

Core effectors of plant-parasitic nematodes and their host targets

Kerry Leslie



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of
Doctor of Philosophy (PhD)
at the University of St Andrews

October 2021

Candidate's declaration

I, Kerry Leslie, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 63,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2016.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 24/04/2022

Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 24/04/2022

Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Kerry Leslie, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 24/04/2022

Signature of candidate

Date 24/04/2022

Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Kerry Leslie, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date 24/04/2022

Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

No embargo on underpinning research data or digital outputs.

Date 24/04/2022

Signature of candidate

Date 24/04/2022

Signature of supervisor

Acknowledgements

Thank you to Prof John Jones, Dr Sebastian Eves-van den Akker, and Dr Sophie Mantelin for their supervision and guidance throughout this project. Thank you to the University of St Andrews and the James Hutton Institute for hosting my PhD and The East of Scotland Bioscience Doctoral Training Partnership (Eastbio) for providing the funding for this project.

Thank you to all the past and present members of the nematology lab, with special mention to my fellow PhD students Dr Shona Strachan, Dr James Price, Dr Kyriakos Varypatakis, and Dr Jamie Orr, without whom I would not have succeeded. Thank you to Sally Myles for being my partner in crime throughout our undergrad and PhDs. To Dave and Vivian for all their help and advice, and a wider thank you to my friends and colleagues at the James Hutton Institute who gave their time to teach me protocols, help with experiments, or lend a sympathetic ear.

To my parents, grandparents, brother, and partner, thank you for your love and support in every facet of the word. Finally, thank you to my uncle, who didn't get to see this thesis printed, but was one of my biggest supporters and would have been one of the happiest to see it completed.

Funding

This work was Funded by the East of Scotland Bioscience Doctoral Training Partnership (Eastbio) (UKRI BBSRC).

This work was supported by the University of St Andrews (School of Biology).

Research Data/Digital Outputs access statement

Research data underpinning this thesis are available at [DOI]

Table of Contents

Declarations	Error! Bookmark not defined.
Acknowledgements.....	5
List of Figures	14
List of Tables.....	16
Abstract	17
Abbreviations	18
1. General Introduction.....	21
1.1 – An introduction to nematodes.....	21
1.1.1 – Plant-parasitic nematodes	23
1.1.1.1 – Migratory ectoparasitism.....	24
1.1.1.2 – Migratory endoparasitism.....	25
1.1.1.3 – Semi-endoparasitism.....	25
1.1.1.4 – Sedentary endoparasitism	25
1.1.1.5 – Control of PPN.....	26
1.1.1.6 – Potato cyst nematodes: <i>Globodera rostochiensis</i> and <i>Globodera pallida</i>	30
1.1.1.7 – <i>Rotylenchulus reniformis</i>	32
1.1.1.8 – <i>Nacobbus aberrans</i>	34
1.2 – The syncytium	37
1.2.1 – Syncytium formation and structure	37
1.2.2 – The role of auxin.....	39
1.2.3 – Syncytia of <i>R. reniformis</i> and <i>N. aberrans</i>	40
1.3 – Effectors	40
1.3.1 – Identification of effectors.....	41
1.3.2 – Plant cell wall degrading enzymes	42
1.4 – Plant defence responses	43
1.4.1 – Pattern-triggered immunity	44
1.4.1.1 – Pattern Triggered immunity and nematodes.....	45
1.4.2 – Effector-triggered susceptibility and immunity	46
1.4.3 – Redefinition of plant immunity models	49
1.4.4 – Primary defence responses.....	50
1.4.5 – Suppression of defence responses by PPN	51

1.4.6 – Core effectors.....	53
1.5 – Thesis outline	54
2. General Materials and Methods	55
2.1 – Nematode collection	55
2.1.1 – Second stage juvenile (J2) collection.....	55
2.1.2 – Mixed parasitic life stage collection.....	56
2.1.3 – Tomato root diffusate	56
2.1.4 – Sucrose floatation	56
2.2 – RNA extraction	57
2.2.1 – RNA extraction from J2 nematodes	57
2.2.2 – DNase treatment of RNA extractions.....	57
2.3 – cDNA synthesis.....	58
2.4 – Polymerase Chain Reaction.....	58
2.4.1 – Gel electrophoresis	60
2.5 – Cloning of cDNA sequences encoding effectors	60
2.5.1 – pCR8/GW/TOPO cloning	60
2.5.2 – pGEM-T easy cloning.....	61
2.6 – Cell transformation	62
2.6.1 – Electroporation	63
2.6.2 – Heat shock transformation	64
2.6.3 – PCR colony screening	64
2.7 – Sample sequencing.....	65
2.8 – <i>In situ</i> hybridisation.....	65
2.8.1 – <i>In situ</i> hybridisation probe production	65
2.8.2 – Fixation and dissection of nematodes	68
2.8.3 – Permeabilization of nematode sections.....	68
2.8.4 – Hybridisation of nematode sections	69
2.8.5 – Staining of nematode sections.....	70
3. Identification of core effectors of syncytia-forming nematodes	71
3.1 – Introduction.....	71
3.1.1 – Effectors	71
3.1.1.1 – The Hs19C07 effector.....	72

3.1.1.2 – CLE peptides	73
3.1.1.3 – The 30D08 effector.....	74
3.1.2 – Genomes and transcriptomes of plant-parasitic nematodes	75
3.1.3 – Effector-associated sequence motifs.....	76
3.1.4 – <i>In-silico</i> pipelines for effector identification	78
3.1.4.1 – Signal peptides and transmembrane helices	79
3.1.4.2 – Gene expression analysis	79
3.1.5 – Chapter aims	80
3.2 – Materials and methods	81
3.2.1 – <i>In silico</i> identification of effectors.....	81
3.2.1.1 – <i>G. rostochiensis</i> effector initial gene set.....	81
3.2.1.2 – BLAST similarity searches for orthologous proteins	81
3.2.1.3 – Candidate sequence filtering	82
3.2.1.4 – Gene expression analysis	83
3.2.1.5 – DOG Box motif analysis.....	83
3.2.2 – Cloning and analysis of candidate core effector genes.....	83
3.3 – Results	88
3.3.1 – BLAST – functional similarity searches.....	95
3.3.2 – BLAST similarity searches against Root-knot nematodes	95
3.3.3 – Protein sequence alignments and RNA-seq analysis	101
3.3.3.1 - 20E03.....	101
3.3.3.2 – GLAND11	104
3.3.3.3 – G23G11.....	104
3.3.3.4 – Gro_DOG_0043	104
3.3.3.5 – Misprediction of further Gro_DOG group members	108
3.3.3.6 – Gro_DOG_0203 and Gro_DOG_0017.....	108
3.3.4 - DOG Box motif analysis.....	108
3.3.5 – Protein localisation and domain predictions	111
3.3.6 – Gene expression profiles of <i>G. rostochiensis</i> candidate core effectors.....	114
3.3.7 – Cloning of <i>G. rostochiensis</i> candidate core effectors.....	115
3.3.8 – <i>In situ</i> hybridisation.....	117
3.4 – Discussion.....	121

3.4.1 – Effector identification pipeline improvement.....	121
3.4.2 – Syncytia-forming species.....	122
3.4.3 – DOG Box motif analysis	123
3.4.4 – <i>In situ</i> hybridisation.....	124
3.5 – Future work.....	127
4. Arabinogalactan endo- β -1,4-galactanases of plant-parasitic nematodes	128
4.1 – Introduction.....	128
4.1.1 – Plant cell walls	128
4.1.1.1 – Pectin.....	129
4.1.2 – Cell wall degrading enzymes	131
4.1.2.1 – Horizontal gene transfer	133
4.1.3 – CWDE modes of action.....	135
4.1.4 – Glycoside hydrolase 53 family.....	135
4.1.4.1 – Arabinogalactan	137
4.1.5 – GH53 family members in nematodes.....	138
4.1.6 – Chapter aims	140
4.2 – Materials and Methods	141
4.2.1 – Cloning.....	141
4.2.2 – Phylogenetic analysis	143
4.2.3 – Protein structure prediction.....	144
4.2.4 – Protein expression and purification	144
4.2.5 – SDS-PAGE gels	145
4.2.6 – Coomassie blue staining.....	146
4.2.7 – Azo-galactan substrate assay	146
4.2.8 – DNS Assay.....	146
4.2.9 – <i>In situ</i> hybridisation & gene expression	147
4.3 – Results	148
4.3.1 – GH53 gene identification and sequence analysis.....	148
4.3.2 – Protein structure models	153
4.3.3 – Phylogenetic analysis	154
4.3.4 – <i>In situ</i> hybridisation and gene expression.....	164
4.3.5 – Biochemical characterisation of nematode GH53 proteins	167

4.3.5.1 – Expression and purification of recombinant proteins.....	167
4.3.5.2 – Azo-galactan substrate assay	168
4.3.5.3 – DNS assay	172
4.4 – Discussion.....	175
4.4.1 – Phylogenetic analysis	175
4.4.2 – <i>In situ</i> hybridisation & gene expression	177
4.4.3 – DNS assay	178
4.4.4 – Azo-galactan assay	180
4.5 – Future work.....	182
5. Cathepsin L-like peptidases of plant-parasitic nematodes	183
5.1 – Introduction.....	183
5.1.1 – Cysteine proteases	184
5.1.2 – Cathepsin L peptidases.....	185
5.1.3 – Chapter aims	187
5.2 – Materials and methods	188
5.2.1 – Gene identification.....	188
5.2.2 – Nematode collection	188
5.2.3 – RNA and cDNA synthesis	188
5.2.4 – Cloning of Gr-cpl-like-2.....	188
5.2.5 – Re-annotation of genome sequences & gene expression data	189
5.2.6 – Bioinformatic and phylogenetic analysis	189
5.2.7 – <i>In-silico</i> structural analysis	190
5.2.8 – <i>In situ</i> hybridisation.....	190
5.2.9 – RNA interference (RNAi)	191
5.2.9.1 – Silencing hairpin construct production	191
5.2.9.2 – Agrobacterium-mediated potato transformation (RNAi transgenic line production).....	194
5.2.9.3 – Callus induction	194
5.2.9.4 – Shoot regeneration	195
5.2.9.5 – Analysis of RNA interference via semi-quantitative reverse transcription PCR (SQRT-PCR)	196
5.2.9.6 – RNAi line screening of females and cysts.....	197

5.3 – Results	198
5.3.1 – Cathepsin L gene identification in PPN	198
5.3.2 – Phylogenetic and gene expression analysis	204
5.3.3 – <i>In situ</i> hybridisation	208
5.3.4 – RNAi silencing of Gr-cpl-like-2	210
5.4 – Discussion	212
5.4.1 – Identification and phylogenetic analysis of nematode Cathepsin L-like peptidases	212
5.4.2 – <i>In situ</i> hybridisation	213
5.4.3 – RNAi silencing of Gr-cpl-like-2	213
5.5 – Future work	214
6. Functional characterisation of core effector genes	216
6.1 – Introduction	216
6.1.1 – Subcellular localisation	216
6.1.2 – Protein-protein interaction identification methods	218
6.1.2.1 – Yeast two-hybrid	219
6.1.2.2 – Co-immunoprecipitation	221
6.1.2.3 – Bimolecular fluorescence complementation	221
6.1.3 – Chapter aims	222
6.2 – Materials and methods	223
6.2.1 – Cloning	223
6.2.2 – Transformation into <i>Agrobacterium tumefaciens</i>	226
6.2.3 – Agroinfiltration	226
6.2.4 – Confocal microscopy	227
6.2.5 – The hypersensitive response in <i>Nicotiana benthamiana</i>	227
6.2.6 – UV imaging	228
6.2.7 – Yeast two-hybrid experiments	228
6.2.7.1 – Yeast cultures	228
6.2.7.2 – Y2H small scale transformation testing	228
6.2.7.3 – Control vectors	230
6.2.7.4 – Small scale Y2H reporter gene assay	230
6.2.7.5 – β -galactosidase assay	231

6.2.7.6 – Y2H screen.....	231
6.2.7.7 – Reporter assays for Y2H screen.....	233
6.2.7.8 – Analysis of Y2H screen.....	233
6.2.8 – Co-immunoprecipitation	234
6.2.8.1 – Preparation of sample leaf material	234
6.2.8.2 – SDS-PAGE gels	235
6.2.8.3 – SDS page gel wet transfer	236
6.2.8.4 – Western blot.....	237
6.3 – Results	238
6.3.1 – The hypersensitive response in <i>Nicotiana benthamiana</i>	238
6.3.2 – Subcellular localisation by confocal microscopy	239
6.3.2.1 – GROS_g05682 localisation	240
6.3.2.2 – GROS_g02469 localisation	241
6.3.2.3 – GROS_g02394 localisation	245
6.3.3 – Yeast Two-Hybrid	248
6.3.4 – Subcellular localisation of GROS_g02394 and the interacting host protein stPRMT1.1	251
6.3.5 – Co-immunoprecipitation for confirmation of Y2H results	254
6.4 – Discussion.....	256
6.4.1 – The hypersensitive response in <i>Nicotiana benthamiana</i>	256
6.4.2 – Subcellular localisation – Confocal microscopy	257
6.4.2.1 – GROS_g05682.....	257
6.4.2.2 – GROS_g02469.....	257
6.4.2.3 – GROS_g02394.....	258
6.4.2.4 – Prediction models versus subcellular localisation assays	258
6.4.3 – Yeast two-hybrid	260
6.4.4 – Co-immunoprecipitation (Confirmation of Y2H results)	264
6.5 – Future work.....	265
7. General Discussion	268
7.1 – The importance of feeding site production	268
7.2 – Understanding the syncytium	269
7.3 – Effectors	272

7.4 – Identification of core effectors of syncytia-forming nematodes	273
7.5 – Arabinogalactan endo- β -1,4-galactanases of plant-parasitic nematodes	277
7.6 – Cathepsin L-like peptidases of plant-parasitic nematodes	278
7.7 – Functional characterisation of core effector genes	279
7.8 – Computational identification and prediction of gene sequences.....	281
7.9 – Future research avenues.....	284
7.10 – Conclusions.....	288
8 – References.....	289
9. Appendix	332
9.1 – Supplementary file 1	332
9.2 – Supplementary file 2	332
9.3 – Supplementary file 3	332
9.4 – Supplementary file 4	332
9.5 – Supplementary file 5	333
9.6 – Supplementary file 6	356
9.7 – Supplementary file 7	358

List of Figures

1.1 – The life cycle of <i>G. rostochiensis</i> and <i>G. pallida</i>	32
1.2 – The life cycle of <i>R. reniformis</i>	34
1.3 – The life cycle of <i>N. aberrans</i>	36
1.4 – The syncytium of <i>G. rostochiensis</i>	37
1.5 – Schematic of potato cyst nematode internal structure.....	41
1.6 – The zig-zag model.....	44
1.7 – The decoy and guard models.....	48
3.1 – Stages of the <i>in-silico</i> pipeline for candidate core syncytial effector identification..	89
3.2 – Alignment of 20E03 effector family.....	103
3.3 – Alignment of <i>N. aberrans</i> sequences present in the G23G11 effector family.....	104
3.4 – RNA-seq analysis of GROS_g09990 from the GROS_DOG_0043 effector family.....	107
3.5 – Gene expression profiles of the top 15 <i>G. rostochiensis</i> candidate core effectors...	115
3.6 – Pairwise alignment of GROS_g09987 sequences.....	116
3.7 – <i>In situ</i> hybridisation of GROS_g05682, GROS_g02469, and GROS_g02394.....	119
3.8 – <i>In situ</i> hybridisation results for candidate effectors GROS_g09987, GROS_g04903, and GROS_g02024.....	120
4.1 – The structure of the plant cell wall.....	129
4.2 – Simplified schematic representation of Rhamnogalacturonan I.....	130
4.3 – Simplified schematic of the metabolism of Rhamnogalacturonan (Type I) arabinogalactan side chain.....	137
4.4 – Phylogenetic tree of nematode species comparing monocotyledon and dicotyledon parasitism.....	139
4.5 – Computationally recapitulated <i>G. pallida</i> sequences (comp4850_c0_seq1 and comp4850_c0_seq4) alignment	149
4.6 – Computationally recapitulated <i>R. reniformis</i> sequences (comp30258_c0_seq1, Comp30258_c1_seq1, and gi 830247817) alignment.....	150
4.7 – A comparison of mature GH53 protein sequences from bacterial and nematode species.....	152
4.8 – Predicted structures of GH53 proteins.....	154
4.9 – Phylogenetic tree of nematode and bacterial GH53 sequences.....	161

4.10 – Phylogenetic tree of nematode GH53 sequences compared to unique BLAST hits from bacterial and fungal species revisited.....	163
4.11 – <i>In situ</i> hybridisation and gene expression profiles of GrGAL1, GpGAL1, and RrGAL1	166
4.12 – Expression and purification of recombinant GrGAL1, GpGAL1, RrGAL1, GalA_Xc and GFP.....	168
4.13 – pH range for GH53 activity during Azo-galactan substrate assay.....	171
4.14 – DNS assay using nematode GH53 enzymes against plant cell wall component substrates.....	174
5.1 – Vector maps of pHANIBAL and pART27 containing the Gr-cpl-like-2 RNAi construct	192
5.2 – Transgenic <i>S. tuberosum</i> line production for RNAi.....	195
5.3 – Alignment and protein structure prediction of Gr-cpl-like-2.....	203
5.4 – Phylogenetic tree of Cathepsin L and L-like peptidase members.....	206
5.5 – Amino acid alignment of Cathepsin L and L-like proteins of PPN.....	207
5.6 – mRNA localisation of Gr-cpl-like-1 to 5, and Gp-cpl-like-1 by <i>in situ</i> hybridisation... ..	209
5.7 – RNA interference of Gr-cpl-like-2.....	211
6.1 – Schematic of yeast two-hybrid reporter gene activation by interacting proteins....	220
6.2 – Yeast two-hybrid plate set up.....	231
6.3 – The hypersensitive response of <i>N. benthamiana</i> induced by the nematode effectors GROS_g05682, GROS_g02469, and GROS_g02394.....	239
6.4 – Subcellular localisation of GROS_g05682	241
6.5 – Co-localisation of the GROS_g02469 effector against plant cytoskeleton components	244
6.6 – Co-localisation of the GROS_g02394 effector against plant cytoskeleton components	246
6.7 – Co-localisation of the GROS_g02394 and GROS_g02469 effectors.....	247
6.8 – stPRMT1.1 amino acid alignment.....	249
6.9 – Y2H screening with the GROS_g02394 effector.....	250
6.10 – Co-infiltration of stPRMT1.1 and the GROS_g02394 effector.....	253
6.11 - Co-immunoprecipitation of the GROS_g02394 effector and stPRMT1.1 following pull down with GFP-trap beads.....	255

6.12 – Arginine N-methyltransferase (PRMT) amino acid alignment.....	262
--	-----

List of Tables

2.1 – Program for GoTaq PCR.....	59
2.2 – Program for KOD PCR.....	59
2.3 – Addition of A-overhangs reaction.....	61
2.4 – pGEM-T Easy ligation reaction for creation of entry clones.....	62
2.5 – Antibiotic concentration usage for cloning.....	63
2.6 – <i>In situ</i> hybridisation primers for candidate effectors.....	67
3.1 – Primers used for cloning & <i>in situ</i> hybridisation of <i>G. rostochiensis</i> and <i>G. pallida</i> ..	85
3.2 – Orthologous groups shared between syncytia-forming species of interest.....	92
3.3 – Top 15 orthologous candidate effector groups put forward for confirmation and functional characterisation.....	97
3.4 – DOG Box motifs in promotor regions of candidate core effectors.....	110
3.5 – Localisation prediction and conserved domains of 15 <i>G. rostochiensis</i> candidate core effectors.....	113
3.6 – <i>In situ</i> hybridisation locations of candidate core effectors from <i>G. rostochiensis</i>	118
4.1 – Primer sets for cloning and <i>in situ</i> hybridisation of GH53 genes.....	142
4.2 – Nomenclature of nematode GH53 sequences.....	151
4.3 – Species included in phylogenetic analysis.....	156
4.4 – Reducing sugars produced on substrate hydrolysis.....	172
5.1 – Primers for cloning of the Gr-cpl-like-2 gene.....	289
5.2 – Primers for production of Cathepsin L peptidase <i>in situ</i> hybridisation probes.....	190
5.3 – Primers for Gr-cpl-like-2 and GFP RNAi construct.....	193
5.4 – Cathepsin L peptidase nomenclature and signal peptide status.....	199
6.1 – Ligation reaction set up.....	223
6.2 – Primer sets for GROS_g05682, GROS_g02469, GROS_g02394, and stPRMT1.1 cloning	224
6.3 – Y2H plasmid combination for screening.....	229
6.4 – Combinations of infiltrated vectors for Co-immunoprecipitation.....	234
6.5 – Summary of predicted vs. confirmed subcellular location of core effectors.....	259

Abstract

Plant parasitic nematodes infect many major food crops worldwide, causing damage valued at approximately 80 billion U.S. dollars per year (Nicol *et al.* 2011). As part of the parasitic process, some nematodes form a feeding site called a syncytium in the roots of their host. Specialised pathogen proteins known as effectors are thought to play critical roles in these processes.

This thesis identifies and characterises a subset of core effectors conserved in the syncytia-forming nematode species *Globodera rostochiensis*, *Globodera pallida*, *Rotylenchulus reniformis*, and *Nacobbus aberrans*, but that are absent from other nematodes. Three of the candidates (GROS_g02394, GROS_g02469, and GROS_g05682) have been validated as effectors using *in situ* hybridisation to confirm expression in the oesophageal gland cells. Further functional characterisation using *in planta* localisation, yeast two-hybrid (Y2H) analysis, and co-immunoprecipitation for host target identification were undertaken. Using Y2H it was possible to identify an arginine N-methyltransferase (stPRMT1.1) from *Solanum tuberosum* as an interacting host protein for GROS_g02394.

In addition, a set of novel GH53 endo- β -1, 4-galactanase effectors has been identified which may assist in invasion of the host and migration through root tissue. These genes have likely been acquired through a horizontal gene transfer event. This has given a greater insight into the invasion process and the co-evolution between the nematode and its host plant. A conserved family of Cathepsin L-like peptidases has also been identified. Analysis using *in situ* hybridisation showed these to be intestinal proteins. Expression analysis suggests conserved functions for different family members across a range of species.

Abbreviations

AA	Amino Acid
AChE	Acetylcholinesterase
ACT	Actin
A. K. A	Also known as
Avr	Avirulence
BiFC	Biomolecular fluorescence complementation
BLAST	Basic Local Alignment Search Tool
Bp	Base pair
CAZY	Carbohydrate Active Enzymes database
cDNA	Complementary DNA
CLE	CLAVATA3 (CLV3)/EMBRYO SURROUNDING REGION-related
CLV	CLAVATA
CM	Chorismate mutase
CWDE	Cell Wall Degrading Enzyme
DEPC	Diethyl pyrocarbonate
DIG	Digoxigenin
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
dpi	Days post infection
DUF	Domain of unknown function
dsDNA	Double stranded DNA
EFR	EF-Tu receptor
EF-Tu	Elongation factor Tu
EF1 α	Elongation factor 1 α
EST	Expressed sequence tag
ETI	Effector triggered immunity
ETS	Effector triggered susceptibility
Flg22	Flagellin
FLP	FMRFamide-like peptide
FLS2	Flagellin Sensitive2
GA3	Gibberellic acid
GFP	Green fluorescent protein
GH	Glycoside hydrolase
HG	Homogalacturonan
HGT	Horizontal gene transfer
HR	Hypersensitive response
HYP	Hyper-variable effector gene family
ID	Identity
ISC	Initial syncytial cell

ISH	<i>In situ</i> hybridisation
ISO	International organisation for standardisation
Kb	Kilobases
kDa	Kilodalton
LDS	Lithium dodecyl sulphate
LGT	Lateral gene transfer
LRR	Leucine-rich repeat
LRR-RLK	Leucine-rich repeat receptor-like kinase
MAMP	Microbial-associated molecular pattern
MAPK	Mitogen-activated protein kinase
mRNA	Messenger RNA
MT	Microtubule
μl	Microlitre
NAA	1-Naphthaleneacetic acid
NPA	N-1-naphthylphtalamic acid
O/N	Overnight
PAMP	Pathogen-associated molecular pattern
PCD	Programmed cell death
PCN	Potato cyst nematode
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDK	Pyruvate orthophosphate dikinase
pDNA	Plasmid DNA
PMSF	Phenylmethylsulphonyl fluoride
PPN	Plant parasitic nematode
PRMT	Protein arginine methyltransferase
PRR	Pattern recognition receptor
PTI	PAMP triggered immunity
PTM	Post-translational modification
PVDF	Polyvinylidene difluoride
PVX	Potato virus X
QTL	Quantitative trait locus
R-	Resistance
rDNA	Ribosomal DNA
RGI	Rhamnogalacturonan I
RGII	Rhamnogalacturonan II
RKN	Root-knot nematode
RLN	Root-lesion nematode
RNAi	RNA interference
ROS	Reactive oxygen species
RPK2	Receptor-like protein kinase 2
SA	Salicylic acid
SAM	Shoot apical meristem
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SMU2	Suppressor of mec-8 and unc-52 2
SOC	Super optimal broth with Catabolite repression
SPRY	Spla and the ryanodine receptor
SSC	Saline-sodium citrate
TAP	Tandem affinity purification
TBE	Tris-borate-EDTA
TDIF	Tracheary element differentiation inhibitory factor
T-DNA	Transfer DNA
TE	Tris-EDTA
TF	Transcription factor
TMM	Trimmed mean of M-values
TRV	Tobacco rattle virus
UBCEP	Ubiquitin carboxyl extension protein
VAP	Venom allergen protein
WT	Wild type
Y2H	Yeast two-hybrid
ZR	Zeatin riboside
%ID	Percentage identity
-LT	-leucine/-tryptophan
-LTH	-leucine/-tryptophan/-histidine
-LHU	-leucine/-tryptophan/-uracil

1. General Introduction

1.1 – An introduction to nematodes

Nematodes are a group of eukaryotic roundworms with diverse life cycles that range from free-living soil-borne species to those which parasitise a wide range of plants or animals. The phylum Nematoda is made up of 27,000 described species as of 2015, however the actual number of total nematode species globally is likely to be much higher (Quist *et al.* 2015). One of the most widely studied species of nematode is *Caenorhabditis elegans*; a free-living nematode which has been used as a model organism for genetic studies since the 1960s (Fatt & Dougherty 1963, Brenner 1974). *Caenorhabditis elegans* is of scientific importance as it was the first multicellular organism to have its genome sequenced (The *C. elegans* Sequencing Consortium 1998).

Morphological classification and phylogenetic analyses of nematodes is extremely challenging due to their conserved morphology and the absence of a fossil record. However, analysis of ribosomal DNA (rDNA) sequences has been used to examine the relationships between different nematode species and allowed for the identification of 12 distinct clades (Holterman *et al.* 2006). Plant-parasitism in nematodes has evolved independently in four of these clades: clades 1, 2, 10, and 12 (Holterman *et al.* 2017). It has been hypothesised that in some cases plant-parasitic nematodes (PPN) have evolved from fungivorous ancestors which may have been associated with plants. These ancestors may have had advantageous genes such as cell wall degrading enzymes, which were then passed on to the PPN descendants. For example *Bursaphelenchus xylophilus*, the pine wilt nematode (found in clade 10) has been observed to consume fungi as a food source as well as the pine tree host (Fukushige 1991). It is also hypothesised that PPN obtained the ability to parasitise

plants from horizontal gene transfer events from bacteria and/or fungi species (Haegeman *et al.* 2011a).

Nematodes have adopted many different lifestyles for survival. The simplest feeding strategies are those used by the free-living nematode species. These nematodes have varied food sources such as bacteria, fungi, decomposing tissue and organic matter. Free-living nematodes are found in marine, fresh-water, and terrestrial environments. Many land-based, free-living species are essentially aquatic and are reliant on the fine layers of water trapped in the soil for their motility (Neher 2010). Feeding behaviour has been studied in detail in *C. elegans*. As this nematode moves through the soil it takes up bacteria and other particles from the environment, passing food into the intestine while filtering liquid back out of the mouth parts (Avery & Shtonda 2003). *Caenorhabditis elegans* selectively consumes bacterial species it encounters in the soil, as it can determine the quality of the food source in terms of which provides the best sustenance for growth (Shtonda & Avery 2006). Some nematode species are fungal feeders which find their food sources either free-living in soil or in the rhizosphere surrounding plant roots (Hasna *et al.* 2007).

Predatory nematode species are also found in soil environments. These species feed upon other animals, including nematodes (Yeates *et al.* 1993). Some predatory nematodes have been observed feeding on plant-parasitic nematode species which may make them potential biocontrol agents (Bilgrami 2008). For example, the predatory nematode *Mononchus aquaticus* has been shown to prey upon many important PPN species such as *Globodera rostochiensis* and *Meloidogyne naasi* (Grootaert & Small 1983).

Many species of nematodes parasitise other organisms to gain the nutrients required to complete their lifecycle. Host organisms vary greatly and include mammals, amphibians, insects, and plants. According to the World Health Organisation (WHO) globally there are approximately two billion people infected with soil-transmitted

helminths such as *Ascaris*, *Trichuris*, and *Ancylostoma* species (World Health Organization 2017). An example of one such human parasitic species is *Ascaris lumbricoides*. *Ascaris lumbricoides* is an intestinal roundworm that can reach up to 14 inches in length and can cause large intestinal blockages in the human host. Although many parasitic nematode species have a negative impact on humans, either directly or indirectly, there are species which can be harnessed for beneficial purposes. Many nematode species that parasitise insects or other pests are used as biocontrol agents. *Phasmarhabditis hermaphrodita* is sold as Nemaslug® and is widely used to control slug pests in residential gardens. The entomopathogenic nematode *Steinernema feltiae* is used to control many insect species including *Sciaridae* (fungus gnat), *Phoridae* (mushroom fly), and *Thripidae* (flower thrips) (Scheepmaker *et al.* 1997, Jagdale *et al.* 2004, Wardlow *et al.* 2001).

1.1.1 – Plant-parasitic nematodes

Plant-parasitic nematodes cause damage valued at ~\$80 billion dollars in global crop losses each year, making them a significant threat to world agriculture and a significant barrier to achieving food security (Nicol *et al.* 2011). Plant-parasitic nematodes can have a restricted host range but, in some cases, such as the tropical root-knot nematodes, host range may be extremely broad. The vast majority of PPN feed exclusively on roots, remaining below soil surface level for the entirety of their life cycle, however there are species of nematodes which live on other parts of the plant as well. The stem and bulb nematode *Ditylenchus dipsaci* is a migratory endoparasite that migrates through and feeds on the host plant (e.g. sugar beet, garlic, alfalfa), creating areas of necrosis and stunting growth (Castillo 2007). This nematode parasitises both above and below ground parts of the plant. *Ditylenchus dipsaci* can also survive outside the host, migrating to new sections of the infected plant or to new hosts through dew or rainwater on the plant surface. Another example are the *Aphelenchoides* spp. which are agriculturally important seed-borne nematodes. The rice white tip nematode *Aphelenchoides besseyi* prefers consuming leaves and young, developing plant material e.g. seedlings (Togashi &

Hoshino 2001). They achieve this by persisting inside the rice seed as a fourth stage juvenile (J4) or adult nematode which emerge in response to the seeds contact with moisture. This allows the nematode to begin feeding on developing tissue as soon as the seed is planted (Tiwari & Khare 2003).

Plant-parasitic nematodes exhibit a wide variety of feeding behaviours ranging from migratory ectoparasites to sedentary endoparasites, as well as those which straddle these definitions. One of the basic differences in nematode lifestyles is the differentiation between migratory and sedentary species. Migratory species are mobile throughout their life cycle and do not establish a fixed feeding site, whereas sedentary species will remain at a fixed feeding site at one or more stages during their life cycle. Nematodes can be ectoparasitic meaning they remain outside of the host while inserting the stylet into the plant tissue to feed on the cytoplasm of cells. Endoparasitic nematode species may insert their head or their whole body inside the host tissue.

1.1.1.1 – Migratory ectoparasitism

Migratory ectoparasitic species are mobile, soil-dwelling nematodes which feed on the exterior epidermal cells of host plant root tissue. Many species of migratory ectoparasitic nematodes cause necrosis in areas of feeding and gall formation in certain cases. Species such as *Belonolaimus longicaudatus* (sting nematode) and *Trichodorus obtusus* (stubby-root nematode) are migratory ectoparasites that feed on the roots of turfgrass, causing the roots to have a stunted, “stubby” appearance (Trenholm *et al.* 2005, Crow 2005). These nematode species are threats to many landscapes, most notably being the golfing industry as they destroy large areas of grass and are easily transmitted by human foot traffic (Crow 2017). Migratory ectoparasites are of agricultural importance as they not only damage the host plant by feeding but are also vectors of plant viruses. *Trichodorus* spp. are vectors for the Tobacco rattle virus (TRV) (Decraemer & Geraert 2006). TRV can infect a range of crop plants including tobacco, potato, and peppers, as well as ornamental flower

species such as daffodils. Common symptoms of TRV include chlorosis, mottling, and necrosis, however symptoms vary dependent on the host plant (Brown *et al.* 1996, Harrison 1968).

1.1.1.2 – Migratory endoparasitism

Migratory endoparasites are like migratory ectoparasites in that they are also mobile throughout the life cycle, but these nematodes can also enter the host plant. This lifestyle can be observed in the *Pratylenchus* genus; the root-lesion nematodes (RLN). *Pratylenchus penetrans* can enter and exit the host plant roots at any life stage between second stage juvenile (J2) and adulthood (Davis & MacGuidwin 2000). This species is a significant threat to agriculture as it has a wide host range and will migrate through the host root, creating large necrotic lesions while feeding (Zunke 1990).

1.1.1.3 – Semi-endoparasitism

Semi-endoparasitic nematodes are those which partially enter the host to feed. The nematodes will migrate along the surface of the host plant before inserting it's head or upper body into the host tissue. The nematode will establish a fixed feeding site at which the nematode will become sedentary. Semi-endoparasitism has been observed in the reniform nematode *Rotylenchulus reniformis* and the citrus nematode *Tylenchulus semipenetrans*. This lifestyle is discussed in further detail in section 1.1.1.7.

1.1.1.4 – Sedentary endoparasitism

Sedentary endoparasites establish a feeding site inside the host plant at which the nematode will remain immobile for part of or (in the case of females) the remainder of their life cycle. Sedentary endoparasitism has be observed to have evolved independently five times across PPN species (Holterman *et al.* 2017). This lifestyle has been observed in all nematode species belonging to the family *Meloidogynidae*. A common ancestor was shown to give rise to sedentary endoparasitism in species

belonging to both the *Heteroderidae* and *Rotylenchulidae* families. There are also individual examples of sedentary endoparasitism present in the suborder of *Criconematina*, for example *Sphaeronema alni* which parasitises tree species such as *Betula pubescens* (Downy birch) and *Alnus incana* (Grey Alder). Sedentary endoparasitism has also evolved in individual species which are phylogenetically closest to other species with distinctly different lifestyles. For example, *Nacobbus aberrans*, the false root-knot nematode is a sedentary endoparasite found amongst a clade of ectoparasitic species. The precise phylogeny of this species is difficult to determine however as its position in relation to other nematodes varies in phylogenetic analyses depending on the details of how the analysis is performed (Eves-van den Akker *et al.* 2014a). Sedentary endoparasitism is also found in the *Fergusobia* nematodes. These nematodes use the *Fergusonina* fly as a vector to infect members of the Myrtaceae family of plants such as eucalyptus and guava (Giblin-Davis *et al.* 2001, Holterman *et al.* 2017).

The most economically important sedentary endoparasitic nematodes are the root-knot nematodes (RKN) and cyst nematodes. Root-knot nematodes often have a broad host range and produce a feeding site, made up of multiple giant cells, by inducing repeated cycles of nuclear division in the absence of cytokinesis. Although these feeding structures share some superficial similarities with the syncytia induced by cyst nematodes (further discussed in section 1.2), they have an entirely different ontogeny. Cyst nematodes, and nematodes in several other closely related genera, form syncytia as their feeding site. Cyst nematodes tend to be host specific (with some exceptions) and frequently cause significant damage to the crops that they infect.

1.1.1.5 – Control of PPN

Control of PPN is vital due to the severe damage they cause to crop plants (Evans & Stone 1977, Gao *et al.* 2001). Good cropping practices must be adopted as a primary defence against PPN infections. These strategies can include crop rotation (where a

narrow host range nematode is present), improved soil practices (tillage), and growth of trap crops (Abawi & Widmer 2000, Bairwa *et al.* 2017). These management strategies are limited because many PPN species, such as the cyst nematodes, can persist in the soil for years after initial infection. The consequence is that extended rotations are required to cause a sufficient reduction in nematode population. Using an initial trap crop (*e.g.*, potato) in the field which is harvested after the nematode has infected the roots, but before the lifecycle is completed is potentially a good way to remove the parasite from soil. A second crop could then be planted, however this is not a financially viable option for many growers (Bairwa *et al.* 2017). Trap crops can also be used as biofumigants. Biofumigants are plant species which produce secondary metabolites that can inhibit pests and pathogens found in the surrounding soil. Members of the *Brassicaceae* family are good biofumigants for a variety of pests, including multiple species of nematode *e.g.*, PCN (Lord *et al.* 2011). Mustard is a well-studied example of a biofumigant which is grown in the field and then mown and incorporated into the soil before it has fully bloomed. This traps the secondary metabolite produced by the mustard (allyl isothiocyanate) in the soil which suppress pests and pathogens from infecting the next crop planted (Evans 2020). Physically changing the environment to make it less hospitable for nematodes is another form of control. Soil solarization is the process of raising the core soil temperature through the application of a plastic film over the soil surface (Stapleton & DeVay 1986). This form of control is difficult to achieve on a large, outdoor scale and is most effective when conducted in glasshouses in areas with warm, manageable climates (Oka *et al.* 2007).

Chemical control of PPN is challenging as many species spend much of their life cycle in the soil or inside the host plant. This means delivery of nematicides must allow for penetration deep enough into the soil or into the plant tissue to reach the nematode (Chitwood 2003). Nematicides frequently only target the invasive J2s, meaning that timing of application is critical. Chemical nematicides are used to control PPN, but these can have a negative impact on the environment and human health, so many

are currently being phased out of use. For example, ethylene dibromide (EDB) was widely used as a nematicide - despite being toxic and carcinogenic to humans - until it was found to have contaminated groundwater around application sites, prompting its use to be banned in the US (Chitwood 2003, United States Environmental Protection Agency 2000). Vydate is the commercial name for the chemical oxamyl which is effective for nematode control but is exceedingly harmful on contact to humans. Due to the harmful nature of this chemical, use of granulated Vydate has been banned in the U.S.A, and 2021 has seen the loss of Vydate in the UK as the government did not reauthorise its approval for use (Agriculture and Horticulture Development Board (AHDB) 2021). Currently there are two chemical nematicides effective against PCN approved for use in the UK: Velum Prime and Nemathorin. The active chemical ingredient in Velum Prime is fluopyram, which is a succinate dehydrogenase inhibitor (Lewis *et al.* 2016). Succinate dehydrogenase is involved in both the electron transport chain and the Krebs (tricarboxylic acid) cycle which means inhibition causes disruption to multiple metabolic pathways such as cellular respiration (Rutter *et al.* 2010). Nemathorin contains the active chemical fosthiazate. Fosthiazate is an acetylcholinesterase (AChE) inhibitor (Lin *et al.* 2007). AChE is an enzyme which breaks down the neurotransmitter acetylcholine at the synapses of neuromuscular junctions. Inhibition of AChE causes a build-up of acetylcholine at the synapse which leads to paralysis and death if not reversed (Trang & Khandhar 2019). The recently banned Vydate also functions as an AChE inhibitor (Lewis *et al.* 2016). None of the targets of these nematicides are specific to nematodes, meaning that application of such chemicals inevitably has a deleterious impact on non-target organisms. There are also organic-based nematicides available for use such as NEMguard. The active agent in NEMguard is a high volume of garlic extract (45%), which theoretically acts as a more “natural” alternative to the nematicides described above (Ecospray Ltd. 2017).

The most economical and environmentally friendly way to control PPN is using natural resistance. Resistant plant cultivars contain resistance (R) genes which can

inhibit the development of a specific pathogen. In terms of nematodes, this is a cultivar which “significantly inhibits the development” of a nematode pathotype or population (EPPO 2006). The H1 gene located on chromosome 5 of *Solanum tuberosum* ssp. is a resistance gene against *Globodera rostochiensis* (Bakker *et al.* 2004). H1 resistance was originally discovered in *S. tuberosum* ssp. *andigena* and has since been bred into popular potato commercial cultivars such as Maris Piper (Ellenby 1952). The presence of H1 in potato cultivars has allowed durable control of *G. rostochiensis* pathotype Ro1 in Europe, although this may be a result of the original *G. rostochiensis* introduction coming from a very restricted gene pool (Gebhardt *et al.* 1993). A second example of natural resistance is that of the *Mi-1* resistance gene widely found across *Lycopersicon* spp. after introgression from *L. peruvianum*. *Mi-1* gives broad resistance to the root-knot nematode species *Meloidogyne incognita*, *M. javanica*, and *M. arenaria*, as well as resistance against other pest species such as the potato aphid; *Macrosiphum euphorbiae* (Vos *et al.* 1998, Rossi *et al.* 1998).

Natural resistance sources must be used strategically because they can cause problems if not managed appropriately. Breeding of H1 resistance into widely used potato cultivars has caused a shift in *Globodera* species abundance. Due to *G. rostochiensis* populations being largely controlled by the presence of H1 there has been strong selection for *G. pallida* which H1 has no effect against. However, the *G. pallida* present in Europe is considerably more genetically heterogenous than the *G. rostochiensis* present, making it challenging to identify a single major resistance gene that provides complete control. Recently the H2 resistance gene has been genetically characterised which confers resistance to *G. pallida* pathotype Pa1 and partial resistance to pathotype Pa2/3 (Blok & Phillips 2012, Strachan *et al.* 2019). Another source of partial resistance against *G. pallida* pathotype Pa2/3 is the *Gpa5* quantitative trait locus (QTL), originally identified in *Solanum vernei*. This has been bred into potato cultivars such as Innovator (Roupe van der Voort *et al.* 2000). In addition, the H3 source derived from *S. andigena* also provides control of *G. pallida*

(Bryan *et al.* 2004). Many breeding programs are now focused on stacking multiple nematode resistance genes to give a more durable source of resistance. Natural resistance sources provide excellent tools for control of nematodes, but they are not always present in cultivars that growers can use. Many natural resistance sources come from wild *Solanum* spp. which are not used in mass production for consumption and need to be bred into popular cultivars of *S. tuberosum*. In addition, careful stewardship of resistance is needed. The overuse of resistant lines in the field has been shown to cause a high selective pressure for the nematode. *Globodera pallida* populations Farcet and Newton have been shown to overcome the resistance sources present in the potato cultivars 'Vales Everest' (H3 resistance) and both 'Innovator' and 'Arsenal' (*S. vernei* derived resistance) (Varypatakis *et al.* 2019).

1.1.1.6 – Potato cyst nematodes: *Globodera rostochiensis* and *Globodera pallida*

Potato cyst nematode (PCN) species make up the genus *Globodera* and are found in the Heteroderidae family of nematodes. The *Globodera* genus includes the well-researched species *G. rostochiensis* and *G. pallida*, as well as new species discovered in the past decade such as *G. ellingtonae*, *G. capensis*, and *G. agulhasensis* (Handoo *et al.* 2012, Knoetze *et al.* 2013, Knoetze *et al.* 2017) The genus also includes *G. tabacum*, a lesser-known complex comprised of three subspecies; *tabacum*, *solanacearum*, and *virginiae* (Syracuse *et al.* 2004, Mota & Eisenback 1993). Although their common name is potato cyst nematodes, *G. rostochiensis* and *G. pallida* can feed on many other host plants found in the Solanaceae family other than potato (*Solanum tuberosum*), such as tomatoes (*Solanum lycopersicum*) and aubergine (*Solanum melongena*) (Sullivan *et al.* 2007). Symptoms observed because of PCN infection are non-specific and include chlorosis, stunted growth, browning, and in extreme cases, premature host death. These non-specific symptoms often mean that farmers fail to recognise the presence of PCN until extremely damaging population levels are established.

Globodera rostochiensis (the golden cyst nematode) and *Globodera pallida* (the pale cyst nematode) are two of the most important and well researched biotrophic sedentary endoparasitic PPN species. PCN are responsible for the loss of 9% of global potato production (Turner & Rowe 2006). *Globodera rostochiensis* and *G. pallida* are highly similar in life cycle and lifestyle. *Globodera pallida* has a slower hatching rate over an extended time scale and a slower usage of stored lipid reserves found inside the hatched juvenile, allowing it to cope with a longer time between hatching and infection of a host plant (Robinson *et al.* 1987, Masler & Perry 2018). The two species also show slightly different temperature optima, with *G. rostochiensis* displaying faster development and reproduction levels at slightly higher temperatures (18 °C) while *G. pallida* was faster at lower temperatures (12 °C) despite both PCN species having similar development rates at the median temperature (15 °C) (Foot 1978, Mugniery 1978).

PCN begin their life cycle as an embryo inside an eggshell which develops into a first-stage juvenile (J1). The nematode will moult to a second-stage juvenile (J2) inside the egg which will not hatch until the presence of root diffusates are detected in the surrounding soil (Perry & Beane 1982). PCN found in temperate regions have a dormancy/diapause period that must be completed before hatching. The J2 develop inside the eggshell where they enter an obligate diapause period, becoming metabolically inactive during a stage of arrested development. Diapause of PCN coincides with the winter months where host plants are absent from the field. Temperature is a large factor in the termination of diapause as it signals the change in season and the likely onset of host plant growth. After hatching in response to root diffusates from a host plant, the J2s will locate and invade the host root where they migrate intracellularly, destroying host tissue. PCN use their stylet to inject virulence factors (effectors) into the root cells to form the syncytium; a feeding site which they stay attached to for the remainder of the life cycle (Mitchum *et al.* 2013). The nematode undergoes rounds of moulting through the J3 and J4 juvenile life stages before it becomes an adult inside the root. The sex of adult PCN is not solely

genetically determined; the proportion of males will increase under conditions of low food availability and high population density (Williamson & Hussey 1996). The sedentary female will be inseminated by a male to produce eggs, which are stored inside her body. The adult female dies after egg production, whereupon the external wall of her body forms a hard cyst to encapsulate and protect the eggs (Figure 1.1). The eggs within the cyst are a survival stage that can persist in soil for over 20 years (Evans & Stone 1977).

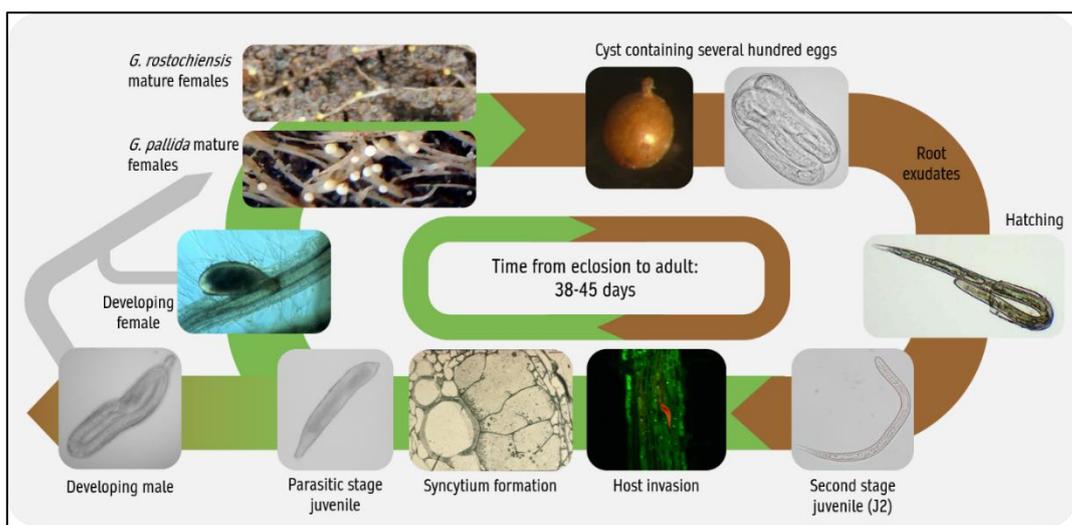


Figure 1.1 - The life cycle of *G. rostochiensis* and *G. pallida*. The juvenile nematode penetrates and moves through the root (lower right). A syncytium is induced that the nematode feeds from and maintains. The juveniles develop into adult parasitic life stages which continue to feed and breed. Successful adult females swell and are yellow or cream in colour depending on species. These can be seen on the root surface. The female produces eggs which are stored inside the protective cyst. Eggs hatch in response to root exudates in the soil where the cycle begins again. Image from (Price et al. 2021).

1.1.1.7 – Rotylenchulus reniformis

The most economically important semi-endoparasitic nematode is *Rotylenchulus reniformis* due to its wide host range and geographical span across tropical and subtropical regions (Van Den Berg et al. 2016). *Rotylenchulus reniformis* is known to parasitise over 350 plant species including many important crops such as soybean, cotton, corn, and sweet potato (Starr 1991). As is the case for many plant-parasitic nematodes, crops infected with *R. reniformis* are often misdiagnosed due to the symptoms being non-distinct and resembling characteristics of drought

and nutrient deficiency. *Rotylenchulus reniformis* is difficult to control in part due to its ability to parasitise ornamental and weed plant species. Parasitising such species allows *R. reniformis* to persist in or adjacent to arable farming land in the absence of host crop plants (Inserra *et al.* 1994, Lawrence *et al.* 2008).

