-105.
- JAOUANNET, M., PERFUS-BARBEOCH, L., DELEURY, E., MAGLIANO, M., ENGLER, G., VIEIRA, P., DANCHIN, E. G., ROCHA, M. D., COQUILLARD, P. & ABAD, P. 2012. A root knot nematode secreted protein is injected into giant cells and targeted to the nuclei. *New Phytologist*, 194, 924-931.
- JAUBERT, S., LAFFAIRE, J. B., ABAD, P. & ROSSO, M. N. 2002. A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. *Febs Letters*, 522, 109-112.
- JENSEN, K. H., MULLENDRE, D. L., HOLBROOK, N. M., BOHR, T., KNOBLAUCH, M. & BRUUS, H. 2012. Modeling the hydrodynamics of phloem sieve plates. *Frontiers in plant science*, 3.
- JENSEN, K. H., VALENTE, A. X. & STONE, H. A. 2014. Flow rate through microfilters: Influence of the pore size distribution, hydrodynamic interactions, wall slip, and inertia. *Physics of Fluids (1994-present)*, 26, 052004.
- JONES, J., REAVY, B., SMANT, G. & PRIOR, A. 2004. Glutathione peroxidases of the potato cyst nematode *Globodera Rostochiensis*. *Gene*, 324, 47-54.
- 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, 946-961.
- JONES, J. T., KUMAR, A., PYLYPENKO, L. A., THIRUGNANASAMBANDAM, A., CASTELLI, L., CHAPMAN, S., COCK, P. J. A., GRENIER, E., LILLEY, C. J., PHILLIPS, M. S. & BLOK, V. C. 2009. Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology*, 10, 815-828.
- JONES, J. T., PERRY, R. N. & JOHNSTON, M. R. L. 1994. Changes in the ultrastructure of the amphids of the potato cyst-nematode, *Globodera rostochiensis*, during development and infection. *Fundamental and Applied Nematology*, 17, 369-382.
- JONES, M. G. K., NOVACKY, A. & DROPKIN, V. H. 1974. Action-potentials in nematode-induced plant transfer cells. *Protoplasma*, 80, 401-405.
- JONES, M. G. K. & PAYNE, H. L. 1977. Structure of syncytia induced by phytoparasitic nematode *Nacobbus aberrans* in tomato roots, and possible role of plasmodesmata in their nutrition. *Journal of Cell Science*, 23, 299-313.
- JURNECZKO, E. & BARRAN, P. E. 2011. How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst*, 136, 20-28.
- KELLNER, R., BHATTACHARYYA, A., POPPE, S., HSU, T. Y., BREM, R. B. & STUKENBROCK, E. H. 2014. Expression profiling of the wheat pathogen *Zymoseptoria tritici* reveals genomic patterns of transcription and host-specific regulatory programs. *Genome Biology and Evolution*.
- KENT, W. J. 2002. BLAT—the BLAST-like alignment tool. *Genome Research*, 12, 656-664.
- KIESE, S., PAPPENBERGER, A., FRIESS, W. & MAHLER, H. C. 2008. Shaken, not stirred: Mechanical stress testing of an IgG1 antibody. *Journal of Pharmaceutical Sciences*, 97, 4347-4366.
- KIMBER, M. J., MCKINNEY, S., MCMASTER, S., DAY, T. A., FLEMING, C. C. & MAULE, A. G. 2007. flp gene disruption in a parasitic nematode reveals motor dysfunction and unusual neuronal sensitivity to RNA interference. *The FASEB Journal*, 21, 1233-1243.
- KOBE, B. & DEISENHOFER, J. 1993. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature*, 751-756.
- KUNSCH, C. & MEDFORD, R. M. 1999. Oxidative stress as a regulator of gene expression in the vasculature. *Circulation Research*, 85, 753-766.
- KUNZE, G., ZIPFEL, C., ROBATZEK, S., NIEHAUS, K., BOLLER, T. & FELIX, G. 2004. The N terminus of bacterial elongation factor Tu elicits innate immunity in *Arabidopsis* plants. *The Plant Cell Online*, 16, 3496-3507.
- KYNDT, T., VIEIRA, P., GHEYSEN, G. & DE ALMEIDA-ENGLER, J. 2013. Nematode feeding sites: unique organs in plant roots. *Planta*, 238, 807-818.