The life cycle of *R. reniformis* begins with the eggs being deposited by the adult female into a gelatinous egg mass, described as a matrix, on the surface of the host plant (Figure 1.2). As seen in PCN, moulting to the J2 occurs within the egg. Hatching of J2s occurs in the soil 8 to 10 days after being deposited in the egg matrix and is not dependent on the presence of host root exudates. This reflects the broad host range of this species, meaning that it has no requirement to link the life cycle to the presence of a specific host plant. The nematodes go through three moults (J2 to J3, J3 to J4, J4 to adult) in the soil without feeding, during a period which lasts between 7 and 9 days (Jones *et al.* 2013). Once matured, the parasitic female inserts its head and upper body into the host root. This is the only stage at which *R. reniformis* enters the host. The nematode then induces the formation of a syncytium from which the female extracts nutrients (Jones 1981a). Morphological changes are observed 2 to 3 days after the initiation of the syncytium as the female nematode enlarges and takes on the shape of a kidney (hence “reniform”). The adult female will begin to produce between 40 to 100 eggs in the matrix outside the root to begin the cycle again, with a full life cycle being completed in as little as 3 weeks (17 to 29 days) depending on soil temperature and host plant species.

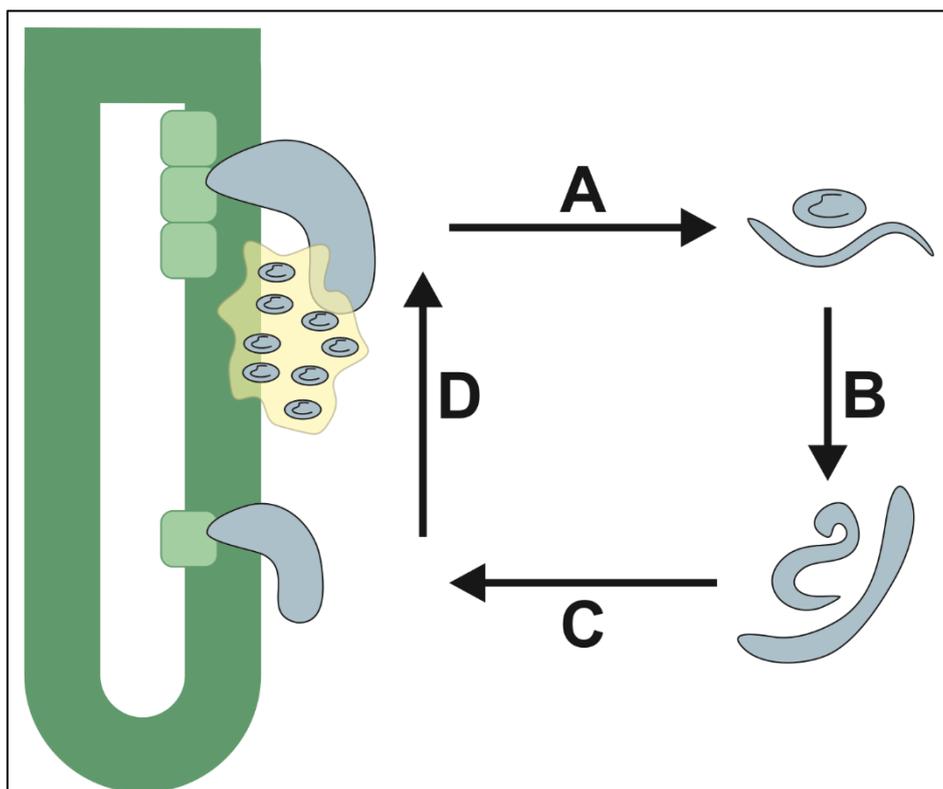


Figure 1.2 - The life cycle of *Rotylenchulus reniformis*. A – Eggs hatch from the gelatinous egg matrix. B – Maturation of juvenile nematodes through a series of moults into males or young infective stage females. C – The female nematode penetrates host root tissue with its head, to establish a feeding site (syncytium), and begins to swell. D – The female lays eggs into the matrix on the root surface where the cycle begins again.

1.1.1.8 – *Nacobbus aberrans*

Nacobbus aberrans, the false root-knot nematode is found in the USA and across both Central and South American countries (e.g., Mexico, Peru, Argentina) where it parasitises important crops such as potato, tomato, and peppers. Yield losses for potato crops due to *N. aberrans* are reportedly as high as 65% (Manzanilla-López *et al.* 2002). Due to the presence of molecular and morphological variation, it is thought that *N. aberrans* is present in a complex (*N. aberrans sensu lato*) alongside other *Nacobbus* species, *N. dorsalis*, *N. batatiformis*, *N. serendipiticus*, and *N. serendipiticus bolivianus*, however the exact species composition of this complex is still debated (Jeger *et al.* 2018). Currently it is agreed that there are three main races present in this complex that are categorised by their differing ability to infect the host species potato, beans, and sugar beet (Manzanilla-López *et al.* 2002).

Nacobbus aberrans is a species which straddles classification boundaries. During the juvenile life stages *N. aberrans* is classed as a migratory endoparasite due to its ability to migrate in and out of host roots (Jones et al., 2013). The migratory stages (J1-J4) cause large lesions of damage on the roots much like those caused by RLN. In contrast, the adult female life stage is a sedentary endoparasite as it becomes sedentary and forms a syncytium from which it feeds (Figure 1.3). The formation of the syncytium also causes the formation of a gall, a phenotype usually associated with infection by root-knot nematodes around their giant cell feeding structures and not observed with other syncytia-forming nematodes (Eves-van den Akker et al. 2014). While feeding the female will begin to produce eggs internally which are then placed into an external gelatinous matrix.

Nacobbus aberrans is an unusual nematode to characterise as it shares multiple traits which are usually exclusive to either cyst nematodes (syncytium formation) or RKN (root galling). Phylogenetically *N. aberrans* appears to be more closely related to RKN than to the cyst nematodes, however it has been noted that the position of *N. aberrans* in phylogenetic trees can vary depending on which other nematode species are included in the analysis and such positions observed are sometimes weakly supported (Holterman *et al.* 2009, van Megen *et al.* 2009, Eves-van den Akker *et al.* 2014). It is still up for debate whether the presence of syncytia in cyst nematodes and *N. aberrans* is an example of convergent evolution or a feature obtained from a common ancestor due to this uncertain phylogenetic position.

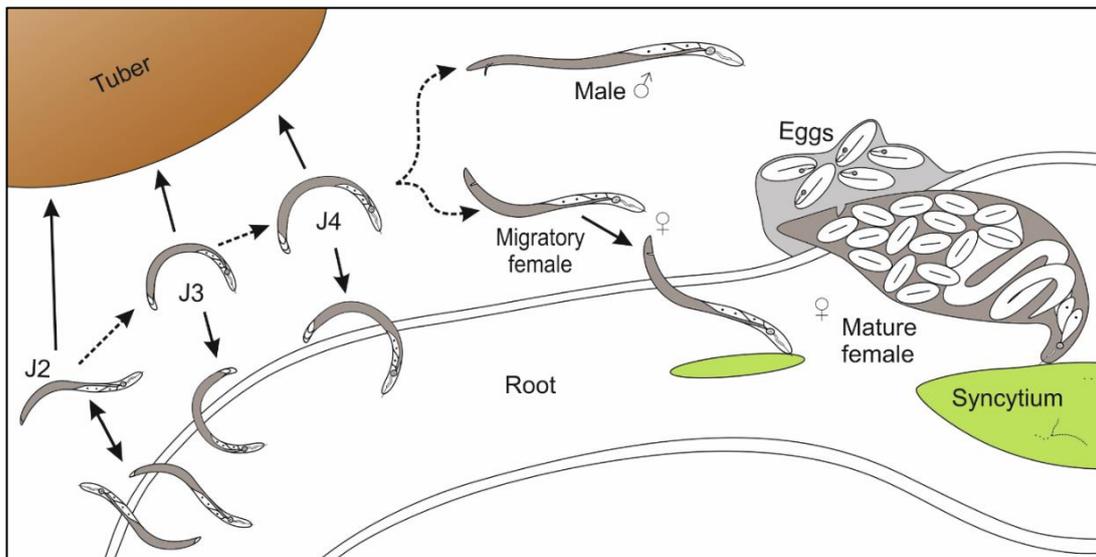


Figure 1.3 – The life cycle of *Nacobbus aberrans*. J2 nematode hatch from eggs into the soil. The J2, J3, and J4 stages are able to enter and exit the host root freely, causing necrosis and lesions. The adult female will enter the host root tissue to establish a syncytium as its feeding site. The nematode stays sedentary feeding at the syncytium where its body will swell and produce eggs. Eggs are placed in an egg matrix on the exterior of the root. Figure from (Eves-van den Akker et al. 2014).

1.2 – The syncytium

The study of biotrophic nematodes that form syncytia is of importance to world agriculture due to the high levels of crop loss and damage they cause. The syncytium is a common denominator observed in a subset of PPN species which display distinctly different stages in their life cycles and adopt different parasitism styles. An in-depth understanding of the strategies (e.g., use of effectors) employed by these nematodes in induction and syncytium maintenance is necessary to complement our knowledge of changes to host cell ultrastructure.

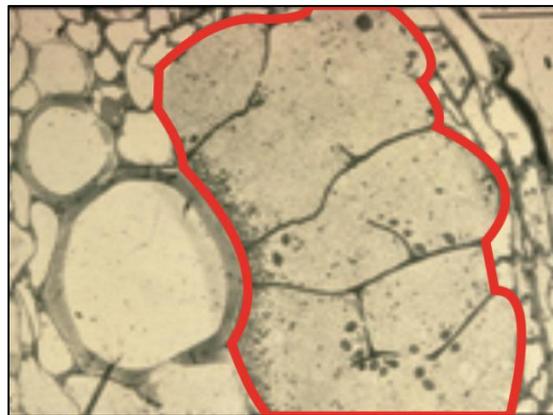


Figure 1.4 - The syncytium of *G. rostochiensis*. The syncytium (outlined in red) formed by the potato cyst nematode in the root of a potato plant. Image from (Jones et al. 2013).

1.2.1 – Syncytium formation and structure

The formation of syncytia by *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans* has been phenotypically described in detail, but the molecular mechanisms driving its formation are still largely unknown. Although these nematodes all induce the formation of a syncytial feeding site, these feeding sites differ in structure, formation, and the host cells from which they are derived. The formation and maintenance of a successful syncytium is crucial for the completion of these nematode's life cycles as they are obligate biotrophs, meaning they are exclusively reliant on the living host as a source of nutrition. Of the four, the processes underlying development of PCN syncytia have been described in the most detail.

Once a cyst nematode has entered the root it migrates towards the vascular cylinder (Davies *et al.* 2012). The juvenile nematode probes cells with its stylet before selecting an initial syncytial cell (ISC). Effectors secreted into the ISC induce cellular changes that cause the syncytium to form (Figure 1.4). The formation of a syncytium begins with the expansion of the plasmodesmata (Bohlmann & Sobczak 2014). Plasmodesmata widening allows for the plasma membrane of neighbouring cells to fuse with the ISC followed by partial cell wall dissolution to form channels between all syncytial cells. The protoplasts of neighbouring cells are then enveloped into the syncytium (Davies *et al.* 2012). The syncytium grows via continued incorporation of neighbouring cells and may eventually consist of approximately 200 cells. As a result of this protoplast fusion the syncytium becomes large, multinucleate, and holds a large volume of highly metabolically active cytoplasm. Syncytia are not only multinucleate, but these nuclei are often much bigger than those observed in unaffected cells neighbouring the syncytium. The nematode can alter the cell cycle to induce rounds of DNA endoreplication while halting mitotic division which results in large polyploid nuclei (Elling *et al.* 2007). In contrast to the hypertrophic nuclei, the vacuoles of host cells incorporated into the syncytium are drastically reduced in size and sometimes lost entirely (Jones & Northcote 1972). There is also an increase in the number of mitochondria observed in the cytoplasm as well as an increase in the number and size of plastids *e.g.* leucoplasts and chloroplasts (Sobczak & Golinowski 2011). Syncytia have been shown to contain high levels of both the cytoskeletal components microtubules (MT) and actin microfilaments (ACT). Both MT and ACT appear to be depolymerised and disrupted or disorganised in the syncytium (de Almeida Engler *et al.* 2004). It has been hypothesised that this cytoskeletal disruption could be initiated by the nematode in order to reduce the viscosity of the cell cytoplasm for easier feeding.

Syncytium production is a complex process that involves substantial changes in host gene expression, cell wall remodelling, and breakdown. It is thought that effectors

are largely responsible for initiating many of these changes. For example it is thought that effectors play a role in causing the nuclei inside the syncytium to enter aberrant cell cycles (Mitchum *et al.* 2013). It is known that plant hormones are also regulated by the nematode upon feeding site formation. Cytokinins promote cell division and their biosynthesis and regulation has been shown to be manipulated by both cyst and root-knot nematode species (Bartlem *et al.* 2014, Siddique & Grundler 2018). It has been shown that *Heterodera schachtii* produce cytokinins which may be released into the host plant to manipulate the cell cycle during syncytium formation (Shanks *et al.* 2016).

1.2.2 – The role of auxin

The plant hormone auxin and the auxin transport pathway is well defined in its role in plant growth, with functions in the initiation of cell elongation in stem and root tips. The syncytia-forming nematode *H. schachtii* has been shown to hijack PIN-FORMED auxin efflux transporter proteins (PIN) in order to redirect the flow of auxin during early syncytium formation (Grunewald *et al.* 2009). PINs were named after the original *A. thaliana* loss of function mutants which failed to make typical floral organs. The PIN mutants instead form abnormal pin-like inflorescences (Okada *et al.* 1991). PIN3 and PIN4 gene expression is upregulated in the syncytium while PIN1 and PIN7 genes are downregulated. The downregulation of PIN1 prevents auxin being transported out of the ISC, causing an accumulation of the hormone. PIN3 is redirected to localise at the lateral boundaries of the cell, directing the auxin into neighbouring cells. This leads to an increase in cell growth and size. The major role that auxin plays in feeding site establishment is confirmed by the N-1-naphthylphtalamic acid (NPA) synthetic auxin transport inhibitor which severely reduces lateral syncytial expansion when applied. These results were also observed using a PIN1 mutant which showed a 52% decrease in nematode development when compared to infection on wildtype *Arabidopsis thaliana* lines (Columbia - *Col-0*) (Goverse *et al.* 2000).

1.2.3 – Syncytia of *R. reniformis* and *N. aberrans*

Less is known about the syncytium formation of *R. reniformis* and *N. aberrans* when compared with cyst nematode species. *Rotylenchulus reniformis* is a semi-endoparasite which means it only inserts its head into the root of the host plant. This means that the ISC of *R. reniformis* is usually an endodermal cell or in rare occurrences a pericycle cell while the ISC of PCN can be endodermal, pericycle, cortical, or a procambial cell (Jones 1981b, Zhang *et al.* 2017). The ISC of *N. aberrans* is unclear, however due to the nematodes location during syncytium formation it is likely to be a cell of the vascular cylinder (Finetti-Sialer 1990). Unlike the syncytia of cyst nematodes which expand laterally to incorporate neighbouring cells, the syncytia of *R. reniformis* expands longitudinally along the root, fusing with the next cell in succession (Rebois *et al.* 1975). Both *R. reniformis* and *N. aberrans* syncytia have different structures to that of cyst nematodes. For example, wall ingrowths have been observed at sites of nutrient exchange across the cell wall in the syncytium of *Heterodera* spp. however, these ingrowths are not observed in the syncytia of *R. reniformis* and *N. aberrans*, meaning their nutrient exchange processes occur via different pathways (Jones & Payne 1977).

1.3 – Effectors

To successfully parasitise their hosts, PPN, like other plant pathogens, have evolved specialised proteins and other molecules, known as effectors, that mediate the interactions with the host. Effectors can be defined as “pathogen proteins and small molecules that alter host-cell structure and function” (Hogenhout *et al.*, 2009). Most effectors are produced in one of three specialised pharyngeal gland cells towards the anterior end of the nematode; one dorsal gland cell and two subventral gland cells (Figure 1.5) (Hussey 1989). In addition, some effectors are also produced in the amphids; anterior chemosensory organs that are surrounded by secretory cells (Figure 1.5) (Melkman & Sengupta 2004, Eves-van den Akker *et al.* 2014). Effectors that are produced in the oesophageal gland cells are released into the plant through the stylet. Generally, effectors produced in the subventral gland cells tend to have

roles in the earlier stages of the plant-nematode interaction, whereas those produced in the dorsal gland cell are important for the later stages of the interaction. There are exceptions to this, however.

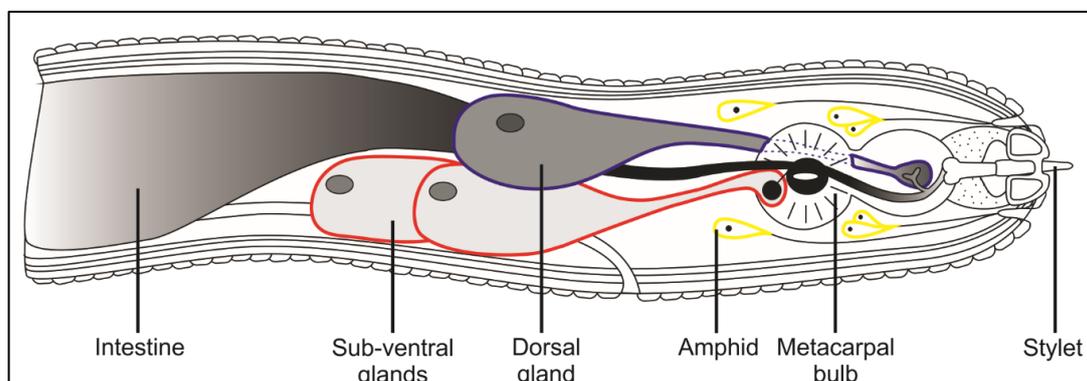


Figure 1.5 - Schematic of potato cyst nematode internal structure. Effector proteins are produced predominantly in the subventral glands (highlighted in red) and the dorsal gland (highlighted in blue) before being secreted outside of the nematode via the stylet. Some effectors are produced in the amphids (highlighted in yellow). Figure adapted from (Eves-van den Akker *et al.* 2014b).

1.3.1 – Identification of effectors

Recent advances in DNA sequencing technology have allowed genome and transcriptome data to be obtained for many nematodes including *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans* (Eves-Van Den Akker *et al.* 2016, Cotton *et al.* 2014, Eves-van den Akker *et al.* 2014). This data allows for the identification of the effector complements of each nematode. Effectors are predicted using a predefined set of criteria including the presence of a signal peptide, a lack of transmembrane helices (TMH), and being expressed in the oesophageal gland cells. Recently there has been a push to identify small sequence motifs which are shared by certain subgroups of effectors for *in-silico* identification purposes. The dorsal gland effector motif (DOG box) was discovered in 2016 and was observed upstream of 26 out of 28 effector families produced in the dorsal gland (Eves-Van Den Akker *et al.* 2016). Identifying the presence of the DOG box motif allows for the future prediction of many new effectors and acts to underpin effectors that have already been identified. Conceptually similar, but sequence unrelated motifs (STATAWAARS

and Mel-DOG) were recently found in the promotor regions of *B. xylophilus* pharyngeal gland cell effectors (Espada *et al.* 2018) and root-knot nematode (*Meloidogyne*) dorsal gland effectors respectively (M. Da Rocha *et al.* 2021). Both STATAWAARS and Mel-DOG are discussed further in section 3.1.3.

1.3.2 – Plant cell wall degrading enzymes

The plant cell wall is a major barrier to infection and consists of a complex set of proteins and carbohydrates (e.g., cellulose, pectin, and hemi-cellulose) which form a dense, branching matrix. Some effectors secreted by nematodes can be classified as cell wall modifying proteins. These include cell wall degrading enzymes (CWDE) as well as proteins that affect the cell wall in a non-enzymatic manner.

PPN produce many CWDE including cellulases and pectate lyases (G. Smant *et al.* 1998, Danchin *et al.* 2010). The pectate lyase Hspel2 from *H. schachtii* produces a reduction in infection efficiency of at least 50% when knocked down by RNAi (Vanholme *et al.* 2007). Some CWDE have also been proven to be biochemically active. *Meloidogyne incognita* produces a β -1, 4-endoglucanase called MI-ENG1 which has been shown to hydrolyse cellulose (Béra-Maillet *et al.* 2000). With a wider availability of genome and transcriptome data it has become clear that CWDE are abundant in PPN species found in clade 12. CWDE are also widespread in plant-pathogenic bacteria and fungi and are absent in almost all metazoans, except for plant-parasitic nematodes. In many cases CWDE deployed by nematodes as effectors have high similarity to bacterial or fungal CWDE, suggesting they may have been obtained by horizontal gene transfer (HGT). Many of these CWDE are conserved across clade 12, e.g. cellulases (for the glycoside hydrolase family 5), which suggests that these CWDE genes were acquired by an HGT event with a common ancestor before diversification of species within the clade (Haegeman *et al.* 2011a).

Many CWDE can be classified as members of different glycoside hydrolase (GH) families based on sequence similarities (Lombard *et al.* 2014). For example the

cellulases found in Clade 12 PPN are from the GH5 family (Danchin *et al.* 2010, Davis *et al.* 2011). Cyst nematodes such as *G. rostochiensis* and *G. pallida* are known to produce arabinogalactan endo- β -1,4-galactanases from the GH53 family, some of which are upregulated during syncytium formation (Thorpe *et al.* 2014). These break down the galactosidic links of arabinogalactan, a pectin sidechain component. Pectin is a major component of the cell wall and is very complex itself. It has both smooth regions which are areas of bare pectin backbone as well as regions which are highly branched with various side chains.

1.4 – Plant defence responses

Plants have sophisticated immune response systems that biotrophic pathogens must circumvent to infect, parasitise, and cause disease. Unlike the immune systems found in mammals, the plant immune system is not adaptive. Plant defence responses have co-evolved alongside pathogen invasion strategies. The different stages of the plant immune system were first depicted to occur as described in the zigzag model (Figure 1.6) (Jones & Dangl 2006). This dynamic model depicts two integral sections: Pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) which work in tandem.

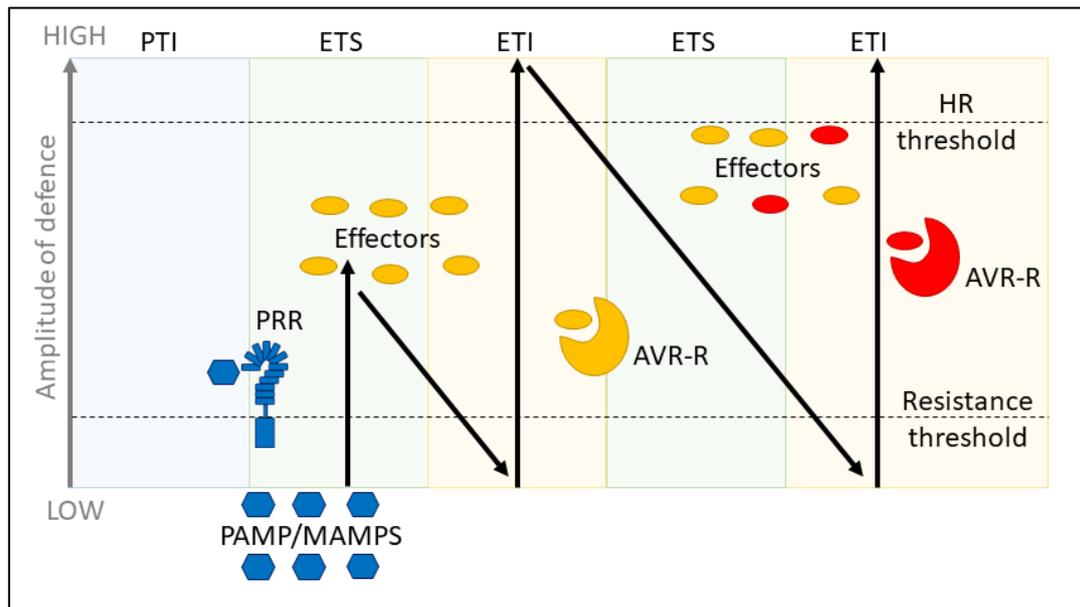


Figure 1.6 - The zig-zag model of plant immunity. This model sees plant immunity split into stages. The first stage is recognition of pathogen/microbial associated molecular patterns (PAMPs/MAMPs) by plant recognition receptors (PRR) resulting in pattern triggered immunity (PTI). The second stage sees effectors produced by the pathogen suppressing the initial immune responses leading to effector triggered susceptibility (ETS). The third stage is when host plant resistance proteins recognise the effectors and trigger effector triggered immunity (ETI). Figure adapted from (Jones & Dangl 2006).

1.4.1 – Pattern-triggered immunity

Pathogen/microbial-associated molecular patterns (PAMPs or MAMPs) are factors that are essential to the pathogen and are absent from the host plant. An example of a PAMP is the bacterial protein flagellin. Many bacteria are motile due to having flagella – long filaments which are rotated to propel the bacterium. The flagella is made up predominantly of repeating subunits of the protein flagellin. Plants recognised the presence of a conserved and essential 22 amino acid portion of the flagellin protein (flg22) and induce defence responses. Other PAMPs include peptidoglycan, lipopolysaccharides, porins (bacterial cell wall components and proteins), and fungal chitin (Livaja *et al.* 2008, Galdiero *et al.* 2004, Wolf & Underhill 2018).

If a successful defence response is mounted in reaction to the presence of a PAMP this is known as pattern-triggered immunity. Plants recognise PAMPs using pattern

recognition receptors (PRRs), which are receptor (-like) kinases that can be either membrane bound or surface-localised (Zipfel 2014). A well-studied PRR is FLS2; a leucine-rich repeat (LRR) receptor-like kinase (RLK) which recognises flagellin on the cell surface. Upon flagellin recognition, FLS2 interacts at the cell membrane with BIK1 and BAK1 proteins to initiate a phosphorylation cascade of downstream mitogen-activated protein kinases (MAPK). This in turn leads to the phosphorylation of transcription factors (TF) which initiate the transcription of defence genes such as Plant defensin 1.2 (Park *et al.* 2012). For example WRKY TFs (named after the amino acid sequence trypsin-arginine-lysine-tyrosine found in the DNA binding domain) activate many defence genes such as phytoalexins – antimicrobials which accumulate at the site of infection alongside reactive oxygen species (ROS) (Phukan *et al.* 2016). Another PRR is the EF-Tu receptor (EFR) found in Brassicaceae such as *Arabidopsis thaliana*, which perceives the bacterial PAMP elongation factor Tu (EF-Tu) (Zipfel 2014). Upon recognition of EF-Tu, EFR induces a common set of responses with FLS2 such as extracellular pH shifts (alkaline) and activation of MAPK signalling cascades to induce defence gene expression (Zipfel *et al.* 2006).

1.4.1.1 – Pattern Triggered immunity and nematodes

Current work is underway to identify PAMPs from nematodes. Ascarosides are small molecules found exclusively in nematodes that have been shown to induce plant defence responses. Ascaroside 18 (ascr18) causes MAPK signalling cascades and upregulation of defence genes (Manosalva *et al.* 2015). Although the pattern recognition receptor responsible for the recognition of ascr18 has not been identified, ascr18 is recognised by multiple different host plants including barley and potato (Sato *et al.* 2019). A second example of nematode PAMPs are found in NemaWater. NemaWater is produced by submerging nematodes (*H. schachtii*, *M. incognita*) in H₂O, which is then incubated for 24 hours before the nematodes are removed. This leaves any potential PAMPs behind in solution. NemaWater has been used to test whether a PTI response from the host plant was induced by the presence of nematode PAMPs or because of the physical wounding produced by nematode

feeding and movement. NemaWater was applied to *A. thaliana* roots which produced a ROS burst and the upregulation of immune response genes. The same immune responses are observed upon the recognition of other PAMPs such as bacterial flagellin (Mendy *et al.* 2017). This suggests that PAMPs left in solution were the cause of this immune response, although specific nematode PAMPs are yet to be characterised.

On the other side of the coin, a PRR specific to a nematode PAMP has also been identified through further research using NemaWater. Although it is not yet known which molecules or proteins found in NemaWater cause activation of the signalling pathway triggered by this receptor, it has been shown that the Nematode-Induced LRR-RLK 1 (LRR-RLK NILR1) and the co-receptor BAK1 are upregulated in response to root contact with NemaWater, and that signalling pathways are induced by the presence of NemaWater that are dependent on these receptors (Mendy *et al.* 2017).

1.4.2 – Effector-triggered susceptibility and immunity

As discussed in section 1.4, pathogens have evolved effector proteins which work to suppress PTI and increase pathogen fitness. Although many effectors suppress host defences, others may have additional roles, particularly in plant-nematode interactions where a feeding site is induced. If the effectors are successful in suppressing PTI, this will result in effector-triggered susceptibility (ETS), allowing the pathogen to infect the host plant (Figure 1.6). Plants have evolved a second layer of defence responses based on the recognition of effectors – effector triggered immunity (ETI). The onset of ETI usually elicits a stronger defence response than that mounted during PTI and is mediated by resistance genes (R-genes). R-genes encode resistance proteins such as nucleotide binding leucine rich repeat (NB-LRR) proteins.

ETI often sees the threshold for initiation of a hypersensitive response (HR) being activated. HR is found in both plants and animal cells. In plant systems it is defined as programmed cell death (PCD), induced by plant resistance (R) proteins (Coll *et al.*

2011). The hypersensitive response sacrifices infected host cells to impede the progress of an invading pathogen and to save the rest of the plant. PCD is observed at different stages of nematode infection; at sites of initial nematode penetration and migration, surrounding the feeding site (syncytium and giant cells) and cell death of the feeding site itself (Sato *et al.* 2019). The death of cells surrounding the syncytium of *G. rostochiensis* in resistant tomato lines containing the *Hero A* gene results in reduced nutrient transfer and the syncytium being cut off from the neighbouring healthy cells. The PCD of surrounding cells can lead to the premature death of the syncytia (Sobczak *et al.* 2005).

The way in which resistance proteins recognise the presence of effectors (often referred to as avirulence proteins) has been the topic of lengthy debate. Historically it was thought that R-proteins act as receptors which recognise and have direct interactions with specific avirulence (Avr) proteins. This was known as the gene-for-gene hypothesis (Flor 1971). Although there are examples which follow the gene-for-gene hypothesis, such as the *Pi-ta* resistance gene from rice and *AVR-Pita* from the rice blast fungus *Magnaporthe grisea*, there are many examples that do not conform to this model (Jia *et al.* 2000). The guard model postulates that R-proteins “guard” host proteins which are targeted by effectors. When the structure or activity of the host effector target is altered, the R-protein is activated, in turn activating the plant immune responses (Figure 1.7 A, B) (Dangl & Jones 2001). The guard hypothesis explains why some resistance genes (*e.g.*, *Mi*) can provide resistance against multiple, unrelated pathogens. A further refinement that has emerged from the guard model is the decoy model. The decoy model proposes that due to the evolutionary strain put upon effector targets as part of the guard model, it is likely that the hosts have evolved decoy proteins which are guarded by R-proteins and are recognised by the effector but that do not function in the same way as the native effector target (Figure 1.7 C) (van der Hoorn & Kamoun 2008). The decoy model now has some strong experimental support including the example of the *Bs3* gene from pepper. *Xanthomonas campestris* pv *vesicatoria* produces *AvrBs3* which binds to the

promotor region of the host *Upa20* gene. This induces hypertrophy of host cells. It has been shown that *AvrBs3* also binds to the promotor region of the R gene *Bs3*. The *Bs3* gene is not expressed in the absence of *AvrBs3* which is consistent with the theory that *Bs3* acts as a decoy for *Upa20* (Kay *et al.* 2007).

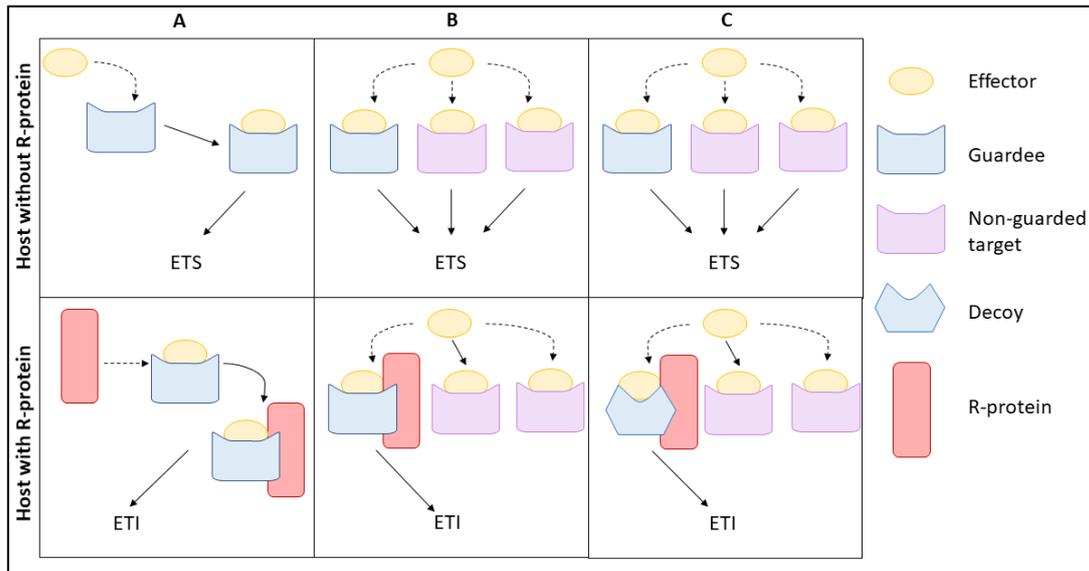


Figure 1.7 – The decoy and guard models. A – The classical guard model where host R-proteins monitor a guardee protein. When an effector interacts with or alters the guardee then the R-protein will begin ETI-related defences. B – The guard model when an effector has multiple targets. C – The decoy model postulates that the host may evolve decoy proteins which the effector in question will interact with, however, these do not act like the native interacting protein. Effector interaction with a decoy is detected by R-proteins and ETI defences are induced. Figure adapted from (van der Hoorn & Kamoun 2008).

1.4.3 – Redefinition of plant immunity models

In recent years the zigzag model has been called into question as being outdated or incomplete (Pritchard & Birch 2014). The components of the model, PTI, ETS, and ETI are in contention and it has been proposed that the zig-zag model is too simplified. This is because it does not consider phytohormones (salicylic acid, jasmonic acid etc.) or RNA silencing which also play a role in disease resistance and should be considered as a third and fourth branch of defence (Robert-Seilaniantz *et al.* 2011). To address this, Andolfo and Ercolano produced the circular model which includes hormone signalling and metabolic pathway modifications occurring alongside PTI and ETI when discussing plant immunity (Andolfo & Ercolano 2015). Host microRNA (miRNA) and phased secondary small interfering RNA (phasiRNA) have also been shown to play a role in regulation of plant defence responses which adds another layer that is not covered by the classic zig-zag model (Fei *et al.* 2016).

Recently there has been a push to redefine the classic zig-zag model which would see the scientific community moving away from describing the actions of the plant immune system occurring as part of PTI or ETI. It has been proposed that there should be a shift towards discussing these defence responses in spatial terms as surface receptor-mediated or intracellular receptor-mediated immunity (He *et al.* 2018, Kanyuka & Rudd 2019, Pok Man Ngou *et al.* 2020). Talking about immunity in terms of receptors allows for a more cohesive narrative where different parts of the immune system work together or in tandem. It has been suggested that ETI responses heighten those seen during PTI. This blurs the line between two previously individual branches of immunity. This was observed using a transgenic *A. thaliana* line containing the AvrRps4 effector from *Pseudomonas syringe* which is oestradiol-inducible. When AvrRps4 is pre-induced ETI is activated by recognition via the Toll/Interleukin receptor/Resistance protein nucleotide binding leucine rich repeat (TIR-NLR) receptors RRS1 and RPS4. This pre-induction of AvrRps4 shows an increase in PTI ROS production when induced using flg22 (Pok Man Ngou *et al.* 2020).

1.4.4 – Primary defence responses

Upon recognition of pathogen there are a vast array of defence responses induced by the host plant. Callose deposition is a well-known basal plant defence response that occurs in response to wounding and puncturing of the cell wall by a pathogen. Callose is a polysaccharide that is deposited at the cell wall-plasma membrane interface as part of papillae; complex structures containing cell wall proteins, reactive oxygen species, and phenolic compounds (Voigt 2014). These papillae are deposited as a physical barrier to prevent further infection by the pathogen.

Plants are well known to produce ROS bursts in response to pathogen recognition. ROS are unstable free radicals which react quickly with other molecules they encounter, causing damage to key components such as DNA, RNA, and lipids as well as overall cell death. ROS such as hydrogen peroxide are produced at the plasma membrane by NADPH oxidase (Bolwell *et al.* 2002). Strong ROS bursts have been observed in resistant (*Mi-1.2*) tomato lines in response to infection attempts by *M. incognita* (Melillo *et al.* 2011). ROS are antimicrobial and toxic to nematodes, however certain nematodes (both cyst and RKN species) have evolved antioxidant enzymes that are present on the nematode surface to protect themselves (Sato *et al.* 2019). Both *G. rostochiensis* and *M. incognita* produce peroxiredoxin enzymes on their surface which are able to remove hydrogen peroxide found in the host apoplast (Dubreuil *et al.* 2011, Robertson *et al.* 2000).

The responses activated when a pathogen is perceived by the host plant include an upregulation of many defence genes, calcium influx, and activation of local ethylene signalling pathways. Infection of *H. schachtii* on *A. thaliana* ethylene signalling mutant lines were observed to establish the initial syncytial cell several hours before those on the wild type (WT) Col-0 lines, suggesting that ethylene signalling plays a role in prevention of syncytium establishment (Marhavý *et al.* 2019). A range of enzymes are also produced by host plants in response to nematode infection. For example, it was observed that chitinases were upregulated by the host plant

(resistant cotton lines) after infection by *M. incognita* – suggesting a link between chitinases and nematode infection (de Deus Barbosa *et al.* 2009).

1.4.5 – Suppression of defence responses by PPN

Like other plant pathogens, PPN have evolved ways to suppress both PTI and ETI host responses. A ubiquitin extension protein from *G. rostochiensis* suppresses a range of PTI-associated immune responses. The GrUBCEP12 gene from *G. rostochiensis* is a ubiquitin carboxyl extension protein (UBCEP) that suppresses PCD (Chronis *et al.* 2013). When Gpa2 (a *G. pallida* resistance gene from potato) alongside its elicitor RBP-1 (*G. pallida* protein) are co-infiltrated into *N. benthamiana* a PCD response is observed, however, when Gpa2/RBP-1 are co-infiltrated alongside GrUBCEP12 the cell death response is suppressed. The same result is observed using the Rx2 potato resistance gene with the coat protein (CP) from potato virus X (PVX). Knock down of GrUBCEP12 also gave a significant reduction in *G. rostochiensis* ability to parasitise RNAi transgenic potato lines.

Cyst nematode genomes encode a family of effectors which contain the SP1a and RYanodine receptor (SPRY) domain. This family of effectors are called the SPRY domain-containing proteins (SPRYSEC) and have been demonstrated to interact with host NB-LRR resistance proteins (Rehman *et al.* 2009). The SPRYSEC-19 effector from *G. rostochiensis* interacts with a NB-LRR from the SW5 resistance gene cluster in tomato, but this interaction does not activate a PCD response (Postma *et al.* 2012). In *N. benthamiana* the co-infiltration of SPRYSEC-19 suppresses the PCD response usually observed with the recognition of RBP-1 by Gpa2 (*G. pallida*, *S. tuberosum*) and CP by Rx1 (PVX, *S. tuberosum*). The SPRYSEC-414-2 from *G. pallida* reduces the production of ROS induced by flagellin (flg22) recognition and also suppresses the PCD response observed by the recognition of RBP-1 by Gpa2 as seen with SPRYSEC-19 (Mei *et al.* 2018). SPRYSEC-414-2 has also been shown to directly interact with a host stCLASP protein which is involved with microtubule cytoskeleton stability and development (Mei *et al.* 2018). The PCD-suppressive function of

SPRYSEC-19 and SPRYSEC-414-2 are clear examples of how nematodes employ effectors to suppress host defence responses. One SPRYSEC protein, RBP-1, has been identified as the avirulence factor recognised by the Gpa2 resistance gene (Sacco *et al.* 2009).

Plant phytohormone pathways play a key role in basal defences against pathogen, insects, and abiotic stresses. The hormone salicylic acid (SA) is involved in regulation of the phenylpropanoid pathway and production of defensive enzymes such as peroxidases and reactive oxygen species (ROS) (War *et al.* 2011). Transient expression of the chorismate mutase (CM) effector Mi-CM-3 from *M. incognita* in *N. benthamiana* causes a reduction in SA levels after infection with *Phytophthora capsici* (Wang *et al.* 2018). Expression of Mi-CM-3 was also shown to increase host susceptibility to infection by *M. incognita*. The effects of Mi-CM-3 may be due to suppression of SA production due to depletion of the chorismate pool available for conversion to SA via an isochorismate intermediate.

The need to suppress host defence responses is common to all biotrophic pathogens and in some cases, the same host defence proteins may be targeted by multiple pathogens. For example, a cysteine protease (Rcr3) which is involved in defence signalling has been shown to be targeted by fungi (*Cladosporium fulvum*), oomycete (*Phytophthora infestans*), and nematodes (*G. rostochiensis*) (Song *et al.* 2009, Lozano-Torres *et al.* 2012). *Globodera rostochiensis* produces an effector, Gr-VAP-1, which interacts with Rcr3^{pim}, this interaction is recognised by Cf-2, a tomato immune receptor which initiates a HR response in cells containing Gr-VAP-1. It has been shown that both Cf-2 and Rcr3^{pim} are essential for resistance with removal of either significantly increasing susceptibility to *G. rostochiensis* infection (Lozano-Torres *et al.* 2012). Knockdown of Gr-VAP-1 by RNAi decreased *G. rostochiensis* infection rates by approximately 50%, demonstrating the importance of this effector to the nematode (Lozano-Torres *et al.* 2014).

1.4.6 – Core effectors

It has now been established why effector proteins are so important to many types of pathogen, as all effectors play important roles in successful infection strategies. Some effectors are conserved across a wide range of species and can be termed “core effectors”. Core effectors are those which are (A) conserved across species in a genus or across multiple genera, and/or (B) effectors which the pathogen would not be able to complete the infection process and life cycle without. The term core effector has been used recently by Seitner *et al.*, to describe the cysteine-rich core effector 1 (Cce1) protein. “Core” is used to describe Cce1 due to it being found in all smut fungi species (*Ustilago ssp.*) which target monocot host plants (Seitner *et al.* 2018). “Core effectors” in terms of this thesis are defined as an effector which has at least one orthologue present in all syncytia-forming PPN, represented in this case by four nematodes species *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans*.

1.5 – Thesis outline

Although *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans* have differing life cycles, each can induce the formation of a syncytium in a series of common hosts (e.g., potato, tomato). Despite the observed differences in syncytium structure, formation, and ontogeny, it is likely that some mechanisms required for feeding site formation in *Solanum* are manipulated by all species. Genome and transcriptome resources are available for these species, as well as tools for predicting effectors. This thesis aims to complete a comparative analysis of effectors produced by *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans* to yield new insights into the induction and maintenance of syncytia. Chapter 3 will discuss the *in-silico* identification of core effectors found to be present across the four syncytia-forming nematode species analysed. The fourth and fifth chapters will cover the functional characterisation work carried out on the arabinogalactan endo- β -1,4-galactanase GH53 family of plant-parasitic nematodes and the conserved cathepsin L-like protein family respectively. The sixth chapter will discuss the functional characterisation of the core effector genes GROS_g05682 (20E03), GROS_g02394 (GLAND11), and GROS_g02496 (GLAND15) from *G. rostochiensis* identified in chapter three.

2. General Materials and

Methods

This chapter contains all protocols which were used across multiple research chapters.

2.1 – Nematode collection

Globodera rostochiensis (Ro1) and *G. pallida* (Pa2/3 Lindley) cysts used in all experiments were obtained from The James Hutton Institute PCN collection. Nematodes were maintained on the potato cultivars Desiree (*G. rostochiensis*) or Maris Piper (*G. pallida*) grown in a glasshouse on a 16 hour day/8 hour night cycle. Cysts were extracted from soil using standard protocols outlined below and were stored at 4 °C for a minimum of 3 months before use.

2.1.1 – Second stage juvenile (J2) collection

Pre-parasitic second stage juveniles (J2) were collected from *G. rostochiensis* cysts (pathotype Ro1, 2012 population). Cysts were placed in a 200 µm sieve inside a large petri dish (140 mm ∅) and incubated in 50 ml of tomato root diffusate (TRD) (Section 2.1.3). Plates were wrapped in a layer of clingfilm with a layer of tin foil on top, before incubation in the dark for 9-14 days at 18 °C to allow J2s to hatch. Nematodes were harvested before a second volume of fresh TRD was applied to the cysts. This was repeated regularly until nematode numbers diminished. J2s were cleaned using the sucrose floatation method (Section 2.1.4) to remove any bacterial and/or fungal contaminants present. Nematodes were centrifuged at 5200 x g for 10 minutes before the supernatant was removed for long-term storage at -80 °C. For use in *in situ* hybridisation (section 2.8.2) J2s were incubated in 2% paraformaldehyde in M9 buffer (pH 7.0, Na₂HPO₄, KH₂PO₄, NaCl, MgSO₄·7H₂O) at 4 °C in lo-bind protein Eppendorf tubes after sucrose floatation.

2.1.2 – Mixed parasitic life stage collection

Mixed parasitic stages of *G. rostochiensis* were obtained from infected potato (cv. *Desiree*). Tubers were warmed and allowed to chit in the glasshouse for two weeks. Chitted potatoes were planted in autoclaved 50:50 sand/loam mix containing 50 cysts per pot. Potatoes were grown in a glasshouse on a 16 hour day/8 hour night cycle with light and gentle daily watering from above to avoid washing J2s out of the sand/loam. After 4-5 weeks roots were extracted from the pot and lightly washed to remove residual sand/loam. Roots were cut into approximately 1-2 inch sections before being blended in sterile distilled water (SDW) using a Kenwood handheld blender. Blended root material was filtered through a series of sieves; 2.8 mm, 500 µm, 250 µm, 90 µm with nematodes being collected on a final 20 µm sieve. Nematodes were cleaned using sucrose floatation (Section 2.1.4), placed in SDW, and were stored at 4 °C. For *in situ* hybridisation use (Section 2.8.2) nematodes were stored in 10% formaldehyde (in M9 buffer) at 4 °C in lo-bind protein Eppendorf tubes.

2.1.3 – Tomato root diffusate

Tomato root diffusate (TRD) was produced using 4-6 week old tomato plants (cv. *MoneyMaker*), grown in a glasshouse using a 16 hour day/8 hour night cycle. Roots were washed to remove soil debris before being submerged in 500 ml of SDW overnight. The resulting TRD was filtered through Whatman paper before storage at 4 °C. TRD was stored for a maximum of four weeks before use.

2.1.4 – Sucrose floatation

Nematodes were separated from root and soil debris and any fungal or bacterial contamination using a sucrose solution (50% w/v). Five millilitres of 50% sucrose was added to 5 ml of nematode solution in TRD in a 15 ml falcon tube, with 500 µl of SDW layered over the top. The solution was centrifuged at 2500 x g or 10 minutes to pellet contaminants at the bottom of the tube while nematodes formed a layer between the water and sucrose. J2s were collected and rinsed with SDW by centrifugation.

For parasitic stage nematodes, the layer of nematodes was added to a 50 µm sieve. SDW was used to wash off excess sucrose from the nematodes which were collected on a 35 µm sieve before storage in lo-bind protein Eppendorf tubes.

2.2 – RNA extraction

2.2.1 – RNA extraction from J2 nematodes

J2s were collected in a 1.5 ml Eppendorf and centrifuged at 2500 x g for 10 minutes to form a pellet. The supernatant was removed from the nematode pellet before freezing in liquid nitrogen. The pellet was crushed into a powder using a sterile mortar and pestle which had been prechilled in liquid nitrogen. One millilitre of Trizol (ThermoFisher) and 200 µl of chloroform was added to the powdered pellet which was then transferred to a fresh Eppendorf tube. The sample was vortexed and incubated at room temperature for 5 minutes with frequent agitation, then centrifuged at 12000 x g at 4 °C for 15 minutes. The upper aqueous layer was removed from the sample and added to 500 µl of 100% isopropanol in a fresh Eppendorf tube. The sample was incubated at room temperature for 10 minutes before being centrifuged at 12000 x g at 4 °C for 10 minutes. The supernatant was removed before washing the pellet in 1 ml of 75% ethanol. The sample was vortexed before centrifugation to form a pellet at 7500 x g at 4 °C for 5 minutes. The supernatant was removed, and the pellet was air dried in a flow cabinet. The pellet was resuspended in 20 µl of RNase-free H₂O and incubated at 60 °C for 10 minutes in order to allow resuspension of the RNA.

2.2.2 – DNase treatment of RNA extractions

RNA was extracted as described above (Section 2.2.1) and was treated with a RQ1 RNase-free DNase kit (Promega) to remove any contaminating DNA from samples following the manufacturer's protocol. Reactions contained 8 µl of RNA, 1 µl of RQ1 10X reaction buffer and 1 µl of RNase-free DNase and were incubated for 30 minutes at 37 °C. The reaction was terminated by the addition of 1 µl of RQ1 DNase Stop

solution followed by incubation for 10 minutes at 65 °C. RNA samples were used for cDNA synthesis immediately after DNase treatment.

2.3 – cDNA synthesis

cDNA was synthesised from DNase-treated RNA using SuperScript™ III reverse transcriptase (Invitrogen) and poly(dT) primers according to the manufacturer's protocol. Eleven microlitres of DNase-treated RNA was added to 1 µl 50 µM oligo(dT)₂₀ and 1 µl 10 mM dNTP mix. The sample was incubated at 65 °C for 5 minutes to denature secondary structure of mRNAs before being transferred onto ice. Four microlitres of 5X First strand buffer, 1 µl 0.1 M DTT, 1 µl of RNaseOUT and 1 µl of Superscript III reverse transcriptase (Invitrogen) were added to the sample before incubation at 50 °C for 1 hour. The reaction was inactivated by incubation for a further 15 minutes at 75 °C.