- LAURENT, N., VOGLMEIR, J. & FLITSCH, S. L. 2008. Glycoarrays—tools for determining protein–carbohydrate interactions and glycoenzyme specificity. *Chemical Communications*, 4400-4412.

- LEBOWITZ, J., LEWIS, M. S. & SCHUCK, P. 2002. Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Science*, 11, 2067-2079.
- LEE, C., CHRONIS, D., KENNING, C., PERET, B., HEWEZI, T., DAVIS, E. L., BAUM, T. J., HUSSEY, R., BENNETT, M. & MITCHUM, M. G. 2011. The novel cyst nematode effector protein 19C07 interacts with the *Arabidopsis* auxin influx transporter LAX3 to control feeding site development. *Plant Physiology*, 155, 866-880.
- LI, X.-Q., WEI, J.-Z., TAN, A. & AROIAN, R. V. 2007. Resistance to root-knot nematode in tomato roots expressing a nematocidal *Bacillus thuringiensis* crystal protein. *Plant Biotechnology Journal*, 5, 455-464.
- LIANG, P.-H., WU, C.-Y., GREENBERG, W. A. & WONG, C.-H. 2008. Glycan arrays: biological and medical applications. *Current Opinion in Chemical Biology*, 12, 86-92.
- LILLEY, C., DAVIES, L. & URWIN, P. 2012. RNA interference in plant parasitic nematodes: a summary of the current status. *Parasitology*, 139, 630-640.
- LILLEY, C. J., BAKHETIA, M., CHARLTON, W. L. & URWIN, P. E. 2007. Recent progress in the development of RNA interference for plant parasitic nematodes. *Molecular Plant Pathology*, 8, 701-711.
- LINFORD, M. & YAP, F. 1940. Some host plants of the reniform nematode in Hawaii. *Proceedings of the Helminthological Society of Washington*, 7, 42-44.
- LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*.
- LOUKAS, A., MULLIN, N. P., TETTEH, K. K. A., MOENS, L. & MAIZELS, R. M. 1999. A novel C-type lectin secreted by a tissue-dwelling parasitic nematode. *Current Biology*, 9, 825-828.
- LUIS GARCIA-GIMENEZ, J., MARKOVIC, J., DASI, F., QUEVAL, G., SCHNAUBELT, D., FOYER, C. H. & PALLARDO, F. V. 2013. Nuclear glutathione. *Biochimica Et Biophysica Acta-General Subjects*, 1830, 3304-3316.
- MACK, E. 1925. Average cross-sectional areas of molecules by gaseous diffusion methods. *Journal of the American Chemical Society*, 47, 2468-2482.
- MARCOTTE, E. M., PELLEGRINI, M., YEATES, T. O. & EISENBERG, D. 1999. A census of protein repeats. *Journal of Molecular Biology*, 293, 151-160.
- MAULE, A. G., MCVEIGH, P., DALZELL, J. J., ATKINSON, L., MOUSLEY, A. & MARKS, N. J. 2011. An eye on RNAi in nematode parasites. *Trends in Parasitology*, 27, 505-513.
- MCCARTER, J., GAO, B. L., BRADLEY, J., HRESKO, M. C., CARUANO-YZERMANS, A. & WILLIAMS, D. J. 2010. Uptake and exclusion of plant-expressed fluorescent proteins by the soybean cyst nematode *Heterodera glycines*. *Journal of Nematology*, 42, 256-256.
- MESLEH, M. F., HUNTER, J. M., SHVARTSBURG, A. A., SCHATZ, G. C. & JARROLD, M. F. 1996. Structural information from ion mobility measurements: Effects of the long-range potential. *Journal of Physical Chemistry*, 100, 16082-16086.
- MILNE, I., LINDNER, D., BAYER, M., HUSMEIER, D., MCGUIRE, G., MARSHALL, D. F. & WRIGHT, F. 2009. TOPALi v2: a rich graphical interface for

evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics*, 25, 126-127.

- MINNIS, S. T., HAYDOCK, P. P. J., IBRAHIM, S. K., GROVE, I. G., EVANS, K. & RUSSELL, M. D. 2002. Potato cyst nematodes in England and Wales - occurrence and distribution. *Annals of Applied Biology*, 140, 187-195.
- MUGLIA, C., COMAI, G., SPEGAZZINI, E., RICCILLO, P. M. & AGUILAR, O. M. 2008. Glutathione produced by *Rhizobium tropici* is important to prevent early senescence in common bean nodules. *Fems Microbiology Letters*, 286, 191-198.
- NOWARA, D., GAY, A., LACOMME, C., SHAW, J., RIDOUT, C., DOUCHKOV, D., HENSEL, G., KUMLEHN, J. & SCHWEIZER, P. 2010. HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen *Blumeria graminis*. *Plant Cell*, 22, 3130-41.
- OSHLACK, A. & WAKEFIELD, M. J. 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, 4.
- PAAL, J., HENSELEWSKI, H., MUTH, J., MEKSEM, K., MENENDEZ, C. M., SALAMINI, F., BALLVORA, A. & GEBHARDT, C. 2004. Molecular cloning of the potato Gro1-4 gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *Plant Journal*, 38, 285-297.
- PAPAYIANNIS, L. C., CHRISTOFOROU, M., MARKOU, Y. M. & TSALTAS, D. 2013. Molecular typing of cyst-forming nematodes *Globodera pallida* and *G. rostochiensis*, using real-time PCR and evaluation of five methods for template preparation. *Journal of Phytopathology*, 161, 459-469.
- PERRY, R. N. 1996. Chemoreception in plant parasitic nematodes. *Annual Review of Phytopathology*, 34, 181-199.
- PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 785-786.
- PICARD, D., PLANTARD, O., SCURRAH, M. & MUGNIERY, D. 2004. Inbreeding and population structure of the potato cyst nematode (*Globodera pallida*) in its native area (Peru). *Molecular Ecology*, 13, 2899-2908.
- PLANTARD, O., PICARD, D., VALETTE, S., SCURRAH, M., GRENIER, E. & MUGNIERY, D. 2008. Origin and genetic diversity of Western European populations of the potato cyst nematode (*Globodera pallida*) inferred from mitochondrial sequences and microsatellite loci. *Molecular Ecology*, 17, 2208-2218.
- PLIEGO, C., NOWARA, D., BONCIANI, G., GHEORGHE, D. M., XU, R., SURANA, P., WHIGHAM, E., NETTLETON, D., BOGDANOVE, A. J., WISE, R. P., SCHWEIZER, P., BINDSCHEDLER, L. V. & SPANU, P. D. 2013. Host-induced gene silencing in barley powdery mildew reveals a class of ribonuclease-like effectors. *Molecular Plant-Microbe Interactions*, 26, 633-642.
- POINAR, G., KERP, H. & HASS, H. 2008. *Palaeonema phyticum* gen. n., sp. n. (Nematoda: Palaeonematidae fam. n.), a Devonian nematode associated with early land plants. *Nematology*, 10, 9-14.
- POSTMA, W. J., SLOOTWEG, E. J., REHMAN, S., FINKERS-TOMCZAK, A., TYTGAT, T. O., VAN GELDEREN, K., LOZANO-TORRES, J. L., ROOSIEN, J., POMP, R. & VAN SCHAIK, C. 2012. The effector SPRYSEC-19 of

- Globodera rostochiensis* suppresses CC-NB-LRR-mediated disease resistance in plants. *Plant Physiology*, 160, 944-954.
- PRINGLE, S. D., GILES, K., WILDGOOSE, J. L., WILLIAMS, J. P., SLADE, S. E., THALASSINOS, K., BATEMAN, R. H., BOWERS, M. T. & SCRIVENS, J. H. 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry*, 261, 1-12.
- QIN, L., OVERMARS, H., HELDER, J., POPEIJUS, H., VAN DER VOORT, J. R., GROENINK, W., VAN KOERT, P., SCHOTS, A., BAKKER, J. & SMANT, G. 2000. An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*. *Molecular Plant-Microbe Interactions*, 13, 830-836.
- QUÉNÉHERVÉ, P., CHABRIER, C., AUWERKERKEN, A., TOPART, P., MARTINY, B. & MARIE-LUCE, S. 2006. Status of weeds as reservoirs of plant parasitic nematodes in banana fields in Martinique. *Crop protection*, 25, 860-867.