2.4 – Polymerase Chain Reaction

Polymerase chain reaction (PCR) was carried out using either KOD Hot Start Polymerase (Merck) or GoTaq G2 Flexi polymerase (Promega). KOD polymerase was used when proofreading capabilities were required *e.g.*, for cloning. For KOD polymerase PCR a total reaction volume of 50 µL was made up of 5 µL 10x KOD buffer, 3 µL MgSO₄, 5 µL 2mM dNTPs, 1.5 µL of sequence specific forward primer, 1.5 µL of sequence specific reverse primer, 1 µL of 5 U/µL KOD polymerase and 1 µL of template DNA in 32 µL H₂O. For GoTaq polymerase PCR a total reaction volume of 20 µL was made up of 4 µL 5x GoTaq buffer, 1.6 µL 25mM MgCl₂, 0.8 µL 10mM dNTPs, 0.6 µL 0.5 µM of sequence specific forward primer, 0.6 µL 0.5 µM of sequence specific reverse primer, 0.08 µL of 5 u/µL GoTaq G2 flexi polymerase and 2 µL of template DNA in 10.32 µL H₂O. Conditions for running both KOD and GoTaq PRC reactions can be found in Table 2.1 and Table 2.2. Elongation time and annealing temperature were subject to change depending on expected product size (kb) and primer pairs used.

Table 2.1 – Program for GoTaq PCR

Temperature	95 °C	95 °C	50-70 °C	72 °C	72 °C	12 °C
Time	10m	45s	30s	Product size dependent	5m	∞
Cycle no.	x1	x35			x1	x1

Table 2.2 – Program for KOD PCR

Temperature	95 °C	95 °C	50-70 °C	72 °C	72 °C	12 °C
Time	3m	30s	30s	Product size dependent	3m	∞
Cycle no.	x1	x35			x1	x1

2.4.1 – Gel electrophoresis

For analysis of PCR products 5 µL of PCR product was mixed with 5 µL SDW and 2 µL 6x Blue/Orange loading dye (Promega) and visualised by gel electrophoresis on a 1.5% agarose gel with 0.8 µl SYBR Safe DNA gel stain in 1x Tris-borate-EDTA (TBE) buffer. Either 1Kb or 100bp DNA ladder (Promega) was used according to expected product band size. Gels were electrophoresed for approximately 20-30 minutes between 75-100 Volts and visualised on a UVIDOC HD2 transilluminator (UVITEC Cambridge).

2.5 – Cloning of cDNA sequences encoding effectors

The coding region of *G. rostochiensis* candidate effector and GH53 genes were amplified by PCR from cDNA using KOD Hot start DNA polymerase (Section 2.4). The full open reading frames of predicted genes were amplified excluding any predicted signal peptide. Forty five microlitres of PCR product was purified by gel electrophoresis on a 1.5% agarose gel with the appropriate bands cut out from the gel under low intensity UV light using a scalpel. A QIAquick gel extraction kit (QIAGEN) was used to purify DNA from excised bands with a final elution of DNA in 30 µL elution buffer (Tris-HCL 10 mM, pH 8.5). In some cases where a single band was obtained following PCR, PCR products were column purified using a QIAquick PCR purification kit (QIAGEN) following the manufacturer's protocol. Quantification of purified products was carried out using a nanodrop spectrophotometer. The purified fragments were cloned into the pCR8/GW/TOPO (Thermofisher) (Section 2.5.1) or pGEM[®]-T Easy vector (Promega) (Section 2.5.2).

2.5.1 – pCR8/GW/TOPO cloning

The pCR8/GW/TOPO vector (Invitrogen) was used for gateway cloning. For genes cloned into the pCR8/GW/TOPO vector, purified PCR products had 3' A-overhangs added by incubating purified PCR product, with GoTaq 5X buffer, 25mM MgCl₂, 10mM dATP, and 0.2 µl 5u/µl GoTaq, H₂O in a PCR machine at 72 °C for 10 minutes (Table 2.3). Four microlitres of the A-overhang reaction product was immediately

used in a TOPO TA-cloning reaction containing 1 μ l 4X diluted salt solution (1.2 M NaCl, 0.06 M MgCl₂), 0.5 μ l TOPO vector and 0.5 μ l H₂O (using ½ of the recommended volume of TOPO vector). The reaction was incubated at room temperature for 30 minutes then placed on ice before transformation.

Table 2.3 – Addition of A-overhangs reaction

Reagent	1x volume (μ l)	Final concentration
H ₂ O	Adjust to 10	-
PCR purified product	Up to 6.7	10-30 ng/ μ l
GoTaq Buffer (5x)	2	1x
MgCl ₂ (25mM)	0.8	2 mM
dATPs (10mM)	0.4	0.4 mM
GoTaq (5u/ μ l)	0.1	0.05 u/ μ l
Total vol.	10	

2.5.2 – pGEM-T easy cloning

The pGEM-T Easy vector was used to clone probe fragments for use in *in situ* hybridisation. For genes cloned into the pGEM-T Easy vector, purified PCR product was added to a ligation reaction (Table 2.4) which was incubated overnight at room temperature and placed on ice before transformation (Section 2.6).

Table 2.4 – pGEM-T Easy ligation reaction for creation of entry clones

Reagent	volume (μl)
H ₂ O	Up to 10
2X rapid ligation buffer	5
Purified PCR product	X*
pGEM-T Easy vector	1
T4 DNA ligase (3 Weiss units/ μ l)	1
Total vol.	10

*X μ l of PCR product was determined using the vector molar ratio 3:1.

$$\frac{ng \text{ of vector } \times kb \text{ of insert}}{kb \text{ size of vector}} \times \text{insert:vector molar ratio} = ng \text{ of insert}$$

2.6 – Cell transformation

pGEM-T Easy and TOPO entry clone plasmids were transformed into electro-competent DH5 α (Section 2.6.1) or chemically competent (heat shock) JM109 *E. coli* cells (Section 2.6.2) and cultured on plates with appropriate antibiotic resistance (Table 2.5). pGEM-T Easy transformations were plated onto LB AIX (ampicillin, isopropyl β -D-1-thiogalactopyranoside, X-gal) agar plates which allows for blue/white colony screening. Colonies which have the insert in the pGEM-T Easy vector were white in colour, whereas negative colonies were blue.

Table 2.5 – Antibiotic concentration usage for cloning

Antibiotic	Stock concentration (mg/ml)	Working concentration (µg/ml)	Destination containing resistance (Vectors, cells)
Ampicillin	100	100	pGEM-T Easy, pHANNIBAL, pOPINS3C, pDEST22
Chloramphenicol	30	7.5	pDONR201, pH7WGR2, pDEST22
Gentamicin	30	7.5	pDEST_32
Kanamycin	50	50	P19, pGRAB_mturq2_GW, pDONR201, pEHISTEV
Rifampicin	25	6.25	AGL1 (<i>A. tumefaciens</i> cells)
Spectinomycin	100	100	pCG8_GW_TOPO, pH7WGR2, pK7WGF2, pK7FWG2, pART27
Tetracycline	5	5	pSOUP

2.6.1 – Electroporation

For each individual transformation reaction an electroporation cuvette was prechilled on ice. Two microlitres of each entry clone reaction was added to 50 µl of cells and incubated on ice for 20 minutes. Reactions were electroporated at 1800 Volts (BioRad MicroPulser, using the EC1 bacteria setting) before the immediate addition of 500 µl of SOC media. Reactions were placed in a sterile 2 ml Eppendorf tube and incubated shaking at 37 °C for 1 hour 30 minutes. One hundred microlitres of the cells were plated on LB + Spectinomycin (100 µg/ml) agar plates for TOPO reactions or LBAIX (100 µg/ml Ampicillin, 32 mg/ml IPTG, 32 mg/ml X-gal) agar plates for pGEM-T Easy reactions and grown overnight at 37 °C (Table 2.5).

2.6.2 – Heat shock transformation

For each individual transformation, 2 µl of each entry clone reaction was added to 50 µl of cells and incubated on ice for 30 minutes. Cells were placed in a 42 °C water bath for 40 seconds before being immediately returned to ice for 2 minutes. Five hundred microlitres of SOC media was added and reactions were then incubated with shaking at 37 °C for 1 hour and 30 minutes. One hundred microlitres of the cells were plated on LB + Spectinomycin (100 µg/ml) agar plates for TOPO reactions or LBAIX (100 µg/ml Ampicillin, 32 mg/ml IPTG, 32 mg/ml X-gal) agar plates for pGEMT-Easy reactions and grown overnight at 37 °C (Table 2.5).

2.6.3 – PCR colony screening

One to eight colonies per construct were each resuspended in 30 µL of SDW. Five microlitres of each bacterial suspension was plated on LB plates with the appropriate antibiotic (Table 2.5). When cloning into the TOPO and pGEM-T Easy vectors it is possible for DNA fragments to be inserted in either orientation. To test for the correct insertion and orientation of the cloned gene into the vector, two PCR reactions were run in tandem using GoTaq conditions detailed above (Section 2.4). PCR one uses M13 forward and reverse primers which amplify the DNA fragment inserted into the vector along with a small amount of flanking DNA sequence. The size of the product amplified in this reaction can therefore be compared to the known size of the original PCR product being cloned. PCR two uses M13 forward and the gene specific reverse primer and will only produce an amplicon if the expected DNA sequence has been incorporated into the vector in the correct orientation. For cloning into the pGEM-T Easy vector, colonies were tested using PCR with M13 forward and reverse primers as orientation of the gene was not important for these cloning reactions. PCR products were visualised by gel electrophoresis on a 1.5% agarose gel alongside DNA molecular weight markers to determine the size of the product (Section 2.4.1).

Colonies containing a correctly orientated insert in the vector were streaked out onto fresh plates with appropriate antibiotic. These colonies were also used to make overnight cultures. Five microlitres of bacterial suspension in 5 ml of LB containing spectinomycin (100 mg/ml) or LBAIX for TOPO and pGEM-T Easy plasmids respectively, were incubated with shaking at 37 °C overnight. Purified plasmid was obtained by miniprep of 3 ml of culture using a QIAprep spin miniprep kit following the manufacturer's protocol (QIAGEN). Purified plasmid concentrations were determined using a NanoDrop spectrophotometer. For long-term storage a 33% glycerol stock was produced for cells containing transformed plasmids. Six hundred microlitres of 60% glycerol were mixed with 1200 µL of the remaining overnight culture put aside before miniprep. These glycerol stocks were stored at -80 °C in 1.8 ml screw top cryotubes (Thermofisher).

2.7 – Sample sequencing

DNA samples were Sanger sequenced in-house, using the service provided at The James Hutton Institute. Universal primers (M13, T7) or gene specific primers were provided to the James Hutton sequencing lab to confirm correct insertion, reading frame and orientation of genes in cloning vectors. Sequencing results were aligned (pairwise) using MUSCLE and analysed using the software Jalview (V 2.10.0b1) (Edgar 2004, Waterhouse *et al.* 2009) and the online tool MultAlin (Corpet 1988). SnapGene (v 4.3.5, GSL Biotech) was used to view and construct vector maps for genes cloned into expression vectors. The online translate tool from ExpASy (Swiss Institute of Bioinformatics, web.expasy.org/translate) was used to translate sequences from nucleotides into amino acids (Artimo *et al.* 2012, Gasteiger *et al.* 2005).

2.8 – In situ hybridisation

2.8.1 – In situ hybridisation probe production

In situ hybridisation probes were designed to span a 200-250 bp region of each *G. rostochiensis* effector candidate gene. Probes of the same approximate size were produced for cathepsin L proteins from *G. rostochiensis*, *G. pallida*, and *R. reniformis*

and GH53 enzymes from *G. rostochiensis* and *G. pallida*. Probe fragments were initially amplified using Taq polymerase PCR (Section 2.4) with sequenced plasmid DNA (pDNA) of each candidate effector as template. *In situ* hybridisation primers used for each candidate effector are detailed in Table 2.6. Digoxigenin labelled probes were produced using an asymmetrical PCR reaction (Section 2.4). These contained GoTaq 5x buffer (1x), MgCl₂ (2 mM), DIG-dNTP mix (10 mM), gene specific reverse or forward primer (10 µM) (for controls), GoTaq (0.05 U/µL) and 2 µL of template from the initial PCR. Gel electrophoresis was used to visualise probes alongside the initial PCR product on a 1.5% agarose gel in 1x Tris-borate-EDTA (TBE) buffer alongside the 100 bp DNA ladder (Promega).

Table 2.6 – In situ hybridisation primers for candidate effectors

Primers	Sequence	Primer Length (bp)	Tm (°c) TAQ	Probe length (bp)
GROS_g05682_F	ACCATGGTTTTAACGAATGATGGTCCGA	22	53.9	202
GROS_g05682_ISH_R	TTTGCACGCCGTTATTGC	18	55.3	
GROS_g01949_F	ACCATGGAGGATGACGACATTCCGATG	21	56.5	234
GROS_g01949_ISH_R	CCGTTACAGGAAGCCAATTGC	20	58.5	
GROS_g02469_F	ACCATGGACGCCGGTGGAAATGGAT	18	59.2	215
GROS_g02469_ISH_R	GTTCTAGTGGACACCCGACG	20	58.8	
GROS_g02470_F	ACCATGGTTATTAACGATTGCCAACCGTG	24	55.7	200
GROS_g02470_ISH_R	TGGCTTGTTGATGTAAAGCGAC	22	56.7	
GROS_g09987_F	ACCATGGAGGAGGACGAACGAATTAACG	22	56.7	250
GROS_g09987_ISH_R	GAAAGCATTCCGGCCCTGC	18	59	
GROS_g11017_F	ACCATGATTGGCTTCCATCCGGTG	19	56.4	250
GROS_g11017_ISH_R	GAATGGCACTGACCGAAGCT	20	58.9	
GROS_g07013_F	ACCATGGACATTCAAACGCAGTGAAAGG	23	55.8	250
GROS_g07013_ISH_R	ACACTTTATTGGTGGCACGA	20	54.9	
GROS_g05985_F	ACCATGGTTGGCAACAATCCCCG	17	55.3	250
GROS_g05985_ISH_R	TGAATTCATGTGCACCTCCG	21	56.5	
GROS_g02394_F	ACCATGGCCAAAGCGTTCAGCAGC	18	59.5	250
GROS_g02394_ISH_R	CCTTTGTTCCGATATTCTCTTTGACC	26	56.8	
GROS_g11020_F	ACCATGACTGGCATGCCAATGCAAAG	20	57.5	250
GROS_g11020_ISH_R	TCTGGTCCGCGAAGCG	16	59.6	
GROS_g04556_F	ACCATGATGGAACGCCGAAATCC	17	53.7	250
GROS_g04556_ISH_R	CGACAGTGACGAAACCG	17	54.1	
GROS_g02024_F	ACCATGCAGTCTTCAAATCGCGATGATGC	23	57.8	250
GROS_g02024_ISH_R	AATGAGCCGCCGTACC	17	59.8	
GROS_g09112_F	ACCATGGACCCTAAAAATCAGTTAGGATTTG	25	52.3	250
GROS_g09112_ISH_R	TTGCTTGTGAAAGTTCGTCC	20	54.2	

GROS_g09671_F	ACCATGGACTTGTCAACCCGACAA	20	56	250
GROS_g09671_ISH_R	TTGCTCGCACATACAGCTC	19	56.2	
GROS_g03615_F	ACCATGGCACCGACCGATCAACAG	18	57.4	242
GROS_g03615_ISH_R	GGTCGAAGCCCACAAATTC	19	55.4	
GROS_g04903_F	ACCATGGCTGTTTCCTGTAACCTTGA	20	53.1	226
GROS_g04903_ISH_R	TGTCCGAACAACCAACG	17	53.2	

2.8.2 – Fixation and dissection of nematodes

Nematodes were pelleted by centrifugation at 2500 x g for 10 minutes. Pelleted J2s were fixed in J2 fixative (2% paraformaldehyde in M9 buffer) and stored at 4 °C for 18 hours. Pelleted mixed parasitic nematodes were fixed in parasitic fixative (10% formaldehyde in M9 buffer) and stored at 4 °C for 18 hours. Fixed nematodes were washed and resuspended in 10X diluted fixative (J2 or parasitic fixative depending on life stage used). One hundred and fifty microlitres of nematodes per probe were pipetted onto a glass slide and cut into small sections using a sterile razor blade. Nematodes were assessed under a light microscope (Olympus) to confirm nematodes had each been cut into approximately 2 to 4 sections. Nematode sections were collected in a 1.5 ml Eppendorf tube in 10X diluted fixative.

2.8.3 – Permeabilization of nematode sections

* All following centrifugation steps were conducted at room temperature, 2 min, 8000 rpm unless otherwise specified.

J2 sections were washed twice in 1 ml M9 buffer. Sections were centrifuged to pellet, and the supernatant was removed. J2 sections were incubated in proteinase-K solution (0.5 mg/ml in M9 buffer) for 30 minutes on a rotator at 22 °C. Mixed parasitic sections were incubated in 0.5 ml proteinase-K solution (2 mg/ml in M9 buffer) for 90 minutes on a rotator at 37 °C. Sections were centrifuged to pellet, and the supernatant was removed. Sections were washed in 1 ml M9 buffer, centrifuged to pellet, and the supernatant was removed. The pellet was chilled for 15 minutes

at -20 °C, resuspended in cold methanol (-20 °C), and incubated for 30 seconds at room temperature. Nematodes were centrifuged to pellet at 13000 rpm for 30 seconds and the supernatant was removed. The pellet was resuspended in cold acetone (-20 °C) and incubated for 1 minute before being centrifuged to pellet at 13000 rpm for 1 minute. Acetone was removed until approximately one hundred microlitres was left in the tube. The nematode sections were rehydrated by adding 100 µl of diethyl pyrocarbonate (DEPC) treated ddH₂O (0.1% DEPC, 1 L double-distilled H₂O) which is RNase free. Sections were centrifuged to pellet and the supernatant was removed.

2.8.4 – Hybridisation of nematode sections

* All following centrifugation steps were conducted at room temperature, 2 min, 8000 rpm unless otherwise specified.

Nematode sections were washed in 500 µl hybridisation buffer (50% formamide, 4x saline-sodium citrate (SSC) buffer, 10% blocking reagent (Roche, description found in section 2.8.5), 2% sodium dodecyl sulphate (SDS), 1x Denhardt's, 1 mM EDTA, 20 µg/ml fish sperm DNA, 2.5 units yeast tRNA/ml) which was preheated to 50 °C. Sections were centrifuged to pellet, and the supernatant was removed. Nematode sections were resuspended in 150 µl of hybridisation buffer per hybridisation probe and distributed into separate 0.5 ml microcentrifuge tubes. Nematode sections were incubated in hybridisation buffer at 50 °C for 15 minutes. DIG-labelled DNA probes were heat-denatured at 100 °C for 10 minutes then cooled directly on ice. Probes (18 µl) were added to the nematode sections and hybridised rotating overnight at 50 °C.

Nematode sections were washed three times for 15 minutes in 4x SSC, rotating at 50 °C. Sections were centrifuged to pellet and the supernatant was removed. Sections were washed three times for 20 minutes in 0.1x SSC/0.1% SDS rotating at 50 °C. Sections were centrifuged to pellet and the supernatant was removed.

2.8.5 – Staining of nematode sections

* All following centrifugation steps were conducted at room temperature, 2 minutes, 8000 rpm unless otherwise specified.

Nematode sections were washed in maleic acid (MA) buffer (0.1 M maleic acid, 0.15 M NaCl, pH 7.5 with 5N NaOH). Sections were centrifuged to pellet, and the supernatant was removed. Nematodes were incubated for 30 minutes in 1X blocking reagent (made by diluting 10X blocking reagent (Roche) in MA buffer). Sections were centrifuged to pellet, and the supernatant was removed. Nematode sections were incubated for 2 hours in alkaline-phosphatase conjugated anti-digoxigenin antibody (150 U, Anti-Digoxigenin-AP, Fab fragments, Roche) diluted 1:1000 in 1X blocking reagent in MA buffer. The sections were washed three times for 15 minutes in MA buffer, washed briefly in alkaline phosphatase (AP) detection buffer (0.1 M Tris-HCl, 0.1 M NaCl, 50 mM MgCl₂.6H₂O, DEPC-ddH₂O, pH 9.5 with 5 N NaOH) and stained in Nitro Blue tetrazolium (NBT), X-phosphate (5-bromo-4-chloro-3-indolyl phosphate p-toluidine salt) staining solution (100 mg/ml NBT, 50 mg/ml X-phosphate, AP detection buffer) at 2-4 °C overnight. Staining was terminated by washing nematode sections twice in 0.01% Tween-20. Staining was observed and imaged using a Leica digital microscope large file storage (DM LFS) light microscope and an Axiocam 560 colour camera (Zeiss).

3. Identification of core effectors of syncytia-forming nematodes

3.1 – Introduction

Many economically important PPN species induce the formation of a syncytium as a feeding site in the roots of their host plants. Although a lot is known about the cellular and molecular changes that take place during syncytium formation, including details of changes in the cell cycle, cell wall, and cytoskeletal remodelling, less is known about how the nematodes initiate these changes in the host. Understanding effector proteins involved in producing a syncytium is important as it may present new avenues for nematode control by highlighting essential protein targets present in multiple nematode species, as well as providing information on how fundamental plant processes can be manipulated by pathogens.

3.1.1 – Effectors

As discussed in section 1.3, effectors can be defined as “pathogen proteins and small molecules that alter host cell structure and or function” (Hogenhout *et al.* 2009). The syncytium is produced from co-opted host plant cells which are induced to physically restructure, and it is known that effectors are involved in this process. Effectors with diverse biological functions have been identified from PPN including cell wall degrading enzymes that aid migration and effectors that suppress host defence responses (Geert Smant *et al.* 1998, Mei *et al.* 2015). Recently there have been novel effectors identified with functions that have been linked to syncytium formation or

maintenance. These include the 19C07, CLE peptides (Hg-4G12), and 30D08 effectors which have been identified in species of *Heteroderidae*.

3.1.1.1 – The Hs19C07 effector

The dorsal gland effector Hg19C07 from *H. glycines* and its ortholog Hs19C07 from *H. schachtii* are likely to influence the development of syncytia via control of auxin signalling (Gao *et al.* 2003, Lee *et al.* 2011). Auxin is an essential phytohormone which is involved in many processes in plant growth and development. Transport of auxin has been shown to have a critical role in the development of a syncytium. The inhibition or reduction of auxin transport can significantly impair development and produce abnormal syncytia, for example those with large, uncharacteristic galling due to disordered cell division (Goverse *et al.* 2000). The failure of *G. rostochiensis* and *H. schachtii* to develop syncytia was also observed in auxin-insensitive *A. thaliana* and tomato mutant lines (Goverse *et al.* 2000). Taken together, these results show that auxin plays a critical role in syncytium development.

Alongside the PIN auxin efflux transport proteins discussed in 1.2.2, there are also auxin influx transporters such as LAX1, LAX2, and LAX3 which are members of the AUX/LAX family. LAX3 has been shown to transport auxin in areas from which lateral root primordia emerge. Auxin influx induces many changes in gene expression including the controlled upregulation of cell wall degrading enzymes causing modification of the surrounding cell walls which allows the lateral root primordia to protrude (Swarup & Péret 2012). LAX3 from *A. thaliana* was identified as an interacting host protein of Hs19C07 by yeast two-hybrid analysis. LAX3 is upregulated in developing syncytia and expression of Hs19C07 in *A. thaliana* led to an accelerated emergence of the lateral root primordia. It is believed that Hs19C07 functions by increasing auxin influx via its interaction with LAX3, causing cell wall modifying enzymes to be active in root cells which aids in syncytium formation and development (Lee *et al.* 2011).

3.1.1.2 – CLE peptides

To establish a successful syncytium a nematode must disrupt the natural cross-talk between the cells of the host plant. CLAVATA3 (CLV3)/EMBRYO SURROUNDING REGION-related (CLE) peptides are small peptide hormones which are integral in signalling pathways for plant cell-to-cell communication (Yamaguchi *et al.* 2016). CLE peptides are classed as either A-type or B-type. A-type CLE functions include the maintenance and suppression of both root and shoot apical meristem (RAM/SAM) stem cell populations (Clark *et al.* 1995). In *A. thaliana* the A-type CLE peptide CLV3 is responsible for initiation of organs from stem cells, and *WUSCHEL* (*WUS*) controls stem cell identity (Schoof *et al.* 2000). CLV3 interacts with the CLV1/CLV2 receptors to restrict the size of the stem cell population. This is done by regulating the expression of the *WUS* gene through a negative feedback loop (Schoof *et al.* 2000). B-type CLEs have been shown to cause division of procambial cells of the meristem as well as repress the differentiation of procambial cells into tracheary elements (Whitford *et al.* 2008). PPN have been shown to produce effectors which mimic plant CLE peptides to influence host cell proliferation and differentiation.

The Hg-4G12 gene (HgCLE2) (formerly described as HgSYV46) was discovered in expressed sequence tags (EST) of transcripts expressed in the gland cells of *H. glycines* and has been shown to contain a CLE domain. *Clv3-1* mutant *A. thaliana* lines display an enlarged shoot apical meristem due to over-proliferation of stem cells, and they also produced more floral organs compared to wildtype lines (Clark *et al.* 1995). It has been shown that expression of HgCLE2 in *A. thaliana* can partially rescue the phenotype displayed by *clv3-1* mutant lines (Wang *et al.* 2005). Furthermore, in *A. thaliana* lines overexpressing CLV3, the apical meristem is halted from producing any organs after the initial leaves (Brand *et al.* 2000). The overexpression of HgCLE2 produces the same phenotype as AtCLV3 expression (Wang *et al.* 2005). HgCLE2 has been shown to have specific binding activity with both the CLV1, CLV2 and Receptor-like protein kinase 2 (RPK2) receptor orthologues from *A. thaliana* and *Glycine max* (Soybean) (Guo *et al.* 2015).

It was demonstrated that a homolog of HgCLE2, HsCLE2 is expressed in the dorsal gland of J3 stage *Heterodera schachtii* (Patel *et al.* 2008), and silencing of HgCLE2 caused a 36% or 40% reduction in production of females on transgenic *A. thaliana* lines through host-induced or J2 ingestion RNA interference (RNAi) assays respectively (Patel *et al.* 2008). HgCLE2 has also been shown to be secreted by the nematode into the cytoplasm of host cells in the syncytia through *in planta* immunolocalization studies on *A. thaliana* (Wang *et al.* 2010). It appears that HgCLE2 is then trafficked to function as a ligand mimic in the plant apoplast (Wang *et al.* 2021). Taken together, this suggests that these nematodes use the HgCLE2/HsCLE2 effectors to influence the host plants signalling pathway which maintains and regulates stem cell populations.

Nematode CLE peptides are still being discovered and the full extent of their function is yet to be elucidated. Most examples of CLE peptides from nematodes identified to date are A-type as B-type CLE peptides have only been described since 2017. The first B-type CLE-like effector from nematodes was identified in *H. glycines* and has been shown as having a similar function to the TDIF (tracheary element differentiation inhibitory factor) B-type CLE found in *A. thaliana* (Guo *et al.* 2017). It is currently unknown how the CLE peptides have evolved in nematode species, however the structural differences between plant and nematode CLEs suggests that it is not from a horizontal transfer event. The CLE peptides from nematodes contain variable domains (VDI and VDII). VDI has been shown to function in translocation of the CLE peptide to the host plant apoplast (Wang *et al.* 2021). These variable domains do not appear in host plant CLE structures. Due to the difference in structure, this is more likely to be an example of convergent evolution rather than horizontal gene transfer.

3.1.1.3 – The 30D08 effector

Manipulation of host gene expression is a critical part of syncytium formation. Changes in host gene expression may be induced due to manipulation of hormone

transport or by secretion of peptide hormone mimics, as described above. Root-knot nematodes have also been shown to secrete an effector that directly interacts with a SCARECROW transcription factor, suggesting a more direct route to manipulation of host gene expression (Huang *et al.* 2006). Cyst nematodes may also target the plant spliceosomal machinery to modulate host gene expression so this process may play a key role in syncytium formation. The 30D08 effector identified in *H. glycines* and *H. schachtii* is secreted from the dorsal gland of parasitic juveniles. Hs30D08 has been shown to interact with the *A. thaliana* protein AtSMU2 (suppressor of *mec-8* and *unc-52* 2), an auxiliary protein of the spliceosome. The spliceosome functions by removing introns from pre-mRNA which is pivotal in the control of gene expression and the production of splice variants (Chung *et al.* 2009). AtSMU2 is thought to interact with and splice pre-mRNA in conjunction with AtSMU1. These changes in splice patterns are thought to play key roles in plant development (Chung *et al.* 2009). 30D08 interacts with atSMU2 in yeast two-hybrid assays (Verma *et al.* 2018). A 25% decrease in the number of J4 females on *smu2-1* mutant lines was observed when compared to Col-0, wildtype *A. thaliana* while RNAi lines for 30D08 silencing showed a marked decrease in susceptibility to nematode infection. Further analysis of changes in gene expression suggested that 30D08 interacts with atSMU2 in order to alter the expression levels of gene clusters involved in syncytium formation (Verma *et al.* 2018).

3.1.2 – Genomes and transcriptomes of plant-parasitic nematodes

With many PPN species responsible for large agricultural economic losses it is important to study the methods used by nematodes to parasitise host plants. Effector proteins are critical in parasitism and therefore major efforts have been made to study these proteins in the last few decades. A key step in effector identification that enabled subsequent functional characterisation was the publication of genomic and transcriptomic data sets for a variety of PPN species. These give a detailed insight into what genes are present and how their expression

profiles change across the life cycle. This can be used to search and predict for genes with specific functions, including effectors.

Nematology has always been at the forefront of genome sequencing. In 1998 *C. elegans* was the first nematode, as well as the first multicellular organism to have its genome fully sequenced (The *C. elegans* Sequencing Consortium 1998) and improvements in technology have since allowed a broader range of nematodes to be studied. The first parasitic nematode genome to be sequenced was the human filarial nematode *Brugia malayi*, published in 2004 (Ghedini *et al.* 2004). The genomes and transcriptomes for the PPN species *M. incognita* and *M. hapla* followed shortly after, with data for both PCN species also now published (Abad *et al.* 2008, Opperman *et al.* 2008, Cotton *et al.* 2014, Eves-Van den Akker *et al.* 2016a). Transcriptome data are considerably less expensive to generate and analyse than most genomes and consequently have been generated from a very wide range of nematodes, including many PPN. The transcriptome of *N. aberrans* was published in 2014 and more recently the transcriptome and genome of *R. reniformis* were published in 2018 and 2019 respectively (Eves-van den Akker *et al.* 2014, Showmaker *et al.* 2018, Showmaker *et al.* 2019). Where RNAseq is available from multiple life stages of an organism detailed information about patterns of gene expression can be determined. In terms of effector identification, the availability of gene expression data enables categorisation of the genes expressed at the life stages where effector proteins are required. In PPN this ranges between the juvenile life stages to the adult female.

3.1.3 – Effector-associated sequence motifs

Identification of effectors from some plant pathogens is facilitated by the presence of sequence motifs that are associated with these proteins. For example, in oomycete pathogens such as the late blight pathogen *Phytophthora infestans*, a highly conserved motif – RXLR (Arg-X-Leu-Arg) – is present in most effectors which can be used for *in silico* effector prediction (Birch *et al.* 2006). The RXLR motif is often

followed by the s/dEER (Ser/Asp-Glu-Glu-Arg) motif ~25 amino acids downstream. These conserved motifs are required for host translocation (Rehmany *et al.* 2005, Whisson *et al.* 2007, Wawra *et al.* 2017). Bacterial effectors secreted into hosts via the Type 3 secretion system are frequently clustered in the genome and have a specific signal peptide associated with Type 3 secretion that can be predicted *in silico* (Toth *et al.* 2006). Identification of effectors from fungal pathogens is less straightforward, although many are small, cysteine rich proteins that function in the apoplast and which can be identified on this basis. By contrast, effectors in nematodes are intrinsically more difficult to identify compared to effectors of these other pathogens as there is no single defining feature that is associated with positive effector status that can be used for confident *in silico* prediction.

Effectors of PPN are restricted in terms of their expression profiles as they are specifically expressed in either the dorsal or subventral gland cells. This offers the prospect of using promoter motifs controlling expression in these tissues to identify effector candidates; a similar approach has been used in *C. elegans* for analysis of tissue specific expression in muscle (GuhaThakurta *et al.* 2004). As part of the analysis of the *G. rostochiensis* genome, the DOG Box motif (ATGCCA) was identified in the upstream promoter region of 77% of 101 known dorsal gland effectors. The dorsal gland effectors tested had an average of 2.54 DOG Box motifs present in the upstream promoter region compared to an average presence of 0.22 in non-effector promoter regions (Eves-Van den Akker *et al.* 2016a, Eves-van den Akker & Birch 2016). This allowed identification of many other putative dorsal gland effectors downstream of the DOG box motif identified elsewhere in the genome.

The process of using promoter regions to identify effector proteins has now been applied to other nematodes. Analysis of the promoter region of a subset of confirmed pharyngeal effector proteins from *Bursaphelenchus xylophilus* led to the identification of the STATAWAARS promoter motif. The presence of the STATAWAARS motif ([C|G]TAT[T|A][T|A]A[T|A][G|A][C|G]) in the 300bp upstream

promotor region of a gene was associated with an enrichment of secreted proteins and presence of the transcript in a cDNA library sequenced from purified gland cells (Espada *et al.* 2018). Effector genes from *B. xylophilus* have been shown to have between 1 and 6 copies of the STATAWAARS motif in their promotor region. The STATAWAARS motif has now been used to identify novel effector genes.

It has become clear that these motifs are by no means universal across all parasitic nematode species. It is evident above how different the STATAWAARS motif is from the DOG box. Current work has now identified a new motif (TGCACTT) from the root-knot nematode *Meloidogyne* genus termed the Mel-DOG (Meloidogyne dorsal gland) (Martine Da Rocha *et al.* 2021). To date versions of this motif have been found upstream of dorsal gland effectors in *Meloidogyne incognita*, *M. arenaria*, *M. javanica*, and *M. eterlobii*. The discovery of the DOG Box, STATAWAARS, and Mel-DOG motifs show that promotor analysis offers the prospect of rapid identification of effector genes in parasitic nematode species.

3.1.4 – In-silico pipelines for effector identification

In-silico approaches have been successfully used for identification of effectors of bacterial and fungal pathogens (McDermott *et al.* 2011, De Jonge 2012). An *in-silico* approach has also been used in many PPN in order to identify candidate effector genes. Using genome and transcriptome data, a list of genes is identified that meet specific criteria relating to what is known about the properties of likely effector proteins. Genome and transcriptome data can be obtained from whole nematode samples or from the isolation of the oesophageal gland cells of the nematode of choice (parasitome). The isolation of gland cell material can reduce the number of non-effector genes from which the candidate effector genes are filtered (Gao *et al.* 2003, Maier *et al.* 2013). The simplest pipeline may be to identify genes/protein sequences via BLAST searches using a query list of previously confirmed effectors from a related species. However, this will not reveal any new effectors. The genes predicted from a full genome sequence can also be used as input for analysis (*e.g.*,

Thorpe *et al.*, 2014), with the genes filtered to identify those proteins which are likely to be secreted by the nematode at parasitic stages.

3.1.4.1 – Signal peptides and transmembrane helices

A set of proteins, either predicted from a genome sequence or the output of a BLAST search, can be assessed for the presence of a signal peptide (SP) and the absence of transmembrane helices (TMH). The SP is a short amino acid sequence found at the N-terminus of proteins destined for secretion outside of the cell during or after translation. The SP is usually cleaved off the protein after translocation in eukaryotes by the signal peptidase complex (Nielsen 2017). When identifying potential effectors, screening genes for those with signal peptides is the first step as those proteins which do not contain a signal peptide are unlikely to be secreted out of the cell and therefore unlikely to function as effectors. The presence of transmembrane helices (TMH) indicate that a protein will be anchored within a cell membrane. As effector proteins are secreted out-with the cell, it is unlikely that candidate effectors would contain any TMH. Those genes that have been identified as having a signal peptide will be filtered to remove candidates that contain TMH. It should be noted that one RXLR effector from *P. infestans* that has a TMH has been identified, which interacts with a NAC transcription factor and prevents its relocalisation from the ER to the nucleus upon stress perception (McLellan *et al.* 2013).

3.1.4.2 – Gene expression analysis

The filtering steps described in 3.1.4.1 provide a list of all potentially secreted proteins and while this will include effectors it will also include many non-effectors, particularly if the starting dataset is a predicted proteome rather than a list of sequences similar to known effectors. A further step in filtering candidate effector proteins is therefore the analysis of gene expression data. Many non-effectors, including housekeeping proteins are likely to show constitutive expression across the lifecycle. By contrast, an effector gene is likely to be expressed most strongly at either the J2 life stage where the nematode is initially identifying and attacking a

plant host and/or during the parasitic stages where the nematode is actively feeding. Applying a filter for those genes upregulated at these life stages can therefore be used for further enrichment of a candidate effector list.

3.1.5 – Chapter aims

The aim of this chapter was to identify a set of candidate core effector proteins present in *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans*. It was hypothesised that such effectors will play a critical and conserved role in the formation and maintenance of a syncytial feeding site. A further aim was to identify a subset of these candidates as effectors for functional characterisation in a later chapter of this thesis (Chapter 6).

3.2 – Materials and methods

3.2.1 – In silico identification of effectors

3.2.1.1 – *G. rostochiensis* effector initial gene set

Effectors from *G. rostochiensis* have been the most extensively studied of the four species chosen for this analysis so this species was selected as a starting point for the research. A gene set of 295 *G. rostochiensis* genes was compiled from proteins previously described in the literature as effectors and genes identified as having an upstream dorsal gland promotor element motif (DOG Box motif) (Gao *et al.* 2001, Thorpe *et al.* 2014, S. Eves-van den Akker *et al.* 2014, Qin *et al.* 2004, Qin *et al.* 2000, Blanchard *et al.* 2007, Noon *et al.* 2015, Eves-Van den Akker *et al.* 2016a). Out of 295 effectors 156 in the initial list were identified as having at least one copy of the DOG Box motif sequence upstream of the coding sequence. The complete initial gene set can be found in supplementary file 1 of the appendix.

3.2.1.2 – BLAST similarity searches for orthologous proteins

The Basic local alignment search tool (BLAST) was used to identify orthologues of the initial effector genes in *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans*. BLAST was also used to analyse whether identified proteins were present in other nematode species e.g., RKN, as the aim was to identify effector candidates specific to syncytium forming nematodes. BLASTN and TBLASTN searches were conducted using BLAST V.2.4.0 with an e-value cut off of -0.00001. These similarity searches were conducted using the following as databases; the *G. rostochiensis* genome and transcriptome (Eves-Van den Akker *et al.* 2016a), the *R. reniformis* genome, gene calls (amino acid and nucleotide sequences), and transcriptome (Eves-Van den Akker *et al.* 2016b, Showmaker *et al.* 2019), the *G. pallida* genome, gene calls (both amino acid and nucleotide sequences), and transcriptome data from the J2 and 7dpi life stages (Cotton *et al.* 2014), and the *N. aberrans* transcriptome (Eves-van den Akker *et al.* 2014). *Script 1* (Sup. File 2) was used to produce a list of names of the genes present in the BLAST similarity search results. *Script 2* (Sup. File 3) or *Script 3* (Sup.

File 4) were used to extract the corresponding amino acid or nucleotide sequences for each gene from the source genome or transcriptome data.

Globodera rostochiensis query gene sequences were also used to identify similar genes present in non-redundant (nr) databases via the BLAST website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Madden 2003). These searches were also used to inform on potential functions for many of the orthologous groups identified across *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans*.

3.2.1.3 – Candidate sequence filtering

Candidate effectors from the BLAST results were filtered by the presence of a signal peptide (SP) using SignalP (Petersen *et al.* 2011). All candidate proteins lacking a signal peptide were removed at this stage of the pipeline. Candidate core effectors were assessed for the absence of any transmembrane helices (TMH) using TMHMM 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) (Krogh *et al.* 2001). All candidate proteins containing any TMH were removed at this stage of the pipeline. Alignments of orthologous groups of gene sequences were prepared using the software Jalview which employs MUSCLE to perform protein sequence alignments (Waterhouse *et al.* 2009, Edgar 2004). Colouration on alignment figures produced are based on the criteria from Clustal X multiple sequencing alignment software (Larkin *et al.* 2007).

Subcellular localisation of candidate effector genes was predicted using two different programs; WoLF PSORT (<https://wolfpsort.hgc.jp/>) and DEEPLOC (<http://www.cbs.dtu.dk/services/DeepLoc/>) (Horton *et al.* 2007, Almagro Armenteros *et al.* 2017). In addition, subcellular localisations *in planta* were predicted for candidate core effectors using LOCALIZER (<http://localizer.csiro.au/>) (Sperschneider *et al.* 2017). Conserved domain searches were conducted using the National Centre for Biotechnology Information Conserved Domain Database (NCBI CDD) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (Marchler-Bauer *et al.* 2015). A schematic representation of the *in-silico* pipeline can be found in Figure

3.1. Smaller protein motif queries were conducted using the MOTIF search by GenomeNet (<https://www.genome.jp/tools/motif/>) (Kanehisa *et al.* 2002).

3.2.1.4 – Gene expression analysis

Normalised expression data for *G. rostochiensis* was obtained from the supplementary data of the genome sequencing paper, additional file 7, file S1 (Eves-Van den Akker *et al.* 2016a). Normalised expression data for *G. pallida*, *R. reniformis*, and *N. aberrans* was obtained from Dr S. Eves-van den Akker.

3.2.1.5 – DOG Box motif analysis

The presence of the DOG box motif (5' A [A|T|G] GCCA 3') and reverse complement (3' TGGC [C|A|T] T 5') sequences were examined in the promotor region, which is the 500 base pairs found immediately upstream of the start codon of each candidate core effector gene from *G. rostochiensis*. The canonical DOG Box motif is ATGCCA, however there is variation at the second base where the thymine can be replaced by an adenine or guanine making the motif (5' A [A|T|G] GCCA 3'). All variants of the DOG Box motif were searched for along with the relevant reverse complement sequences.

3.2.2 – Cloning and analysis of candidate core effector genes

The coding regions of *G. rostochiensis* candidate effector genes identified through the pipeline described above were amplified by PCR from cDNA using KOD Hot start DNA polymerase (Merck) as detailed in section 2.4. Effectors were amplified using the forward (F) and reverse (R) or forward/reverse-no stop codon (_Rns in table) primer pairs detailed in Table 3.1. Effectors were amplified using the _Rns primers which contain no stop codon in order to subsequent cloning into vectors with C-terminal tags e.g., GFP. Cloning and sequencing of amplified genes encoding effectors was carried out as described in sections 2.5 to 2.7. Expression profiles of candidate effectors were examined by *in situ* hybridisation as described in section

2.8, with probes amplified using the forward and *in situ* hybridisation reverse (ISH_R) primers detailed in Table 3.1.

Table 3.1 - Primers used for cloning and in situ hybridisation of *G. rostochiensis* candidate effectors

Candidate orthologous group	<i>G. rostochiensis</i> gene name	Primer name	Sequence	Primer function	Length of gene (bp)	<i>In situ</i> hybridisation probe length (bp)
20_E_03	GROS_g05682	GROS_g05682_F	ACCATGGTTTTAACGAATGATGGTCCGA	Cloning / <i>In situ</i> hybridisation	909	202
		GROS_g05682_Rns	AGCGCAATAGGGCCAGA	Cloning		
		GROS_g05682_R	TCAAGCGCAATAGGGCC	Cloning		
		GROS_g05682_ISH_R	TTTGCACGCCGTTATTGC	<i>In situ</i> hybridisation		
Gro_DOG_0116	GROS_g01949	GROS_g01949_F	ACCATGGAGGATGACGACATTCCGATG	Cloning / <i>In situ</i> hybridisation	1056	234
		GROS_g01949_Rns	GAGTTCGTCTCTCGTCTC	Cloning		
		GROS_g01949_R	TCAGAGTTCGTCTCTCGTCG	Cloning		
		GROS_g01949_ISH_R	CCGTTCAAGCAATGTC	<i>In situ</i> hybridisation		
G23G11/GLAND15	GROS_g02469	GROS_g02469_F	ACCATGGACCCGGTGAATGGAT	Cloning / <i>In situ</i> hybridisation	1416	215
		GROS_g02469_Rns	AGCCTTCCGTCAAGCTTTCC	Cloning		
		GROS_g02469_R	TTAAGCCTTCCGTCAAGCTTTC	Cloning		
		GROS_g02469_ISH_R	GTTCTAGTGGACACCCGACG	<i>In situ</i> hybridisation		
Gro_DOG_0043	GROS_g09987	GROS_g09987_F	ACCATGGAGGAGGACGAACGAATTAACG	Cloning / <i>In situ</i> hybridisation	342	250
		GROS_g09987_Rns	TTGGGCCATCACCAATTGC	Cloning		
		GROS_g09987_R	TCATTGGGCCATCACC	Cloning		
		GROS_g09987_ISH_R	GAAAGCATTGCGCCCTGC	<i>In situ</i> hybridisation		
Gro_DOG_0203	GROS_g11017	GROS_g11017_F	ACCATGATTGGCTTTCCATCCGGTG	Cloning / <i>In situ</i> hybridisation	426	250
		GROS_g11017_Rns	CGCGCTCAGTGCAACG	Cloning		
		GROS_g11017_R	TCACGCGCTCAGTGCAAC	Cloning		
		GROS_g11017_ISH_R	GAATGGCACTGACCGAAGCT	<i>In situ</i> hybridisation		
Gro_DOG_0199	GROS_g07013	GROS_g07013_F	ACCATGGACATTCAAAACGCAGTGAAAGG	Cloning / <i>In situ</i> hybridisation	648	250
		GROS_g07013_Rns	GTAGCCGGCCATTCCGG	Cloning		
		GROS_g07013_R	TCAGTAGCCGGCCATTCCG	Cloning		

		GROS_g07013_ISH_R	ACACTTTATTGGTGGCACGA	<i>In situ</i> hybridisation		
Gro_DOG_0149	GROS_g05985	GROS_g05985_F	ACCATGGTTGGCAACAATCCCCG	Cloning / <i>In situ</i> hybridisation	1509	250
		GROS_g05985_Rns	ATGATGAATTTGTCGGGAGTTTGT	Cloning		
		GROS_g05985_R	TCAATGATGAATTTGTCGGGAGT	Cloning		
		GROS_g05985_ISH_R	TGAATTCATGTGCACCTTCCG	<i>In situ</i> hybridisation		
GLAND11	GROS_g02394	GROS_g02394_F	ACCATG GCCAAAGCGTTCAGCAGC	Cloning / <i>In situ</i> hybridisation	1815	250
		GROS_g02394_Rns	AATGCTCTCCAGCGAATTAAGA	Cloning		
		GROS_g02394_R	TCAAATGCTCTCCAGCGAA	Cloning		
		GROS_g02394_ISH_R	CCTTTGTTCCGATATTCTCTTGACC	<i>In situ</i> hybridisation		
Gro_DOG_0017	GROS_g11020	GROS_g11020_F	ACCATGACTGGCATGCCAATGCAAAG	Cloning / <i>In situ</i> hybridisation	396	250
		GROS_g11020_Rns	CGCGCTCAGCGCAAC	Cloning		
		GROS_g11020_R	TTACGCGCTCAGCGCAAC	Cloning		
		GROS_g11020_ISH_R	TCTGGTCCGGAAGCG	<i>In situ</i> hybridisation		
Gro_DOG_0015	GROS_g04556	GROS_g04556_F	ACCATGATGGAACGCCGAAATCC	Cloning / <i>In situ</i> hybridisation	456	250
		GROS_g04556_Rns	CAGCACTCTTGTTAGATGTTCA	Cloning		
		GROS_g04556_R	TCACAGCACTCTTGTTAGATGT	Cloning		
		GROS_g04556_ISH_R	CGACAGTGACGAAACCG	<i>In situ</i> hybridisation		
Gro_DOG_0028	GROS_g02024	GROS_g02024_F	ACCATGCAGTCTTCAAATCGCGATGATGC	Cloning / <i>In situ</i> hybridisation	822	250
		GROS_g02024_Rns	TTCACTTTTTGCTCTTTCAATGCA	Cloning		
		GROS_g02024_R	TCATTCACTTTTTGCTCTTTCAATGC	Cloning		
		GROS_g02024_ISH_R	AATGAGCCGCCGTCACC	<i>In situ</i> hybridisation		
Gro_DOG_0169	GROS_g09112	GROS_g09112_F	ACCATGGACCCTAAAAATCAGTTAGGATTTG	Cloning / <i>In situ</i> hybridisation	1053	250
		GROS_g09112_Rns	CAAAAATTTTCATGAATTGCATATATGT	Cloning		
		GROS_g09112_R	TCACAAAAATTTTCATGAATTGCATAT	Cloning		
		GROS_g09112_ISH_R	TTGCTTGTGAAAGTTCGTCC	<i>In situ</i> hybridisation		
Gro_DOG_0187	GROS_g09671	GROS_g09671_F	ACCATGGACTGTGCAACCCGACAA	Cloning / <i>In situ</i> hybridisation	1191	250

		GROS_g09671_Rns	AAGAACCCCTTTCCCATTTGC	Cloning		
		GROS_g09671_R	TTAAAGAACCCCTTTCCCATTTG	Cloning		
		GROS_g09671_ISH_R	TTGCTCGCACATACAGCTC	<i>In situ</i> hybridisation		
Gro_DOG_0201	GROS_g03615	GROS_g03615_F	ACCATGGCACCGACCGATCAACAG	Cloning / <i>In situ</i> hybridisation	957	242
		GROS_g03615_Rns	AACAACGGGATATGATGCCA	Cloning		
		GROS_g03615_R	TTAAACAACGGGATATGATGCCA	Cloning		
		GROS_g03615_ISH_R	GGTCGAAGCCCACAAATTC	<i>In situ</i> hybridisation		
Gro_DOG_0200	GROS_g04903	GROS_g04903_F	ACCATGGCTGTTTCCTGTAACCTTGA	Cloning / <i>In situ</i> hybridisation	366	226
		GROS_g04903_Rns	CAGAAGTATGAAGAGGCCG	Cloning		
		GROS_g04903_R	TCACAGAAGTATGAAGAGGCC	Cloning		
		GROS_g04903_ISH_R	TGTCCGAACAACCAACG	<i>In situ</i> hybridisation		

3.3 – Results

In order to identify a subset of core effectors in syncytia-forming nematode species a gene set of 295 *G. rostochiensis* genes was compiled from proteins previously described in the literature as effectors. These genes from *G. rostochiensis* formed a list of query sequences for BLAST similarity searches. The BLAST similarity searches and subsequent pipeline identified a total of 1455 candidate effector genes present in *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans* (Figure 3.1). Comparisons of the presence/absence of these sequences across all four species were performed and are summarised in Table 3.2. All genes included at this stage of analysis were sorted into groups (also referred to as families). Each group contains genes from each of the four species analysed which share high sequence similarity to the one of the original *G. rostochiensis* query gene sequences. As expected, the species that shared the most orthologous groups of genes (65 groups, 33.2% of groups identified) between them were the cyst nematodes, *G. rostochiensis* and *G. pallida*, due to them being the most closely related species included in the analysis. There were 31 groups (15.8%) of genes shared between *G. rostochiensis*, *G. pallida*, and *R. reniformis* which were missing orthologues in *N. aberrans*. This was also expected as *N. aberrans* is the most genetically dissimilar of the four species. This analysis initially revealed that 37 groups of orthologous candidate effector genes were present in all four species (Table 3.2).

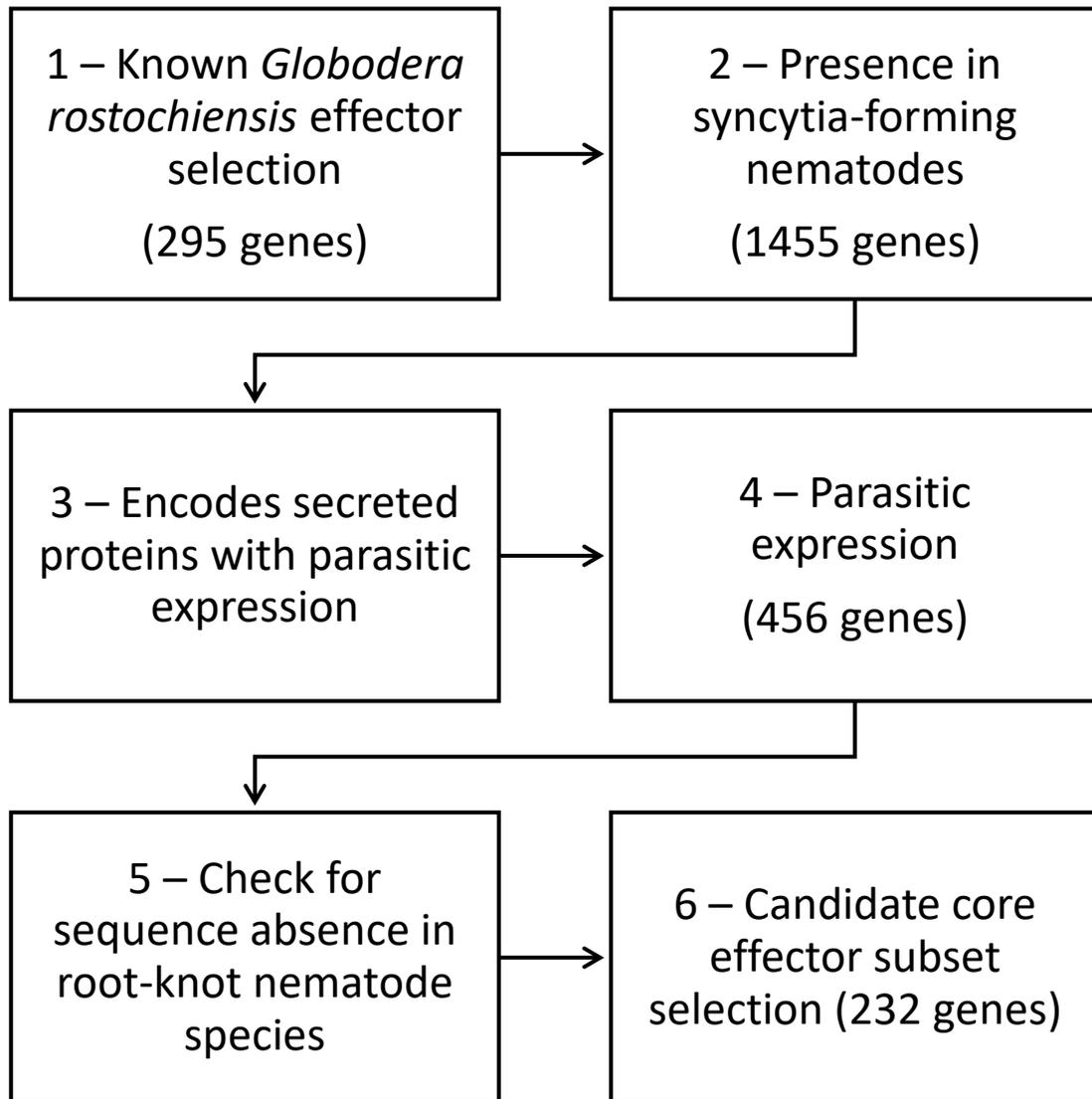


Figure 3.1 - Stages of the in-silico pipeline used for candidate core syncytial effector identification. Stage 1 – An initial set of 295 *G. rostochiensis* genes were selected as a basis for sequence comparison analysis. Stage 2 – BLAST analysis to determine the presence of 1455 orthologous genes in the syncytia-forming species *G. pallida*, *R. reniformis*, and *N. aberrans*. Stage 3 – Candidate gene sequences were filtered to select for genes containing a signal peptide and the absence of transmembrane domains. Stage 4 – Candidates were filtered to select for genes with expression during parasitic life stages. At this stage candidates with previously well studied functions were also filtered out to give a total of 456 genes. Stage 5 – Candidates were filtered for the absence of orthologs in non-syncytia-forming nematode species (root-knot nematodes). Stage 6 – A subset of candidate effector gene groups containing a total of 232 genes across the four syncytia-forming nematodes species were produced for confirmation in *G. rostochiensis* via *in situ* hybridisation.

Originally there was an additional column in Table 3.2 titled “Orthologous groups in PCN and *N. aberrans*”. It was considered very unlikely that these groups did not have an orthologous gene present in *R. reniformis*, as this would imply multiple cases of specific gene loss within *R. reniformis* given the phylogenetic relationships between the nematodes. For example, members of the pectate lyase group are found across many nematode species and so the absence in *R. reniformis* would imply an improbable gene loss. It was subsequently determined that *R. reniformis* genes from the 6 orthologous groups present in the additional column were originally missed due to the stringency used when conducting the initial BLAST searches. By dropping the E-value from 1e-5 to 1e-4, orthologues from *R. reniformis* were identified. These additional 6 orthologous groups (Pectate lyase, Gro_DOG_0057, Gro_DOG_0067, Gro_DOG_0073, Gro_DOG_0079, Gro_DOG_0211) can be seen highlighted in green in Table 3.2. Subsequent results in this chapter continue to reference the 37 orthologous groups as these were the original results identified at the time of conducting the pipeline, but the subsequent discovery of the *R. reniformis* gene sequences takes this to a total of 43 groups (21.9%) (Table 3.2). These additional six groups were not assessed or included in lab analysis but may pose an avenue for further work in the future. It should also be noted that the original total of 1455 genes identified across the four species in the second stage of the pipeline (Figure 3.1) was taken to 1480 with the addition of the 25 new *R. reniformis* gene sequences identified across the 6 additional groups.

The remaining 29.1% of groups (not included in Table 3.2) were sequences only found in *G. rostochiensis* with no homologs in other species identified. These were not analysed any further in this thesis, however it appears odd to have so many gene sequences that were not found in *G. pallida*. It is suspected that if this group was investigated more closely in future there would turn out to be genes in this group that would actually be in the PCN section.

As part of the *in silico* pipeline any gene sequences lacking a signal peptide or that had a transmembrane domain were removed. In addition, candidates were removed if their expression patterns did not align with the life stages where effector proteins are expressed and secreted. Candidates with expression patterns that were exclusively high at the cyst, egg, late sedentary female, or male life stages were removed. Candidates displaying constitutive expression across all life stages were also removed. The resulting 37 orthologous groups were therefore candidate core effectors present in all four syncytia-forming species that are targeted for secretion, and that show elevated expression at parasitic life stages (J2 and adult females). Each group was given a name derived from the initial *G. rostochiensis* query gene sequence (Table 3.3). The initial candidate orthologous groups were named 20E03, Gro_DOG_0116, G23G11, Gro_DOG_0043, Gro_DOG_0203, Gro_DOG_0199, Gro_DOG_0049, Gro_DOG_0149, GLAND11, Gro_DOG_0017, Gro_DOG_0031, Gro_DOG_0015, Gro_DOG_0028, Gro_DOG_0169, Gro_DOG_0187, Gro_DOG_0201, Gro_DOG_0200, Gro_DOG_0069, Gro_DOG_0197, Gro_DOG_0213, Gro_DOG_0218, Gro_DOG_0221, Gro_DOG_0002, Gro_DOG_0003, DG_Pioneer_from_GPLIN_000834600, Gro_DOG_0154, Gro_DOG_0027, Gro_DOG_0156, 10A06, VAP (Venom Allergen-like proteins), β -1,4-endoglucanase, chitinase, CM (chorismate mutase), GLAND10, GrEXP (expansins), Gro_DOG_0217, and Gro_DOG_0220.

Table 3.2 - Orthologous groups shared between syncytia-forming species of interest

	Orthologous groups in potato cyst nematodes (PCN) only (33.2% of groups identified)	Orthologous groups in PCN and <i>R. reniformis</i> only (15.8% of groups identified)	Orthologous groups present in PCN, <i>R. reniformis</i> , and <i>N. aberrans</i> (21.9% of groups identified)
1	19C07	10C02	10A06
2	20_E_12	24A12	20E03
3	Gr1106/32E03	33A09	Venom allergen-like proteins (VAP)
4	4G05/30G12/25A01	3H07_Ubiquitin_extension	β -1,4-endoglucanase
5	A4	4D06	Chitinase
6	E9	CLE	Chorismate mutase (CM)
7	HYP	GLAND1	GLAND10
8	IA7	GLAND12	GLAND11
9	IVG9	GLAND3	G23G11
10	SVG_pioneer_from_GPLIN_000333000	Glutathione Synthase (GS)	Expansins (GrEXP)
11	Gro_DOG_0001	Invertase (INV)	DG_pioneer_from_GPLIN_000834600
12	Gro_DOG_0005	SPRY-SEC	Gro_DOG_0002
13	Gro_DOG_0008	Gro_DOG_0004	Gro_DOG_0003
14	Gro_DOG_0019	Gro_DOG_0011	Gro_DOG_0015
15	Gro_DOG_0022	Gro_DOG_0030	Gro_DOG_0017
16	Gro_DOG_0023	Gro_DOG_0045	Gro_DOG_0027
17	Gro_DOG_0024	Gro_DOG_0047	Gro_DOG_0028
18	Gro_DOG_0029	Gro_DOG_0074	Gro_DOG_0031
19	Gro_DOG_0036	Gro_DOG_0075	Gro_DOG_0043
20	Gro_DOG_0038	Gro_DOG_0093	Gro_DOG_0049
21	Gro_DOG_0039	Gro_DOG_0096	Gro_DOG_0069

22	Gro_DOG_0040	Gro_DOG_0108	Gro_DOG_0116
23	Gro_DOG_0078	Gro_DOG_0110	Gro_DOG_0149
24	Gro_DOG_0080	Gro_DOG_0133	Gro_DOG_0154
25	Gro_DOG_0089	Gro_DOG_0142	Gro_DOG_0156
26	Gro_DOG_0094	Gro_DOG_0143	Gro_DOG_0169
27	Gro_DOG_0098	Gro_DOG_0145	Gro_DOG_0187
28	Gro_DOG_0099	Gro_DOG_0146	Gro_DOG_0197
29	Gro_DOG_0101	Gro_DOG_0177	Gro_DOG_0199
30	Gro_DOG_0107	Gro_DOG_0178	Gro_DOG_0200
31	Gro_DOG_0109	Gro_DOG_0185	Gro_DOG_0201
32	Gro_DOG_0113		Gro_DOG_0203
33	Gro_DOG_0115		Gro_DOG_0213
34	Gro_DOG_0117		Gro_DOG_0217
35	Gro_DOG_0120		Gro_DOG_0218
36	Gro_DOG_0121		Gro_DOG_0220
37	Gro_DOG_0122		Gro_DOG_0221
38	Gro_DOG_0124		Pectate lyase
39	Gro_DOG_0127		Gro_DOG_0057
40	Gro_DOG_0129		Gro_DOG_0067
41	Gro_DOG_0131		Gro_DOG_0073
42	Gro_DOG_0134		Gro_DOG_0079
43	Gro_DOG_0141		Gro_DOG_0211
44	Gro_DOG_0147		
45	Gro_DOG_0148		
46	Gro_DOG_0151		
47	Gro_DOG_0152		

48	Gro_DOG_0153		
49	Gro_DOG_0157		
50	Gro_DOG_0159		
51	Gro_DOG_0160		
52	Gro_DOG_0161		
53	Gro_DOG_0162		
54	Gro_DOG_0163		
55	Gro_DOG_0166		
56	Gro_DOG_0168		
57	Gro_DOG_0170		
58	Gro_DOG_0172		
59	Gro_DOG_0173		
60	Gro_DOG_0174		
61	Gro_DOG_0195		
62	Gro_DOG_0196		
63	Gro_DOG_0206		
64	Gro_DOG_0207		
65	Gro_DOG_0216		

*Groups highlighted in green are those additional 6 orthologous groups missed in the original BLAST searches due to stringency of E-value used

3.3.1 – BLAST – functional similarity searches

The potential functions of the 37 orthologous groups were initially investigated by carrying out BLAST similarity searches against the non-redundant (nr) database. Following on from this, six groups were excluded from further analysis. These groups were identified as containing genes with functions that have already been categorised extensively and therefore did not present viable candidates for study during this project (Hewezi *et al.* 2010, Lozano-Torres *et al.* 2014, Wubben *et al.* 2010). The groups discarded at this time were 10A06, Venom Allergen-like proteins (VAPs), beta-1,4-endoglucanase, chitinases, chorismate mutase (CM), and expansins. As a result of this filtering the 37 groups were reduced to 31 groups (consisting of 952 genes in total).

3.3.2 – BLAST similarity searches against Root-knot nematodes

The remaining 31 orthologous groups were subsequently used in BLAST similarity searches to determine whether any orthologues were present in root-knot nematodes. Root-knot nematodes also form feeding sites in the host plant roots but these feeding sites – giant cells – are different in structure and ontogeny to the syncytium. The rationale behind this project was to identify the presence of core effectors that are required for creating a syncytium between phylogenetically distinct syncytia-forming nematode species. It follows that if orthologues of candidate effectors were present in RKN then they are unlikely to be required specifically for syncytium formation. This analysis allowed two other groups of orthologues with matches in RKN to be removed; Gro_DOG_0031 was similar to transthyretin-like proteins from *Meloidogyne javanica*, and GLAND10 was removed due to similarity to sequences from both *M. javanica* and *M. incognita*.

Further filtering was performed on the remaining 29 orthologous groups. First, genes were excluded that were expressed at very low levels across the life cycle as there may have been issues with cloning such genes. Finally, genes were chosen that displayed high levels of sequence conservation among the four species, as this was

thought to be indicative of a likely conserved function. This led to a final list of 15 groups of candidate effector gene families (Table 3.3) that contained 232 genes in total across the four species. At this stage of the analysis the decision was taken to work predominantly on the initial *G. rostochiensis* representatives from each of the 15 effector gene families detailed in Table 3.3.

Table 3.3 – Top 15 orthologous candidate effector groups put forward for confirmation and functional characterisation

Candidate orthologous group name	<i>G. rostochiensis</i> gene	<i>G. rostochiensis</i> BLAST hits	<i>G. pallida</i> BLAST hits	<i>R. reniformis</i> BLAST hits	<i>N. aberrans</i> BLAST hits	BLASTp (Nr) top hits
20E03	GROS_g05682	None	GPLIN_000926600 GPLIN_000662500 GPLIN_000962200	g30282.t1 g30282.t2 g33065.t1 g21202.t1	cds.Nab_22156_c0_seq1_m.16877	G20E03 (<i>H. glycines</i>) & hits against a transcriptional regulator in multiple streptomyces species
Gro_DOG_0116	GROS_g01949	GROS_g10784	GPLIN_000806500 GPLIN_000344300	g12045.t1	cds.Nab_28487_c0_seq1_m.28420	NO HIT - PIONEER
GLAND15/G23G11*	GROS_g02469	GROS_g02470	GPLIN_000763000	g951.t1 g25265.t1 g25264.t1 g952.t1 g33471.t1	cds.Nab_57241_c0_seq1_m.50052 cds.Nab_57240_c0_seq1_m.50041 cds.Nab_57239_c0_seq1_m.50031 cds.Nab_57238_c0_seq1_m.50020 cds.Nab_57237_c0_seq1_m.50010 cds.Nab_28879_c0_seq1_m.29300	Putative gland protein G23G11 / Oesophageal gland-localised secretory protein 15 (GLAND15) both <i>H. glycines</i>
Gro_DOG_0043	GROS_g09987	GROS_g09990 GROS_g05703	GPLIN_001140700 GPLIN_001140900 GPLIN_000476700 GPLIN_000919800	g27892.t1 g14819.t1 g3594.t1	cds.Nab_21641_c0_seq1_m.16185	Uncharacterised in many worms / ADP-ribose pyrophosphatase mitochondrial (Trichinella)
Gro_DOG_0203**	GROS_g11017	None	GPLIN_000183000 GPLIN_000182800 GPLIN_000182900 GPLIN_001020100 GPLIN_000183200 GPLIN_000183400 GPLIN_000183300	g14106.t1 g19010.t1 g30766.t1 g19878.t1 g29037.t1	cds.Nab_23010_c0_seq1_m.18132 cds.Nab_25727_c0_seq1_m.22943	Hit with a hypothetical protein from <i>Ruegeria atlantica</i>

			GPLIN_000713100 GPLIN_001396600 GPLIN_000225800 GPLIN_000968000 GPLIN_001043700			
Gro_DOG_0199	GROS_g07013	None	GPLIN_001036900	g30503.t1	cds.Nab_17897_c0_seq1_m.12768	Wrt-10 (warthog protein 10) (<i>A. suum</i> , <i>Loa</i> , <i>C. elegans</i>)
Gro_DOG_0149	GROS_g05985	GROS_g13738 GROS_g09111 GROS_g09869 GROS_g03837 GROS_g12421	GPLIN_001355500	g21345.t1 g21349.t1 g17738.t1	cds.Nab_59091_c0_seq1_m.56063	Phosphatidylinositol 4- phosphate 5-kinase type-1 alpha-like (<i>Fundulus heteroclitus</i>)
GLAND11	GROS_g02394	None	GPLIN_000714100	g22583.t1 g26304.t1 g24562.t1	cds.Nab_59725_c0_seq1_m.58226 cds.Nab_59727_c0_seq1_m.58236 cds.Nab_59726_c0_seq1_m.58231	Oesophageal gland-localised secretory protein 11 (GLAND11) (<i>H. glycines</i>)
Gro_DOG_0017**	GROS_g11020	None	GPLIN_000182900 GPLIN_001020100 GPLIN_000182800 GPLIN_000183000 GPLIN_000183200 GPLIN_000183400 GPLIN_000183300 GPLIN_000713100 GPLIN_001396600 GPLIN_000225800 GPLIN_000968000 GPLIN_001043700	g14106.t1 g30766.t1 g19878.t1 g29037.t1 g19010.t1	cds.Nab_23010_c0_seq1_m.18132 cds.Nab_25727_c0_seq1_m.22943	NO HIT - PIONEER

Gro_DOG_0015	GROS_g04556	GROS_g10456 GROS_g05089	GPLIN_001083100 GPLIN_000258300 GPLIN_001268900	g35960.t1 g20182.t1 g30627.t1 g16213.t1	cds.Nab_23179_c0_seq1_m.18402	Hypothetical proteins containing nitrobindin heme-binding domain
Gro_DOG_0028	GROS_g02024	GROS_g10752 GROS_g07487 GROS_g02955	GPLIN_001456600	g30388.t1 g14062.t1 g13920.t1 g33880.t1 g1805.t1 g20430.t2 g1119.t1	cds.Nab_23130_c0_seq1_m.18312	Serine proteinase <i>M. incognita</i> & trypsin <i>D. viviparus</i>
Gro_DOG_0169	GROS_g09112	GROS_g09110 GROS_g13625 GROS_g13626 GROS_g07878 GROS_g05684 GROS_g02733 GROS_g01721 GROS_g13400	GPLIN_001355500 GPLIN_000962700	g33126.t1 g21345.t1 g34102.t1	cds.Nab_29527_c0_seq1_m.30990 cds.Nab_29366_c0_seq1_m.30587 cds.Nab_58262_c0_seq1_m.53270	Predicted: tartrate-resistant acid phosphatase type 5 (<i>Austrofundulus limnaeus</i> , <i>Stegastes partitus</i>)
Gro_DOG_0187	GROS_g09671	GROS_g03808 GROS_g13673 GROS_g02007 GROS_g01693 GROS_g09164 GROS_g12317 GROS_g05767 GROS_g02879	GPLIN_001406900 GPLIN_001195200 GPLIN_000618300 GPLIN_000508200	g34866.t1 g29945.t1 g598.t1 g24286.t1 g16589.t1 g17037.t1	cds.Nab_27332_c0_seq1_m.25942 cds.Nab_24620_c0_seq1_m.20957	Acid phosphatase (<i>H.</i> <i>avenae</i>)

Gro_DOG_0201	GROS_g03615	GROS_g03960 GROS_g06676 GROS_g11484 GROS_g02883 GROS_g09263 GROS_g06777 GROS_g10895 GROS_g11524 GROS_g12051 GROS_g13246 GROS_g07942 GROS_g05847 GROS_g08721	GPLIN_001009100 GPLIN_000168400 GPLIN_001099500 GPLIN_000983400 GPLIN_000402300	g18199.t1 g25539.t1 g14349.t1 g2797.t1 g19341.t1 g23306.t1 g17659.t1 g12778.t1 g32695.t1 g30861.t1 g21071.t1 g611.t1 g19307.t1 g19694.t1 g30861.t2	cds.Nab_32144_c0_seq1_m.35269 cds.Nab_37164_c0_seq1_m.36703 cds.Nab_59225_c0_seq1_m.56546 cds.Nab_11237_c0_seq1_m.9474 cds.Nab_15927_c0_seq1_m.11543 cds.Nab_28809_c0_seq1_m.29088 cds.Nab_23144_c0_seq1_m.18337 cds.Nab_21101_c0_seq1_m.15529 cds.Nab_25601_c0_seq1_m.22724 cds.Nab_19861_c0_seq1_m.14190 cds.Nab_28125_c0_seq1_m.27586	Cathepsin S-like cysteine proteinase (<i>H. glycines</i> , <i>C. elegans</i>)
Gro_DOG_0200	GROS_g04903	GROS_g08379 GROS_g08377 GROS_g08380 GROS_g08646 GROS_g08378 GROS_g08649 GROS_g13812	GPLIN_000913700 GPLIN_000913800 GPLIN_000913500 GPLIN_000562200 GPLIN_000158600 GPLIN_000913300 GPLIN_000913400 GPLIN_000310800 GPLIN_000922300	g1257.t1 g18097.t1 g14551.t1 g11381.t1	cds.Nab_26040_c0_seq1_m.23527	Hypothetical proteins (<i>Caenorhabditis latens</i> , <i>C. remanei</i>)

* GROS_g02469 was originally identified and labelled as a member of the GLAND15 orthologous group due to its sequence similarity with the GLAND15 effector identified by Noon *et al.* (Noon *et al.* 2015). However, when GROS_g02469 was used as a query sequence in subsequent BLAST searches, it became apparent that it has a much higher percentage identity (ID) (69%) with the putative gland protein G23G11 from *H. glycines* than GLAND15. Therefore, the orthologous group is termed G23G11 from this point onwards.

3.3.3 – Protein sequence alignments and RNA-seq analysis

All alignments to accompany the effector gene families discussed in Table 3.3 can be found in supplementary file 5. The alignment for each family was assessed in case it could inform on function. Some of the alignments gave unexpected results, such as large insertions. A manual screen of RNA-seq data was therefore undertaken to determine their authenticity.

3.3.3.1 - 20E03

The orthologous sequences which make up the 20E03 group are well conserved except for a region of amino acids which are only found in the *G. rostochiensis* protein sequence GROS_g05682 (residues 45-161) and the *G. pallida* protein sequence GPLIN_000962200 (residues 41-160) (red boxes in Figure 3.2A). It was first thought that these regions had been included due to the misprediction of an intron/exon boundary, however when RNAseq data was manually assessed these regions appear to be real and not a misprediction by software. This was further confirmed to be the true amino acid sequence after GROS_g05682 was cloned. When this region of the gene was examined independently, it was demonstrated that there is 64.2 percentage identity between *G. rostochiensis* and *G. pallida*. BLAST searches of this region showed no significant similarity with other proteins and no conserved domains were present. At present the significance of this additional section of gene remains unknown.

There were no domains identified in the 20E03 family of effectors which could inform on function (discussed in detail in section 3.3.5). Due to this, the protein was manually scanned for smaller motifs. No predetermined motifs were identified using the MOTIF search (GenomeNet); however, it was evident through a manual search that there were two glycine-proline-glycine-cystine (GPGC) repeats present. In *G. rostochiensis* the first GPGC repeat is amino acids 227-230, where the repeat is conserved across all four nematode species. The second GPGC repeat is at amino acids 285-288 in *G. rostochiensis* and is also present in the *G. pallida* and *N. abberans* proteins. However, in three of the four *R. reniformis* proteins the proline has been substituted with an alanine residue (Figure 3.2). Glycine residues adjacent to cysteine residues allow for disulphide bonds to be formed. Prolines attach to the protein backbone twice making these inflexible compared to other amino acids. Proline residues introduce kinks to this region (Lodish *et al.* 2000). The GPGC repeats observed are possibly required to maintain structural integrity of the protein.

3.3.3.2 – GLAND11

When analysing the amino acid sequences in the GLAND11 group, the *N. aberrans* proteins were observed to contain a large proline-rich stretch which is not present in the sequences from the other species. Poly-proline motifs/repeats can be associated with different structural conformation and protein functions. As well as causing kinks in the structure, runs of six or more proline residues have also been linked to functions such as DNA and actin processing (Morgan & Rubenstein 2013).

3.3.3.3 – G23G11

The G23G11 (formerly identified as GLAND15) orthologous group shows variation between the amino acid sequences from *G. rostochiensis*, *G. pallida*, and *R. reniformis* when compared to those identified from *N. aberrans*. Two poly-proline repeats are present in the *N. aberrans* proteins that are not present in those from the other three species, however this cannot currently be confirmed due to the lack of genome data for *N. aberrans* (Figure 3.3).

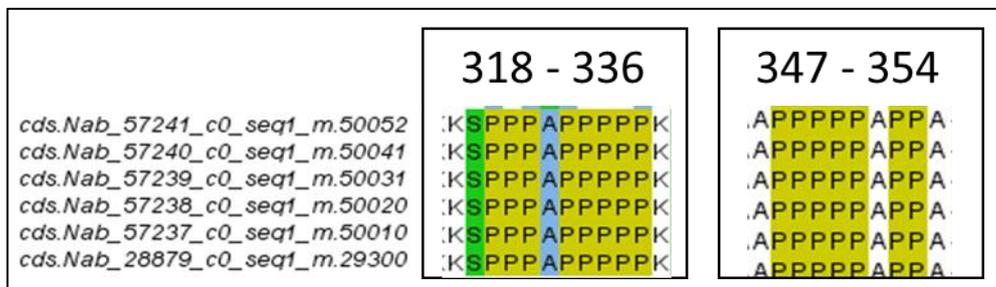


Figure 3.3 – Amino acid alignment of *N. aberrans* amino acid sequences present in the G23G11 effector family. Boxes indicate the repeated, conserved stretches of proline residues.

3.3.3.4 – Gro DOG 0043

The sequences in the Gro_DOG_0043 effector group all appear to be well conserved and relatively short at approx. 140 amino acids long, apart from GROS_g09990. GROS_g09990 was perfectly conserved with the original *G. rostochiensis* protein sequence GROS_g09987 up to the 136th amino acid, after which it has an additional 934 amino acids. This additional stretch of amino acids is quite substantial in size compared to the other orthologs from *G. pallida*, *R. reniformis*, and *N. aberrans*. The

additional 934 amino acids present in GROS_g09990 was run through conserved domain and motif searches which displayed sequence similarity with a DNA helicase domain (RecQ). This domain was not present in any of the other proteins found in this group. In order to check whether the software used to attribute RNA-seq data to individual genes (gene calls) had possibly mispredicted two separate genes and labelled them as one, a manual search of RNA-seq data was undertaken. When looking at GROS_g09990 RNA-seq data there are multiple reasons which would indicate that this gene had been mispredicted. Firstly, there is a substantial difference in expression levels between the first section of GROS_g09990 (the conserved section with the other family members) which is much higher when compared to the second additional section (Figure 3.4 A). This suggests there are two genes with different levels of expression. Secondly, if this was one gene it would be unusual as the pattern of RNA-seq mapping does not follow the canonical shape that we expect to see such data to assume. The typical shape is represented by the trapezium in Figure 3.4 B, where expression rises at the start of the gene, plateaus, then decreases at the end of the gene. If this were one continuous gene the expected shape would look like the red trapezium, however it is evident that there are two of these canonical shapes present which are represented by the blue and yellow trapezia (Figure 3.4 B). This is not a parameter which the software can assess, so is one reason that it may have been mispredicted.

The software has predicted a large intron present between two sections of the GROS_g09990 gene (black lines in red box – Figure 3.4 A). Manually it is possible to show this is not likely due to the shape of the RNA-seq reads. When an intron starts there is a sharp decrease in the number of reads which is represented by a clear vertical halt at the end of an exon. This is different from the end of a gene which decreases and tapers off gradually. In the instance of GROS_g09990 there is a large intron predicted between the two sections however there is no sharp halt in reads to coincide with the beginning and end of the intron. From the alignment file it is possible to see that all the other genes in this family end with the amino acid

sequence methionine – alanine or threonine – glutamine (M A/T Q) before the stop codon (*). Looking at the amino acid sequence from the RNA-seq data of GROS_g09990 in the different reading frames, it was clear that this sequence was present but the correct stop codon for this gene had not been selected (red box – Figure 3.4 C). When the amino acid sequence is corrected to end at the stop codon highlighted in Figure 3.4 C, the alignment between GROS_g09990 and the entries from the other nematode species become more highly conserved (Figure 3.4 D). With all evidence considered, the current annotation of GROS_g09990 is likely to be incorrect.

3.3.3.5 – Misprediction of further Gro DOG group members

When assessing the alignments of the Gro_DOG_0116, Gro_DOG_0149, Gro_DOG_0015, Gro_DOG_0028, Gro_DOG_0169, Gro_DOG_0201, and Gro_DOG_0200 groups it is evident that there are genes in all of these families which may have fallen foul of the same software misprediction of sequences as described above with the Gro_DOG_0043 family (Section 3.3.3.4). At least one sequence in each of these families contains additional fragments which do not appear in line with the rest of the amino acid sequences in the alignment (these can be viewed in Sup. File 5). They are likely mispredictions of intron/exon boundaries or misprediction of stop codons. Members of these families were not chosen for further characterisation in the pipeline process, meaning they were not scrutinised at the same level of detail as described above for the proteins that formed the focus of the further work. This would be required before any future characterisation work is carried out on these proteins.

3.3.3.6 – Gro DOG 0203 and Gro DOG 0017

Inexplicably Gro_DOG_0203 (GROS_11017) and Gro_DOG_0017 (GROS_g11020) were originally identified as two separate orthologous families, however it has since been determined that they are part of the same family. It is possible that the E-value was set too high on initial searches. For the alignment found in sup. file 5, these orthologous groups have been combined. Interestingly all sequences have a variable poly-glycine stretch followed by a furin protease RRKR (or RKKR/RRRR/RRRK) cleavage site motif. This indicates that these proteins could be synthesised in an inactive state and must have this poly-glycine section removed by a protease before they can function (Wise *et al.* 1990).

3.3.4 - DOG Box motif analysis

As discussed in 3.2.1.5, there have been new advancements in the discovery of motifs which can be used to signpost effector genes. Each *G. rostochiensis* candidate effector in Table 3.3 was examined to determine whether the DOG Box motif

(ATGCCA) was present in the 500 bp region promoter region, upstream of the start codon. This analysis showed that the genes encoding all 15 *G. rostochiensis* candidate core effectors contain at least one copy of the DOG box motif in the ~500 bp upstream region (Table 3.4), with an average of 2.3 DOG Box motifs present. This is compared to an average of 0.2 DOG Box motifs present upstream when 15 known housekeeper genes were analysed. This is the same background occurrence rate that was observed in the original study.

Table 3.4 - DOG Box motifs in promoter regions of candidate core effectors

<i>G. rostochiensis</i> seq. ID	DOG box motif sequences identified	No. of motifs identified in promotor region	Total DOG Box motifs present
GROS_g05682	ATGCCA	4	4
GROS_g01949	ATGCCA TGGCCT	3 1	4
GROS_g02469	TGGCAT AAGCCA	1 1	2
GROS_g09987	TGGCAT ATGCCA	3 1	4
GROS_g11017	ATGCCA AAGCCA	1 1	2
GROS_g07013	ATGCCA AGGCCA	1 1	2
GROS_g05985	TGGCAT TGGCCT	1 1	2
GROS_g02394	ATGCCA TGGCAT	1 1	2
GROS_g11020	AAGCCA TGGCAT	1 1	2
GROS_g04556	AAGCCA TGGCCT	1 1	2
GROS_g02024	TGGCAT TGGCCT	1 1	2
GROS_g09112	TGGCAT	2	2
GROS_g09671	TGGCCT	2	2
GROS_g03615	TGGCCT	1	1
GROS_g04903	AAGCCA TGGCCT	1 1	2

3.3.5 – Protein localisation and domain predictions

Multiple models now exist which allow for the target location of proteins to be computationally predicted. Some models predict where the protein will target inside the nematode cells, while others predict where these will be secreted and localised inside a host organism. As effectors are secreted by the nematode into the host plant, it was hypothesised that the 15 *G. rostochiensis* candidate effectors would all return a result of “extracellular” when analysed by localisation prediction programs. Two different programs; WoLF PSORT and DEEPLOC for prediction of protein localisation were utilised in this analysis and it was anticipated that they would return the same predictions (Horton *et al.* 2007, Almagro Armenteros *et al.* 2017). As observed in Table 3.5, only six of the candidate effectors (GROS_g05682, GROS_g11017, GROS_g07013, GROS_g02394, GROS_g11020, and GROS_g02024) returned the same extracellular localisation prediction from WoLF PSORT and DEEPLOC analysis.

The subcellular localisation *in planta* was predicted for the 15 candidate core effectors using LOCALIZER (Sperschneider *et al.* 2017). This analysis showed that out of the 15 effectors, 8 were predicted to target the plant nucleus, 2 predicted in the mitochondria, and 2 were predicted to target the chloroplast (Table 3.5). This was compared to 15 *G. rostochiensis* protein sequences which contain signal peptides but are not secreted as effectors (GROS_g04939, GROS_g5352, GROS_g05707, GROS_g10807, GROS_g06238, GROS_g12890, GROS_g02714, GROS_g06830, GROS_g05089, GROS_g01995, GROS_g13538, GROS_g09510, GROS_g12744, GROS_g06312 and GROS_g10196). None of these should return a result from LOCALIZER, however only 9 returned a non-applicable (N/A) result (GROS_g04939, GROS_g5352, GROS_g05707, GROS_g10807, GROS_g06238, GROS_g06830, GROS_g13538, GROS_g09510, and GROS_g06312). Three returned results of chloroplast targeting, 1 was nuclear, 1 was both mitochondrial and nuclear and 1 returned results of chloroplast, mitochondrial, and nuclear. This reflects the fact that the software simply seeks known motifs associated with particular subcellular

structures. The number of effectors which have been assigned by LOCALIZER as nuclear is above the random rate, however the 4 predicted to be either mitochondrial or chloroplast targeting may not be greater than expected by chance. Data from prediction models need to be tested in plants and subcellular localisation studies of a subset of the *G. rostochiensis* core effector list can be found in chapter 6.

As a predicted function for some of the 15 effector groups could not be established based on sequence similarity as a whole, a conserved domain search was conducted on the candidate core effectors using the National Centre for Biotechnology Information Conserved Domain Database, NCBI CDD (Marchler-Bauer *et al.* 2015). This analysis showed that six of the fifteen *G. rostochiensis* core effectors had no conserved domains identified (ND in Table 3.5), two have conserved domains of unknown function (DUF) and seven have conserved domains of known function (Table 3.5). The DUF1794 domain conservation identified in GROS_g04556 is annotated as being associated with nematode larval development (pfam08768) based on *C. elegans* data. Some of the predictions were consistent with roles likely to be associated with effectors. For example, domains potentially associated with disrupting plant signalling pathways such as phosphatidylinositol phosphate kinases (PIPKc) were present.

Table 3.5 – Localisation prediction and conserved domains of 15 *G. rostochiensis* candidate core effectors

Candidate orthologous group name	<i>G. rostochiensis</i> seq. ID	WOLF PSORT localisation prediction	DEEPLOC localisation prediction	Predicted localisation in <i>planta</i>	Conserved domains (ND – not detected)
20E03	GROS_g05682	Extracellular	Extracellular	Mitochondria and Nucleus	ND
Gro_DOG_0116	GROS_g01949	Endoplasmic reticulum	Lysosome/Vacuole	No prediction	ND
G23G11	GROS_g02469	Extracellular	Endoplasmic reticulum	Chloroplast and Nucleus	Alanyl-tRNA synthetase
Gro_DOG_0043	GROS_g09987	Extracellular	Extracellular	Nucleus	Crustacean CHH/MIH/GIH neurohormone family
Gro_DOG_0203	GROS_g11017	Plasma membrane	Extracellular	Nucleus	ND
Gro_DOG_0199	GROS_g07013	Extracellular	Extracellular	No prediction	DUF4106 (unknown function)
Gro_DOG_0149	GROS_g05985	Endoplasmic reticulum	Cytoplasm	Nucleus	Phosphatidylinositol phosphate kinases (PIPKc)
GLAND11	GROS_g02394	Extracellular	Extracellular	Nucleus	ND
Gro_DOG_0017	GROS_g11020	Extracellular	Extracellular	Nucleus	ND
Gro_DOG_0015	GROS_g04556	Extracellular	Mitochondrion	No prediction	DUF1794 (unknown function – putative function in larval development)
Gro_DOG_0028	GROS_g02024	Extracellular	Extracellular	Chloroplast	Trypsin-like serine protease
Gro_DOG_0169	GROS_g09112	Endoplasmic reticulum	Lysosome/Vacuole	Nucleus	Metallophosphatase
Gro_DOG_0187	GROS_g09671	Extracellular	Endoplasmic reticulum	No prediction	Histidine phosphatase
Gro_DOG_0201	GROS_g03615	Extracellular	Lysosome/Vacuole	Mitochondria	Peptidase_C1 (papain family cysteine protease)
Gro_DOG_0200	GROS_g04903	Extracellular	Cell Membrane	No prediction	ND

3.3.6 – Gene expression profiles of *G. rostochiensis* candidate core effectors

The RNAseq data available for *G. rostochiensis* was used to determine the expression profiles across the life cycle for each of the fifteen *G. rostochiensis* candidate core effectors. As previously discussed, effectors have peak expression at either the J2 life stage when the nematode is attempting to enter the host, the adult parasitic life stages where the nematode is actively feeding or expressed across both J2 and adult parasitic stages. Four effectors displayed peak expression at J2 (GROS_g01949, GROS_g09987, GROS_g05985, and GROS_g09112), while the remaining eleven candidate effectors showed peak expression at 14 days post infection (dpi) (Figure 3.5). It was hypothesised that GROS_g01949, GROS_g09987, GROS_g05985, and GROS_g09112 would likely display subventral gland staining during *in situ* hybridisation assays and the remaining eleven candidate effectors would display dorsal gland expression.

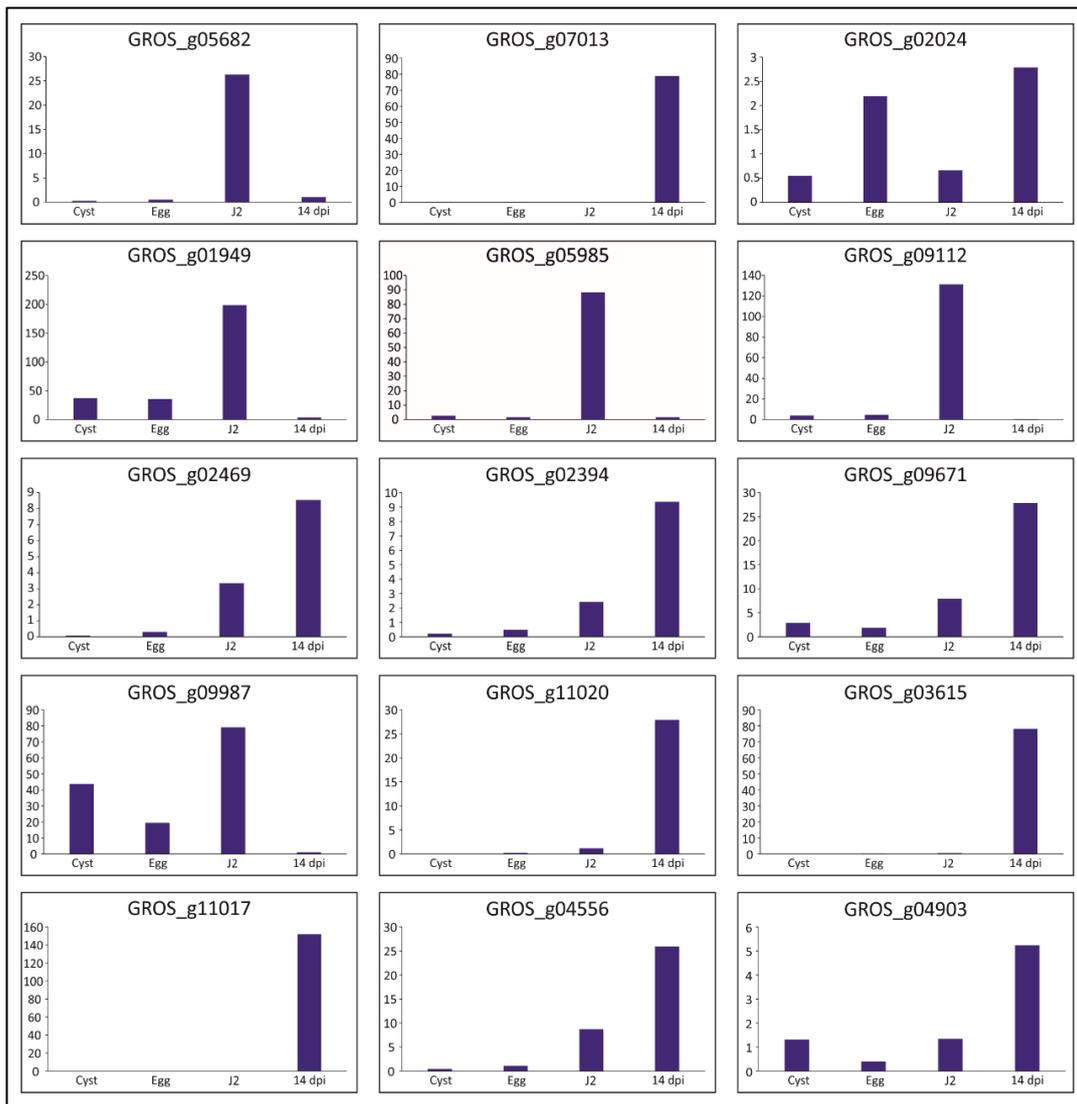


Figure 3.5 - Gene expression profiles of the top 15 *G. rostochiensis* candidate core effectors. All graphs are measured in Trimmed mean of M-values (TMM) normalised expression on the Y axis. Graphs were produced using normalised expression data from (Eves-Van den Akker et al. 2016a)

3.3.7 – Cloning of *G. rostochiensis* candidate core effectors

Primers were designed for cloning of all 15 *G. rostochiensis* core effector candidates (Table 3.1). The full length (including start and stop codon, excluding signal peptide), coding sequences for five of the effectors candidates - GROS_g05682 (20E03), GROS_g01949 (Gro_DOG_0116), GROS_g07013 (Gro_DOG_0199), GROS_g09987 (Gro_DOG_0043), and GROS_g03615 (Gro_DOG_0201) were successfully cloned into the pCR8/GW/TOPO vector while a further two GROS_g02469 (G23G11) and GROS_g02394 (GLAND11) were cloned into the pGEMT-Easy vector. All cloned gene

were identical to those predicted from the *G. rostochiensis* genome data with the exception of GROS_g09987, which was cloned with 4 nucleotide substitutions. One of these substitutions resulted in a change in amino acid; an arginine residue in the sequence from the genomic data was observed as a proline residue in the cloned gene sequence (Figure 3.6). This change of amino acid was observed in multiple cloning attempts of this gene.

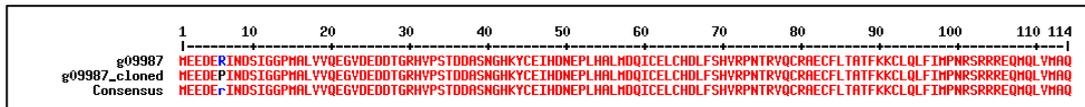


Figure 3.6 – Pairwise alignment of GROS_g09987 amino acid sequences. A pairwise alignment between the amino acid sequence of GROS_g09987 predicted from the genome data and the sequence of GROS_g09987 when cloned into the pCR8/GW/TOPO vector. At position 6 there is an arginine residue in the predicted sequence and a proline in the cloned sequence.

Despite repeated attempts, it was not possible to clone the full-length gene of the remaining eight effector candidates in either vector system (GROS_04556, GROS_11017, GROS_04903 and GROS_02024, GROS_g05985, GROS_g11020, GROS_g09112, and GROS_g09671). For four genes (GROS_04556, GROS_11017, GROS_04903, and GROS_02024) it was possible to amplify and clone a short fragment that could be used for *in situ* hybridisation. For these, the amplified gene sequences were the same as those predicted in the *G. rostochiensis* genome.

3.3.8 – *In situ* hybridisation

In situ hybridisation (ISH) was used to determine the spatial expression of the cloned *G. rostochiensis* orthologues of the candidate core effectors. In PPN, *in situ* hybridisation remains the final determination of the status of each gene encoding an effector as this is dependent on the mRNA expression being observed in the oesophageal gland cells (dorsal or sub ventral glands) or the amphids. The gene expression data found in Figure 3.5 was used to determine which nematode life stage the ISH for each candidate effector was conducted on. As displayed in Table 3.6 *In situ* hybridisation was conducted on eleven of the fifteen candidate effectors. Three of the eleven (GROS_g02394, GROS_g02469, and GROS_g05682) gave consistent gland cell expression signals confirming these candidate genes as effectors. Dorsal gland staining was observed for GROS_g02394 and GROS_g02469, while GROS_g05682 showed subventral gland expression (Figure 3.7). Seven of the candidates (GROS_g09987, GROS_g07013, GROS_g04556, GROS_g011017, GROS_g04903, GROS_g02024, and GROS_g01949) remain potential effector candidates but the hybridisation signals from these were weak or difficult to interpret. These seven candidates produced high levels of background staining which prevented a confident confirmation of gene expression location. Examples of three of these unclear *in situ* hybridisation assays can be seen in Figure 3.8. GROS_g09987 consistently displayed a two-point localisation which is possibly the nuclei of the subventral glands but is yet to be definitively confirmed. GROS_g04903 gave inconsistent coverage in its localisation making it difficult to determine whether this was expressed in the intestines or a different organ which was being obscured by diffuse background staining. GROS_g02024 did display what could have been dorsal gland staining however, it could not be replicated consistently, and indiscriminate background staining was again an issue. GROS_g03615 gave clear and consistent intestinal/gut localisation, excluding it as a candidate effector, however further work on GROS_g03615 is described in chapter 5. The result of this analysis produced three spatially confirmed effectors (GROS_g02394, GROS_g02469, and GROS_g05682) present in all four syncytium forming species for further study.