- QUENTIN, M., ABAD, P. & FAVERY, B. 2013. Plant parasitic nematode effectors target host defense and nuclear functions to establish feeding cells. *Frontiers in plant science*, 4.
- RAFFAELE, S., WIN, J., CANO, L. M. & KAMOUN, S. 2010. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genomics*, 11, 637.
- RAHMAN RAZAK, A. & EVANS, A. 1976. An intracellular tube associated with feeding by *Rotylenchulus reniformis* on cowpea root. *Nematologica*, 22, 182-189.
- RANSON, H., NIKOU, D., HUTCHINSON, M., WANG, X., ROTH, C. W., HEMINGWAY, J. & COLLINS, F. H. 2002. Molecular analysis of multiple cytochrome P450 genes from the malaria vector, *Anopheles gambiae*. *Insect Molecular Biology*, 11, 409-418.
- RAPAPORT, F., KHANIN, R., LIANG, Y., PIRUN, M., KREK, A., ZUMBO, P., MASON, C. E., SOCCI, N. D. & BETEL, D. 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, 14.
- RASO, S. W., ABEL, J., BARNES, J. M., MALONEY, K. M., PIPES, G., TREUHEIT, M. J., KING, J. & BREMS, D. N. 2005. Aggregation of granulocyte-colony stimulating factor *in vitro* involves a conformationally altered monomeric state. *Protein Science*, 14, 2246-2257.
- RAZAK, A. R. & EVANS, A. A. F. 1976. An intracellular tube associated with feeding by *Rotylenchulus reniformis* on cowpea root. *Nematologica*, 22, 182-189.
- REBOIS, R. V. 1980. Ultrastructure of a feeding peg and tube associated with *Rotylenchulus reniformis* in cotton. *Nematologica*, 26, 396-405.
- REBOIS, R. V., MADDEN, P. A. & ELDRIDGE, B. J. 1975. Some ultrastructural changes induced in resistant and susceptible soybean roots following infection by *Rotylenchulus reniformis*. *Journal of Nematology*, 7, 122-139.
- REHMAN, S., POSTMA, W., TYTGAT, T., PRINS, P., QIN, L., OVERMARS, H., VOSSSEN, J., SPIRIDON, L.-N., PETRESCU, A.-J. & GOVERSE, A. 2009. A secreted SPRY domain-containing protein (SPRYSEC) from the plant-parasitic nematode *Globodera rostochiensis* interacts with a CC-NB-LRR

- protein from a susceptible tomato. *Molecular Plant-Microbe Interactions*, 22, 330-340.
- REPLOGLE, A., WANG, J., BLECKMANN, A., HUSSEY, R. S., BAUM, T. J., SAWA, S., DAVIS, E. L., WANG, X., SIMON, R. & MITCHUM, M. G. 2011. Nematode CLE signaling in *Arabidopsis* requires CLAVATA2 and CORYNE. *The Plant Journal*, 65, 430-440.
- ROBERTS, C. J. 2007. Non-native protein aggregation kinetics. *Biotechnology and Bioengineering*, 98, 927-938.
- ROBERTSON, L., ROBERTSON, W. M., SOBCHAK, M., HELDER, J., TETAUD, E., ARIYANAYAGAM, M. R., FERGUSON, M. A., FAIRLAMB, A. & JONES, J. T. 2000. Cloning, expression and functional characterisation of a peroxiredoxin from the potato cyst nematode *Globodera rostochiensis*. *Molecular and Biochemical Parasitology*, 111, 41-49.
- ROBINSON, A. F., INSERRA, R. N., CASWELL-CHEN, E. P., VOVLAS, N. & TROCCOLI, A. 1998. Rotylenchulus species: Identification, distribution, host ranges, and crop plant resistance. *Nematropica*, 27, 127-180.
- ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.
- RODLAND, E. A. 2013. Compact representation of k-mer de Bruijn graphs for genome read assembly. *BMC Bioinformatics*, 14.
- ROSSO, M. N., JONES, J. T. & ABAD, P. 2009. RNAi and functional genomics in plant parasitic nematodes. *Annual Review of Phytopathology*, 47, 207-232.
- RUMPENHORST, H. J. 1984. Intracellular feeding tubes associated with sedentary plant parasitic nematodes. *Nematologica*, 30, 77-85.