Table 3.6 – In situ hybridisation locations of candidate core effectors from *G. rostochiensis*

<i>G. rostochiensis</i> seq. ID	Was ISH conducted in <i>G. rostochiensis</i>?	Localisation pattern observed
GROS_g05682	Yes	Subventral glands
GROS_g01949	Yes	Unconfirmed
GROS_g02469	Yes	Dorsal gland
GROS_g09987	Yes	Unconfirmed
GROS_g11017	Yes	Unconfirmed
GROS_g07013	Yes	Unconfirmed
GROS_g05985	No	N/A
GROS_g02394	Yes	Dorsal gland
GROS_g11020	No	N/A
GROS_g04556	Yes	Unconfirmed
GROS_g02024	Yes	Unconfirmed
GROS_g09112	No	N/A
GROS_g09671	No	N/A
GROS_g03615	Yes	Intestines
GROS_g04903	Yes	Unconfirmed

*N/A – non-applicable due to no ISH assay being conducted

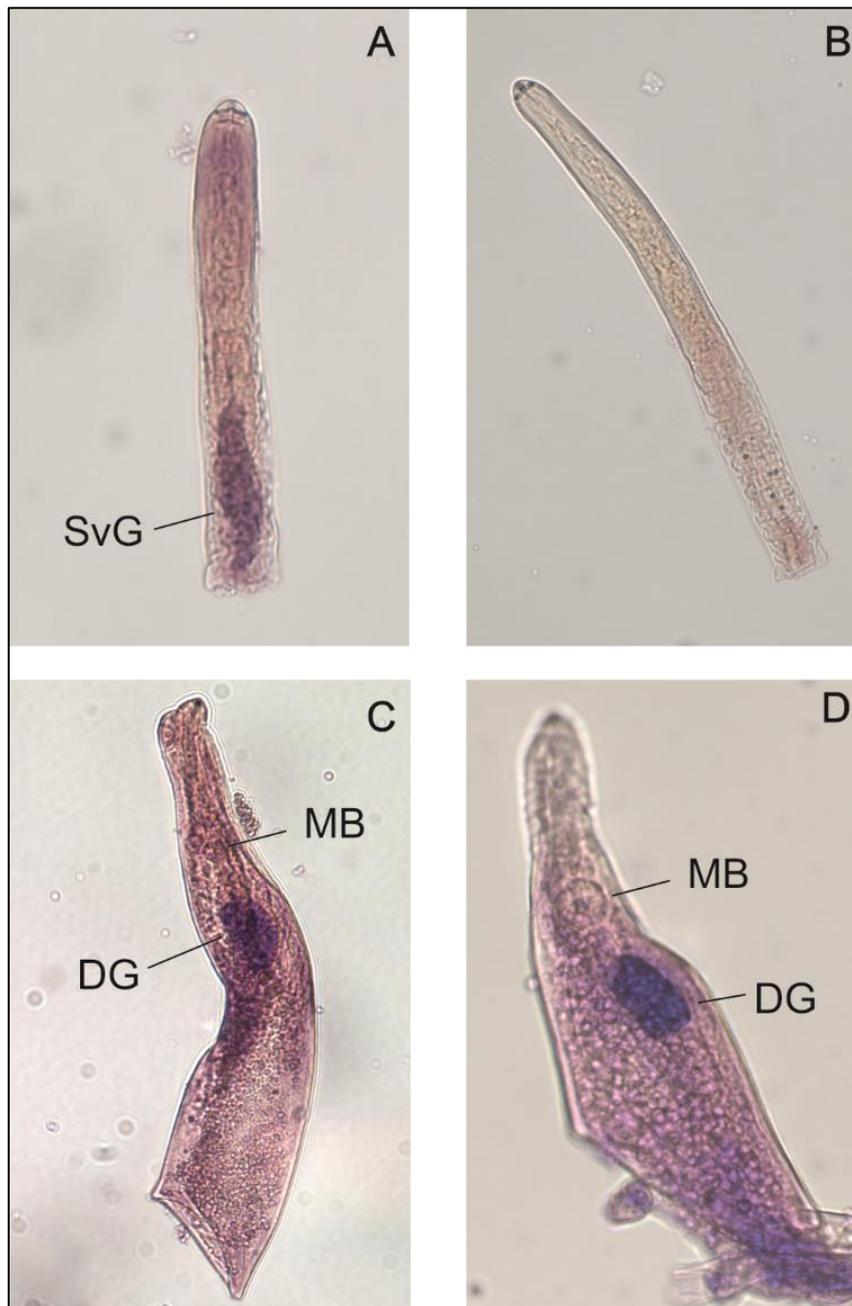


Figure 3.7 - In situ hybridisation of GROS_g05682, GROS_g02469, and GROS_g02394. A – ISH of GROS_g05682 displaying staining of the subventral glands of a J2 nematode. B – Negative control for GROS_g05682 using the forward primer probe in a J2 nematode. C – ISH of GROS_g02469 displaying staining of the dorsal gland. D – ISH of GROS_g02394 displaying staining of the dorsal gland. SvG – subventral gland, DG – dorsal gland, MB – metacarpal bulb.



Figure 3.8 – In situ hybridisation results for candidate effectors *GROS_g09987*, *GROS_g04903*, and *GROS_g02024*. A and B – Potentially subventral gland nuclei ISH staining of *GROS_g09987*. C and D – Diffuse ISH staining of *GROS_g04903*. E and F - Diffuse ISH staining of *GROS_g02024*. *GROS_g02024* potentially displays dorsal gland staining as seen in E, however the background staining is too diffuse across all repeats to confirm. MB – metacarpal bulb.

3.4 – Discussion

In this research chapter an *in silico* pipeline was developed that has successfully identified 37 (which later increased to 43) orthologous groups of candidate effectors present in all four of the syncytia-forming nematode species *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans*. This represents a valuable resource for future studies on these nematodes. Members of three of these groups (GROS_g05682, GROS_g02394, and GROS_g02469) were confirmed as effectors through *in situ* hybridisation and subsequently prioritised for functional characterisation in chapter 6.

3.4.1 – Effector identification pipeline improvement

This research is not likely to have identified the full effector suite of all nematode species analysed. Beginning with an initial gene set of validated effectors from *G. rostochiensis* gave a firm base to begin the research, however, setting the initial criteria for inclusion on the starting list at such a high stringency (E value) is likely to have excluded some genuine effectors. For example, there are many proteins in *G. rostochiensis* predicted to be effectors due to high sequence similarity to validated effectors from other nematode species. For this analysis it was decided to exclude any genes that had not been functionally confirmed as effectors in *G. rostochiensis* to minimise false positives. This also extends to previously undiscovered, uncharacterised, or pioneer effectors. Effectors which were unrepresented in the initial gene set are likely to be entirely excluded from the candidate core effector lists as they have not appeared in the BLAST similarity searches. Following on from this, it is evident that there are missing data that could have been obtained if this research had been repeated with initial confirmed effector lists from *G. pallida*, *R. reniformis* and *N. aberrans*. However, for *R. reniformis* and for *N. aberrans* in particular, there are very few confirmed effectors from which to start this work.

It should also be noted that there were 6 additional effector groups that were identified after the initial data were processed. As discussed in section 3.3, the effector groups pectate lyase, Gro_DOG_0057, Gro_DOG_0067, Gro_DOG_0073, Gro_DOG_0079, and Gro_DOG_0211 were originally missed in the set of resulting genes identified from the *R. reniformis* genome and transcriptome data. This was due to BLAST parameters originally being set too stringently. Therefore, the genes present in each of these groups were not subjected to further study. This means that there are potentially interesting effectors for future study in these groups. The function of the pectate lyase has been characterised in detail and wasn't investigated further here. The *G. rostochiensis* amino acid sequence in each of the effector groups Gro_DOG_0057, Gro_DOG_0067, Gro_DOG_0073, Gro_DOG_0079, and Gro_DOG_0211 (GROS_g04735, GROS_g08471, GROS_g10784, GROS_g08534, and GROS_g05526 respectively) were used as queries for further BLAST searches. GROS_g04735, GROS_g10784, GROS_g08534, and GROS_g05526 all returned BLAST results of hypothetical or unnamed proteins from a host of nematode species. These groups of effectors (Gro_DOG_0057, Gro_DOG_0073, Gro_DOG_0079, and Gro_DOG_0211) would have been removed at subsequent steps in the pipeline due to having similarity with root-knot nematode proteins. GROS_g08471 returned similarity to FMRamide (H-Phe-Met-Arg-Phe-NH₂)-related and -like peptides (FLPs) from multiple different nematode species of both cyst and root-knot families. FLPs are a diverse group of neuropeptides involved in reproductive processes as well as mobility in nematodes, however FLPs are also involved in cardiovascular function in higher organisms (Peymen *et al.* 2014). FLPs would certainly be an interesting set of proteins to study but they would have also been removed from this pipeline due to homologs existing in root-knot nematodes.

3.4.2 – Syncytia-forming species

Globodera rostochiensis, *G. pallida*, *R. reniformis*, and *N. aberrans* were chosen for analysis because they all produce syncytia as their feeding site, although they are phylogenetically diverse species. These four species are clearly not the only species

that produce syncytia however, so potentially there are core effectors identified in this research that may not be present in 100% of syncytium forming species. 2019 saw the release of the soybean cyst nematode, *Heterodera glycines* genome which is another nematode which forms a syncytium as its feeding site (Masonbrink *et al.* 2019). Access to this data came too late to be included in the initial research conducted in this thesis. A BLAST (BLASTp) similarity search against *H. glycines* using the *G. rostochiensis* protein sequences from the top 15 orthologous groups (full amino acid sequences including signal peptides) returned 7 hits; GROS_g05682, GROS_g02469, GROS_g11017, GROS_g02394, GROS_g04556, GROS_g02024, and GROS_g03615 (Sup. File 6). This shows that the three effectors chosen for functional analysis: GROS_g05682, GROS_g02469, and GROS_g02394 are present in five syncytia-forming species instead of the initial four included in the pipeline. It was unlikely that there would be eight effector candidates (GROS_g01949, GROS_g09987, GROS_g07013, GROS_g05985, GROS_11020, GROS_g09112, GROS_g09671, and GROS_g04903) which were not present in *H. glycines*. This prompted a retrospective BLAST search without a set E-value to identify whether any candidates were present but diverged to an extent where they were excluded in the original search. At least one *H. glycines* ortholog was discovered for each of the top 15 *G. rostochiensis* candidates using this approach; the results of this analysis can be found in Sup. File 6.

3.4.3 – DOG Box motif analysis

The average effector produced in the dorsal gland has 2.54 DOG Box motifs present in their upstream promotor region whereas non-effector proteins only have an average of 0.22 (Eves-Van den Akker *et al.* 2016a). Eves-van den Akker *et al.* also showed that a higher number of DOG box motifs associated with a sequence was correlated with an increased likelihood of the presence of a signal peptide. This is in line with the results found in this study where an average of 2.3 DOG Box motifs were present in the 500 bp upstream promotor region of the 15 *G. rostochiensis* candidate core effectors. All genes identified in this analysis had the DOG box motif

in their promotor regions. However, GROS_g05682 has four DOG box motifs, but has been shown to be expressed in the subventral glands (Figure 3.7). This is much higher than the average seen in non-effector proteins (0.22) which suggests this has not occurred by random chance, but it is not clear as to why this effector has these upstream motifs at present.

Typically, DNA is compacted around histones in the nucleus into dense (heterochromatin) or loose (euchromatin) regions of chromatin (Annunziato 2008). This condensed storage of DNA allows for the regulation of gene transcription and expression alongside protecting the DNA from damage. In chromatin dense regions genes are inaccessible to transcription factors which means they are less likely to be expressed (Baker 2011). Enhancers are motifs that can be found upstream of promoters for specific genes. These can act as regions for transcription factors to attach to. Enhancers can be found hundreds of kilobases upstream of the promoters and genes they regulate (Spicuglia & Vanhille 2012). Due to this lack of proximity, it is often difficult to identify new enhancers and easy to miss known motifs which lay out with the upstream region being observed. It is thought that the DOG Box could play a role as an enhancer. Although the DOG box is proving a useful effector identification tool, it is worth noting that 23% of 101 dorsal gland expressed effectors tested in the original study did not have a DOG box motif found in the 500 bp upstream promotor region (Eves-Van den Akker *et al.* 2016a). It is therefore possible that some of the candidate effectors included in this thesis that were reported as not having a DOG box motif may still have the motif associated with it, but its location was out with the parameters of the search. It is also possible that they do not have a motif associated with them, however this doesn't count them out as effectors based on this criterion alone.

3.4.4 – *In situ* hybridisation

Performing *in situ* hybridisation on gene sequences of interest is important to visualise where genes are being expressed in the nematode. This can be used

practically to identify whether a protein of interest could be an effector based on localisation to the dorsal gland, subventral glands, or amphids. The ISH results from this analysis confirm that the *in silico* pipeline used for effector identification will inevitably give rise to false positive results. This is made evident by GROS_g03615 which was one of the 15 candidate core effectors until the ISH analysis showed that this gene is expressed in the nematode intestine and therefore not an effector. The pipeline used has the potential to let through proteins that are not effectors but still meet the pipeline requirements, such as gene expression profile and presence of a signal peptide. GROS_g03615 is a Cathepsin L-like endopeptidase with high gene expression during adult parasitic (14 dpi) life stages, is targeted to the secretory pathway, and was found in the four syncytium species analysed. Despite being expressed in the intestine, GROS_g03615 also contained a single DOG Box motif (TGGCCT) in its upstream promoter region. This example shows the importance of candidate validation by *in situ* hybridisation. The intestinal gene expression suggests that this gene could be involved in nutritional uptake. Further work on GROS_g03615 can be found in chapter 5.

In situ hybridisation results for seven of the genes were inconclusive. Further analysis of candidate GROS_g01949 suggests that this is unlikely to be an effector. The hybridisation signal was not clear enough to confirm the location of this candidate. However, it was evident on further inspection of the amino acid sequence that it contained an endoplasmic reticulum (ER) retention signal at the C-terminus. Endoplasmic reticulum retention signals prevent proteins from being secreted and transported to the Golgi apparatus (Pelham 1990). Common ER retention signals are KDEL (Lys-Asp-Glu-Leu) seen in both mammals and plants, and HDEL (His-Asp-Glu-Leu) in yeast. However there is considerable variation within recognised retention signals with RDEL, KNEL, and HTEL all observed as retention signals in mammalian or yeast species (Potter *et al.* 1992, Pelham 1990). The presence of the RDEL (Arg-Asp-Glu-Leu) sequence at the terminus of GROS_g01949 means it is unlikely that this protein is secreted and therefore unlikely to be an

effector. Another example of this is calretuculin (Mi-CRT) from *M. incognita* (Jaouannet *et al.* 2013). Mi-CRT is termed an effector by RNAi studies used to determine its importance in parasitism, however an HDEL (His-Asp-Glu-Leu) endoplasmic reticulum retention signal is present in this protein sequence which was removed in the cloning process prior to functional studies. Due to this it is unclear whether this protein would be secreted into the host plant by the nematode in its native state, although immunolocalization has shown that an antiserum raised against calretuculin binds to the apoplast during invasion.

There is still a possibility that the remaining 6 candidate core effector genes GROS_g09987, GROS_g04556, GROS_g011017, GROS_g04903, GROS_g02024, and GROS_g01949 are effectors but unfortunately the ISH reactions were inconclusive. There are several different reasons why ISH may have been unsuccessful for these candidates. In the case of GROS_g04903 and GROS_g02024, these candidate effectors had extremely low expression levels and this may have contributed to the weak, diffuse signals observed that were difficult to interpret. Several attempts were made to refine the ISH process to identify the expression sites of these genes. However, after repeated failures a decision was made to stop conducting ISH on GROS_g09987, GROS_g04556, GROS_g011017, GROS_g04903, GROS_g02024, and GROS_g01949 and to focus analysis on the genes for which a signal had been obtained. It is possible that with the correct combination of new probe design, probe concentration and stain incubation times, that some of the candidates could be confirmed as effectors in the future. It is also possible that GROS_g05985, GROS_g11020, GROS_g09112, and GROS_g09671 are also effectors, but the failure to clone these genes meant that *in situ* hybridisation assays could not be conducted so the location of their expression in the nematode remains unknown. Finally, GROS_g07013 (Gro_DOG_0199) has been removed from the list as a potential candidate as recent repeats of the BLAST searches have identified similar proteins in *C. elegans*. As *C. elegans* is a soil-dwelling nematode and not parasitic to plants it is

unlikely that GROS_g07013 has a function based around syncytium formation or maintenance.

A subset of core effectors, GROS_g05682, GROS_g02469, and GROS_g02394, were chosen to be put forward for functional characterisation which would include experiments with yeast two-hybrid screening. These three effectors were chosen on the strength of the *in situ* hybridisation results confirming their status as effectors as well as the bioinformatics analysis.

3.5 – Future work

When BLAST searches were conducted on the candidate core effectors a small subset were identified as pioneers containing domains of unknown function. These effectors (e.g., those in the Gro_DOG_0116, and Gro_DOG_0017/0203 groups) weren't chosen for further analysis in this project but could be starting points for future studies. This chapter stands as a starting point for many avenues of characterisation work which unfortunately could not have been completed in the time scale of this thesis. Between the identification of mispredicted sequences which may have hindered the cloning process of some candidates, as well as failed attempts at optimisation of *in situ* hybridisation, there are several candidates which may well be effectors ready for further characterisation in the future.

4. Arabinogalactan

endo- β -1,4-galactanases of

plant-parasitic nematodes

4.1 – Introduction

4.1.1 – Plant cell walls

For endo- and semi-endoparasitic nematode species to feed upon their plant host they must overcome the initial barrier posed by plant cell walls. The basic function of the cell wall is to protect the internal components of the cell and to provide a stable structure. In addition, the cell wall acts as one of the first lines of defence against infection by pathogens. The cell wall is made up of an interconnected network of proteins and polysaccharides with the most abundant being cellulose, hemi-cellulose, and pectin (Figure 4.1). This network provides a strong physical barrier due to the properties of the components, most notably the covalent bonding between polysaccharides (Iiyama *et al.* 1994). Cellulose is the most abundant polysaccharide and consists of large, unbranched strands of β -1,4-linked glucose molecules. Hemicelluloses are also abundant in the cell wall with the most common being xylan, arabinoxylan, and mannan. The hemicelluloses are intertwined with the cellulose fibres via hydrogen bonds and together these are enmeshed within layers of pectin (Figure 4.1).

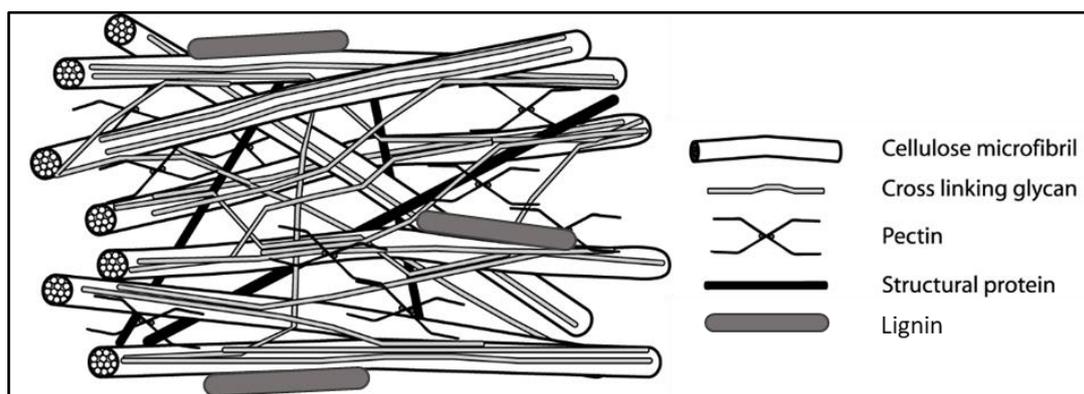


Figure 4.1 – The structure of the plant cell wall. The interconnecting network of proteins and polysaccharides observed in the cell wall highlighting cellulose, glycans, pectin, lignin, and structural proteins. Based on image provided by Prof. J. Jones (pers. comm.)

4.1.1.1 – Pectin

Pectin is the collective name given to the third main component of the cell wall. Pectin is abundant in areas such as root caps and in soft fruits (Yang & Anderson 2020). There are three main types of pectin: homogalacturonan (HG), and rhamnogalacturonan I and II (RGI and RGII) with HG being the most abundant. Pectin can be categorised into smooth or hairy regions based on the level of branching sidechains that are present. HG comprises large chains of 1,4-linked α -D-galacturonic acid and as there are no branching chains extending from the HG backbone, it is classified as smooth (Ochoa-Villarreal *et al.* 2012). RGI differs from HG due to having repeating units of L-rhamnose and D-galacturonic acid in its backbone. Branching chains such as galactan, arabinan, and arabinogalactan attach to the L-rhamnose molecules of the RGI backbone (Figure 4.2) (Ridley *et al.* 2001). RGII has a backbone of galacturonic acid units (monosaccharides), like HG, but also has branching side chains like RGI. These side chains are often complex and made up of multiple different types of carbohydrates including apiose, fucose, and aceric acid (3-C-carboxy-5-deoxy-L-xylofuranose) (Pérez *et al.* 2003). Due to the presence of branching side chains on RGI and RGII these are referred to as hairy regions of pectin.

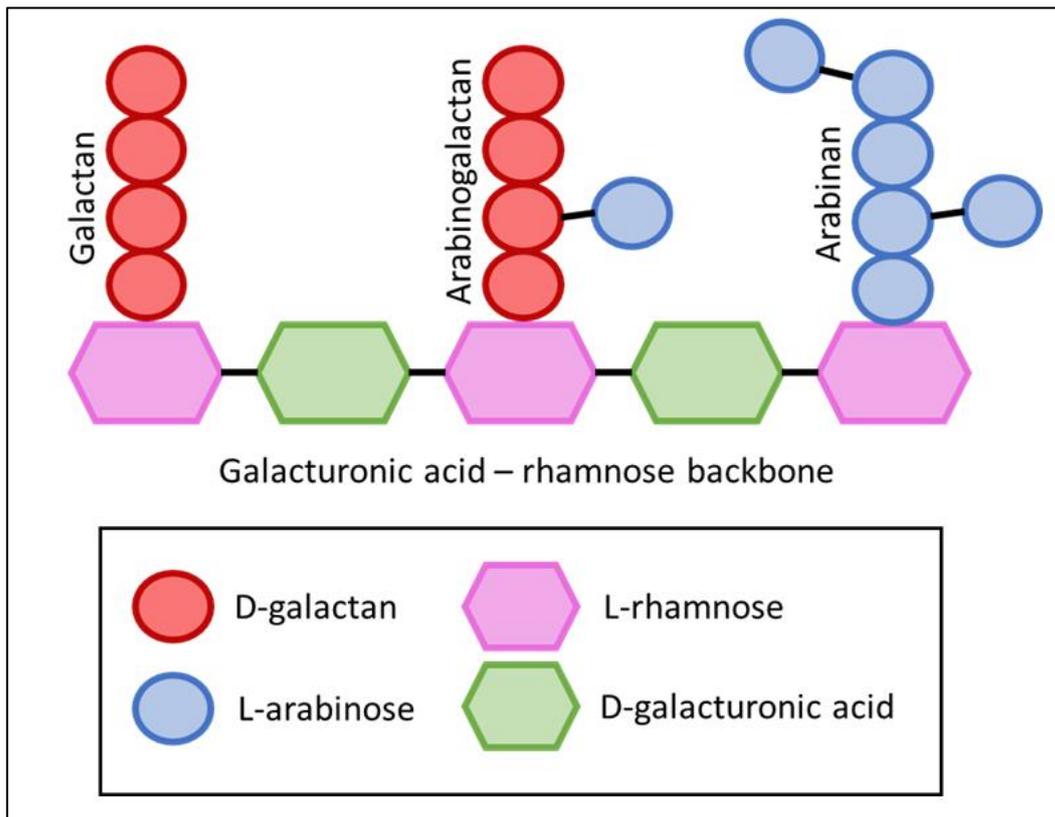


Figure 4.2 – Simplified schematic representation of rhamnogalacturonan I (RGI). Branching side chains of RGI: arabinogalactan (galactose backbone with arabinose sidechains), arabinan (repeating subunits of arabinose), and galactan (repeating subunits of galactose). Backbone of repeating L-rhamnose and galacturonic subunits. Figure adapted from (Pérez *et al.* 2003).

The composition of the cell wall varies between different species of plant. Differences in the cell wall composition of host and non-host plants may play a role in the success of infection by certain pathogens (Vorwerk *et al.* 2004, Carapito *et al.* 2013). Many pathogens need to overcome the plant cell wall to be successful in parasitising the host plant. Some species have evolved ways in which they can physically puncture the cell wall to infect their host. Puncturing the cell wall creates an entry point for the pathogen, whether they stay on the plant surface to feed or then migrate inside the host tissue. Approaches to puncturing the cell wall are varied and include fungal appressoria and nematode stylets. Appressoria are specialised fungal structures which apply turgor pressure to burst through the cell wall (Ryder & Talbot 2015). The stylet is a hollow needle-like appendage used by nematodes to physically pierce the cell wall as well as deliver effectors.

4.1.2 – Cell wall degrading enzymes

Pathogens use cell wall degrading enzymes (CWDE) and cell wall modifying proteins in addition to physical means to break the cell wall. CWDE are found in many species of plant-parasitic bacteria, fungi, and nematodes. Pathogens frequently have an arsenal of CWDE which are used in parallel to break down different components of the cell wall. Certain bacteria, including some species of *Clostridium* such as *C. acetobutylicum* and *C. cellulovorans*, produce large enzyme complexes called cellulosomes, which are used predominantly to break down the cellulose layer of the cell wall (Doi & Kosugi 2004). These cellulosomes can include cellulases, hemi-cellulases, and pectinases held together by scaffoldin fibrils. CWDE of fungal species have been well researched with examples including the Bcpgl endopolygalacturonase from *Botrytis cinerea* (grey mould). Bcpgl mutant *B. cinerea* lines were shown to have a decrease in secondary infection abilities, showing that Bcpgl is necessary for infection (Ten Have *et al.* 1998).

In nematode species CWDE are produced and secreted into the host plant via the stylet to break down complex cell wall components at the same time as probing and puncturing occurs during migration (see section 1.3.2). Increased availability of both genome and transcriptome data has shown that many PPN species produce a wide range of CWDE (Section 1.3.2). The first CWDE to be characterised in nematodes were cellulases (β -1,4-endoglucanases) from *G. rostochiensis* (GR-ENG-1 & 2) and *H. glycines* (HG-ENG-1 & 2) (G. Smart *et al.* 1998). This study showed that HG-ENG-2 was expressed in the subventral glands, which are known to produce effector proteins involved in parasitism. Following on from this, several other CWDE were identified in PPN species. These include pectate lyases, which were originally identified in *G. rostochiensis*, with homologs subsequently identified in *H. glycines*, *M. incognita*, and *M. javanica* (Popeijus *et al.* 2000, Kudla *et al.* 2007, De Boer *et al.* 2002, Huang *et al.* 2005, Doyle & Lambert 2002). Like the cellulases, the pectate lyase Gr-Pel2 from *G. rostochiensis* was shown to be expressed in the subventral glands of the nematode. Transient expression of Gr-Pel2 in *Nicotiana benthamiana* for

functional analysis showed conserved structural and biochemical properties with bacterial cellulases (Kudla *et al.* 2007). Further CWDE were subsequently identified from PPN, including the polygalacturonase *Mi-pg-1* from *M. incognita* (Jaubert *et al.* 2002). These enzymes hydrolyse polygalacturonic acid (also known as pectic acid). *Mi-pg-1* is also expressed in the subventral glands (Rosso *et al.* 2005).

As well as CWDE, nematodes produce proteins classed as cell wall modifying proteins. Cell wall modifying proteins do not have any detectable enzymatic activity, but target cell wall components, often making them more susceptible to degradation by CWDE. One example of cell wall modifying proteins is the expansin family. Expansins are proteins produced by plant species that act to “loosen” the cell wall in order to facilitate expansion and growth of tissue (Sampedro & Cosgrove 2005). Expansins are thought to disrupt non-covalent interactions between cell wall components, thus enhancing access by cell wall degrading enzymes. Plant-parasitic species of bacteria, fungi, and nematodes have all been shown to produce expansin or expansin-like proteins that aid in parasitism (Cosgrove 2017). *Globodera rostochiensis* has been shown to produce multiple expansin-like proteins including Gr-EXP1 and GrEXPB2. Gr-EXP1 was discovered in 2004 and is produced in the subventral glands of the nematode (Qin *et al.* 2004). Homogenates of J2 nematodes displayed cell wall loosening activity when applied to wheat shoot tips and cucumber hypocotyls. When GrEXPB2 was expressed in *N. benthamiana* there was chlorosis of the leaf tissue alongside dwarfing of the plant when compared with the empty vector control lines (Ali *et al.* 2015). Necrosis of leaf tissue was observed around toothpick inoculation sites on both potato and tomato lines. These results have prompted the hypothesis that GrEXPB2 is secreted into the plant alongside other nematode effectors to promote cell wall loosening and facilitate nematode migration.

As previously discussed, the availability of genome and transcriptome data for different PPN species has allowed for the discovery of many different cell wall degrading enzymes and modifying proteins that were previously unknown to be

produced by nematodes. Examples of CWDE identified this way include GH45 and GH5 family members from *B. xylophilus*, GH16, GH30, and GH43 proteins from *Radopholus similis*, and GH32 proteins from *G. pallida* and *G. rostochiensis* (Palomares-Rius *et al.* 2014, Huang *et al.* 2019, Cotton *et al.* 2014, Eves-Van den Akker *et al.* 2016a). Glycoside hydrolase families and functions are covered in more detail in section 4.1.3.

4.1.2.1 – Horizontal gene transfer

Horizontal gene transfer (HGT), also known as lateral gene transfer (LGT), is the transfer of genetic material from one species to another through asexual means (Burmeister 2015). It is commonly seen between prokaryotic organisms, however it can also occur more rarely between eukaryotic species (Andersson 2005). HGT is a mechanism for rapid evolution of a species as it often results in the acquisition of genes with functions which are novel and advantageous to the species which take up the genetic material. A notable example of this is the acquisition of antibiotic resistance genes by previously susceptible bacterial species (Dzidic & Bedekovic 2003).

Cell wall degrading enzymes are abundant in plant-pathogenic bacteria and fungi but are absent in almost all metazoans except for plant-parasitic nematodes. Many animal species rely on symbiotic relationships with micro-organisms in their digestive tracts to break down cellulose and other plant matter in their diet. For example the microbiota of ruminant digestive tracts is well studied and shown to contain diverse protozoan and fungal species (Morrison *et al.* 2009). Due to the high sequence similarity between genes encoding CWDE in bacteria, fungi, and PPN, it is hypothesised that these genes may have been acquired through a horizontal gene transfer event. HGT candidate genes in nematodes can be defined by having an Alien index (AI) > 0. The AI score is calculated by taking a gene of interest from the nematode (recipient) and similar sequences from bacteria and/or fungi (donor). Using BLAST E-values it is determined whether the donor sequence is either more

similar to sequences from donor species (in this case fungal or bacterial origin) or to other sequences found in the recipient (nematode). If the recipient sequence is more like the donor sequences it will have an AI > 0 (Rancurel *et al.* 2017). Using the alien index 519 candidate HGT-acquired genes were identified from *G. rostochiensis*, of which 91 proteins are robust candidates with an AI > 30 (Eves-Van den Akker *et al.* 2016a). The list of HGT candidates includes many CWDE, as well as candidates involved in processes such as nutrient processing, feeding site induction, and host defence manipulation.

It has been noted that many of the genes thought to have been obtained through HGT exist in large gene families (Haegeman *et al.* 2011b). This could be due to multiple HGT event occurring, but it is more likely to be due to gene duplication events occurring following their initial acquisition. When a new beneficial gene is acquired, there may be positive selective pressure on those offspring that had multiple copies present in their genome. Over time this would lead to a nematode population with multiple copies of the same or similar genes (Danchin *et al.* 2010). With the presence of multiple gene copies there is the opportunity to diversify through evolution within these gene families. This can result in entirely new gene functions or specialisation of functions allowing similar genes to perform differently or on different targets (Lynch & Conery 2000). This may also lead to functional redundancy in certain cases. Over time homogenisation of genes acquired by HGT in nematodes has taken place. Examples of this homogenisation include changes to the GC content and codon usage (Danchin *et al.* 2010). It has also been shown that genes acquired by HGT in nematodes now contain introns. As bacterial species do not contain introns in their genomes this is a further example of adaptation of genes to their host genome. The presence of introns can increase the transcription level of a gene so this adaptation may be critical for gain of function (Husnik & McCutcheon 2018).

4.1.3 – CWDE modes of action

Many cell wall degrading enzymes have been categorised into different glycoside hydrolase (GH) families. Glycoside hydrolases are defined as proteins which hydrolyse the glycosidic bonds of glycosides. GH are classified into the various families on the basis of sequence similarity (Henrissat & Bairoch 1993, Henrissat & Bairoch 1996). Currently there are 167 individual GH (GH1-GH167) families categorised in the Carbohydrate Active Enzymes database (CAZY) (<http://www.cazy.org/>) (Lombard *et al.* 2014). Some GH families have been categorised into clans which have been determined by similarity of tertiary folding structures and are likely to share a common ancestry (Henrissat & Bairoch 1996). Currently there are clans GH-A through to GH-R.

In the majority of cases the GH enzymes catalyse the hydrolysis of the glycosidic bond in the substrate using two catalytic amino acids, usually glutamate or aspartate residues (Davies & Henrissat 1995). Hydrolysis of glycosidic bonds is achieved by either the inverting or retaining stereochemistry at the anomeric position. The inverting method is a single-displacement mechanism with one catalytic residue acting as a general acid and the other as a general base. The retaining method is double-displacement mechanism, which proceeds through a covalent intermediate; one catalytic residue acts as a general acid/base and the other as the nucleophile (Ardèvol & Rovira 2015). Glycoside hydrolases can also be classed as *endo*- or *exo*-acting, the former being able to cut glycosidic bonds in the middle of the chain and the latter cutting at the end of the chain.

4.1.4 – Glycoside hydrolase 53 family

Arabinogalactan endo- β -1,4-galactanases are members of the glycoside hydrolase 53 family and will be referred to as GH53s from this point onwards. The GH53 family resides in clan GH-A alongside 22 other GH families (Lombard *et al.* 2014). According to the CAZY database, GH53 enzymes from 6 species have been structurally characterised to date. Two of these species are bacteria: *Bacillus licheniformis* and

Bacteroides thetaiotaomicron, and four are fungi: *Aspergillus aculeatus*, *Aspergillus nidulans*, *Humicola insolens*, and *Thermothelomyces thermophilus*. All show a $(\beta/\alpha)_8$ barrel structure and this is presumed to be shared by other members of the family (Le Nours *et al.* 2003, Ryttersgaard *et al.* 2004, Böger *et al.* 2019, Ryttersgaard *et al.* 2002, Otten *et al.* 2013, Christgau *et al.* 1995). At the time of writing there are 3286 enzymes classified as GH53 family members across Archaea (3), Bacteria (3213), Eukaryota (67), as well as 3 “unclassified” sequences from “uncultured organisms” (Lombard *et al.* 2014).

The key catalytic residues of the GH53 family were first experimentally confirmed in the protein GalA (EMBL accession no. X91885) from *Cellvibrio japonicus* (previously known as *Pseudomonas fluorescens*). Both catalytic residues are glutamates: [E161] acting as the acid-base residue and [E270] as the nucleophilic catalytic residue, and was shown to use the retaining method of hydrolysis (Braithwaite *et al.* 1997). The function of GH53 enzymes has been defined as hydrolysing the β -1,4-glycosidic bonds found in type I arabinogalactan and was first determined in *Bacillus subtilis* (Emi & Yamamoto 1972). In the context of the plant cell wall, the GH53 endo- β -1,4-galactanases break down the glycosidic bonds of the backbone of arabinogalactan side chains found in the RGI regions of pectin. These work in tandem with other enzymes such as α -L-arabinofuranosidases which remove the L-arabinose subunits from the galactan backbone (Figure 4.3) (Seiboth & Metz 2011).

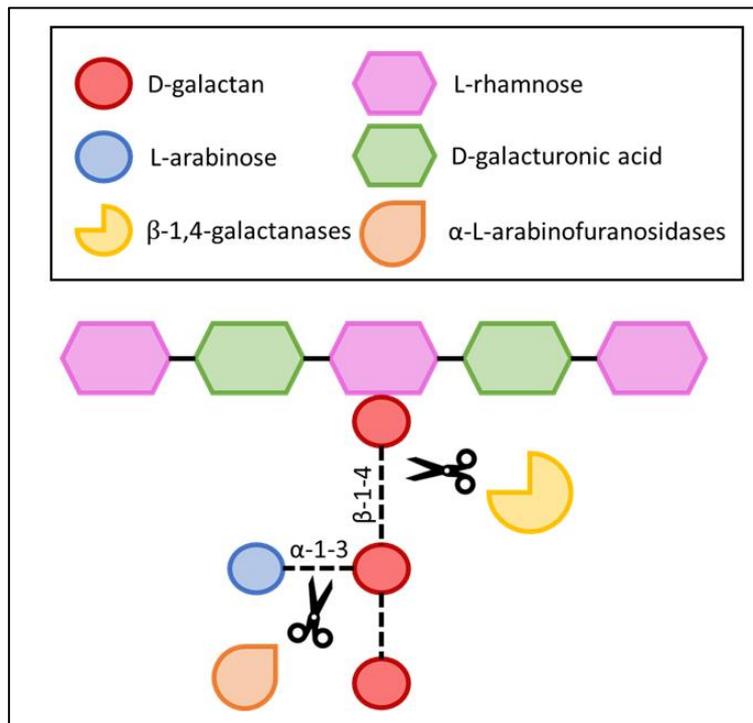


Figure 4.3 – Simplified schematic of the metabolism of rhamnogalacturonan I (type I) arabinogalactan side chain. β -1,4-glycosidic bonds in the backbone of arabinogalactan are broken down by β -1,4-galactanases (GH53) while L-arabinose subunits are debranched by other enzymes such as α -L-arabinofuranosidases.

4.1.4.1 – Arabinogalactan

Arabinogalactan can be classified into two classical types which are defined by differing positions of glycosidic bonds present in the backbone. Type I (AG-I) are arabino-1, 4-galactan meaning they have 1, 4 glycosidic bonds, while type II (AG-II) are arabino-3, 6-galactan, so have a galactose backbone with β -1, 3 glycosidic bonds linked with branches of galactose with β -1, 6 glycosidic bonds. AG-I and AG-II vary in abundance across different plant species. AG-I is more abundant in potato for example, while AG-II is present in high abundance in tree species such as the larch tree. AG-II from the larch tree is commonly found in dietary supplements to boost the immune system (Dion *et al.* 2016). Although similar in structure, the differing positions of the glycosidic bonds mean that different enzymes are required for hydrolysis of AG-I and AG-II arabinogalactan. For example, GH53 (endo- β -1,4-galactanases) break down the bonds of AG-I, while enzymes from the GH30 and

GH35 families (endo- β -1,6-galactanases and exo- β -1,3- or 1,6-galactosidase respectively) are required to hydrolyse AG-II (Knoch *et al.* 2014).

4.1.5 – GH53 family members in nematodes

Previously published genome and transcriptome data analysis have shown that GH53 proteins are present in PPN. The GH53 proteins present in nematodes are good candidates for having been obtained through HGT with the GH53 from *G. rostochiensis* having an AI > 349.3. The function of GH53 family members have only been experimentally characterised in 22 species; 10 bacterial, 10 eukaryotic (all fungal), and 2 from an unclassified “uncultured organism” according to the CAZY database at the time of writing (Lombard *et al.* 2014).

Vanholme *et al.* showed that *H. schachtii* contains a full length GH53 gene, termed HsGAL1, and a fragment of another gene, HsGAL2 (Vanholme *et al.* 2009). Currently available genome and transcriptome data show that GH53 genes are also present in the PPN species: *G. rostochiensis*, *G. pallida*, *R. reniformis*, *G. ellingtonae*, and *H. glycines*, as well as *H. schachtii* (Eves-Van den Akker *et al.* 2016a, Thorpe *et al.* 2014, Fosu-Nyarko *et al.* 2016, Pokhare *et al.* 2020). Figure 4.4 shows a phylogenetic tree constructed using Core Eukaryotic Genes Mapping Approach (CEGMA) genes by Pokhare *et al.* (2020). Looking at the genomes of species included in this study it was determined that each of those species included in the green box contained GH53 encoding genes. The GH53 genes in these nematodes are thought to have derived from a single HGT event to a common ancestor (indicated by a star in Figure 4.4). No GH53 proteins have been reported in any root-knot nematode species to date. These nematode species in which GH53 genes have been identified are all parasites of dicotyledon host plant species, except for *R. reniformis* which can parasitise both dicotyledon and monocotyledon plants. The arabinogalactan branches of pectin are found in high abundance in dicotyledon plant species, whereas there are few to no arabinogalactan side chains present in the pectin of monocots (Pattathil *et al.* 2015, Wefers *et al.* 2014). Phylogenetic analysis shows that parasitism of monocots has

arisen secondarily and that in this lineage GH53 encoding genes are absent, possibly reflecting the difference in composition of the cell walls of monocots (Pokhare *et al.*, 2020)

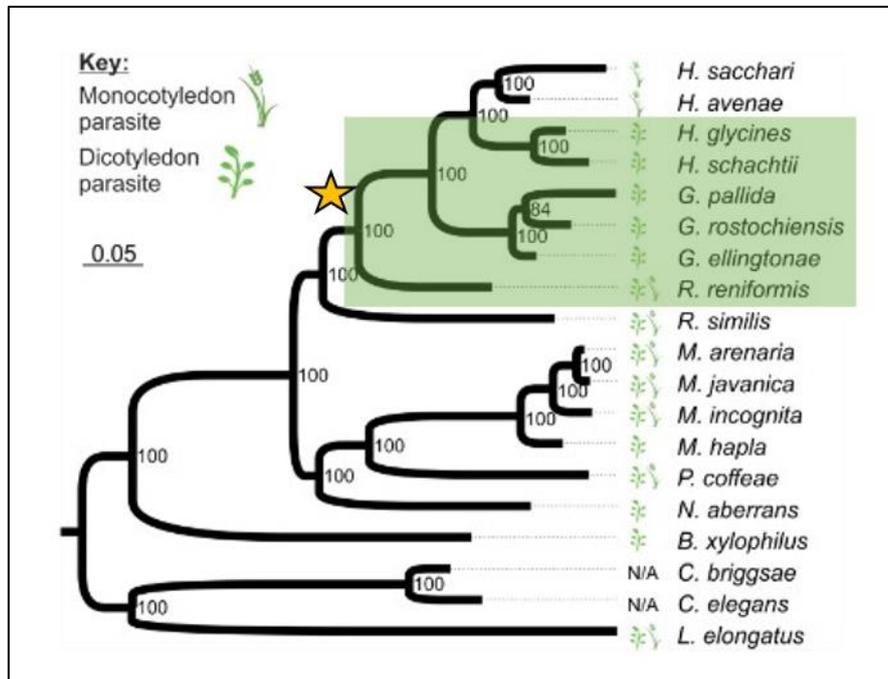


Figure 4.4 - Phylogenetic tree of nematode species comparing monocotyledon and dicotyledon parasitism. Species situated within the green highlighted box contain β -1,4-galactanase genes in their genome/transcriptome. Star (yellow) represents the likely occurrence of the HGT event where a common ancestor acquired the initial β -1,4-galactanase gene which led to the ability to break down dicotyledon cell wall material more easily. Figure adapted from (Pokhare *et al.* 2020).

4.1.6 – Chapter aims

A set of novel endo- β -1,4-galactanases, GrGAL1, GpGAL1, and RrGAL1, have been identified in the PPN species: *Globodera rostochiensis*, *Globodera pallida*, and *Rotylenchulus reniformis*. These proteins may assist in invasion of the host and migration through root tissue. Further characterisation of these effectors will give a greater insight into the invasion process and the co-evolution between the nematode and its host plant. The first aim of this chapter was to clone GrGAL1, GpGAL1, and RrGAL1 and use *in situ* hybridisation to ascertain whether these proteins are expressed in the gland cells and thus likely to act as effectors. The second aim is to conduct a phylogenetic analysis between nematode endo- β -1,4-galactanases and similar genes from fungi and bacteria. Finally, the GrGAL1, GpGAL1, and RrGAL1 proteins were purified and expressed to determine if they have the predicted biochemical activity.

4.2 – Materials and Methods

4.2.1 – Cloning

The coding regions of putative GH53 genes from *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *X. campestris* (GROS_g08150 (GrGAL1), Pal _4850 (GpGAL1), Ren30258 (RrGAL1) and XC_0587 (Gala_Xc)) were amplified by PCR from cDNA using the proof-reading KOD Hot Start DNA polymerase (Merck), as described in section 2.5 using primer sets shown in Table 4.1. The open reading frame of each of the genes was cloned from start to stop codon excluding the endogenous signal peptide. PCR products were separated on 1.5% agarose gels, excised, and purified using a QIAquick gel extraction kit (Qiagen). Direct thymine-alanine (TA)-cloning was conducted into the Gateway-compatible TOPO entry vector pCR8/GW/TOPO (Invitrogen). TOPO constructs were transformed into *E. coli* DH5 α competent cells by electroporation. Transformants were selected on LB medium (tryptone, yeast extract, NaCl, pH 7.0 with 5N NaOH) agar plates supplemented with spectinomycin (100 μ g/ml) and grown overnight at 37 °C. Colonies containing inserts of the anticipated size in the correct orientation were identified by colony PCR using a combination of M13 and gene specific primers. The isolated bacteria were used to seed 5 ml overnight cultures in LB supplemented with spectinomycin (100 μ g/ml). Plasmids were extracted from cultures using a GeneJet plasmid preparation kit (ThermoFisher) following the manufacturer's instructions. Plasmid clones were sequenced at The James Hutton Institute sequencing facility. 60% glycerol stocks were produced for long-term storage at -80 °C. Alignments of the cloned genes were conducted using CLUSTAL multiple sequence alignment (MUSCLE). It should be noted that three truncated GH53-like sequences from *N. aberrans* have also been identified (51041_c0_seq1, 6113_c0_seq1, and 23294_c0_seq1). These could not be cloned and were not investigated further.

GROS_g08150 (GrGAL1), Pal _4850 (GpGAL1), Ren30258 (RrGAL1), and XC_0587 (Gala_Xc) were subsequently cloned into the protein expression vector pOPIN_S3C. The pOPIN_S3C vector contains a 6x Histidine (His) tag, a SUMO domain, and a 3C

protease cleavage site. GFP, previously cloned into the bacterial expression vector pKC026, was also cloned into the pOPIN_S3C vector for use as a negative control. Cloning into the pOPIN_S3C vector was carried out using the NEBuilder HiFi DNA assembly kit following the manufacturer's protocol (New England Biolabs). Primer sets for pOPIN_S3C cloning are shown in Table 4.1.

Table 4.1 – Primer sets used for cloning & in situ hybridisation of GH53 genes

Species	Primer name	Sequence	Length (bp)	Function
<i>G. rostochiensis</i>	08150_F	ACCATGCTGTACAAAGGTGC	20	Cloning into TOPO vector
<i>G. rostochiensis</i>	08150R_nostop	TTGGTAATTGAACGCTGCATC	22	Cloning (absence of stop codon) into TOPO vector
<i>G. rostochiensis</i>	08150R_stop	TTATTGGTAATTGAACGCTGTC	22	Cloning (presence of stop codon) into TOPO vector
<i>G. rostochiensis</i>	IF-EXTEND-GrGH53-F	AAGTTCTGTTTCAGGGCCCG ACCATGCTGTACAAAGGTGC	40	Cloning
<i>G. rostochiensis</i>	IF-EXTEND-GrGH53-R	GACAGCGTTCAATTACCAA ATGGTCTAGAAAGCTTTA	37	Cloning
<i>G. rostochiensis</i>	Rosg08150_ISUHF	TTTGTGCTTCTGAAGTCGTTTG	22	<i>In situ</i> hybridisation
<i>G. rostochiensis</i>	Rosg08150_ISUHR	GACGAGTTGTGACAGCGAAT	20	<i>In situ</i> hybridisation
<i>G. rostochiensis</i>	Rosg08150_ISHF2	TTCTGTTCACTGGTGGAGGC	20	<i>In situ</i> hybridisation
<i>G. rostochiensis</i>	Rosg08150_ISHR2	TTTGGCCTTGTGCAACGTG	19	<i>In situ</i> hybridisation
<i>G. rostochiensis</i>	Rosg08150_ISHF3	ACTACCTGAAGAGCAACGGC	20	<i>In situ</i> hybridisation
<i>G. rostochiensis</i>	Rosg08150_ISHR3	GAACCACGTGAAATCGGC	18	<i>In situ</i> hybridisation
<i>G. rostochiensis</i>	Rosg08150_ISHF4	GGAGTGAAGAAGGCCGGTG	19	<i>In situ</i> hybridisation

<i>G. rostochiensis</i>	Rosg08150_ISHR4	CTCACGCGTTTCAGCATGTC	20	<i>In situ</i> hybridisation
<i>G. pallida</i>	FOR_4850	ACCATGCTGTACAAAGGTGCCGA TGTC	27	Cloning
<i>G. pallida</i>	REV_4850_nostop	ACTGAACGCTTTCATCGCCT	20	Cloning (absence of stop codon)
<i>G. pallida</i>	REV_4850_stop	ATGTTAACTGAACGCTTTCATCG	23	Cloning (presence of stop codon)
<i>G. pallida</i>	Pal4850_ISUHF	AACTGATCCATTTGTGCTTCTG	22	<i>In situ</i> hybridisation
<i>G. pallida</i>	Pal4850_ISUHR	ACACGTCCGTGACGAGTTG	19	<i>In situ</i> hybridisation
<i>G. pallida</i>	Pal4850_ISHF2	ACGGCAAGAAGGTGATGGTG	20	<i>In situ</i> hybridisation
<i>G. pallida</i>	Pal4850_ISHR2	CATCGCCTCCGTGAATTTGC	20	<i>In situ</i> hybridisation
<i>G. pallida</i>	Pal4850_ISHF3	GCACACCTACGGCATTTTGA	20	<i>In situ</i> hybridisation
<i>G. pallida</i>	Pal4850_ISHR3	TGTTGATCAGACTCGCCAGG	20	<i>In situ</i> hybridisation
<i>R. reniformis</i>	FOR_830247817	ACCATGCTCACAAACGGGTGCCG	22	Cloning
<i>R. reniformis</i>	830247817_REV_nos top	TAGCGCACTCAATGCCTC	18	Cloning (absence of stop codon)
<i>R. reniformis</i>	830247817_REV_sto p	TCTGATCATAGCGCACTCAA	20	Cloning (presence of stop codon)
<i>R. reniformis</i>	Ren30258_ISUHF	TTGTGAACCCGCCAGATG	18	<i>In situ</i> hybridisation
<i>R. reniformis</i>	Ren30258_ISUHR	AACCTGCACCCAATCCAC	18	<i>In situ</i> hybridisation

4.2.2 – Phylogenetic analysis

BLAST similarity searches were conducted using the BLASTp function with GrGAL1, GpGAL1, and RrGAL1 as query sequences (Altschul *et al.* 1990). The top 100 results from each of the BLAST searches were combined and filtered to remove duplicates and low confidence hits, resulting in a unique list of 78 sequences. Pairwise

alignments of all genes were created using Jalview (Waterhouse *et al.* 2009). A phylogenetic tree was produced using IQ-TREE (Nguyen *et al.* 2015). All amino acid sequences were gap trimmed before a Bayesian tree was constructed using a LG (general matrix) + F (empirical base frequency) + G4 (rate heterogeneity gamma model) model as determined by ModelFinder (Kalyaanamoorthy *et al.* 2017). This phylogenetic tree was produced with 1000 bootstraps (Hoang *et al.* 2018). The tree was produced using FigTree v1.4.3.

4.2.3 – Protein structure prediction

Predicted structures of GrGAL1, GpGAL1, and RrGAL1 proteins were produced using a 1-to-1 thread model based on the amino acid sequence of β -1,4-galactanase from *Bacteroides thetaiotaomicron*. The sequence and structure from *B. thetaiotaomicron* were identified using BLAST similarity searches with an E-value cut-off of 1 and mask low complexity settings applied. Searches were completed using BLAST and the RCSB PDB (Berman *et al.* 2000) (www.rcsb.org). The 1-to-1 thread model was achieved using Protein Homology/analogy Recognition Engine V 2.0 (PHYRE2) (Kelley *et al.* 2015). Rendered images of predicted protein structures were produced using CCP4 molecular graphics (CCP4mg) (V2.10.10) (McNicholas *et al.* 2011).

4.2.4 – Protein expression and purification

Ten millilitres LB was inoculated with a single colony from either GrGAL1, GpGAL1, RrGAL1, GalA_Xc, or the GFP control in Shuffle *E. coli* cells (New England Biolabs). Cultures were grown overnight at 37 °C with shaking. One hundred microlitres of the overnight cultures were added to 100 ml of fresh LB media which were incubated at 30 °C with shaking until an OD₆₀₀ of 0.7 was reached. Cultures were cooled to 18 °C. Forty microlitres of each culture was taken out and stored as a pre-induction control. Pre-induction control (1) samples were mixed with 15 μ l 4x loading dye and 3 μ l 500 mM DTT and boiled for 5 minutes before storage at 4 °C overnight (or -20 °C for long term storage). Expression of protein in the remaining cultures was induced by

the addition of 1 M IPTG to a final concentration of 1 mM, before incubation overnight at 18 °C with shaking. A 20 µl sample of each culture was taken as a post-induction control (2) and stored at 4 °C overnight (or -20 °C for long term storage). Cultures were centrifuged at 4000 g for 10 minutes to pellet the remaining cells. Supernatant was removed and the pellet was resuspended in 1 ml of ice cold lysis and wash buffer (50 mM Tris-HCL, 500 mM NaCl, 50mM glycine, 5% glycerol, 20 mM imidazole, EDTA-free protease inhibitor tablets pH 8.0). Cells were sonicated (15 µm) for 30 seconds followed by 30 seconds cooling on ice – this cycle was repeated 6 times. The lysate was centrifuged at 13000 g for 2 minutes. Twenty microlitres of the supernatant was taken and stored at 4 °C as a soluble protein sample (3) (or -20 °C for long term storage). The rest of the supernatant was added to a new Eppendorf tube with 100 µl of Ni-NTA Superflow resin (Qiagen) and incubated for 1 hour at room temperature on a rotator. The sample was centrifuged twice at 13000 g for 1 minute and the supernatant was removed. Twenty microlitres of the flow through was retained and termed wash sample before storage at 4 °C overnight (or -20 °C for long term storage). Beads were washed in 1 ml lysis and wash buffer, centrifuged at 13000 g x 1 minute before removal of supernatant. 2.5 ml of elution buffer (50 mM Tris-HCL, 500 mM NaCl, 50 mM glycine, 5% glycerol, 500 mM imidazole, EDTA-free protease inhibitor tablets, pH 8.0) was added to the beads. Beads were incubated at room temperature for 10 minutes and then centrifuged twice at 13000 g for 1 minute. The supernatant, containing the purified protein of interest, was stored at 4 °C overnight (or -20 °C for long term storage).

4.2.5 – SDS-PAGE gels

Gel electrophoresis was used to visualise 10 µl of pre-induction (cell lysate) (1), post-induction (cell lysate) (2), soluble protein (post lysis) (3), wash (flow through from Ni-NTA beads) (4), and purified protein (5) samples for GrGAL1, GpGAL1, RrGAL1, GalA_Xc, and the GFP control on polyacrylamide gels (Precast NuPAGE 4-12% Bis-Tris gel, Invitrogen) following the manufacturer's instructions. Samples

were electrophoresed alongside the prestained PageRuler Plus Protein Standard ladder (ThermoScientific) at 150 V for 45 minutes.

4.2.6 – Coomassie blue staining

Gels were washed in SDW for 5 minutes twice to remove any remaining Lithium dodecyl sulphate (LDS) detergent. Gels were then incubated in Coomassie blue (0.1% Coomassie Blue R-250, 40% methanol, 10% acetic acid) for 2 hours at room temperature with gentle shaking. Gels were transferred into destain solution (20% methanol, 10% acetic acid) for 1-2 hours until bands were visible with minimal background staining. Destaining was conducted at room temperature with gentle shaking.

4.2.7 – Azo-galactan substrate assay

Five hundred microlitres of enzyme solution (in elution buffer as described in 4.2.4) was added to 500 μ l (2 % w/v) azo-galactan substrate solution (2 g azo-galactan (Megazyme), 90 ml SDW, 5 ml elution buffer). Samples were mixed using a vortex before incubation at 40 °C for 10 minutes. Temperature, time, and pH were all altered to ascertain optimum conditions (10 - 60 °C, 10 - 120 mins and pH 2-8) in subsequent experiments. Two thousand five hundred microlitres of 95% ethanol was added followed by brief vortex mixing to terminate the reaction. Samples were left to equilibrate for 10 minutes at room temperature and centrifuged at 3000 rpm for 10 minutes to pellet high molecular weight substrate. The supernatant was added to a cuvette and absorbance of each sample was measured at 590 nm using a spectrophotometer. A blank sample was prepared by adding an additional 500 μ l of ethanol to 500 μ l of substrate solution, instead of 500 μ l of enzyme sample.

4.2.8 – DNS Assay

This assay was based on the “enzymatic assay of β -amylase (EC3.2.1.2)” protocol from Sigma Aldrich. Volumes were adjusted to adapt the assay for use in 96 well plates from cuvettes. Twenty five microlitres of each enzyme: GrGAL1, GpGAL1,

RrGAL1, GalA_Xc, or the GFP control (500 µg/ml), was incubated with 25 µl of 0.1% w/v galactan substrate solution in 50 mM sodium acetate (pH 5) at room temperature for 1 hour. A blank sample was also set up containing 25 µl SDW and 25 µl of galactan solution. Fifty microlitres of DNS reagent (60 mM 3,5-dinitrosalicylic acid, 500 mM NaOH, 150 mM potassium sodium tartrate tetrahydrate) was added to each sample. Samples were boiled for 15 minutes before cooling for 3 minutes on ice. Two hundred microlitres of SDW was added to each sample. Absorption readings for each sample were taken at 540 nm using a Promega GloMax multi+ multiplate reader.

The additional substrates xylan (from beechwood, Sigma), saccharose (VWR chemicals), pectin (from apple, Sigma), arabinogalactan (AG-II) (from Larchwood, Sigma) and polygalacturonic acid (Sigma) were tested using the same protocol as above. All substrates were used at 0.1% w/v solutions in 50 mM sodium acetate pH 5 as described above. A standard curve was produced using 1 mg/ml (0.1% w/v) galactose (the reducing sugar produced on hydrolysis of galactan polymer). Seven samples were used containing 0.05, 0.2, 0.4, 0.6, 0.8, 1, and 2 mg galactose plus DNS reagent. A blank sample containing SDW and DNS reagent only was also tested.

4.2.9 – *In situ* hybridisation & gene expression

In situ hybridisation was conducted as described in section 2.8. Primers used for *in situ* hybridisation are shown in Table 4.1. Information on the gene expression data used can be found in section 3.2.1.4. Average gene expression graphs for each species were produced from 2 replicates of each life stage for *G. rostochiensis* and *G. pallida*, and 3 replicates per life stage for *R. reniformis*.

4.3 – Results

4.3.1 – GH53 gene identification and sequence analysis

The *G. rostochiensis* GrGAL1 (GROS_g08150) gene (identified as part of the genome project for this species and which showed high similarity to other GH53 proteins (Eves-Van den Akker *et al.* 2016a)) was used as a query sequence for BLAST similarity searches to identify putative GH53 proteins in *G. pallida* and *R. reniformis*. The search against the *G. pallida* transcriptome returned two incomplete but overlapping sequences: comp4850_c0_seq1 and comp4850_c0_seq4. These were computationally recapitulated and given the working title Pal_4850 before being renamed GpGal1 upon successful cloning of the gene (Figure 4.5). The putative GH53 from *R. reniformis* was identified as three incomplete overlapping sequences (from transcriptome and genome databases); comp30258_c0_seq1, comp30258_c1_seq1 and gi|830247817|gb|LDKF01001503.1. All three sequences were computationally recapitulated and given the working title of Ren30258 before being renamed RrGal1 (Figure 4.6). Based on these computational predictions, primers were designed to amplify the full-length sequence of each gene from nematode cDNA. After cloning and the gene sequence confirmed, the proteins were named GrGAL1, GpGAL1, and RrGAL1 (*G. rostochiensis*, *G. pallida*, and *R. reniformis* respectively) as shown in Table 4.2.

comp4850_c0seq4	1	CCGGCATTCAAAAAATTAATAATATTTAAATTAATTTCTTCATGCATTTTTTGAAC TTC
GpGAL1	1	-----
comp4850_c0seq1	1	-----
comp4850_c0seq4	61	ACATTTTTATGCC TTTCTGTTCACTGGTGGGGCAGCGCATGTTTGTGCACTGTACAAA
GpGAL1	1	-----ACCATGCTGTACAAA
comp4850_c0seq1	1	-----
comp4850_c0seq4	121	GGTGGCGATGTCAGCTGGGTGACCCAACAGGAATTCGAAAGGCCAATCATTTTACACAGT
GpGAL1	16	GGTGGCGATGTCAGCTGGGTGACCCAACAGGAATTCAGGAGGCCAATCATTTTACACAGT
comp4850_c0seq1	1	-----
comp4850_c0seq4	181	CCGGCACC AAACTGATCCATTTGTGCTTCTGAAGTCGTTTGGCATCAACGGGTGGGA
GpGAL1	76	CCGGCACC AAACTGATCCATTTGTGCTTCTGAAGTCGTTTGGCATCAACGGGTGGGA
comp4850_c0seq1	1	-----
comp4850_c0seq4	241	TTGCGCGTTTGGGTGAACCCGTCGCGGGATGGAATGACGGGGCGGACAGTTGCACAAG
GpGAL1	136	TTGCGCGTTTGGGTGAACCCGTCGCGGGATGGAATGACGGGGCGGACAGTTGCACAAG
comp4850_c0seq1	1	-----
comp4850_c0seq4	301	GCCAAACGGGCGATGGCCAGGGCATGGCCATCATGATCGACTTCCACTACAGCGACACG
GpGAL1	196	GCCAAACGGGCGATGGCCAGGGCATGGCCATCATGATCGACTTCCACTACAGCGACACG
comp4850_c0seq1	1	-----
comp4850_c0seq4	361	TGGGCAGACCCGGCCACACAGCGGTCCCGTCCGCATGGGAAAGCCATTGCGTGTCAA
GpGAL1	256	TGGGCAGACCCGGCCACACAGCGGTCCCGTCCGCATGGGAAAGCCATTGCGTGTCAA
comp4850_c0seq1	1	-----
comp4850_c0seq4	421	CTCGTCACGGACGTGTACCAGCACACTACGGCATTTTGAAC TACCTGAAAAGCAACGGC
GpGAL1	316	CTCGTCACGGACGTGTACCAGCACACTACGGCATTTTGAAC TACCTGAAAAGCAACGGC
comp4850_c0seq1	1	-----
comp4850_c0seq4	481	ATCAGAGTCACATGGGTCCAGGTCCGCAACGAGATCAACGGTGG-----
GpGAL1	376	ATCAGAGTCACATGGGTCCAGGTCCGCAACGAGATCAACGGTGGCATGCTTTGGCCAAAT
comp4850_c0seq1	1	-----GCTCACATGGGTCCAGGTCCGCAACGAGATCAACGGTGGCATGCTTTGGCCAAAT
comp4850_c0seq4		-----
GpGAL1	436	CGGAAAACACCAATTTTGCAAACTTCCCAAGTCTGATCAACAGTGGCTACAAGGCATCC
comp4850_c0seq1	57	CGGAAAACACCAATTTTGCAAACTTCCCAAGTCTGATCAACAGTGGCTACAAGGCATCC
comp4850_c0seq4		-----
GpGAL1	496	AAAGCGGTGTACCCAAATGCACCGGTGATTTTACATTTGGCCAACGGCTACAAAATGCC
comp4850_c0seq1	117	AAAGCGGTGTACCCAAATGCACCGGTGATTTTACATTTGGCCAACGGCTACAAAATGCC
comp4850_c0seq4		-----
GpGAL1	556	GATTTCAAGTGGTTCCTTTGACGGGGTGAAGAAGCCGGTGGCAGTTGGGACGTCATTGGC
comp4850_c0seq1	176	GATTTCAAGTGGTTCCTTTGACGGGGTGAAGAAGCCAAATGGCAGTTGGGACGTCATTGGC
comp4850_c0seq4		-----
GpGAL1	616	ATTTCCCATTTATCCGACATCCGCAAAATGGAAGACACTGAACGCGCAGGCGGCAACAACG
comp4850_c0seq1	236	ATTTCCCATTTATCCGACATCCGCAAAATGGAAGACACTGAACGCGCAGGCGGCAACAACG
comp4850_c0seq4		-----
GpGAL1	676	TTGAGAACAATGATCAGTCGGTACGGCAAGAAGGTGATGGTGGCCGAAATCGGCATGCCA
comp4850_c0seq1	296	TTGAGAACAATGATCAGTCGGTACGGCAAGAAGGTGATGGTGGCCGAAATCGGCATGCCA
comp4850_c0seq4		-----
GpGAL1	736	TGGGACGAGCCAGTCTCTGCAAGTCAATGTTCAGGACATGCTGAAACGCGTGAAGGCA
comp4850_c0seq1	356	TGGGACGAGCCAGTCTCTGCAAGTCAATGTTCAGGACATGCTGAAACGCGTGAAGGCA
comp4850_c0seq4		-----
GpGAL1	796	TTGGGCAGTAATGGCATCGGCGTGTCTACTGGGAGCCACAGGCGACGCCCGGCTGGAAT
comp4850_c0seq1	416	TTGGGCAGTAATGGCATCGGCGTGTCTACTGGGAGCCACAGGCGACGCCCGGCTGGAAT
comp4850_c0seq4		-----
GpGAL1	856	CACTACCAGTGGAGCGCATTTGGACAAC TCCGCAAAATTCACGGAGGCGATGAAAGCGTTC
comp4850_c0seq1	476	CACTACCAGTGGAGCGCATTTGGACAAC TCCGCAAAATTCACGGAGGCGATGAAAGCGTTC
comp4850_c0seq4		-----
GpGAL1	916	AGT TAA
comp4850_c0seq1	536	AGT TAA CATTAAAATGACTGGATTTGGATTGAAATGTTCAATAAATACCCCTGAGATAGTA

Figure 4.5 - Computationally recapitulated *G. pallida* sequences (comp4850_c0_seq1 and comp4850_c0_seq4) alignment with the confirmed cloned gene sequence for GpGAL1. Signal peptide is highlighted in green. Stop codon highlighted in red. 3' Untranslated region (UTR) is highlighted in blue, and the 5' UTR is highlighted in yellow.

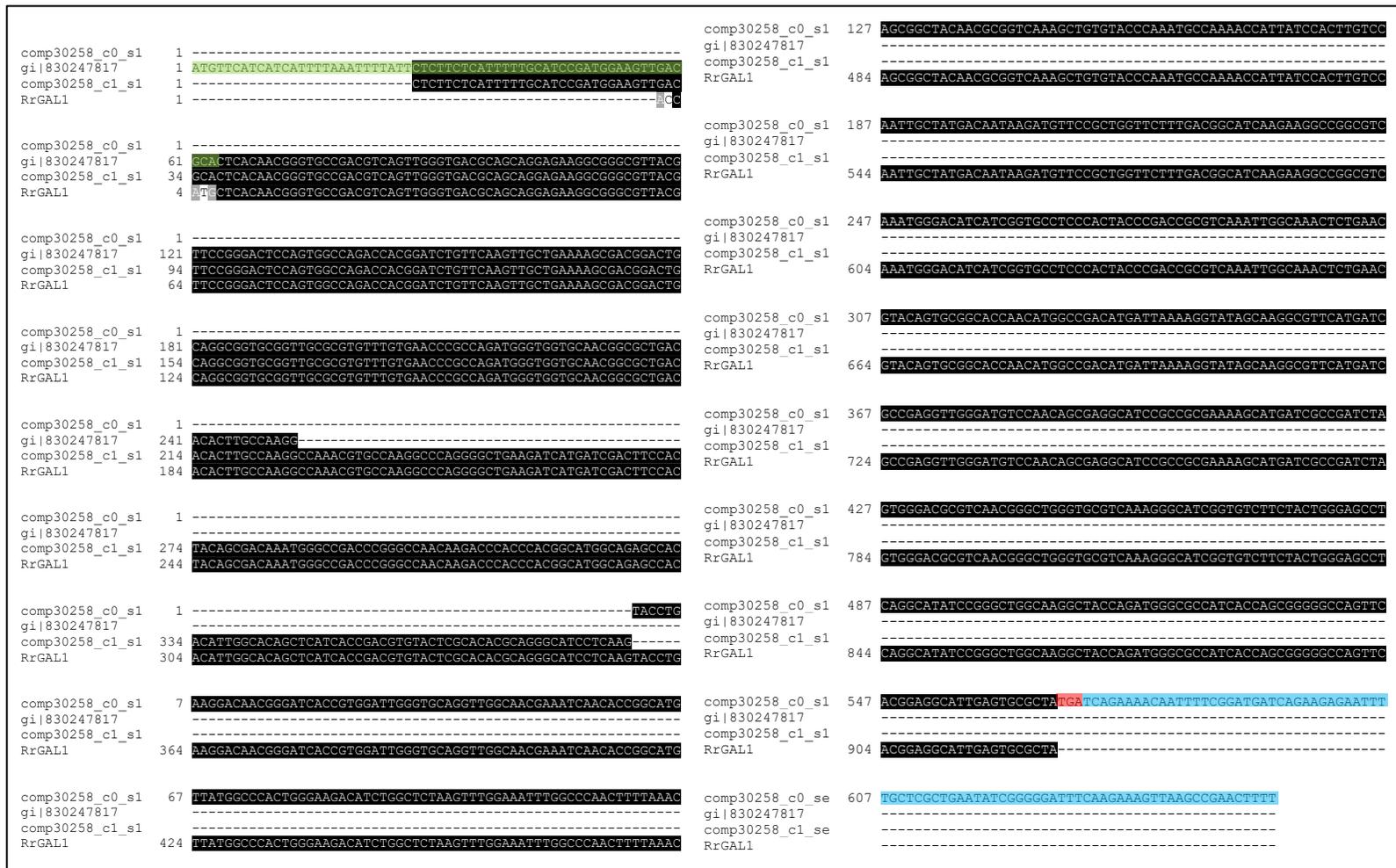


Figure 4.6 - *R. reniformis* computationally recapitulated sequences. Comp30258_c0_seq1, Comp30258_c1_seq1, and gi|830247817 alignment with the confirmed cloned gene sequence for RrGAL1. Signal peptide is highlighted in green. Stop codon highlighted in red. 3' Untranslated region (UTR) is highlighted in blue.

Table 4.2 – Nomenclature of nematode GH53 sequences

Species	Sequence IDs	Final name	Formerly known as – (matching primer names)
<i>G. rostochiensis</i>	GROS_g08150	GrGAL1	g08150
<i>G. pallida</i>	comp4850_c0_seq1, comp4850_c0_seq4	GpGAL1	Pal_4850
<i>R. reniformis</i>	comp30258_c0_seq1, comp30258_c1_seq1, gi 830247817 gb LDKF01001503.1 <i>Rotylenchulus reniformis</i> RREN_1503	RrGAL1	Ren30258

GrGAL1, GpGAL1, and RrGAL1 were used as a query for BLAST sequence similarity searches (BLASTp – non-redundant (nr) database) and pairwise alignment analysis (BLOSUM62). GrGAL1 and GpGAL1 share high sequence percentage identity with a GH53 protein from *Duganella sacchari* (67.11%) (NCBI seqID: WP_072787792.1). *D. sacchari* is a Gram-negative, soil dwelling bacteria, strains of which have been isolated from the rhizosphere of sugar cane plants (Madhaiyan *et al.* 2013). GrGAL1 also shared high percentage identity with GalA, a GH53 protein from *Xanthomonas campestris* (55.37%) (NCBI seqID: WP_011038708.1). *Xanthomonas campestris* is a Gram-negative bacterial pathogen of plant species such as tomato and peppers (Potnis *et al.* 2015). RrGAL1 had the highest percentage identity with an arabinogalactan endo-1,4- β -galactosidase from the *sp.* YR242 subspecies of *Roseateles* bacteria (65.68%) (NCBI seqID: WP_092947600.1). GrGAL1, GpGAL1, and RrGAL1 were aligned against the GH53 mature protein sequences from *X. campestris*, *D. sacchari*, and *R.sp.* YR242 as well as HsGAL1 from the nematode species *H. schachtii* (Vanholme *et al.* 2009). This alignment showed that the key catalytic glutamate (E) residues are conserved across PPN and bacterial species (Figure 4.7).

HsGAL1	1	-----M ^V YK ^G AD ^S SE ^V TQ ^O E ^K SG-- ^Q SFLD ^N NG ^K K ^T D ^L F ^A L ^M K ^S Y ^G M ^N AV ^R L ^R V ^V W ^N PA
RrGAL1	1	-----ML ^T T ^G AD ^V SW ^V TQ ^O E ^K AG-- ^V TFR ^D SS ^C Q ^T T ^D L ^F K ^L L ^K SD ^G L ^Q AV ^R L ^R V ^V W ^N PP
Gala_Xc	1	----MQ ^G LAK ^G AD ^V SW ^I N ^Q QAA ^N PP ^Q V ^F QD ^A SG ^K T ^T D ^F L ^K L ^F K ^D V ^G GN ^A L ^R L ^R V ^V W ^N P ^Q
GrGAL1	1	-----M ^L YK ^G AD ^V SW ^V TQ ^O E ^F FRG-- ^Q L ^F Y ^N S ^A G ^T K ^T D ^P F ^V L ^L K ^S F ^G I ^N AV ^R L ^R V ^V W ^N PS
GpGAL1	1	-----M ^L YK ^G AD ^V SW ^V TQ ^O E ^F FRG-- ^Q L ^F Y ^N S ^A G ^T K ^T D ^P F ^V L ^L K ^S F ^G I ^N AV ^R L ^R V ^V W ^N PS
Gala_Ds	1	----MAT ^F ANG ^A D ^V SW ^V S ^Q EE ^S SG-- ^Y AF ^Y NS ^S SV ^K T ^D P ^F V ^L L ^K N ^L Q ^V NA ^T R ^L R ^V V ^V W ^N PS
Gala_Rsp.YR	1	MPALAAT ^L LK ^G AD ^V SW ^V S ^Q EE ^S AG-- ^Y S ^F Y ^N S ^A G ^T K ^T D ^P F ^K L ^L SD ^L GV ^N T ^I R ^L R ^V V ^V SP ^S
HsGAL1	53	GGWC ^N KV ^D TL ^N KAK ^R AKA ^Q GMA ^V MID ^F HY ^A DSW ^A DPG ^K Q ^P IP ^S AW ^K G ^H SL ^D Q ^L V ^T DV ^Y K ^H
RrGAL1	53	DGWC ^N GAD ^T LAKAK ^R AKA ^Q G ^L KIM ^I D ^F HY ^S D ^K WAD ^P GQ ^Q DE ^P TAW ^O SH ^I L ^A Q ^L I ^T DV ^Y SH
Gala_Xc	57	GGW ^N DGR ^D TL ^D KAK ^R AAA ^Q GMR ^I MID ^F HY ^S DSW ^A DPG ^K Q ^T KPA ^A WASH ^S V ^A Q ^L N ^T DV ^Y NH
GrGAL1	53	GGW ^N DGAD ^T LHKAK ^R AMA ^Q GMA ^I MID ^F HY ^S D ^I WAD ^P AH ^Q T ^V PSAW ^E SH ^S LS ^Q L ^V T ^D V ^Y QH
GpGAL1	53	GGW ^N DGAD ^T LHKAK ^R AMA ^Q GMA ^I MID ^F HY ^S D ^I WAD ^P AH ^Q T ^V PSAW ^E SH ^S LS ^Q L ^V T ^D V ^Y QH
Gala_Ds	55	GGW ^N DGAD ^V L ^Y KAK ^R AAA ^Q GQ ^R I ^I DF ^H Y ^S DSW ^A DPG ^Q Q ^T KPA ^A WASH ^T L ^I Q ^L N ^S DV ^Y SH
Gala_Rsp.YR	59	GGW ^C DGAD ^T L ^Y KAK ^R AVA ^Q GQ ^K I ^M S ^F HY ^S DSW ^A DPG ^K Q ^T KPA ^A W ^S SH ^I LS ^Q L ^V T ^D V ^Y SH
HsGAL1	113	TYE ^V L ^S YL ^K S ^N G ^I S ^V LW ^V Q ^V GN ^E INN ^G ML ^W P ^T L ^K R-- ^P NFA ^A ISK ^L INS ^G Y ^K ASK ^A V ^Y PN
RrGAL1	113	TQGI ^L KYL ^K D ^N G ^I TVD ^W VQ ^V GN ^E INN ^G ML ^W P ^T G ^K TSG ^S K ^F EN ^L AQ ^L INS ^G Y ^N AW ^K AV ^Y PN
Gala_Xc	117	TQGI ^L KYL ^K D ^N G ^I TVT ^W VQ ^V GN ^E INS ^G ML ^W DQ ^G KT-- ^P NFAN ^L CQ ^F INS ^G Y ^N AW ^K AV ^Y PN
GrGAL1	113	TYGI ^L NYL ^K S ^N G ^I R ^V TW ^V Q ^V GN ^E ING ^G ML ^W P ^N GKT-- ^I NFAN ^L AS ^L INS ^G Y ^K ASK ^A V ^Y PN
GpGAL1	113	TYGI ^L NYL ^K S ^N G ^I R ^V TW ^V Q ^V GN ^E ING ^G ML ^W P ^N GKT-- ^I NFAN ^L AS ^L INS ^G Y ^K ASK ^A V ^Y PN
Gala_Ds	115	TYGI ^L NYL ^K ING ^I TV ^S W ^V Q ^V GN ^E INS ^G ML ^W PE ^G KT-- ^T SE ^S NLAG ^L IN ^N GY ^S AA ^K AV ^Y PN
Gala_Rsp.YR	119	TQGI ^L SYL ^K ING ^I TVD ^Y VQ ^V GN ^E INS ^G ML ^W P ^T QAS ^G SS ^F AN ^L VQ ^L INS ^G Y ^D AW ^K AV ^Y PN
HsGAL1	171	ALVI ^L HLAS ^G Y ^R TA ^Q FNA ^V FE ^A LKK ^A G ^T NY ^D AV ^G ISH ^Y PN ^A KN ^W Q ^L NA ^O SE ^I TM ^K Q ^M IS
RrGAL1	173	AKTI ^L HL ^S NC ^Y DN ^K M ^F R ^W FFD ^G IK ^K AG ^V K ^W D ^I GASH ^Y PTAS ^N W ^Q T ^L NV ^Q CG ^T NM ^A DM ^I K
Gala_Xc	175	AKVI ^L HLANG ^Y DNAN ^F R ^W FFD ^N LR ^S AG ^G K ^W D ^V IG ^M SH ^Y PP ^V GS ^W D ^Q NR ^L D ^L RD ^M IS
GrGAL1	171	APVI ^L HLANG ^Y KTAD ^F K ^W FFD ^G V ^K KAG ^A SW ^D VIG ^I SH ^Y PT ^S AN ^W K ^T LNA ^O AA ^T L ^R T ^M IS
GpGAL1	171	APVI ^L HLANG ^Y KTAD ^F K ^W FFD ^G V ^K KAG ^A SW ^D VIG ^I SH ^Y PT ^S AN ^W K ^T LNA ^O AA ^T L ^R T ^M IS
Gala_Ds	173	AQVI ^L HLANG ^Y DNAN ^F R ^W FFD ^G V ^K AG ^A K ^W D ^V IG ^M SH ^Y PP ^A AD ^W AS ^Y N ^S K ^L S ^T NM ^W DM ^V A
Gala_Rsp.YR	179	AKVI ^L HL ^S NG ^Y DNAL ^F R ^W FFD ^G M ^K ANS ^A K ^Y D ^V IG ^M SH ^Y PT ^S SN ^W S ^T LNA ^O LS ^T NM ^A DM ^I K
HsGAL1	231	KFG ^K K ^V V ^A E ^T GM ^K WT ^D ADACE ^A ML ^A DM ^F K ^R V ^V AL ^G - ^N NG ^I GV ^F Y ^E W ^E F ^N S ^A F-- ^Y GH ^Q MG
RrGAL1	233	RYSK ^A E ^M TA ^E VG ^M SN ^S E ^A AA ^K SM ^L AD ^L VGR ^V N ^L LG ^A SK ^G I ^G V ^F Y ^E W ^E PQ ^A Y ^P GW ^Q Y ^Q MG
Gala_Xc	235	RYG ^S DV ^V VA ^E VG ^M DW ^Q QA ^T TRAM ^L EN ^L IT ^R SN ^A IG-SR ^M IG ^I F ^Y W ^E F ^N AY ^P GW ^Q Y ^T MG
GrGAL1	231	RYG ^K K ^V M ^V VA ^E IG ^M PW ^D EP ^S V ^C K ^S ML ^Q D ^M L ^K R ^V AL ^G - ^S NG ^I GV ^F Y ^E W ^E PQ ^A T ^P GW ^N HY ^Q WS
GpGAL1	231	RYG ^K K ^V M ^V VA ^E IG ^M PW ^D EP ^S V ^C K ^S ML ^Q D ^M L ^K R ^V AL ^G - ^S NG ^I GV ^F Y ^E W ^E PQ ^A T ^P GW ^N HY ^Q WS
Gala_Ds	233	RYAK ^P V ^M V ^T E ^T VG ^M DW ^T QA ^A T ^S KAM ^L S ^D L ^L T ^K T ^R AL ^G - ^A NG ^I GV ^F Y ^E W ^E PQ ^A Y ^P GW ^Q Y ^T MG
Gala_Rsp.YR	239	RYAK ^P V ^V VA ^E ET ^G MDW ^Q QA ^A AA ^K AML ^A D ^V IT ^R V ^S AL ^G - ^T NG ^I GV ^L Y ^E W ^E PQ ^A Y ^P GW ^Q Y ^T MG
HsGAL1	288	AMDS ^S GK ^F TKA ^L TS ^Y NR ^Y
RrGAL1	293	AIT ^S CG ^Q FTE ^A LS ^A L---
Gala_Xc	294	AVD ^N NG ^F LTE ^A LQ ^A F---
GrGAL1	290	ALD ^N SGK ^F TEA ^L KAF ^S --
GpGAL1	290	ALD ^N SGK ^F TEA ^L KAF ^S --
Gala_Ds	292	ALD ^A SGK ^F TVA ^L DP ^F ---
Gala_Rsp.YR	298	ALD ^G SG ^F F ^T SAL ^D DP ^F ---

Figure 4.7 - A comparison of mature GH53 protein sequences from bacterial and nematode species. Alignment of HsGAL1 (*H. schachtii*), RrGAL1 (*R. reniformis*), Gala_Xc (*X. campestris*), GrGAL1 (*G. rostochiensis*), GpGAL1 (*G. pallida*), Gala_Ds (*D. sacchari*), and Gala_Rsp.YR (*R. sp.* YR242). Key catalytic glutamate residues (E) are highlighted in red.

4.3.2 – Protein structure models

Protein sequences of GrGAL1, GpGAL1, and RrGAL1 were used for BLAST similarity searches against the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) to identify hits of GH53 proteins which have had their crystal structure solved. In all three cases the highest identity hit (GrGAL1 & GpGAL1 40%, RrGAL1 39% %ID) was with the β -1,4-galactanase BTGH53 from *Bacteroides thetaiotaomicron* (NCBI seqID: 6GP5_A) (Böger *et al.* 2019). A 1-to-1 thread model for GRGAL1, GpGAL1, and RrGAL1 was produced using the protein structure and sequence alignment with the solved structure from *B. thetaiotaomicron* (Figure 4.8 D). The 1-to-1 thread model was achieved using Protein Homology/analogy Recognition Engine V 2.0 (PHYRE2). The predicted protein structures suggest that all three nematode proteins follow the same folding pattern - $(\beta/\alpha)_8$ barrel. Furthermore, the conserved catalytic glutamates of the GH53 proteins are in the centre of the β -barrel in each structural prediction which matches their placement in bacterial proteins (Figure 4.8 A, B, C, and D).

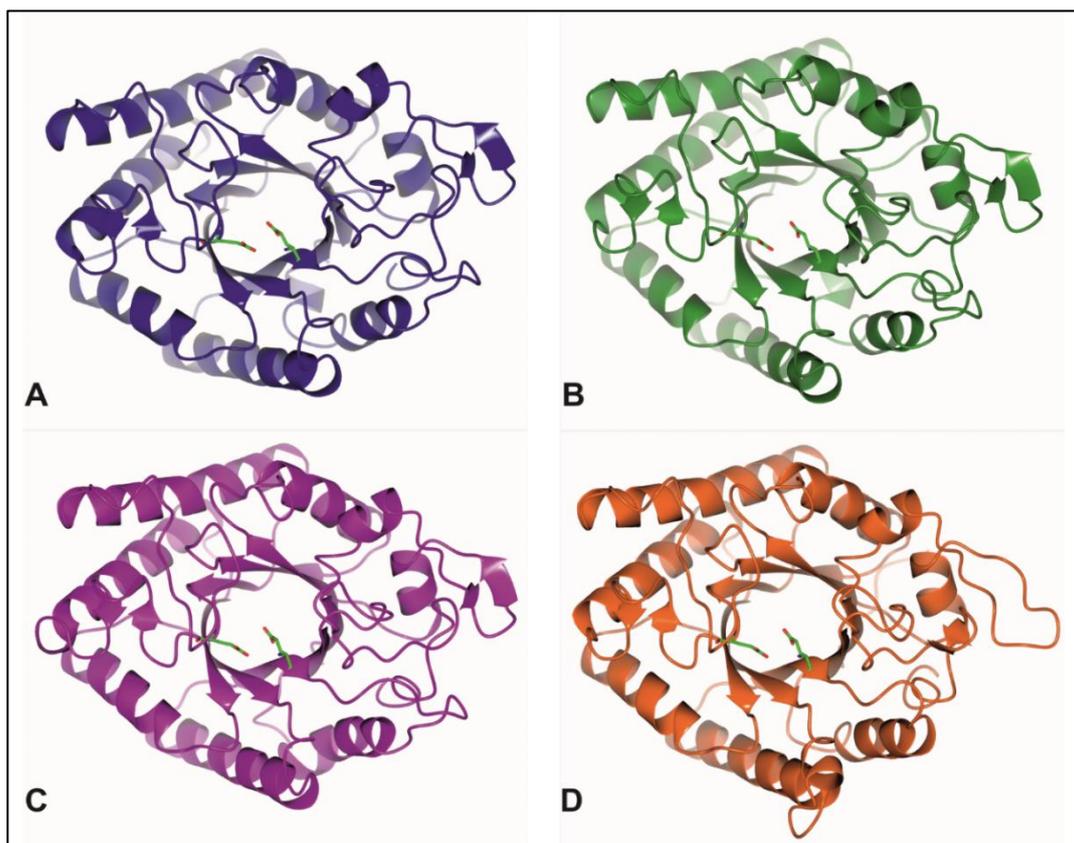


Figure 4.8 – Predicted structures of nematode GH53 proteins using a 1-to-1 threaded model. A – GrGAL1 (*G. rostochiensis*), B – GpGAL1 (*G. pallida*), C – RrGAL1 (*R. reniformis*), D – Confirmed structure of BTGH53 from *Bacteroides thetaiotaomicron* (6gp5_A). Catalytic glutamates in the active site are present in all four structures, in virtually the same positions (highlighted green/red, central). Structures were created using a 1-to-1 threaded model using PHYRE2, based on the solved structural sequence of BTGH53. All images taken at 20Å on ccp4mg.

4.3.3 – Phylogenetic analysis

A phylogenetic tree was produced to assess where the nematode GH53 proteins sat amongst those from bacteria and fungi (Figure 4.9). **As GH53 proteins are not produced by organisms other than these**, this was done to ascertain whether the nematode GH53s were more similar to fungal or bacterial GH53 proteins. GrGAL1, GpGAL1, and RrGAL1 were used in a BLAST similarity search against the non-redundant database. The top 100 results against GrGAL1, GpGAL1, and RrGAL1 were gathered, combined, and any results from other nematodes were removed. All results returned were bacterial on origin. Duplicate results were removed, for example there were more than 40 hits in the original data attributed to *Xanthomonas campestris campestris* which were filtered down to 8 based on

sequence identity (ID%). There were also several entries titled “multispecies” which were filtered out at this stage. After this filtering stage there was a total of 78 unique protein sequences present (Table 4.3).

Table 4.3 – Species included in phylogenetic analysis

No. In tree	Sequence name in tree	Species name	Protein name (link to NCBI)	Species type
1	RrGAL1	<i>R. reniformis</i>	RrGAL1	Nematode
2	GpGAL1	<i>G. pallida</i>	GpGAL1	Nematode
3	GrGAL1	<i>G. rostochiensis</i>	GrGAL1	Nematode
4	WP_092947600.1	<i>Roseateles sp. YR242</i>	glycosyl hydrolase 53 family protein	Bacteria
5	WP_058935312.1	<i>Roseateles depolymerans</i>	glycosyl hydrolase 53 family protein	Bacteria
6	ALV07139.1	<i>Roseateles depolymerans</i>	arabinogalactan endo-1,4-beta-galactosidase	Bacteria
7	SEL35718.1	<i>Roseateles sp. YR242</i>	arabinogalactan endo-1,4-beta-galactosidase	Bacteria
8	WP_056157311.1	<i>Duganella sp. Leaf126</i>	glycosyl hydrolase 53 family protein	Bacteria
9	WP_090322982.1	<i>Duganella sp. CF517</i>	glycosyl hydrolase 53 family protein	Bacteria
10	WP_048492385.1	<i>Xanthomonas sp. NCPPB 1128</i>	glycosyl hydrolase 53 family protein	Bacteria
11	WP_010343123.1	<i>Xanthomonas sacchari</i>	glycosyl hydrolase 53 family protein	Bacteria
12	WP_043094721.1	<i>Xanthomonas sacchari</i>	glycosyl hydrolase 53 family protein	Bacteria
13	OYT87800.1	<i>Burkholderiales bacterium PBB6</i>	arabinogalactan endo-1,4-beta-galactosidase	Bacteria
14	WP_020701442.1	<i>Oxalobacteraceae bacterium AB_14</i>	glycosyl hydrolase 53 family protein	Bacteria
15	WP_017909090.1	<i>Xanthomonas sp. SHU 199</i>	glycosyl hydrolase 53 family protein	Bacteria
16	SFR90348.1	<i>Mitsuaria sp. PDC51</i>	arabinogalactan endo-1,4-beta-galactosidase	Bacteria
17	WP_017912982.1	<i>Xanthomonas sp. SHU 166</i>	glycosyl hydrolase 53 family protein	Bacteria

18	WP_017915880.1	<i>Xanthomonas sp. SHU 308</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
19	WP_052198190.1	<i>Methylibium sp. CF059</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
20	WP_035053143.1	<i>Andreprevotia chitinilytica</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
21	WP_072787792.1	<i>Duganella sacchari</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
22	WP_047507752.1	<i>Methylibium sp. CF468</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
23	WP_018609111.1	<i>Uliginosibacterium gangwonense</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
24	WP_050911238.1	<i>Xanthomonas campestris</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
25	WP_040940540.1	<i>Xanthomonas campestris</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
26	WP_095575148.1	<i>Xanthomonas hortorum</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
27	WP_056141038.1	<i>Duganella sp. Leaf61</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
28	WP_012437258.1	<i>Xanthomonas campestris</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
29	WP_014509121.1	<i>Xanthomonas campestris</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
30	WP_055828484.1	<i>Xanthomonas sp. Leaf131</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
31	WP_011038708.1	<i>Xanthomonas campestris</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
32	WP_023905279.1	<i>Xanthomonas hortorum</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
33	KLD77750.1	<i>Xanthomonas hyacinthi DSM 19077</i>	<u>arabinogalactan endo-1,4-beta-galactosidase</u>	Bacteria
34	WP_057671413.1	<i>Xanthomonas campestris</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
35	WP_064507484.1	<i>Xanthomonas floridensis</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
36	WP_039434208.1	<i>Xanthomonas vasicola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
37	WP_076053575.1	<i>Xanthomonas campestris</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
38	WP_064629974.1	<i>Xanthomonas nasturtii</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
39	WP_039443498.1	<i>Xanthomonas vasicola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria

40	GAE52748.1	<i>Xanthomonas arboricola</i> pv. <i>pruni</i> str. MAFF 311562	<u>arabinogalactan endo-1,4-beta-galactosidase</u>	Bacteria
41	WP_039531138.1	<i>Xanthomonas arboricola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
42	WP_039515579.1	<i>Xanthomonas arboricola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
43	GAE55131.1	<i>Xanthomonas arboricola</i> pv. <i>pruni</i> MAFF 301420	<u>arabinogalactan endo-1,4-beta-galactosidase</u>	Bacteria
44	WP_047128019.1	<i>Xanthomonas arboricola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
45	WP_039564616.1	<i>Xanthomonas cannabis</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
46	WP_039405212.1	<i>Xanthomonas cannabis</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
47	WP_006452505.1	<i>Xanthomonas hortorum</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
48	WP_047693475.1	<i>Xanthomonas cannabis</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
49	WP_070249534.1	<i>Duganella phyllosphaerae</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
50	WP_047126143.1	<i>Xanthomonas arboricola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
51	WP_018608135.1	<i>Uliginosibacterium gangwonense</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
52	WP_005997364.1	<i>Xanthomonas vesicatoria</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
53	WP_022972646.1	<i>Xanthomonas maliensis</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
54	WP_054393996.1	<i>Xanthomonas vasicola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
55	WP_010379559.1	<i>Xanthomonas vasicola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
56	WP_082569018.1	<i>Rhizobacter</i> sp. Root1221	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
57	KQV81237.1	<i>Rhizobacter</i> sp. Root1221	<u>arabinogalactan endo-1,4-beta-galactosidase</u>	Bacteria
58	WP_065469824.1	<i>Xanthomonas bromi</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
59	WP_046963529.1	<i>Xanthomonas pisi</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
60	EWC50953.1	<i>Xanthomonas citri</i> pv. <i>glycines</i> str. 8ra	<u>arabinogalactan endo-1,4-beta-galactosidase</u>	Bacteria

61	EEF26962.1	<i>Ricinus communis</i>	<u>Arabinogalactan endo-1,4-beta-galactosidase, putative</u>	Plant
62	WP_078567753.1	<i>Xanthomonas campestris</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
63	WP_040260234.1	<i>Xanthomonas citri</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
64	KGE51643.1	<i>Xanthomonas axonopodis pv. vasculorum</i>	<u>arabinogalactan endo-1,4-beta-galactosidase</u>	Bacteria
65	CDN18356.1	<i>Xanthomonas citri pv. Viticola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
66	WP_007962641.1	<i>Xanthomonas citri</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
67	WP_099803554.1	<i>Xanthomonas citri</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
68	WP_018609122.1	<i>Uliginosibacterium gangwonense</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
69	CEJ49178.1	<i>Xanthomonas citri pv. bilvae</i>	<u>Arabinogalactan endo-1,4-beta-galactosidase</u>	Bacteria
70	WP_033836215.1	<i>Xanthomonas citri</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
71	WP_029818548.1	<i>Xanthomonas euvesicatoria</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
72	WP_078590862.1	<i>Pseudomonas cissicola</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
73	WP_017158928.1	<i>Xanthomonas phaseoli</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
74	WP_089094626.1	<i>Xanthomonas citri</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
75	OQP76910.1	<i>Xanthomonas phaseoli pv. syngonii LMG 9055</i>	<u>arabinogalactan endo-1,4-beta-galactosidase</u>	Bacteria
76	WP_017165126.1	<i>Xanthomonas phaseoli</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
77	WP_017155887.1	<i>Xanthomonas phaseoli</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
78	WP_013634183.1	<i>Pseudopedobacter saltans</i>	<u>glycosyl hydrolase 53 family protein</u>	Bacteria
79	AAA32692.1	<i>Aspergillus aculeatus</i>	<u>arabinogalactan endo-1,4-beta-galactosidase</u>	Fungus
80	4BF7_A	<i>Aspergillus nidulans</i>	<u>Chain A, Arabinogalactan endo-1,4-beta-galactosidase A</u>	Fungus
81	1HJQ_A	<i>Humicola insolens</i>	<u>Chain A, beta-1,4-galactanase</u>	Fungus
82	1HJS_A	<i>Thermothelomyces thermophilus</i>	<u>Chain A, beta-1,4-galactanase</u>	Fungus

It quickly became clear that the nematode GH53 proteins are most like GH53 proteins of bacterial origin. No fungal GH53 proteins were represented in the top 100 BLAST results from each of the similarity searches against GrGAL1, GpGAL1, and RrGAL1. Out of the 78 unique proteins there were 52 from species (and subspecies) of the *Xanthomonas* genus. One protein of note included in the unique sequence list was EEF26962.1 (number 61 in table 4.3). This is from *Ricinus communis* (the castor bean plant) so is the only entry which is not bacterial in origin other than the three nematode proteins. EEF26962.1 is defined as an arabinogalactan endo-1,4- β -galactosidase, however this is a putative description so has not been confirmed through gene cloning or enzyme assays as a functional protein.

As previously discussed in section 4.1.2.1, an alien index score was given to 519 *G. rostochiensis* candidates to determine the likelihood of acquisition through a horizontal gene transfer event (Eves-Van den Akker *et al.* 2016a). GrGAL1 (GROS_g08150) has an AI of 349.2997869, the 5th highest AI score of the 519 candidates assessed. The AI score alongside the phylogenetic analysis conducted in section 4.3.3 indicate a strong likelihood that these nematode GH53 genes originate from a horizontal gene transfer event with genetic material that was bacterial in origin.

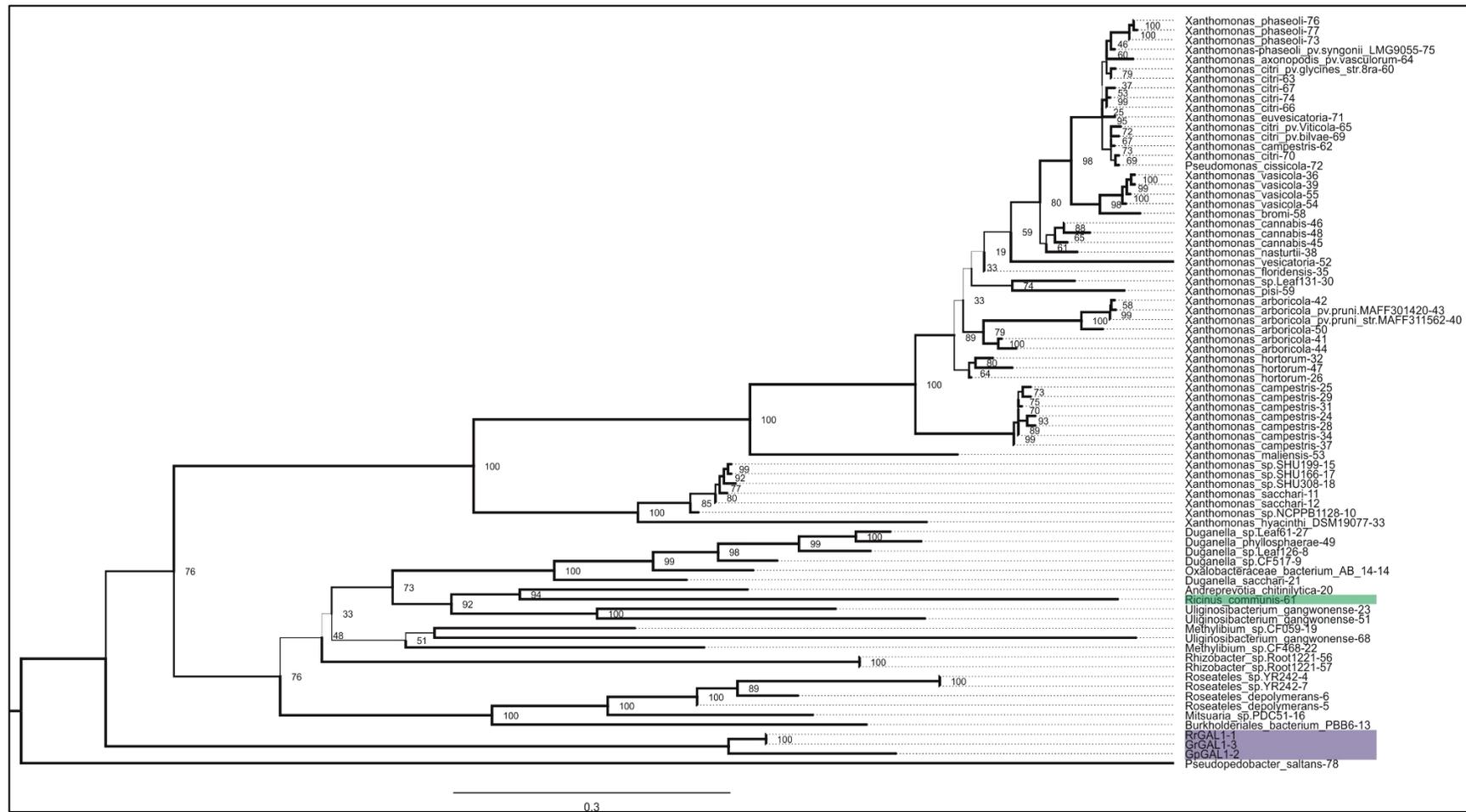


Figure 4.9 - Phylogenetic tree of nematode GH53 proteins compared to unique BLAST hits from bacterial species. Nematode GH53 species are highlighted in blue (numbers 1-3). GH53 identified from *Ricinus communis* is highlighted in green (number 61). Tree is midpoint re-rooted and based on 1000 bootstraps. Each entry is followed by “-XX” which is the identification number attributed in Table 4.3.