- RUTSCHOW, H. L., BASKIN, T. I. & KRAMER, E. M. 2011. Regulation of solute flux through plasmodesmata in the root meristem. *Plant Physiology*, 155, 1817-1826.
- RUTTER, W. B., HEWEZI, T., MAIER, T. R., MITCHUM, M. G., DAVIS, E. L., HUSSEY, R. S. & BAUM, T. J. 2014. Members of the *Meloidogyne* avirulence protein family contain multiple plant ligand-like motifs. *Phytopathology*, 104, 879-885.
- SANABRIA, N., GORING, D., NÜRNBERGER, T. & DUBERY, I. 2008. Self/nonself perception and recognition mechanisms in plants: a comparison of self-incompatibility and innate immunity. *New Phytologist*, 178, 503-514.
- 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, 5463-5467.
- SASAGAWA, Y., NIKAIDO, I., HAYASHI, T., DANNO, H., UNO, K. D., IMAI, T. & UEDA, H. R. 2013. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biology*, 14.
- SCOTT, J. A., COLLINS, F. H. & FEYEREISEN, R. 1994. Diversity of cytochrome-p450 genes in the mosquito, *Anopheles albimanus*. *Biochemical and biophysical research communications*, 205, 1452-1459.
- SEMBLAT, J.-P., ROSSO, M.-N., HUSSEY, R. S., ABAD, P. & CASTAGNONE-SERENO, P. 2001. Molecular cloning of a cDNA encoding an amphid-secreted putative avirulence protein from the root-knot nematode *Meloidogyne incognita*. *Molecular Plant-Microbe Interactions*, 14, 72-79.

- SHELIMOV, K. B., CLEMMER, D. E., HUDGINS, R. R. & JARROLD, M. F. 1997. Protein structure in vacuo: Gas-phase confirmations of BPTI and cytochrome c. *Journal of the American Chemical Society*, 119, 2240-2248.
- SHVARTSBURG, A. A. & JARROLD, M. F. 1996. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters*, 261, 86-91.
- SHVARTSBURG, A. A., MASHKEVICH, S. V., BAKER, E. S. & SMITH, R. D. 2007. Optimization of algorithms for ion mobility calculations. *Journal of Physical Chemistry A*, 111, 2002-2010.
- SHVARTSBURG, A. A., SCHATZ, G. C. & JARROLD, M. F. 1998. Mobilities of carbon cluster ions: Critical importance of the molecular attractive potential. *Journal of Chemical Physics*, 108, 2416-2423.
- SIDDIQUE, S., MATERA, C., RADAKOVIC, Z. S., SHAMIM HASAN, M., GUTBROD, P., ROZANSKA, E., SOBCZAK, M., TORRES, M. A. & GRUNDLER, F. M. 2014. Parasitic worms stimulate host NADPH oxidases to produce reactive oxygen species that limit plant cell death and promote infection. *Science signaling*, 7, ra33.
- SIMPSON, D. J. & LEE, T. H. 1976. Fine-structure and formation of fibrils of *Capsicum annum* L chromoplasts. *Zeitschrift Fur Pflanzenphysiologie*, 77, 127-138.
- SINHA, D., GUPTA, M. K., PATEL, H. K., RANJAN, A. & SONTI, R. V. 2013. Cell wall degrading enzyme induced rice innate immune responses are suppressed by the type 3 secretion system effectors XopN, XopQ, XopX and XopZ of *Xanthomonas oryzae* pv. *oryzae*. *PLoS ONE*, 8, e75867.
- SKIBBE, D. S., DOEHLEMANN, G., FERNANDES, J. & WALBOT, V. 2010. Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. *Science*, 328, 89-92.
- SLOOTWEG, E. J., SPIRIDON, L. N., ROOSIEN, J., BUTTERBACH, P., POMP, R., WESTERHOF, L., WILBERS, R., BAKKER, E., BAKKER, J. & PETRESCU, A.-J. 2013. Structural determinants at the interface of the ARC2 and leucine-rich repeat domains control the activation of the plant immune receptors Rx1 and Gpa2. *Plant Physiology*, 162, 1510-1528.
- SMITH, D. P., KNAPMAN, T. W., CAMPUZANO, I., MALHAM, R. W., BERRYMAN, J. T., RADFORD, S. E. & ASHCROFT, A. E. 2009. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *European Journal of Mass Spectrometry*, 15, 113-130.
- SOBCZAK, M. & GOLINOWSKI, W. 2009. Structure of cyst nematode feeding sites. In: BERG, R. & TAYLOR, C. (eds.) *Cell Biology of Plant Nematode Parasitism*. Springer Berlin / Heidelberg.
- SOBCZAK, M. & GOLINOWSKI, W. 2011. Cyst Nematodes and Syncytia. In: JONES, J., GHEYSEN, G. & FENOLL, C. (eds.) *Genomics and Molecular Genetics of Plant-Nematode Interactions*.
- SOBCZAK, M., GOLINOWSKI, W. & GRUNDLER, F. M. 1997. Changes in the structure of *Arabidopsis thaliana* roots induced during development of males of the plant parasitic nematode *Heterodera schachtii*. *European Journal of Plant Pathology*, 103, 113-124.
- SOBCZAK, M., GOLINOWSKI, W. A. & GRUNDLER, F. M. W. 1999. Ultrastructure of feeding plugs and feeding tubes formed by *Heterodera schachtii*. *Nematology*, 1, 363-374.

- SONESON, C. & DELORENZI, M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14, 1-18.
- STAM, R., HOWDEN, A. J., DELGADO-CEREZO, M., AMARO, T. M., MOTION, G. B., PHAM, J. & HUITEMA, E. 2013a. Characterization of cell death inducing *Phytophthora capsici* CRN effectors suggests diverse activities in the host nucleus. *Frontiers in plant science*, 4.
- STAM, R., JUPE, J., HOWDEN, A. J., MORRIS, J. A., BOEVINK, P. C., HEDLEY, P. E. & HUITEMA, E. 2013b. Identification and characterisation CRN Effectors in *Phytophthora capsici* shows modularity and functional diversity. *PLoS ONE*, 8, e59517.
- SUZUKI, S., ONO, N., FURUSAWA, C., YING, B.-W. & YOMO, T. 2011. Comparison of sequence reads obtained from three next-generation sequencing platforms. *Plos One*, 6, e19534.
- SZAKASITS, D., HEINEN, P., WIECZOREK, K., HOFMANN, J., WAGNER, F., KREIL, D. P., SYKACEK, P., GRUNDLER, F. M. W. & BOHLMANN, H. 2009. The transcriptome of syncytia induced by the cyst nematode *Heterodera schachtii* in Arabidopsis roots. *Plant Journal*, 57, 771-784.
- TAKKEN, F. L. & GOVERSE, A. 2012. How to build a pathogen detector: structural basis of NB-LRR function. *Current Opinion in Plant Biology*, 15, 375-384.
- TAMANO, K., AIZAWA, S. & SASAKAWA, C. 2002. Purification and detection of *Shigella* type III secretion needle complex. *Bacterial Pathogenesis, Pt C*, 358, 385-392.
- TENCER, J., FRICK, I. M., OQUIST, B. W., ALM, P. & RIPPE, B. 1998. Size-selectivity of the glomerular barrier to high molecular weight proteins: Upper size limitations of shunt pathways. *Kidney International*, 53, 709-715.
- TERRY, B. R. & ROBARDS, A. W. 1987. Hydrodynamic radius alone governs the mobility of molecules through plasmodesmata. *Planta*, 171, 145-157.
- THIERY, M., FOUVILLE, D. & MUGNIERY, D. 1997. Intra- and interspecific variability in *Globodera*, parasites of *Solanaceous* plants, revealed by random amplified polymorphic DNA (RAPD) and correlation with biological features. *Fundamental and Applied Nematology*, 20, 495-504.
- TOMALOVA, I., IACHIA, C., MULET, K. & CASTAGNONE-SERENO, P. 2012. The map-1 gene family in root-knot nematodes, *Meloidogyne spp.*: a set of taxonomically restricted genes specific to clonal species. *Plos One*, 7, e38656.
- TOMILOV, A. A., TOMILOVA, N. B., WROBLEWSKI, T., MICHELMORE, R. & YODER, J. I. 2008. Trans-specific gene silencing between host and parasitic plants. *Plant J*, 56, 389-97.
- TORRES, M. A. 2010. ROS in biotic interactions. *Physiologia Plantarum*, 138, 414-429.