Upon reflection of the phylogenetic analysis done, it could be argued that the non-redundant database for BLAST searches of the nematode GH53 proteins was not the most appropriate. Due to the high number of bacterial genes and proteins compared to fungal sequences represented in this database the “top 100” hits list produced may have been skewed towards bacterial results. Additionally, the inclusion of so many *Xanthomonas* species made this original tree highly weighted which needed to be addressed. The phylogenetic analysis was redone to reduce the number of proteins included in the analysis from 78 to 26 in total and incorporate fungal GH53 proteins (Figure 4.10). The 40 *Xanthomonas* species entries was reduced to 7 and all duplicate species were removed, keeping only the highest percentage identity hits in cases where there were multiple sequences included from one species. The fungal protein sequences incorporated were from the *Aspergillus aculeatus*, *Aspergillus nidulans*, *Humicola insolens*, and *Thermothelomyces thermophilus* species which were identified through BLAST searches of the Protein Data Bank (PDB) database using GrGAL1, GpGAL1 and RrGAL1 as query sequences.

The four fungal species above have been added to Table 4.3 as numbers 79-82. When the phylogenetic tree was reproduced to include fungal proteins there was no change to the overall placement of the nematode GH53s, meaning the nematode species did not shift to be closer to these fungal species. The original tree was kept as Figure 4.9 to demonstrate the original research process; however, the new tree has been created to demonstrate the inclusion of the fungal species (Figure 4.10).

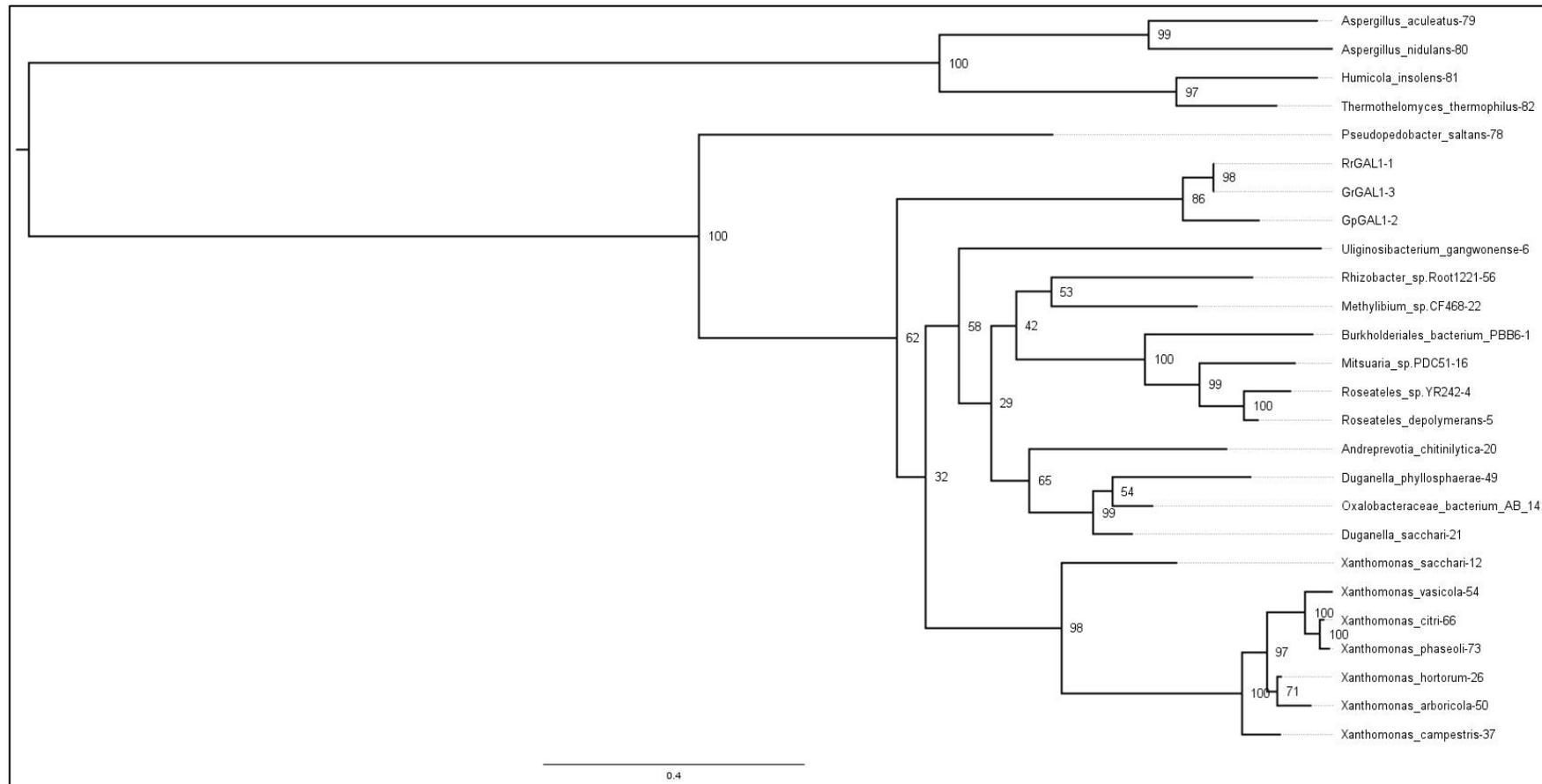


Figure 4.10 - Phylogenetic tree of nematode GH53 proteins compared to unique BLAST hits from bacterial and fungal species revisited. Nematode species are numbers 1-3. Fungal species are numbers 79-82, all other numbers are bacterial species. The tree is midpoint re-rooted and based on 1000 bootstraps. Each entry is followed by "-XX" which is the identification number attributed in Table 4.3

4.3.4 – In situ hybridisation and gene expression

In situ hybridisation was conducted on GrGAL1, GpGAL1, and RrGAL1 to identify localisation of the mRNA transcripts for each gene at the J2 nematode life stage. This was done with the knowledge that the majority of effectors are expressed in either the dorsal or subventral gland cells. GrGAL1 (Figure 4.11 A, B) and RrGAL1 were localised in the subventral gland cells (Figure 4.11 C, D). Analysis of GpGAL1 from *G. pallida* produced unexpected ISH results. It was hypothesised that much like GrGAL1 and RrGAL1, GpGAL1 would display subventral gland staining. GpGAL1 consistently produced a condensed spherical staining pattern in the region of the oesophageal glands (Figure 4.11 E, F, G, and H). This structure is too small to be either the full dorsal gland or the subventral glands, although it is possible this structure is the nucleus of a gland cell being stained or possibly the nerve ring. It is unclear why the transcript appears to localise only at the nucleus and not across the full gland cell. ISH negative controls for each species were run using the appropriate sense primer probe. Controls for all three nematode species displayed no specific signals with minimal background staining around the cut site.

To determine the life stages at which the GH53 proteins are required, the expression profiles of each gene were determined using available RNAseq data for each species. In most cases expression was higher in the invasive stages than in parasitic stages. Average gene expression data for GrGAL1, GpGAL1, and RrGAL1 can be found in Figure 4.11. RrGAL1 shows expression at the J2 life stage with a comparatively small reduction in mature females. Although this follows the general expression pattern expected for GH53s, a lower expression in mature females was expected. Unfortunately, other life stages e.g., cyst and male expression are not available for *R. reniformis*. GpGAL1 is expressed highly at the J2 life stage only, while GrGAL1 is most highly expressed during both the egg and J2 life stages before dropping in expression at the 14 dpi stage. It is common to see effector genes required at the earliest stages of infection being highly expressed at the egg stage. This is because the egg contains the J2 and may reflect the nematode preparing and stockpiling

effector proteins for use in the early parasitic life stages. This aligns with the nematode locating a host plant and initiating entry into the host and migration through the root, which requires cell wall degradation and is consistent with the predicted function of the GH53 proteins.

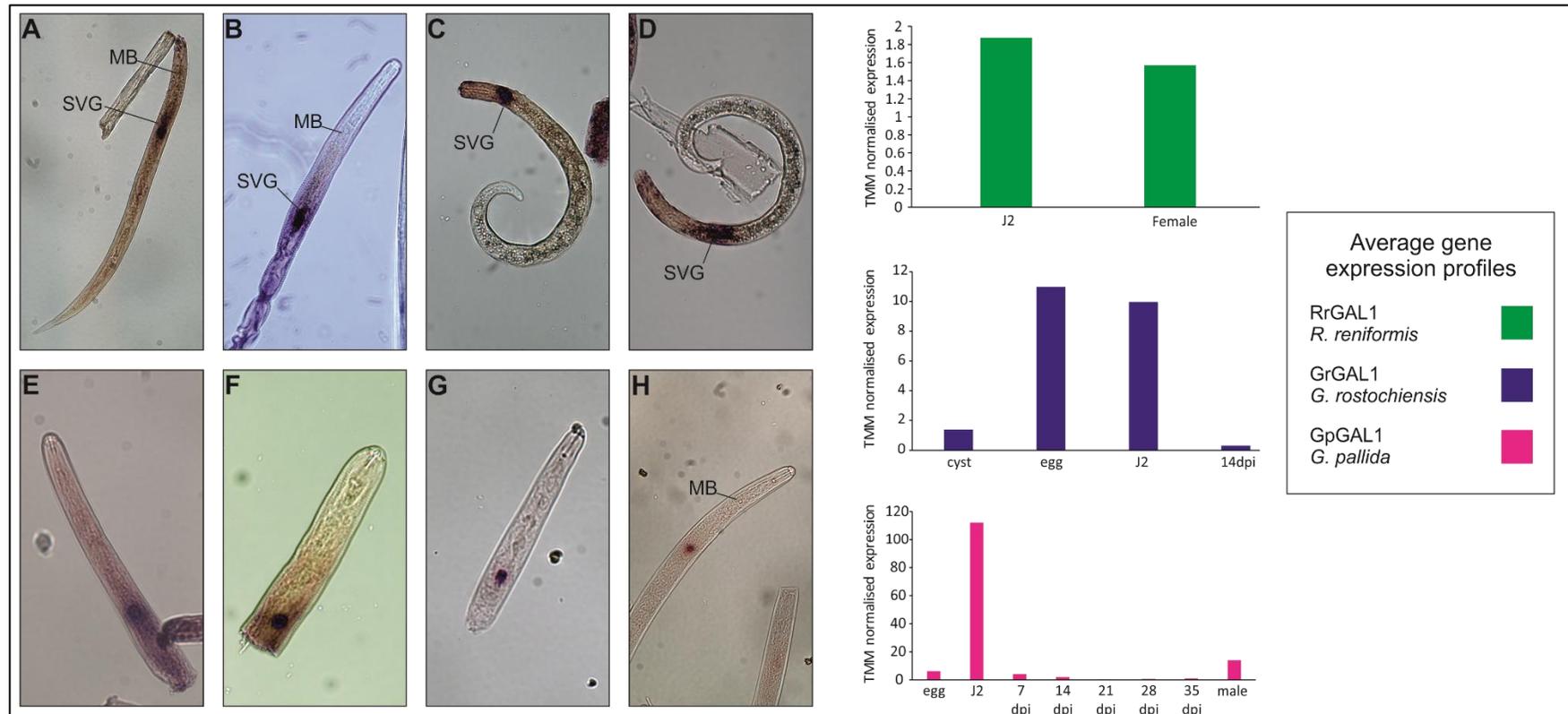


Figure 4.11 - In situ hybridisation and gene expression profiles of GrGAL1, GpGAL1, and RrGAL1 in J2 *G. rostochiensis*, *G. pallida*, and *R. reniformis*. A - B - GrGAL1 displaying subventral gland staining, C - D - RrGAL1 displaying subventral gland staining, E - H - GpGAL1 displaying undefined staining localisation. Average gene expression data from multiple life stage repeats. *G. rostochiensis* and *G. pallida* have two replicates for each life stage and *R. reniformis* has three replicates per life stage. Image brightness increased 20% for publication purposes. SvG - subventral gland, MB - metacarpal bulb.

4.3.5 – Biochemical characterisation of nematode GH53 proteins

4.3.5.1 – Expression and purification of recombinant proteins

In order to confirm the function of the nematode GH53 proteins, it was important to express, purify, and biochemically assay GrGAL1, GpGAL1, and RrGAL1. Many previous functional studies of proteins more generally have failed at this stage due to difficulties obtaining purified protein samples for *in vitro* analysis. GrGAL1, GpGAL1, RrGAL1, GalA_Xc (endo-1,4- β -galactosidase from *X. campestris*), and the GFP negative control were expressed in overnight cultures of T7 Shuffle *E. coli* cells (New England Biolabs). Cells were induced to express the recombinant proteins via isopropyl-B-D-thiogalactoside (IPTG). Uninduced and induced cells expressing GrGAL1, GpGAL1, RrGAL1, GalA_Xc, and GFP can be observed in lanes 1 and 2 respectively, for each of the gel images (Figure 4.12). In each case, the presence of a protein band of the size expected was enriched after induction. Cultures containing the individual recombinant proteins were sonicated and supernatant was collected to observe the soluble proteins in the sample. This soluble protein fraction is present in lane 3 for each gel (Figure 4.12). Each recombinant protein was captured using Ni-NTA resin while the rest of the supernatant was washed away. The wash was retained and is shown in lane 4 (Figure 4.12). It is possible to see a band of each of the proteins of interest in the wash sample. This indicates that, in future, optimisation could be done to prevent loss of sample and increase the final yield of purified protein. Finally, GrGAL1, GpGAL1, RrGAL1, GalA_Xc, and GFP were eluted from the beads and are shown in lane 5 of each gel (Figure 4.12). GrGAL1, GpGAL1, RrGAL1, and GalA_Xc can all be observed at approximately 55 kilodaltons (kDa) while the GFP negative control is slightly below this, reflecting its smaller size. It was possible to express and purify all recombinant GH53s consistently, but it should be noted that the RrGAL1 GH53 always expressed at a lower level than GH53 originating from the PCN species. The purified recombinant protein samples were used for biochemical analysis using azo-galactan substrate and DNS assays.

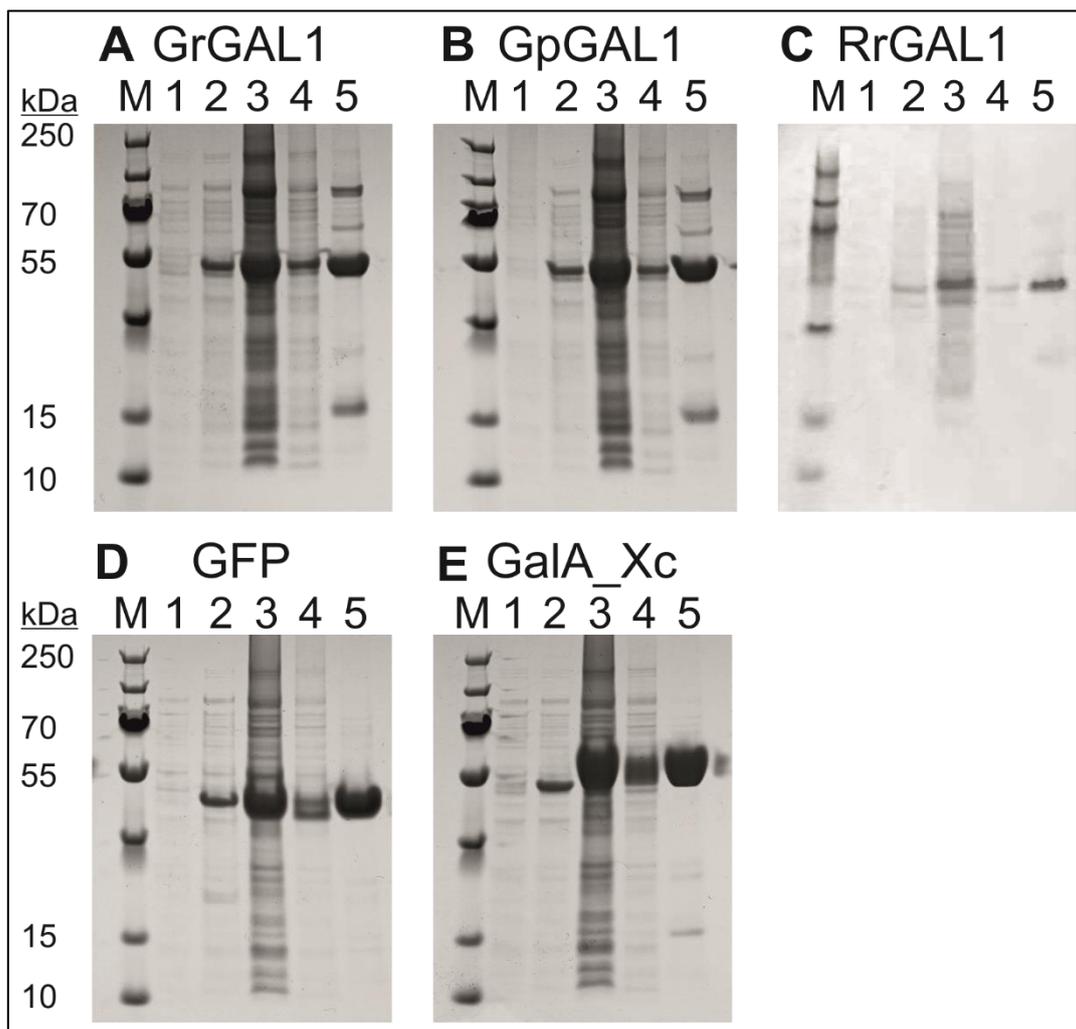


Figure 4.12 – Expression and purification of recombinant GrGAL1, GpGAL1, RrGAL1, GalA_Xc, and GFP. M – PageRuler Plus protein ladder (marker), 1 – pre-induction sample, 2 – post-induction (IPTG) sample, 3 – soluble protein, 4 – resin wash, 5 – purified protein of interest

4.3.5.2 – Azo-galactan substrate assay

The azo-galactan substrate assay was first used to determine whether the nematode GH53 enzymes would hydrolyse the predicted substrate. Azo-galactan comprises galactan polymers obtained from potato cell wall pectin that have been stripped of arabinose by preincubation with an arabinofuranosidase by the manufacturer (Megazyme), and subsequently dyed with Remazolbrilliant Blue R. When azo-galactan is hydrolysed by an endo- β -1,4-galactanase, the low molecular weight dyed molecules are released into solution and can be quantified using a spectrophotometer after termination of the reaction.

A blank sample was used which contained no enzyme, only substrate and ethanol. This produced a clear supernatant (all dye molecules remain at a high molecular weight in the pellet due to there being no enzyme present to hydrolyse the substrate) following centrifugation and was used to blank the spectrophotometer. The GH53 from *X. campestris* hydrolysed the substrate, producing enough dyed fragments in solution that a 1:10 dilution in elution buffer was prepared to get an accurate spectrophotometer reading. The *X. campestris* GH53 consistently produced the same results across all replications of this assay, making it a highly reliable positive control. The GFP sample, used as a negative control, showed no endo- β -1,4-galactanase activity and displayed a clear supernatant; this matched the results seen for the blank sample when incubation periods of enzyme with the substrate were relatively short periods of time between 10 minutes and an hour. However, when the assay was repeated with much longer incubation periods or at alkaline pH there was faint activity detected by the spectrophotometer in the GFP sample. Using this assay, no activity was detected for the three nematode GH53s when using the 10 minute incubation time as advised in the manufacturer's protocol. Therefore, the assay was conducted over a time course with samples taken at 10 minutes, 30 minutes, 60 minutes, 90 minutes, and 120 minutes. The assay was repeated at multiple temperatures (4 °C, 20 °C, 30 °C, 40 °C, 50 °C, and 60 °C) however no activity was detected for the nematode GH53s at any of the temperatures initially tested.

Subsequent replications were conducted using an overnight incubation of the enzyme with the substrate (approximately 16 hours). This extended incubation time saw the *G. pallida* and *R. reniformis* GH53s display activity along with the *X. campestris* GH53. There was slight activity observed with the GFP negative control, but this was much lower than the absorbance readings obtained from the GH53-containing samples. Replicates of the assay were conducted at a range of pH values (2, 3, 4, 5, 6, and 7, all at 28 °C) to ascertain the pH optima for GpGal1, RrGal1, and GalA_Xc. These experiments were also designed to determine whether activity

could be detected from GrGAL1 at a different pH. This test showed that both GpGal1 and RrGal1 alongside the GalA_Xc control all showed activity at pH 4 (Figure 4.13). Unfortunately, no activity was observed by GrGAL1 at any pH or temperature tested. These data indicated that the GpGAL1 and RrGAL1 proteins function as endo- β -1,4-galactanases *in vitro*, albeit with relatively low activity rates.

Several further assays using a range of temperatures (18 °C, 28 °C, and 37 °C) were conducted as well as further technical replicates of the experiments described above. Unfortunately, despite multiple independent replicates being carried out, the results were extremely inconsistent meaning optimum conditions could not be determined. A different assay was therefore selected which could be conducted in a microtiter plate and thus more readily lent itself to analysis of multiple conditions and substrates.

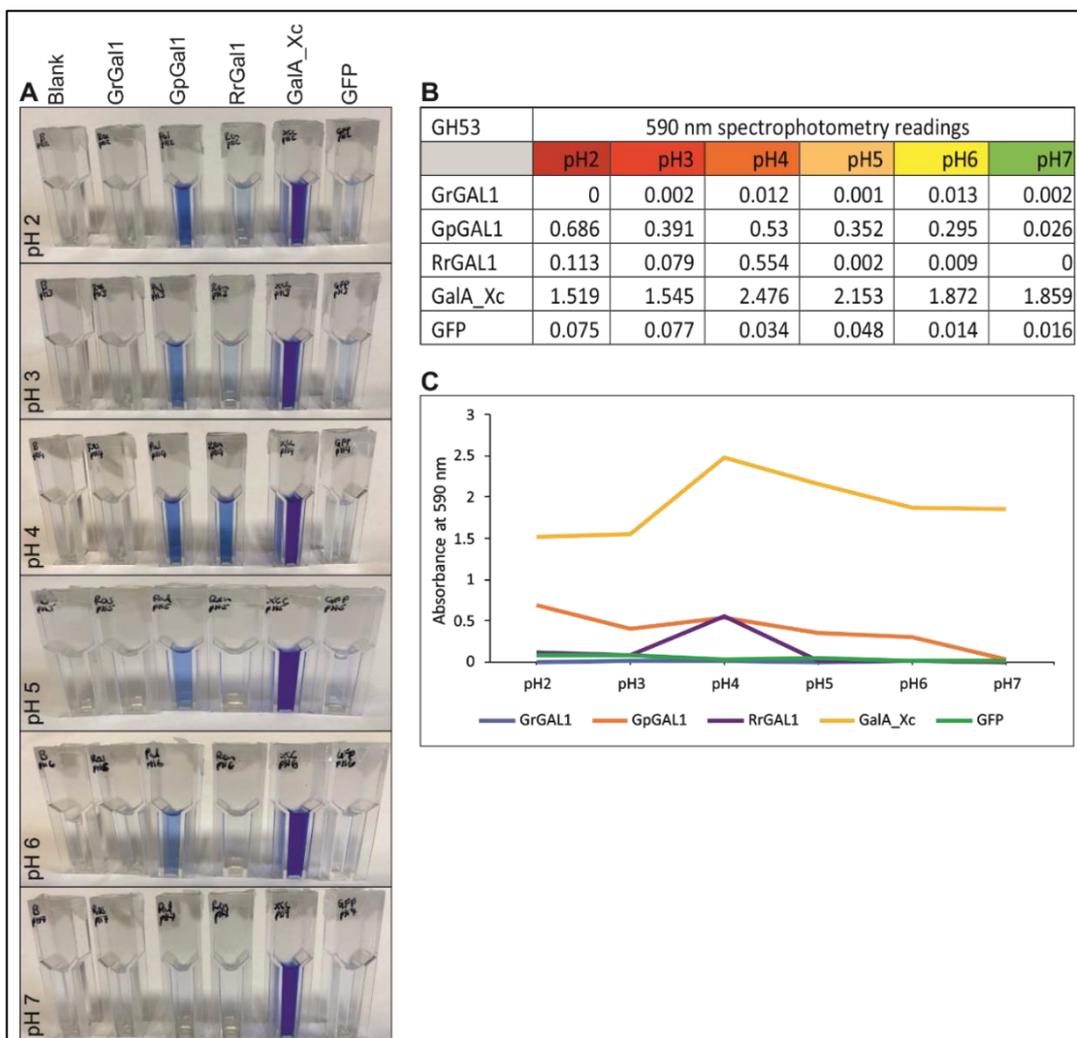


Figure 4.13 - pH range for GH53 activity using azo-galactan substrate assay. A – Images of cuvettes giving a visual representation of colour change observed for GH53 samples at pH 2 to pH7. B – Absorbance readings of each GH53 enzyme with azo-galactan at 590 nm. C – Graphical representation of data displayed in B.

4.3.5.3 – DNS assay

Although the data obtained using the azo-galactan substrate assay showed that two of the GH53 proteins (GpGAL1 and RrGAL1) were likely to be functional, the activity detected in these assays (at pH 4, room temperature) was much lower than expected. The design of this assay also precluded wider testing of conditions, therefore the 3,5-dinitrosalicylic acid (DNS) assay was used to further examine the function of the GH53 proteins. This assay functions by the DNS reacting with reducing sugars to produce a colour change (yellow to red/brown) measurable at 540 nm with a spectrophotometer. Reducing sugars are released when polysaccharide substrates are hydrolysed. The GH53 enzymes should hydrolyse the galactan substrate and release shortened galactan oligosaccharides with reducing ends. The GH53s were also tested against other common polysaccharides found in the plant cell wall: xylan, saccharose, pectin, type II arabinogalactan (AG-II), and polygalacturonic acid (pectic acid). Reducing sugars produced by degradation of each of these substrates are detailed in Table 4.4. It was anticipated that no reducing sugars would be released from substrates other than the galactan as there would be no activity from the GH53 enzymes and consequently no detectable colour change would be observable. It was not possible to obtain AGI for use as a substrate at this time.

Table 4.4 – Reducing sugars produced on substrate hydrolysis

Substrate	Reducing sugars produced
Galactan	shortened galactan oligosaccharides with reducing ends
Xylan	Xylose
Saccharose	Glucose and fructose
Pectin	* Arabinose, galactose, glucose, xylose, and mannose. ** Shortened Galacturonic acid subunits with reducing ends
Arabinogalactan (AG-II)	Galactose
Polygalacturonic acid	** Shortened Galacturonic acid subunits

*There is the potential for multiple reducing sugars to be released from pectin depending on side chain composition and enzymes (Garna *et al.* 2004). ** Much like pectin the reducing sugars produced vary depending on the enzyme – polygalacturonic acid can be hydrolysed to produce shortened galacturonic acid subunits/chains of varying lengths with reducing terminal sugars. The same can be said for shortened galactan oligosaccharides produced from the hydrolysis of galactan.

The DNS assay showed that recombinant GrGAL1, GpGAL1, and RrGAL1 all consistently showed the anticipated hydrolytic enzymatic activity when galactan was used as the substrate at room temperature and pH 4 (Figure 4.14 A). As expected, no activity was detected using the substrates xylan, polygalacturonic acid, and pectin (Figure 4.14 B, C, D). Similarly, no activity was detected using saccharose or arabinogalactan (AG-II); all enzymes with both substrates gave a reading of zero at 540 nm (Figure 4.14 E, F). Figure 4.14 G displays the colour change observed in galactan substrate samples with GH53 enzymes when compared to the other substrates tested.

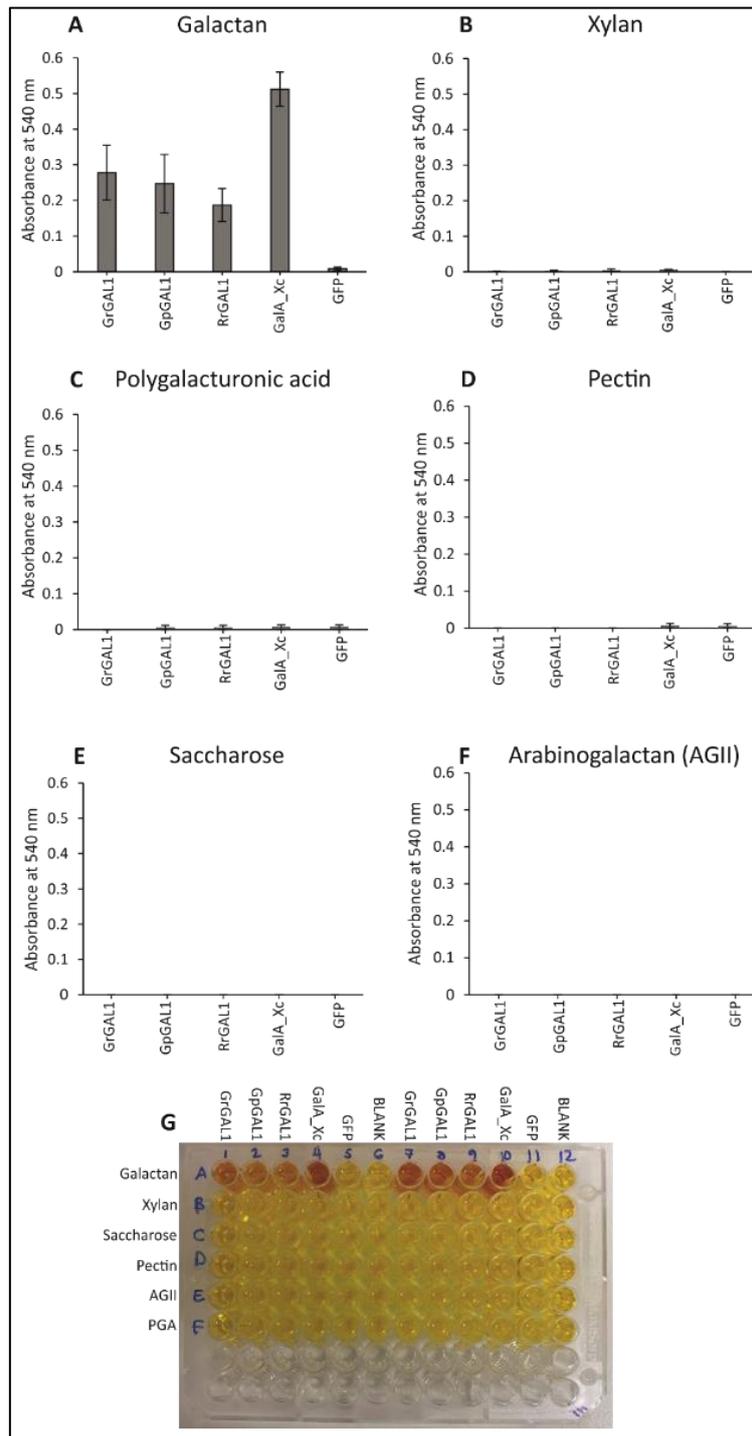


Figure 4.14 – DNS assay using nematode GH53 enzymes against plant cell wall component substrates. A – GrGAL1, GpGAL1, RrGAL1, GaIA_Xc, and GFP hydrolysis of galactan substrate producing absorbance readings at 540 nm. B – GrGAL1, GpGAL1, RrGAL1, GaIA_Xc, and GFP displaying little to no activity using xylan as a substrate. C – GrGAL1, GpGAL1, RrGAL1, GaIA_Xc, and GFP displaying little to no activity using polygalacturonic acid as a substrate. D – GrGAL1, GpGAL1, RrGAL1, GaIA_Xc, and GFP displaying little to no activity using pectin as a substrate. E – GrGAL1, GpGAL1, RrGAL1, GaIA_Xc, and GFP displaying no activity using saccharose as a substrate. F – GrGAL1, GpGAL1, RrGAL1, GaIA_Xc, and GFP displaying no activity using arabinogalactan type II (AGII) as a substrate. G – image of 96-well plate assay depicting the colour change observed when GrGAL1, GpGAL1, RrGAL1, GaIA_Xc, and GFP were tested using different substrates. All error bars were calculated using standard error.

4.4 – Discussion

The identification and cloning of the genes encoding GrGAL1, GpGAL1, and RrGAL1 have shown that GH53 proteins exist in both the *Globodera* and *Rotylenchulus* nematode genera as well as those previously studied in *Heterodera* (Vanholme *et al.* 2009). This aligns with the work done by Pokhare *et al.* in their 2020 paper, which shows these nematodes cluster together phylogenetically as those which feed on dicotyledon plant species. Although there is still a lack of confirmed protein structure, the high conservation displayed by sequence alignment and phylogenetic analysis (Figure 4.7, 4.9), alongside the 1-to-1 thread models (Figure 4.8) demonstrate the high likelihood that the GH53 proteins in nematodes are derived by HGT from bacteria. This is more likely than the alternative hypotheses. The first alternative hypothesis is that the common ancestor between the nematode species tested and bacteria would have had GH53 genes present. This ancestor would have subsequently gone on to independently lose these genes on many occasions across species who do not have genes encoding GH53 proteins today. The second alternative hypothesis is that the GH53 proteins seen today have independently evolved in both bacteria and nematodes. This seems equally unlikely due to the high similarity of structure and sequence observed between the two proteins. Independent evolution would also likely be evident in the Alien index scores discussed in section 4.1.2.1 and 4.3.3. If the GH53 proteins in nematodes had evolved independently of the bacterial equivalents then their sequence similarity would be likely reduced, meaning their alien index score would have been much closer to zero. The higher an AI score, the higher the difference between E-values of the donor species (bacteria, fungi) and the recipient species (nematode). This indicates a higher likelihood of a HGT event having occurred.

4.4.1 – Phylogenetic analysis

The phylogenetic analysis showed that all the nematode GH53 protein sequences clustered more closely with the bacterial GH53 proteins included than those from fungi, providing a good indication that the GH53 genes from the nematodes analysed

in this thesis are originally bacterial in origin. This supports the hypothesis that these genes would have been obtained through horizontal gene transfer between an ancestral nematode species and bacterial genomic material in the surrounding environment. These data are supported by the phylogenetic analysis conducted by Vanholme *et al.* In their 2009 paper they demonstrated the presence of putative arabinogalactan endo-1,4- β -galactosidase genes (HsGal1, HsGal2) in the PPN species *H. schachtii* (Vanholme *et al.* 2009). A phylogenetic tree also showed the *H. schachtii* proteins to be more like bacterial than fungal enzymes. HsGal1 and HsGal2 appear to be closely related to those from *X. campestris* and *X. axonopodis*. Both bacterial species feature in the phylogenetic tree conducted for GrGal1, GpGal1, and RrGal1 in section 4.3.3.

As discussed in Section 4.3.3, the presence of a putative arabinogalactan endo-1,4- β -galactosidase from the castor bean plant *Ricinus communis* in the results of the BLAST searches conducted with the nematode GH53s was unexpected. No functional characterisation of the *Ricinus* GH53 proteins has been reported and this sequence was identified as part of a genome shotgun sequencing project. While it is possible that this is a genuine plant GH53 which is similar to bacterial proteins, it is more likely to be derived from bacterial DNA contaminating the sample used for sequencing of this plant species genome and thus incorrectly attributed as a plant gene. The alternative to this would be that this is a genuine plant GH53 enzyme which is the result of a horizontal gene transfer from a bacterial species to a plant. Horizontal gene transfer from bacteria to plants is possible as highlighted by the transfer of transfer DNA (T-DNA) from *Agrobacterium* species to host plants such as *N. benthamiana* (Quispe-Huamanquispe *et al.* 2017). In line with this the amino acid sequence of GrGAL1 was used as a query to BLAST against plant species. This returned 9 results: two sequences (one partial) from *Ricinus communis* with percentage sequence identity of 62.7% and 48.5%. The other 7 hits from this similarity search came from *Ceratodon purpureus* (fire moss), *Physcomitrium patens* (spreading leaved earth moss), and *Salix suchowensis* (riparian shrub willow),

however none of these exceeded a percentage sequence identity of 34% so it is unlikely that these are of significance. Finally, none of these GH53 proteins of plant origin are experimentally validated. All are hypothetical proteins identified in large scale sequencing projects.

Three GH53-like sequences from *N. aberrans* have also been identified (51041_c0_seq1, 6113_c0_seq1, and 23294_c0_seq1) (Eves-van den Akker *et al.* 2014). It should be noted that all three of these amino acid sequences presently appear to be truncated and their nucleotide sequence cannot be confirmed given the current lack of genome data for this species. In future, if genome data were available for this species, it may be possible to analyse the truncated sequences to see if they have been mispredicted and undergo manual recapitulation to produce the full protein sequences. If these proteins do exist, in future the phylogenetic tree in Figure 4.4 may change. *Nacobbus aberrans* currently resides in a clade alongside several *Meloidogyne* species in Figure 4.4, but this species is known to shift in phylogenetic analyses depending on the other species included in the alignment. It is difficult to pinpoint where *N. aberrans* sits definitively due to its production of a syncytial feeding site and similarity of effector suite to other cyst nematode species. This is at odds with the formation of root galling around the feeding site of *N. aberrans* which is usually associated with root-knot nematode species. It is possible that *N. aberrans* could shift upwards towards the current position of *R. similis* in Figure 4.4 with the confirmation and addition of full length GH53 gene sequences.

4.4.2 – In situ hybridisation & gene expression

In situ hybridisation is used to determine where the mRNA of a gene is expressed in a nematode. This information can be used to determine whether a protein is likely to be secreted out of the nematode as an effector. *In situ* hybridisation results showed that GrGAL1 and RrGAL1 were expressed in the subventral gland cells. A more complex expression pattern was observed for GpGAL1 as the mRNA was

consistently detected in the region where the oesophageal glands are found in J2, however the staining was never widespread across the full gland cells, thus making it difficult to determine its location with certainty. The condensed spherical staining pattern observed could be the nucleus of a gland cell. It is currently unclear why only the nucleus of a gland cell would be stained, especially considering the gene expression levels observed for GpGAL1 in section 4.3.4. In general, effectors that function in the earlier stages of parasitism are expressed in the subventral glands, while effectors involved in later parasitic stages, such as syncytium maintenance, are expressed in the dorsal gland. As seen in section 4.3.4, GpGAL1 has high expression exclusively at the J2 life stage while GrGal1 is most highly expressed at both the egg and J2 life stages. RrGAL1 shows a slightly different expression profile than the *Globodera* species. RrGAL1 is not as abundantly expressed as the GH53 proteins from the cyst nematode species and in addition RrGAL1 is also expressed in the adult female life stage, although there was a small decrease between the J2 and female life stages. This indicates that RrGAL1 is functional during both life stages. Currently there is only expression data for *R. reniformis* at J2 and adult female life stages. In future if the data become available for other life stages such as cyst, egg, male etc. it may shed more light on the function of this gene. The ISH results taken alongside expression data at the J2 life stage, and the presence of a signal peptide, all indicate that GrGAL1, GpGAL1, and RrGAL1 are effectors that the nematode secretes into the host plant during migration.

4.4.3 – DNS assay

The DNS assay showed that GrGAL1, GpGAL1, RrGAL1, and the GalA_Xc positive control hydrolysed the galactan substrate, releasing shortened galactan oligosaccharides with reducing ends, to prove that they are functional endo- β -1,4-galactanases. This was further confirmed by the lack of activity when using the substrates xylan, saccharose, pectin, and polygalacturonic acid. This shows that GrGAL1, GpGAL1, RrGAL1 all have a specific function and cannot break down

other cell wall components. This indicates that these GH53 proteins are effectors that are secreted by the nematode as part of a cocktail of CWDEs.

It is however acknowledged that the galactan used here was not the optimal substrate. It was not possible to source type I arabinogalactan for use in this assay. As discussed in section 4.1.4.1, GH53 proteins hydrolyse AG-I but not AG-II due to the difference in the type of glycosidic bonds which form the galactan backbone. AG-I have β -1, 4 glycosidic bonds while AG-II have both β -1, 3- and β -1, 6- glycosidic bonds. When identifying manufacturers for substrate purchase it became clear that AG-I was not available. Most manufacturers who stock arabinogalactan do not state which type it is specifically on packaging, however the majority state that it is sourced from larch wood. Larch wood produces high levels of AG-II which the nematode GH53 proteins have now been shown not to hydrolyse (Figure 4.13). It was equally not possible to produce the substrate in house due to lack of resources/equipment. The most important test for these data in future work would be to identify a reliable, high quality source of AG-I to repeat the DNS assay with. It is hypothesised that the use of AG-I would produce similar results as those using galactan as the substrate (Figure 4.13).

All replicates of the DNS assay were conducted between pH 4 and pH 5. Like the azo-galactan substrate assays this initial pH range was chosen as it replicates the pH observed in the plant apoplast. A pH assay was conducted but apparent activity was detected in the negative (GFP) control at more alkali pH levels. Alongside this there appeared to be an unidentified “cloudy” aggregate formed in certain samples at acidic pH (pH 3). These aggregates were most likely protein in origin and gave artificially high readings in the spectrophotometer, meaning it was not possible to get accurate replicates of this assay at a wide range of pH levels. It was not explored whether the aggregate was the enzyme, the substrate, or a mixture of both, however there are published data on other catalytic enzymes behaving in a similar way. For example, α -Amylase has been shown to be inactivated and formed a “pH-induced aggregation” in acidic conditions below pH 4.5 (Yadav & Prakash 2011). Additionally,

this assay was only conducted at room temperature however it is possible there would be differences in absorbance levels if this was repeated at both higher and lower temperatures.

The galactan substrate used in the DNS assay are long chains of galactose subunits with β -1, 4 glycosidic bonds. This replicates the bond linkage pattern that GH53 enzymes recognise and hydrolyse in AG-I samples, however, it is missing the arabinose units. Although it is encouraging to witness no GH53 activity when AG-II was used as a substrate - which confirmed their inability to catalyse β -1, 3- or β -1, 6 glycosidic bonds - the results of this experiment would be more robust if the nematode GH53 proteins could have been assessed for enzymatic activity against their natural AG-I substrate. To summarise, this assay has proven that GrGAL1, GpGAL1, and RrGAL1 are β -1,4-galactanases, however it is still currently outstanding if they are true arabinogalactan endo- β -1,4-galactanases due to lack of substrate for functional testing. It is highly likely due to the sequence similarity conserved between the nematode GH53 proteins and other functionally characterised GH53 family members that they do function on arabinogalactan.

4.4.4 – Azo-galactan assay

Using the azo-galactan assay, it was not possible to reliably detect enzyme activity from the *G. rostochiensis* GH53 GrGAL1, although initial experiments did detect activity from the GH53 proteins from GpGAL1 and RrGAL1. This is likely due to a failure to replicate the correct conditions required for this enzyme to function. It is possible that the substrate used was not optimal. Azo-galactan has had the arabinose units removed from the backbone by the manufacturer. Although it was anticipated that the GH53 enzymes tested would act on the galactan backbone of arabinogalactan, it was unclear whether the prior removal of the arabinose would alter or inhibit the enzymes from functioning normally at the time of running this assay. Subsequent successfully runs of the DNS assay has shown this is unlikely to be the case, however. Unfortunately, it was not possible to buy an azo-dyed

arabinogalactan to use as a substrate in this assay. The pH was an important factor for this assay. The elution buffer which the preliminary assays were conducted in was at pH 8. This has now been shown to be too basic for the nematode GH53s to function. As shown in Figure 4.13, pH 4 was optimal for GpGAL1, RrGAL1, and GalA_Xc. pH 4 is more acidic than first predicted as the optimum pH for the nematode GH53 proteins. On average the plant apoplast has an approximate pH of 5, however this can alter due to stresses such as pathogen interaction and drought (Geilfus 2017). Optimum pH levels may have been off due to the mixture of substrate and enzyme in elution buffer and may have been a factor in some of the negative results, or those positive results which were difficult to replicate reliably.

There are many different buffers that proteins can be purified in and it is possible that the optimum buffer for the nematode GH53s to function was not used in this assay. It is interesting that the GalA_Xc shows abundant activity across a wide variety of temperatures and pH after as little as 10 minutes incubation time while the nematode GH53s had relatively low activity across extended periods of time (1h - overnight). It is possible that the nematode GH53s may have shown a higher activity level if they had been purified in a different buffer. In addition, GalA_Xc is a bacterial protein that is being expressed in a bacterial cell line. It is possible that higher activity levels of GrGAL1, GpGAL1, and RrGAL1 would be observed if it was possible to express them in a way which would replicate their native environment more closely. Although it is possible to transform and express proteins in the model organism *C. elegans*, it is not currently possible to replicate this in parasitic nematode species (Kranse *et al.* 2021a).

Detection of activity in the GFP negative control was not expected as this sample did not contain any enzyme that would have the ability to degrade the galactan substrate. This result could have been due to contamination from another GH53 sample such as the *X. campestris* positive control. It could also be that the azo-galactan substrate begins to degrade naturally after extended time periods;

however, this is unlikely as no such degradation was detected in later replicates of the nematode GH53 reactions.

After the pH range assay was completed, a second assay to determine an optimum temperature for the GH53 was attempted. Unfortunately, an optimum temperature was never ascertained because no further activity could be detected. The protein was purified and confirmed by SDS-PAGE gel on three separate occasions and all buffers and substrate solutions used in this assay were made fresh for each replicate, however none of these changes saw a return of the activity observed in previous azo-galactan assays. Ultimately this assay proved too unpredictable and non-replicable which led to the switch to using the DNS assay to confirm preliminary results.

4.5 – Future work

This chapter shows that the GH53 gene sequences observed in *G. rostochiensis*, *G. pallida*, and *R. reniformis* are expressed and the proteins are functional. In future, more in-depth study is required to determine accurate activity levels of the GrGAL1, GpGAL1, and RrGAL1 enzymes. As all work so far has been conducted *in vitro*, it would be useful to examine the function of these proteins inside the host plant. For example, the impact of silencing the expression of the GH53 genes using RNAi might be useful in determining whether these proteins play a key role during migration, as has been shown for other nematode CWDEs *e.g.* (Peng *et al.* 2016). Crystallography could also be used to confirm the structural predictions made using the 1-to-1 thread models produced in this chapter. These models gave a good overview; however, it is acknowledged that in practice the nematode GH53 proteins potentially have structural differences *e.g.*, amino acid side chains.

5. Cathepsin L-like

peptidases of plant-

parasitic nematodes

5.1 – Introduction

The potato cyst nematode species *Globodera rostochiensis* and *Globodera pallida* are sedentary endoparasites which feed and sequester nutrients from the host plant, causing yellowing and stunted growth. The longevity of PCN due to their hardy, dormant cysts in the field causes significant problems for the potato growing industry. With the withdrawal of many nematicides due to environmental concerns, it is important to understand the underlying biology of these species to discover possible new control methods.

As discussed in chapter 4, nematodes produce cell wall degrading enzymes in order to break down host tissues during invasion and migration. Nematodes also produce digestive enzymes that break down ingested host tissue into useful nutrients. Many of these enzymes are proteinases that can be categorised into 4 classes: cysteine, serine, aspartic, and metalloproteinases (Coomb & Mottram 1997). It should be noted that the words proteinases, proteases, and peptidases are widely used across the literature to denote the same function. Targeting of these digestive enzymes has been explored as a control method for PPN. For example, digestive cysteine proteinases were targeted for the control of *G. pallida*. Transgenic tomato lines expressing variants of the cysteine proteinase inhibitor Oryzacystatin-I (Oc-I) from *Oryza sativa* (rice) showed a significant decrease in development and reproductive

ability of *G. pallida* (Urwin *et al.* 1995). Similar studies have shown that transgenic plants expressing such proteinase inhibitors can give useful control of other PPN including RKN and migratory plant-parasitic nematodes. A modified version of the Oc-I gene (OC-IΔD86) in *A. thaliana* transgenic lines was subsequently shown to reduce reproductive fitness in both *H. schachtii* and *M. incognita* (Urwin *et al.* 1997). A 78.3% reduction in fecundity was also observed in *M. incognita* on transgenic OC-IΔD86 eggplant lines (Papolu *et al.* 2016). OC-IΔD86 was transiently expressed in banana lines which resulted in reduced reproductive fitness of the burrowing nematode *Radopholus similis* (Howard J. Atkinson *et al.* 2004).

5.1.1 – Cysteine proteases

Cysteine proteases (also known as thiol proteases) function during protein degradation by hydrolysing peptide bonds. Cysteine proteases are found in all organisms and have been reported as having functions in a wide range of processes such as basal protein processing, digestion, and turn over. Cysteine proteases have more complex roles in different species such as programmed cell death and protein storage in plants, and roles in embryogenesis and parasitism in nematodes (Grudkowska & Zagdańska 2004, Caffrey *et al.* 2018). These proteases are ancient and have been shown to be present as far back as the common ancestor between fungi and bacteria (Barrett & Rawlings 2014).