- TRAPNELL, C., HENDRICKSON, D. G., SAUVAGEAU, M., GOFF, L., RINN, J. L. & PACHTER, L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotech*, 31, 46-53.
- TREUHEIT, M. J., KOSKY, A. A. & BREMS, D. N. 2002. Inverse relationship of protein concentration and aggregation. *Pharmaceutical Research*, 19, 511-516.
- TRIAN, D. A. & WHITEHEAD, A. 2009. Simultaneous extraction of high-quality RNA and DNA from small tissue samples. *Journal of Heredity*, 100, 246-250.

- TRIPLETT, B. A., MOSS, S. C., BLAND, J. M. & DOWD, M. K. 2008. Induction of hairy root cultures from *Gossypium hirsutum* and *Gossypium barbadense* to produce gossypol and related compounds. *In Vitro Cellular & Developmental Biology-Plant*, 44, 508-517.
- TRUDGILL, D. L. 1997. Parthenogenetic root-knot nematodes (*Meloidogyne* spp); How can these biotrophic endoparasites have such an enormous host range? *Plant Pathology*, 46, 26-32.
- TYTGAT, T., VANHOLME, B., DE MEUTTER, J., CLAEYS, M., COUVREUR, M., VANHOUTTE, I., GHEYSEN, G., VAN CRIEKINGE, W., BORGONIE, G. & COOMANS, A. 2004. A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes. *Molecular Plant-Microbe Interactions*, 17, 846-852.
- URWIN, P. E., LILLEY, C. J. & ATKINSON, H. J. 2002. Ingestion of double-stranded RNA by preparasitic juvenile cyst nematodes leads to RNA interference. *Molecular Plant-Microbe Interactions*, 15, 747-752.
- URWIN, P. E., LILLEY, C. J., MCPHERSON, M. J. & ATKINSON, H. J. 1997a. Characterization of two cDNAs encoding cysteine proteinases from the soybean cyst nematode *Heterodera glycines*. *Parasitology*, 114, 605-613.
- URWIN, P. E., MCPHERSON, M. J. & ATKINSON, H. J. 1998. Enhanced transgenic plant resistance to nematodes by dual proteinase inhibitor constructs. *Planta*, 204, 472-479.
- URWIN, P. E., MOLLER, S. G., LILLEY, C. J., MCPHERSON, M. J. & ATKINSON, H. J. 1997b. Continual green-fluorescent protein monitoring of cauliflower mosaic virus 35S promoter activity in nematode-induced feeding cells in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions*, 10, 394-400.
- VALENTINE, S. J., ANDERSON, J. G., ELLINGTON, A. D. & CLEMMER, D. E. 1997. Disulfide-intact and -reduced lysozyme in the gas phase: Conformations and pathways of folding and unfolding. *Journal of Physical Chemistry B*, 101, 3891-3900.
- VALENTINE, T. A., RANDALL, E., WYPIJEWSKI, K., CHAPMAN, S., JONES, J. & OPARKA, K. J. 2007. Delivery of macromolecules to plant parasitic nematodes using a tobacco rattle virus vector. *Plant Biotechnology Journal*, 5, 827-834.
- VAN DUIJN, E., BARENDREGT, A., SYNOWSKY, S., VERSLUIS, C. & HECK, A. J. R. 2009. Chaperonin complexes monitored by ion mobility mass spectrometry. *Journal of the American Chemical Society*, 131, 1452-1459.
- VAN MEGEN, H., VAN DEN ELSSEN, S., HOLTERMAN, M., KARSSSEN, G., MOOYMAN, P., BONGERS, T., HOLOVACHOV, O., BAKKER, J. & HELDER, J. 2009. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology*, 11, 927-950.
- VIEIRA, P., DANCHIN, E. G. J., NEVEU, C., CROZAT, C., JAUBERT, S., HUSSEY, R. S., ENGLER, G., ABAD, P., DE ALMEIDA-ENGLER, J., CASTAGNONE-SERENO, P. & ROSSO, M.-N. 2011. The plant apoplast is an important recipient compartment for nematode secreted proteins. *Journal of Experimental Botany*, 62, 1241-1253.
- WAGNER, G. P., KIN, K. & LYNCH, V. J. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, 131, 281-285.

- WANG, J., LEE, C., REPLOGLE, A., JOSHI, S., KORKIN, D., HUSSEY, R., BAUM, T. J., DAVIS, E. L., WANG, X. & MITCHUM, M. G. 2010. Dual roles for the variable domain in protein trafficking and host specific recognition of *Heterodera glycines* CLE effector proteins. *New Phytologist*, 187, 1003-1017.
- WANG, J., REPLOGLE, A., HUSSEY, R., BAUM, T., WANG, X., DAVIS, E. L. & MITCHUM, M. G. 2011. Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode *Heterodera schachtii*. *Molecular Plant Pathology*, 12, 177-186.
- WANG, X. H., MITCHUM, M. G., GAO, B. L., LI, C. Y., DIAB, H., BAUM, T. J., HUSSEY, R. S. & DAVIS, E. L. 2005. A parasitism gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR (CLE) of *Arabidopsis thaliana*. *Molecular Plant Pathology*, 6, 187-191.
- WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10, 57-63.
- 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, 1189-1191.
- WEIBERG, A., WANG, M., LIN, F.-M., ZHAO, H., ZHANG, Z., KALOSHIAN, I., HUANG, H.-D. & JIN, H. 2013. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science*, 342, 118-123.
- WESLEY, S. V., HELLIWELL, C. A., SMITH, N. A., WANG, M., ROUSE, D. T., LIU, Q., GOODING, P. S., SINGH, S. P., ABBOTT, D. & STOUTJESDIJK, P. A. 2001. Construct design for efficient, effective and high-throughput gene silencing in plants. *The Plant Journal*, 27, 581-590.
- WESTWOOD, J. H., RONEY, J. K., KHATIBI, P. A. & STROMBERG, V. K. 2009. RNA translocation between parasitic plants and their hosts. *Pest Management Science*, 65, 533-539.
- WHISSON, S. C., BOEVINK, P. C., MOLELEKI, L., AVROVA, A. O., MORALES, J. G., GILROY, E. M., ARMSTRONG, M. R., GROUFFAUD, S., VAN WEST, P. & CHAPMAN, S. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*, 450, 115-118.
- WONDJI, C. S., IRVING, H., MORGAN, J., LOBO, N. F., COLLINS, F. H., HUNT, R. H., COETZEE, M., HEMINGWAY, J. & RANSON, H. 2009. Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles funestus*, a major malaria vector. *Genome Research*, 19, 452-459.
- WOODS, L. A., PLATT, G. W., HELLEWELL, A. L., HEWITT, E. W., HOMANS, S. W., ASHCROFT, A. E. & RADFORD, S. E. 2011. Ligand binding to distinct states diverts aggregation of an amyloid-forming protein. *Nature Chemical Biology*, 7, 730-739.
- WUBBEN, M. J., CALLAHAN, F. E., TRIPLETT, B. A. & JENKINS, J. N. 2009. Phenotypic and molecular evaluation of cotton hairy roots as a model system for studying nematode resistance. *Plant Cell Reports*, 28, 1399-1409.
- WULFF, B. B., HEESE, A., TOMLINSON-BUHOT, L., JONES, D. A., DE LA PEÑA, M. & JONES, J. D. 2009. The major specificity-determining amino acids of the tomato Cf-9 disease resistance protein are at hypervariable solvent-exposed positions in the central leucine-rich repeats. *Molecular Plant-Microbe Interactions*, 22, 1203-1213.

- WYSS, U. 1992. Observations on the feeding behaviour of *Heterodera schachtii* throughout development including events during moulting. *Fundamental and Applied Nematology*, 15, 75-89.
- WYSS, U. & GRUNDLER, F. M. W. 1992. Feeding-behavior of sedentary plant parasitic nematodes. *Netherlands Journal of Plant Pathology*, 98, 165-173.
- WYSS, U., JANKLADWIG, R. & LEHMANN, H. 1979. Formation and ultrastructure of feeding tubes produced by Trichodorid nematodes. *Nematologica*, 25, 385-&.
- YANG, Y. & SMITH, S. A. 2013. Optimizing *de novo* assembly of short-read RNA-seq data for phylogenomics. *Bmc Genomics*, 14.
- ZACHARIAS, D. A., VIOLIN, J. D., NEWTON, A. C. & TSIEN, R. Y. 2002. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science*, 296, 913-916.
- ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research*, 18, 821-829.