Due to the hydrolytic nature of these enzymes, the majority are produced as zymogen with a pro-domain to inhibit unwanted activity before the protease reaches its target. The pro-domain must be removed or have a conformation change to prevent blocking of the active site for the cysteine protease to perform its catalytic function. Many cysteine proteases contain sites for N-glycosylation which are thought to play a role in stabilisation during the activation process of these proteases (Goettig 2016). The active site of cysteine proteases is made up of a conserved catalytic triad of a cysteine, histidine, and a third residue, usually an asparagine or aspartic acid. The hydrolysis of proteins by cysteine proteases is conducted using the

thiol group of the triad cysteine residue which is nucleophilic (electron donor) while the histidine acts as a proton donor (Buttle & Mort 2013, Verma *et al.* 2016). Cysteine proteases are divided into 14 superfamilies (or clans) which are primarily defined by evolutionary background, but are also categorised by structural similarity (Barrett & Rawlings 2014). There is also a set of “unassigned” families that do not currently fit into these 14 superfamilies. One of the largest clans is clan CA which contains the papain family – papain-like cysteine proteases were the first to be sequenced – and is also where the cathepsin subfamilies reside (Barrett & Rawlings 1996). Other large clans of note include clan CD which contains caspase-like proteases. The caspase-like proteases have roles in programmed cell death (Atkinson *et al.* 2009).

Cysteine proteases are found in the genomes of all living organisms, including parasites. It has been shown that cysteine proteases are integral for virulence and host invasion of some tropical parasites including *Trypanosoma brucei* (the causal agent of sleeping sickness), *Leishmania* subspecies (leishmaniasis), *Plasmodium* subspecies (malaria) and *Toxoplasma gondii* (toxoplasmosis) (Siqueira-Neto *et al.* 2018). Many of the cysteine proteases from these parasitic species have been assessed as chemotherapeutic and vaccine targets due to the ability to design inhibitors of these proteins (Rosenthal *et al.* 2005). As discussed above (Section 5.1), it has been shown that proteinase inhibitors can be used for PPN reduction and control. Both cysteine and serine proteinase inhibitors can be targeted using a similar strategy. Using the serine proteinase inhibitor cowpea trypsin inhibitor (CpTI), the sexual fate of *G. pallida* was skewed towards males when CpTI was expressed in transgenic potato lines (Hepher & Atkinson 1992). For *M. incognita* a reduction in female egg production was recorded in the presence of CpTI (Hepher & Atkinson 1992).

5.1.2 – Cathepsin L peptidases

Cathepsins are a diverse family of cysteine proteases found in clan CA. Cathepsins can be further divided into multiple subfamilies including Cathepsin L peptidases.

The subfamily of cathepsin L peptidases (Cpl) mostly consists of endopeptidases and have been identified in many diverse species. This includes many species of nematodes, including free-living nematodes such as *Caenorhabditis elegans*, and many PPN species such as *Heterodera glycines* (Hashmi *et al.* 2002, Urwin *et al.* 1997).

There are many examples of cathepsin L peptidases playing an integral role in nematode egg development. The *cpl-1* cathepsin L gene (also referred to as *Ce-cpl-1*) from *C. elegans* has been functionally characterised using RNAi as well as production of mutant lines (Hashmi *et al.* 2002). RNAi resulted in the arrest of early embryonic development between the 100-200 cell stage. Aberrant processing of yolk proteins was also observed. Taken together, these results suggest that *cpl-1* plays a vital role during embryogenesis. *Cpl-1* also plays a role in digestion and fat storage in *C. elegans* (Lin *et al.* 2019). This study showed when *cpl-1* expression is inhibited the synthesis of serotonin is upregulated to induce fat loss. This result has been replicated in mouse models as cathepsin L peptidases have been linked to obesity in both mice and humans (Lin *et al.* 2019). Cathepsin L peptidases found in animal parasitic nematodes also have digestive roles. The cathepsin L peptidase SmCL3 from the human blood fluke *Schistosoma mansoni* localises in the gastrodermis (the inner cell lining of invertebrate alimentary tract) and has been shown to hydrolyse haemoglobin and serum albumin, blood components ingested from the human host (Dvorák *et al.* 2009). Many examples of cysteine proteinases that are important for digestion in PPN have also been described, and their potential use as control targets has been demonstrated, as described above. Targeting *Mi-cpl-1* from the root-knot nematode *Meloidogyne incognita* using RNAi showed an almost 60% reduction in egg-producing females (Shingles *et al.* 2007). Recently three new cathepsin L-like peptidases, *Bx-cpl-1*, *Bx-cpl-2*, and *Bx-cpl-3*, were identified in the pine wood nematode, *Bursaphelenchus xylophilus*. RNAi silencing of *Bx-cpl-1*, *Bx-cpl-2*, and *Bx-cpl-3* showed varying degrees of reduction in pathogenicity and reproduction and

one of these was specifically expressed in the intestine, implying a role in digestion of food (Xue *et al.* 2019).

5.1.3 – Chapter aims

This chapter reports the characterisation of the Gr-cpl-like-2 protein which was identified via the effector pipeline used in chapter 3. With a diverse range of functions for these proteases noted in the literature, Gr-cpl-like-2 was analysed to ascertain whether its function was to assist in parasitism. This work was conducted to complement the work produced for the unpublished manuscript “Cathepsin L cysteine proteinases are conserved in diverse plant-parasitic nematode species” by C. J. Lilley, J. Shingles, H. J. Atkinson, and P. E. Urwin, which details the identification and subsequent characterisation of gp-cpl-1 from *G. pallida*.

5.2 – Materials and methods

5.2.1 – Gene identification

Gr-cpl-like-2 (GROS_g03615) from *G. rostochiensis* was identified from the effector pipeline detailed in chapter 3. Gr-cpl-like-2 was used alongside two genes from *H. glycines* (Hgcp-I and Hgcp-II (Urwin *et al.* 1997)) as query sequences for BLAST similarity searches against *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *H. glycines* genome and transcriptome data. All BLAST searches were conducted with an E-value threshold of 1e-05.

5.2.2 – Nematode collection

Pre-parasitic (J2) and mixed parasitic nematodes were collected from cysts of *G. rostochiensis* (pathotype Ro1) and *G. pallida* (pathotype Pa2/3) populations as described in sections 2.1.1 and 2.1.2 respectively.

5.2.3 – RNA and cDNA synthesis

RNA and cDNA was synthesised as described in sections 2.2 and 2.3.

5.2.4 – Cloning of Gr-cpl-like-2

Multiple genes were identified from multiple nematode species as a result of the pipeline analysis. It was decided that all further work would be conducted using Gr-cpl-like-2 as a representative of these cathepsin L peptidases. The coding region of the *G. rostochiensis* gene Gr-cpl-like-2 (GROS_g03615) was amplified by PCR from cDNA using KOD Hot start DNA polymerase (Merck). The Gr-cpl-like-2 gene was cloned with open reading frame from start to stop codon, excluding the endogenous signal peptide. The PCR product was electrophoresed on a 1.5% agarose gel and excised using the QIAquick gel extraction kit (Qiagen). The amplified fragment was cloned into the pGEM[®]-T Easy vector (Promega) following manufacturer's protocol. The constructs were transformed into *E. coli* (DH5 α) competent cells and following miniprep were sequenced using M13 and gene specific primers (Table 5.1). An in

depth description of the protocols used for cloning of genes can be found in sections 2.5 to 2.7.

Table 5.1 – Primers for cloning of the *Gr-cpl-like-2* gene

Primer name	Sequence	Function
Gr-cpl-like-2_F	ACCATGGCACCGACCGATCAACAG	Cloning
Gr-cpl-like-2_R	TTAAACAACGGGATATGATGCCA	Cloning
M13 Forward	GTAAAACGACGGCCAG	Sequencing
M13 Reverse	CAGGAAACAGCTATGAC	Sequencing

5.2.5 – Re-annotation of genome sequences & gene expression data

Gene re-annotation was conducted to ensure predicted gene sequences used from *G. rostochiensis*, *G. pallida*, *H. glycines*, and *R. reniformis* were full length and that gene expression data was correctly attributed. Transcriptome reads were trimmed of poor quality sequences using Trimmomatic (Bolger *et al.* 2014). Gene sequences were mapped against the trimmed reads using bowtie2 (Langmead & Salzberg 2012). Read alignments were sorted and indexed using SAMtools (Li *et al.* 2009). Indexed and sorted files were entered into Bedtools to create count and gene length data (Quinlan & Hall 2010). Differential expression of RNA-seq data was analysed using EdgeR (McCarthy *et al.* 2012). All gene expression data were normalised using the Trimmed means of M values (TMM) normalisation method prior to use in this thesis (Robinson & Oshlack 2010).

5.2.6 – Bioinformatic and phylogenetic analysis

Genes from *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *H. glycines* were assessed for the presence of a signal peptide using SignalP 4.1. Amino acid sequences were aligned using MUSCLE (version 3.8.31) with cathepsin L peptidase protein from *H. glycines*, *C. elegans*, and *H. sapiens* from Genbank for comparison. The alignment was trimmed using TrimAL (-strict). A phylogenetic tree was created using Bayesian inference (MrBayes) with a Whelan and Goldman (WAG), proportion of invariable sites (+I), rate of variation across sites (+G) substitution model with 500000

generations, a burn in rate of 25%, and a sample frequency of 10%. The tree was constructed using TOPALi v2.5.

5.2.7 – In-silico structural analysis

The predicted protein structure of Gr-cpl-like-2 was modelled using a 1-to-1 threading model based on *H. sapiens* procathepsin L protein CTSL1 (RCSB-PDB: 1CS8) (Coulombe *et al.* 1996). CTSL1 has 46.08 percentage identity (%ID), making this the most conserved cathepsin L peptidase compared to Gr-cpl-like-2 with a solved crystal structure. Modelling was conducted using protein homology/analogy recognition engine V2.0 (Phyre2) (Kelley *et al.* 2015). Protein models were annotated using CCP4mg molecular graphics software (McNicholas *et al.* 2011).

5.2.8 – In situ hybridisation

In situ hybridisation was conducted as described in section 2.8 using primers described in Table 5.2.

Table 5.2 – Primers for production of Cathepsin L peptidase in situ hybridisation probes

Primer name	Sequence
Gp-cpl-like-1_ISHF	CACGCTCGTCAGACGGGACAG
Gp-cpl-like-1_ISHR	GGTGTCCGTTGCGCCACATC
Gr-cpl-like-1_ISHF	TGGCGTGCTCGTGATCGT
Gr-cpl-like-1_ISHR	TGTCGATGAACTGCTTAGCG
Gr-cpl-like-2_ISHF	ACCATGGCACCGACCGATCAACAG
Gr-cpl-like-2_ISHR	GGTCGAAGCCCACAAATTC
Gr-cpl-like-3_ISHF	ATGCCCAAAGTGAGGACAAA
Gr-cpl-like-3_ISHR	CCGACATTTCTTTGGACAAATT
Gr-cpl-like-4_ISHF	GAACCCTTTGAACAGCTCTCTG
Gr-cpl-like-4_ISHR	AATCGCCTCAGTTTTCTCA
Gr-cpl-like-5_ISHF	TTGGACGGTAAAAGCAGTCG

Gr-cpl-like-5_ISHR	ATTACGCCCAAACGCTTCCA
--------------------	----------------------

5.2.9 – RNA interference (RNAi)

5.2.9.1 – Silencing hairpin construct production

Primers were used to amplify a 522 bp region of Gr-cpl-like-2 and a 549 bp region of GFP genes (Table 5.3). The RNAi constructs were made using restriction digest cloning into pHANNIBAL as the entry vector for generation of the RNAi hairpin for structure to silence genes in plants. The pART27 vector was subsequently used as the destination vector due to its high copy number replication in both *E. coli* and *Agrobacterium* (Figure 5.1). In order for the gene of interest to be silenced a dsRNA hairpin structure was created in the pHANNIBAL vector. This required two sections of the gene of interest, one sense and one anti-sense, to be inserted on either side of a linker sequence. In the case of the pHANNIBAL vector this linker sequence is the pyruvate orthophosphate dikinase (PDK) intron. The PDK intron will be spliced out to produce a section of double stranded RNA which will cause silencing of the gene of interest in the nematode. To produce this hairpin structure XhoI or XbaI restriction sites were integrated into the forward primer for addition to the 5' end of the gene. KpnI or HindIII restriction sites were incorporated into the reverse primer for addition to the 3' end of the gene. Through PCR this produced two amplified sections: XhoI_Gr-cpl-like-2_KpnI and XbaI_Gr-cpl-like-2_HindIII; the Gr-cpl-like-2 region chosen for RNAi was flanked by the restriction site pairs XhoI-KpnI and XbaI-HindIII. The pHANNIBAL vector was digested with XhoI and KpnI before ligation with the XhoI_Gr-cpl-like-2_KpnI fragment. A second digest of this vector with KpnI and HindIII was performed before ligation with the XbaI_Gr-cpl-like-2_HindIII fragment. Correct insertion was confirmed by PCR. The hairpin construct was removed from the pHANNIBAL vector by restriction digest with NotI. The pART27 vector was also digested with the NotI restriction enzyme. The hairpin construct was inserted into the pART27 destination vector via ligation catalysed by a T4 DNA ligase reaction (New England Biolabs). All enzymes used in this process were from Promega.

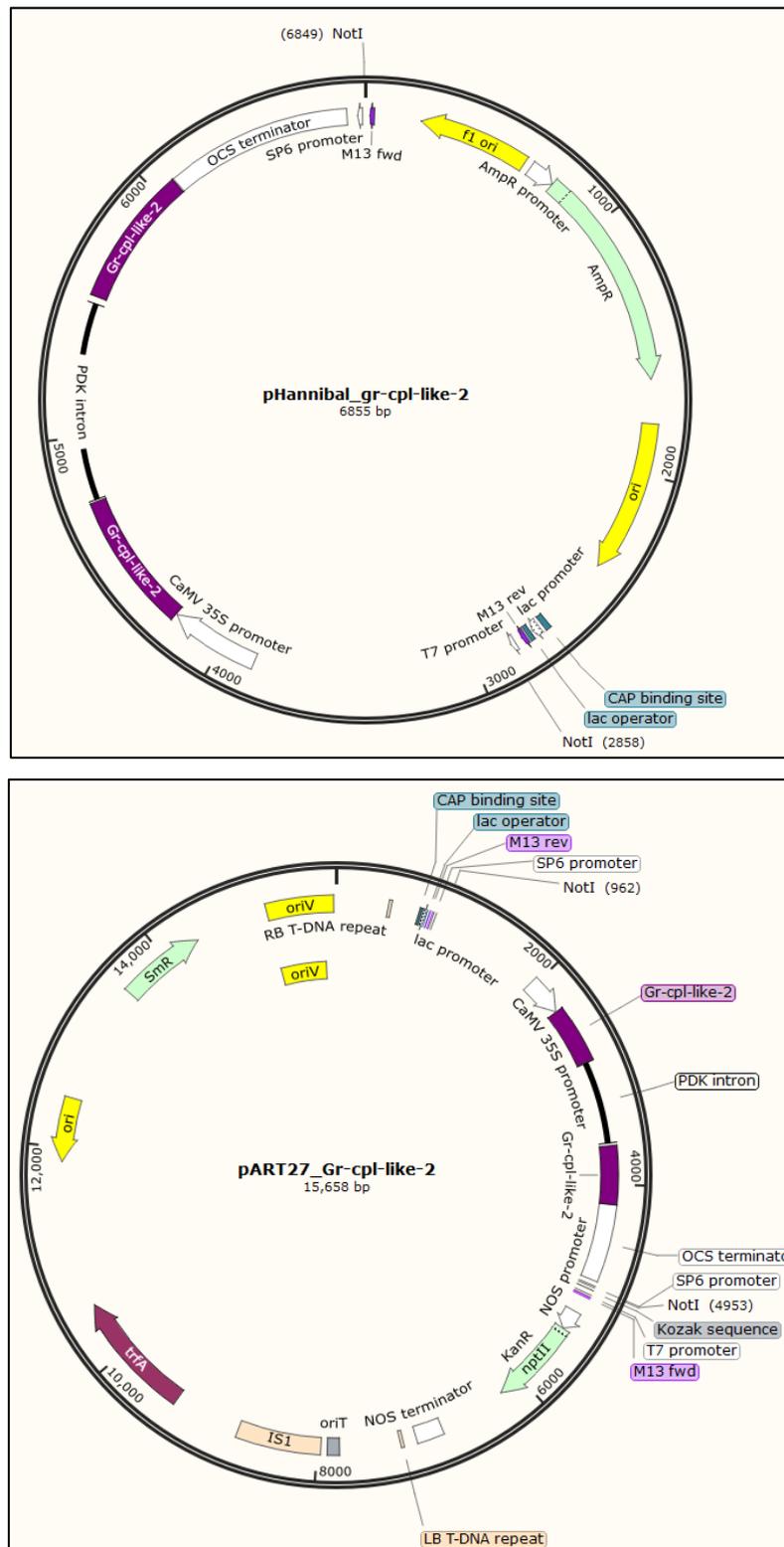


Figure 5.1 - Vector maps of pHANNIBAL and pART27 containing Gr-cpl-like-2 RNAi construct. Upper map – pHANNIBAL entry vector containing the RNAi fragments of Gr-cpl-like-2 flanking the PDK intron for hairpin formation using XhoI-KpnI and XbaI-HindIII restriction enzyme pairs. Lower map – pART27 entry vector containing the Gr-cpl-like-2 hairpin inserted via NotI restriction digest and ligation.

Table 5.3 – Primers for Gr-cpl-like-2 and GFP RNAi constructs

Primer name	Sequence	Function
Gr-cpl-like-2_XhoI_F	ACA CTC GAG GTTTCATTGTCCGAGCAAATCTG	Cloning RNAi region with restriction site addition
Gr-cpl-like-2_KpnI_R	ACA GGT ACC TTAAACAACGGGATATGATGCCA	Cloning RNAi region with restriction site addition
Gr-cpl-like-2_XbaI_F	ACA TCT AGA GTTTCATTGTCCGAGCAAATCTG	Cloning RNAi region with restriction site addition
Gr-cpl-like-2_HindIII_R	ACA AAG CTT TTAAACAACGGGATATGATGCCA	Cloning RNAi region with restriction site addition
RNAi GFP XhoI_F	ACA CTC GAG TGGCCAACACTTGTCCTACT	Cloning RNAi region with restriction site addition
RNAi GFP KpnI_R	ACA GGT ACC TTATTTGTAGAGCTCATCCATGCC	Cloning RNAi region with restriction site addition
RNAi GFP XbaI_F	ACA TCT AGA TGGCCAACACTTGTCCTACT	Cloning RNAi region with restriction site addition
RNAi GFP HindIII_R	ACA AAG CTT TTATTTGTAGAGCTCATCCATGCC	Cloning RNAi region with restriction site addition
Kana-nptII-F	ATGACTGGGCACAACAGACAATCGGCTG CT	confirmation of RNAi construct presence in transgenic lines
Kana-nptII-R	CGGGTAGCCAACGCTATGTCCTGATAGC GG	confirmation of RNAi construct presence in transgenic lines
StEF1 α _F	CCAGAAGAAGGGAAAAGTGAA	SQRT PCR control to confirm equal loading of cDNA
StEF1 α _R	CAACAAAAGCAAAGAAAACAG	SQRT PCR control to confirm equal loading of cDNA

5.2.9.2 – Agrobacterium-mediated potato transformation (RNAi transgenic line production)

The pART27 constructs were transformed into *Agrobacterium tumefaciens* strain AGL1. Overnight cultures of AGL1 containing either the *Gr-cpl-like-2* hairpin or GFP control pART27 constructs were grown at 28 °C with shaking at 250 rpm in LB supplemented with kanamycin (50 mg/ml). Cultures were centrifuged at 200 rpm for 10 minutes 2-4 hours before transformation and resuspended in 10 ml MS30 broth. Cultures were assessed using a spectrophotometer for an OD₆₀₀ between 0.5 and 0.8 before 100 mM acetosyringone was added.

Stem segments (approximately 5-10 mm in length), internodes, and petioles of 4 to 6 week old potato plantlets (*cv. Desiree*) were cut and halved lengthwise down the centre. Sections were stored in 25 ml of MS30 until approximately 100 explants per transformation were cut. Explants were then placed on HB1 media plates with no antibiotic selection (HB1: MS30 medium, HB1 stock solution (1L H₂O, 2.5g zeatin riboside (ZR), 0.2 mg 1-Naphthaleneacetic acid (NAA), 0.02 mg Gibberellic acid (GA3))). One hundred microlitres of AGL1 MS30 broth culture containing either the RNAi construct or GFP control was added to the explants before incubation for 20 minutes, shaking at 50 rpm, at room temperature in the dark. Explants were blot dried onto sterile Whatman filter paper. Explants were transferred onto HB1 plates (maximum 100 per plate) with the cut surface of the explant facing down onto the media. Explants were grown in low light conditions on potato callus induction medium (HB1) plates for 2 to 3 days.

5.2.9.3 – Callus induction

Explants (25 per plate) were transferred onto HB1 plates with Agrobacterium and transgene selection (HB1, timentin (160 mg/ml), kanamycin (50 mg/ml)) and grown for 7 days in full light conditions between 18-24 °C. Explants were transferred to HB1 plates with Agrobacterium and transgene selection every 14 days until the production of well-developed calli (3-6 weeks).

5.2.9.4 – Shoot regeneration

Twelve to thirteen explants with callus were transferred onto potato regeneration medium (HB2) plates (HB2: MS30 medium (MS30: Murashige & Skoog with vitamins (MO222)), HB2 stock solution ((1L H₂O, 2g ZR, 0.02 mg NAA, 0.02 mg GA3) + kanamycin (50 mg/ml) and spectinomycin (50 mg/ml)) (Figure 5.2 A). Explants were transferred to fresh HB2 plates every 14 days until the development of shoots. Shoots over 1 cm in length were excised from the explant and transferred to MS30 agar plates with selection for agrobacterium and construct: sucrose, pH 5.7, kanamycin (50 mg/ml), spectinomycin (50 mg/ml)), and allowed to root (Figure 5.2 B). Shoots were labelled with the corresponding number of the “mother” explant as multiple shoots were harvested from the same explant. If callus formation was visible on new shoot plantlets the callus was excised and placed onto fresh MS30 plates containing kanamycin (50 mg/ml) and spectinomycin (50 mg/ml). Roots were observed approximately 2 weeks into this process.

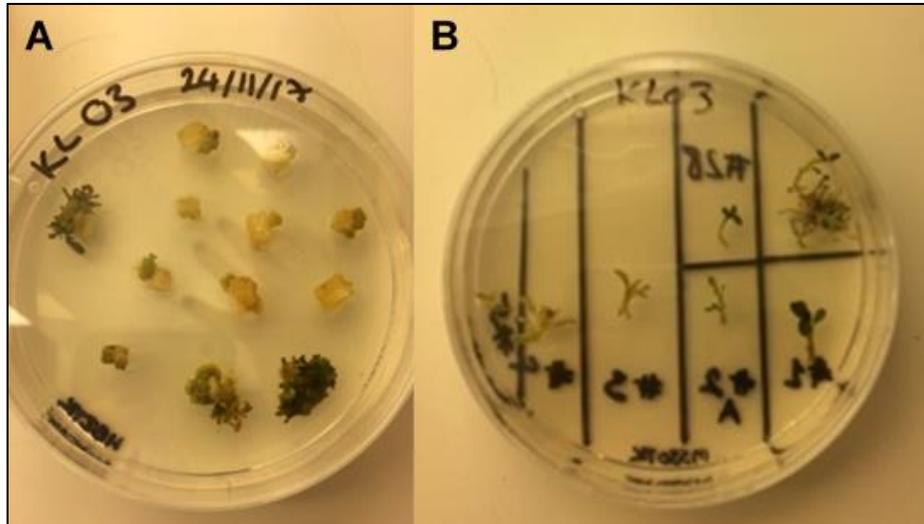


Figure 5.2 – Transgenic *S. tuberosum* line production for RNAi. A – Callus formation on HB2 plates. B – Excised shoot material on MS30 plates for root production.

Plants were assayed for the presence of the constructs using PCR. Leaf material was excised using a scalpel and DNA was extracted from the tissue using an Aquagenomic DNA extraction kit (MultiTarget Pharmaceuticals). Tissue was homogenized in 200 µl

of aquagenomic solution using a sterile micropestle. Homogenate was transferred to a fresh 1.5 ml Eppendorf tube and incubated for 20 minutes at 75 °C. Samples were centrifuged at 12000 rpm for 5 minutes. The supernatant was transferred to a new 0.5 ml Eppendorf tube. One volume of isopropanol was added and the sample was mixed with a vortex for 1 minute. The samples were centrifuged at 12000 rpm for 5 minutes to pellet the DNA and the supernatant was discarded. The tube was gently rinsed with 70% ethanol twice before air drying for 5 to 10 minutes. The DNA pellet was resuspended in 100 µl Tris-EDTA (TE) buffer and incubated at 22 °C for 15 minutes. Samples were centrifuged at 12000 rpm for 5 minutes to pellet any insoluble components. The supernatant was transferred to a fresh 0.5 ml Eppendorf tube and stored at -20 °C before use in PCR. A PCR to confirm construct presence (Gr-cpl-like-2 or GFP in pART27) using GoTaq polymerase was carried out as described in section 2.4 using Kana-nptII-F and Kana-nptII-R primers (Table 5.3). The PCR was run for 30 cycles with an annealing temperature of 64 °C and extension time of 1 minute. Transgenic plantlets that contained the RNAi construct were transferred to soil in a propagator and grown until large enough for transfer into 3.5-inch, and subsequently 12-inch pots. Plantlets were grown in a glasshouse on 16 hour light/ 8 hour dark cycle between 18-24 °C.

5.2.9.5 – Analysis of RNA interference via semi-quantitative reverse transcription PCR (SQRT-PCR)

Leaf material was excised from each plant line and frozen in liquid nitrogen. RNA was extracted from each line using the RNeasy plant mini kit (Qiagen) following the manufacturers protocol. Purity was established based on a 260 nm/280 nm ratio using a nanodrop spectrophotometer. A ratio of approximately 2.0 was considered pure (Thermo Scientific 2012). A PCR was used to check for DNA contamination of RNA. Two PCRs were simultaneously run, one using the newly extracted RNA as a template and the other using gDNA. Gr-cpl-like-2_XbaI_F and Gr-cpl-like-2_HindIII_R primers were used. PCRs were run for 40 cycles to ensure no DNA was present in

RNA samples. cDNA was synthesised using the protocol stated in sections 2.2 and 2.3. Equal volumes of RNA were added to each cDNA synthesis reaction.

A semi-quantitative reverse transcription polymerase chain reaction (SQRT-PCR) was performed to compare construct expression levels between plant lines. A PCR using the Gr-cpl-like-2_XbaIF and Gr-cpl-like-2_HindIIIR primers was carried out for 28 cycles. Gel bands were imaged using a UVIdoc HD2 gel imager (UVITEC Cambridge). Primers for the housekeeping gene Elongation factor 1 α (stEF1 α) were used in a control PCR to confirm equal loading of cDNA. Primers can be found in Table 5.3.

5.2.9.6 – RNAi line screening of females and cysts

Uniform stem cuttings of each RNAi line (Gr-cpl-like-2 lines 12, 14, 20, 23, and GFP control) had rooting growth hormone powder (Doff) applied to the base and any excessively large foliage was removed. Cuttings were grown in Jiffy-7[®] peat pellets for 14 days until roots were visible. Cuttings were moved to soil in root trainers for a further 14 days to establish stable root structures. Approximately 1000 *G. rostochiensis* J2s were applied to each of 8 replicates per transgenic line and allowed to grow for 7 weeks before females on root systems were manually counted. Screens were conducted in a glasshouse on a 16 h light, 8 h dark cycle between 18 - 24 °C. Female count data was normalised before significance testing using One-Way ANOVA with post hoc (Tukey) test using IBM SPSS statistics software (IBM Corp 2017). Approximately 20 cysts from each replication were randomly selected and imaged using an Amscope stereo microscope and camera. The area of each cyst (mm²) was measured using ImageJ (Schneider *et al.* 2012). Cyst area data were assessed for normality before significance testing using One-Way ANOVA with post hoc (Tukey) test using IBM SPSS statistics software (IBM Corp 2017). Significance in graphs produced is indicated by a system of asterisks: * = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$.

5.3 – Results

5.3.1 – Cathepsin L gene identification in PPN

Gr-cpl-like-2 (GROS_g03615) is a *G. rostochiensis* gene which was initially identified through the effector pipeline described in chapter 3. Gr-cpl-like-2 alongside Hgcp-I and Hgcp-II (previously identified by (Urwin *et al.* 1997)) were subsequently used as query sequences in order to identify other similar genes from *G. rostochiensis*, *G. pallida*, and *R. reniformis* that may have been missed during the pipeline process. Using these similarity searches, five cathepsin L-like peptidases were identified in each of the PPN species *G. rostochiensis*, *G. pallida*, and *R. reniformis* (*Gr-cpl-like-1 to 5*, *Gp-cpl-like-1 to 5* and *Rr-cpl-like-1 to 5*). Three new cathepsin L-like peptidases were also identified from *H. glycines* which have been named Hgcp-like-III, Hgcp-like-V and Hgcp-like-VI respectively to complement the previously published nomenclature (Table 5.4) (Urwin *et al.* 1997). Gr-cpl-like-2 (*G. rostochiensis*) and Gp-cpl-like-1 (*G. pallida*) were shown to be orthologs of Hgcp-II and Hgcp-I (*H. glycines*) respectively. Gp-cpl-like-2 was discovered to be the *G. pallida* homolog of Gr-cpl-like-2 and Hgcp-II. All newly identified genes were allocated a name e.g., Gr-cpl-like-1, for consistency (Table 5.4). Due to the circumstances in which Gr-cpl-like-2 was identified, it was decided to investigate the clade it resides in more closely (clades defined in Figure 5.4).

Table 5.4 – Cathepsin L peptidase nomenclature and signal peptide status

Nematode species	Sequence name from genome/transcriptome data	Signal peptide (Yes/No)	Given gene name
<i>G. rostochiensis</i>	GROS_g03960	Y	Gr-cpl-like-1
<i>G. rostochiensis</i>	GROS_g03615	Y	Gr-cpl-like-2
<i>G. rostochiensis</i>	GROS_g06676	Y	Gr-cpl-like-3
<i>G. rostochiensis</i>	GROS_g11484	N	Gr-cpl-like-4
<i>G. rostochiensis</i>	Comp51505_c0_seq8	N	Gr-cpl-like-5
<i>G. pallida</i>	comp392_c0_seq1 (AY999065.1)	Y	Gp-cpl-like-1
<i>G. pallida</i>	7dpi_comp1973_c0_seq1	Y	Gp-cpl-like-2
<i>G. pallida</i>	GPLIN_000168400	Y	Gp-cpl-like-3
<i>G. pallida</i>	GPLIN_000543500	N	Gp-cpl-like-4
<i>G. pallida</i>	GPLIN_000876700	N	Gp-cpl-like-5
<i>R. reniformis</i>	comp43411_c0_seq3 (AY999066)	Y	Rr-cpl-like-1
<i>R. reniformis</i>	comp45376_c0_seq1	Y	Rr-cpl-like-2
<i>R. reniformis</i>	g2797.t1	Y	Rr-cpl-like-3
<i>R. reniformis</i>	c12773_g1_i1	N	Rr-cpl-like-4
<i>R. reniformis</i>	comp46192_c0_seq1	N	Rr-cpl-like-5
<i>H. glycines</i>	CAA70693.1	Y	Hgcp-I
<i>H. glycines</i>	CAA70694.1	Y	Hgcp-II
<i>H. glycines</i>	Hetgly.G000011538.t1	Y	Hgcp-like-III
<i>H. glycines</i>	Hetgly.G000003719.t1	N	Hgcp-like-VI
<i>H. glycines</i>	Hetgly.G000002279.t1	N	Hgcp-like-V
<i>C. elegans</i>	CCG28194.1	Y	Cpl-1
<i>H. sapiens</i>	CR457053.1	Y	CTSL1

BLAST similarity searches identified genes from both *G. pallida* and *R. reniformis* that aligned against Hgcp-II and Gr-cpl-like-2 (Figure 5.3 A). It became apparent that some of the sequences (GPLIN_0001480600 and GPLIN_000084500) had been

mispredicted meaning these were annotated incorrectly in the genome data as individual partial sequences. These were recapitulated into one full length gene sequence using the corresponding sequence (7dpi_comp1973_c0_seq1) from the transcriptome data and renamed Gp-cpl-like-2. The opposite issue was present with one of the original genes identified from the *R. reniformis* genome data. G29606.t1 originally had an additional 140 amino acids incorrectly attributed to the end of this gene sequence. G29606.t1 was replaced on further analysis with the corresponding sequence (comp43411_c0_seq3) from the transcriptome data and renamed Rr-cpl-like-1. Due to this, all gene expression data attributed to all the genes identified above were reattributed to the newly corrected gene sequences.

Sequence similarity searches using BLASTp (nr) and pairwise alignments (BLOSUM62) show that Gp-cpl-like-2 has the highest sequence identity (%ID) with Hgcp-II at 64.1%, followed closely Gr-cpl-like-2 at 63.82%, and Rr-cpl-like-2 at 54.33%. These proteins are conserved across nematodes more broadly as Hgcp-II and Cpl-1 from *C. elegans* have 56% ID. The alignment shows that the amino acids which form the catalytic triad observed in cathepsin L peptidases, cysteine (CYS151), histidine (HIS291), and asparagine (ASN312) are highly conserved across these PPN species. The majority of the identified cathepsin L-like peptidases contain a signal peptide (SP) to target them to the secretory pathway (Table 5.4). It is interesting to note that two genes from each nematode species analysed do not contain a SP; Gr-cpl-like-4, Gr-cpl-like-5, Gp-cpl-like-4, Gp-cpl-like-5, Rr-cpl-like-4, Rr-cpl-like-5, Hgcp-like-VI, and Hgcp-like-V. This may indicate a different role to that of the other cathepsin L-like peptidases. N-glycosylation sites have been identified in the pro-domains of Gr-cpl-like-2, Gp-cpl-like-2, and Rr-cpl-like-2. Canonically N-glycosylation sites occur at the amino acid motif NXS/T. The -NSS N-glycosylation motif is conserved across Hgcp-II, Gr-cpl-like-2, and Gp-cpl-like-2, while Rr-cpl-like-2 has an alternative N-glycosylation sequence, -NVS. A conserved proline residue can also be identified promptly after the N-glycosylation site which is likely the first residue of the mature protein. This complements what is known about many

cysteine proteases being translated with a pro-domain that inhibits inappropriate activity by preventing substrate access to the active site. This pro-domain is then removed in an auto-catalytic process. The cathepsin L peptidase pro-domain is approximately 100 amino acids in length and contains two conserved domains: ERFNIN and the GNFD motif which both play a role in protein folding (Vernet *et al.* 1995, Karrer *et al.* 1993, Aich & Biswas 2018). The ERFNIN (Ex_{2/3}Rx₂Fx₂Nx₃lx₃N) motif is present in the nematode species analysed here (Figure 5.3 A), however it is not as conserved as observed in other species (Karrer *et al.* 1993). The human cathepsin L peptidase CTSL1 has the motif ERWNIN where the conserved phenylalanine (F) is replaced with a tryptophan (W) (Karrer *et al.* 1993). In nematode cathepsin L peptidase it appears that the majority of the motif e.g., the glutamic acid (E), isoleucine (I), and second asparagine (N) are conserved across *G. rostochiensis* (Residues E56, I71, N75), *G. pallida* (E53, I68, N72), *R. reniformis* (E40, I55, N59), and *H. glycines* (E64, I78, N82). The arginine (R) is conserved in *G. rostochiensis* (R60), *G. pallida* (R57), and *H. glycines* (R68), however is replaced by a glutamine (Q) in *R. reniformis* (Q44). The phenylalanine (F) position in the motif appears to be the least conserved as it is only present in *H. glycines* (F72) and substituted with a different amino acid in each of the other nematode amino acid sequences, *G. rostochiensis* (Tyrosine, Y64), *G. pallida* (Isoleucine, I61), *R. reniformis* (Leucine, L48). The first asparagine residue has been substituted with an alanine in all nematode amino acid sequences analysed, *G. rostochiensis* (A67), *G. pallida* (A64), *R. reniformis* (A51), and *H. glycines* (A74). The ERFNIN motif functions to restrict the folding pattern of the pro-domain (Groves *et al.* 1998). The GNFD motif (GxNxFx_D) is also identifiable in these nematodes but it is rarely present in this form, rather instead they use the alternative amino acids proposed and observed in other studies of cysteine proteases (Vernet *et al.* 1995, Musyoka *et al.* 2019). In the four nematodes analysed the GNFD motif actually appears as ANLX (Figure 5.3 A). The alanine (A), asparagine (N), and leucine (L) are all conserved but the aspartic acid (D) is substituted for tyrosine (Y) in *R. reniformis*, histidine (H) in *H. glycines*, and asparagine (N) in both *G. rostochiensis* and *G. pallida*. *G. rostochiensis* (A88, N90,

L92, N94), *G. pallida* (A85, N87, L89, N91), *R. reniformis* (A72, N74, L76, Y78), and *H. glycines* (A95, N97, L99, H101). It is currently unclear what effect the observed differences in these motifs have in the pro-domain folding of these nematodes.

A 1-to-1 threaded model was produced to predict the protein structure of Gr-cpl-like-2 (Figure 5.3 B, C). This was modelled on the crystal structure of pro-cathepsin L peptidase, CTSL from *H. sapiens*. The model demonstrates that Gr-cpl-like-2 most likely shares the same canonical structure of the cathepsin L family (Figure 5.3 C). It is apparent that the amino acids in and positioning of the catalytic triad, cysteine, histidine, and asparagine that form the active site are also likely to be highly conserved (Figure 5.3 A, C).

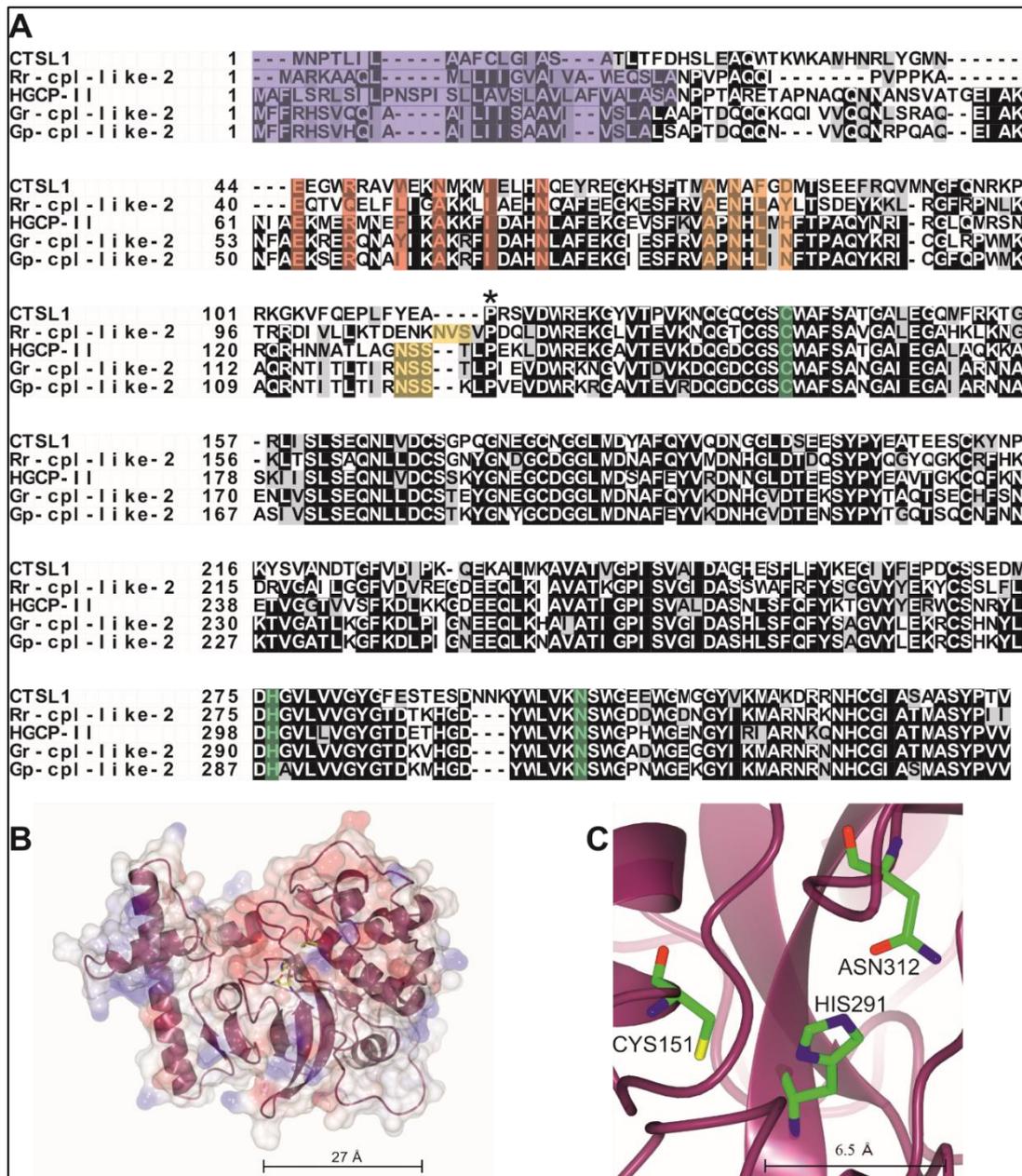


Figure 5.3 – Alignment and protein structure prediction of *Gr-cpl-like-2*. A - Protein alignment of Cathepsin L-like peptidases from *G. rostrchiensis* (*Gr-cpl-like-2*), *G. pallida* (*Gp-cpl-like-2*), *R. reniformis* (*Rr-cpl-like-2*), *H. glycines* (*Hgcp-II*), and *H. sapiens* (*CTSL1*). The residues of the ERFNIN and GNFD motifs are highlighted in red and orange respectively. B - Predicted protein structure of *Gr-cpl-like-2* constructed using a 1-to-1 thread based on *H. sapiens* cathepsin L peptidase *CTSL1*. C - Cysteine, histidine and asparagine forming the catalytic triad of *Gr-cpl-like-2*. Signal peptides are highlighted in blue, N-glycosylation sites are highlighted in yellow, catalytic triad residues cysteine, histidine, and asparagine highlighted in green. * - Predicted primary residue of mature protein after pro-domain removal

5.3.2 – Phylogenetic and gene expression analysis

Newly identified cathepsin L-like peptidases were aligned against previously characterised Hgcp-I and Hgcp-II from *H. glycines*, cpl-1 from *C. elegans*, and CTSL1 from *H. sapiens*. A phylogenetic tree was constructed using a Bayesian inference model. The tree split into five distinct clades, each containing one of the cathepsin L-like peptidases from *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *H. glycines* (Figure 5.4).

Gene expression data from *G. rostochiensis*, *G. pallida*, and *R. reniformis* was analysed to determine at which life stages this cathepsin L-like family of peptidases were expressed across the five clades (Figure 5.4). As Gr-cpl-like-2 was initially identified through the effector identification pipeline and shown to be expressed during parasitic life stages, it was expected that the orthologs from other nematode species would also share this expression pattern. In Figure 5.4, gene expression data collected from nematodes in the J2 stage of their lifecycle were classed as pre-parasitic while data collected from nematodes at 14 dpi were classed as post-parasitic (Eves-Van den Akker *et al.* 2016a). It is evident that although these proteins are conserved, they are deployed at different times during the life cycle of the nematode which may indicate differences in function. This is strengthened by the absence of a signal peptide in all of the proteins in clades 4 and 5. It also shows that differing gene expression between clades are conserved across different species. The cathepsin L-like peptidase members of clades 1 and 4 are expressed at similar levels during both pre- and post-parasitic life stages of the PPN (Figure 5.4). Interestingly, the genes in clade 2 are highly expressed at the post-parasitic life stage with little to no expression in the pre-parasitic stage. This implies that Gr-cpl-like-2, Gp-cpl-like-2, Rr-cpl-like-2, and Hgcp-II have a specific role after the nematode has successfully entered the host plant and established a feeding site. In contrast to this, the genes in clade 3 are more highly expressed during the pre-parasitic stage inferring their importance while the nematode is moving towards and initiating infection of a host plant. Genes in clade 5 show a similar expression profile to those

in clade 2, although the upregulation at parasitic stages is not as strong. In addition, although Gr-cpl-like-5 and Gp-cpl-like-5, both from PCN, increase in expression during post-parasitic stages, Rr-cpl-like-5 is more evenly expressed across pre- and post-parasitic stages. An alignment to show sequence conservation of all genes used in the phylogenetic tree can be found in Figure 5.5.

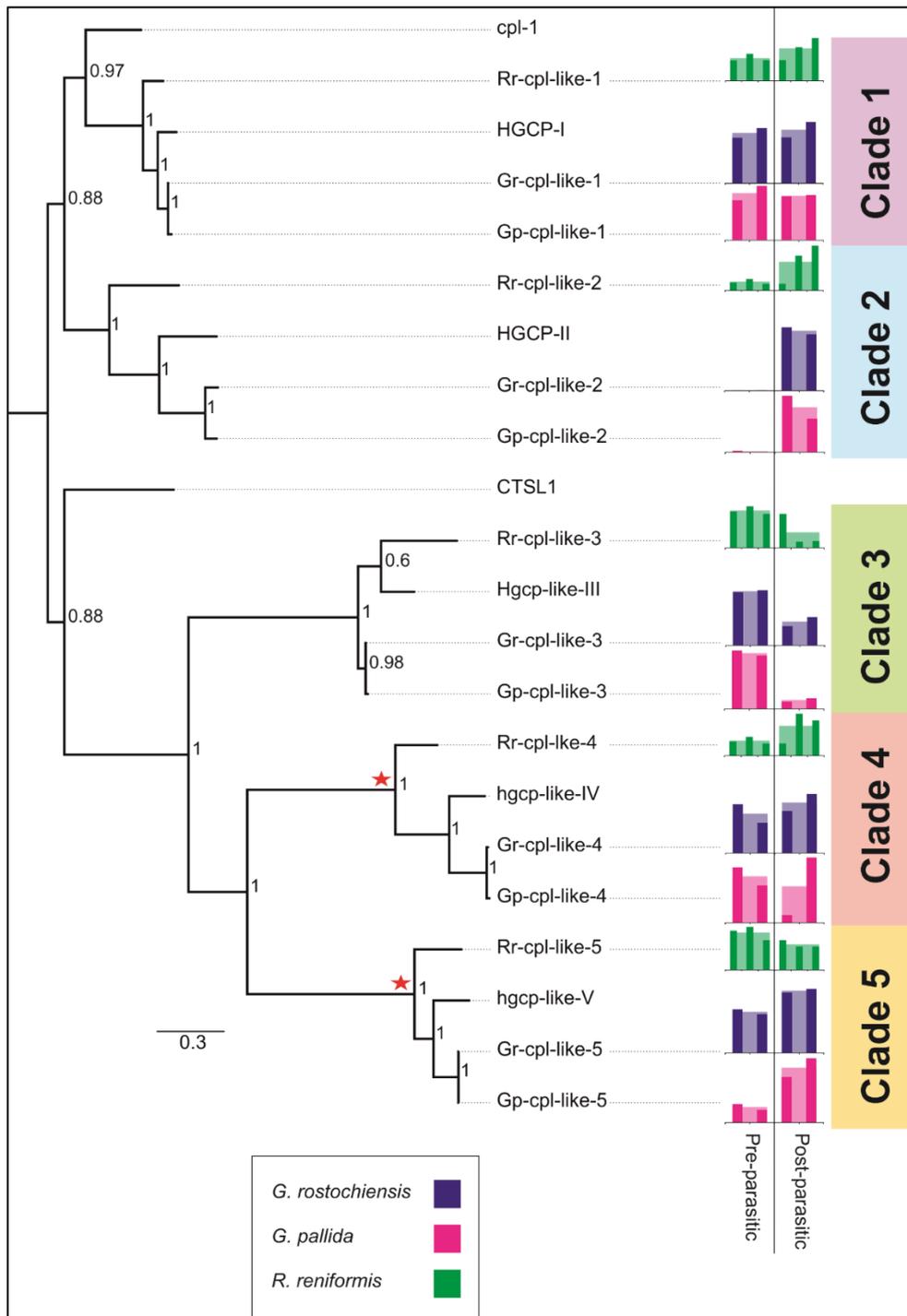


Figure 5.4 - Phylogenetic tree of cathepsin L and L-like peptidase members from *G. rostochiensis* (*Gr-cpl-like-1* to *Gr-cpl-like-5*), *G. pallida* (*Gp-cpl-like-1* to *Gp-cpl-like-5*), *R. reniformis* (*Rr-cpl-like-1* to *Rr-cpl-like-5*), *H. glycines* (*Hgcp-I*, *Hgcp-II*, *Hgcp-like-III* to *Hgcp-like-V*), *C. elegans* (*cpl-1*) and *H. sapiens* (*CTSL1*). Gene expression (TMM normalised expression) is split into pre- and post-parasitic life stages, individual replicates are in bold with the repeat average expression as a transparent overlay. Blue – *G. rostochiensis*, Pink – *G. pallida* and Green – *R. reniformis*. Branch length (0.3) indicates substitutions per sequence site. Branches containing genes which do not have signal peptides are indicated with a red star.



Figure 5.5 – Amino acid alignment of Cathepsin L and L-like peptidase proteins of PPN. All 5 genes of clades 1-5 from *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *H. glycines* are aligned against *Cpl-1* from *C. elegans* and *CTSL1* from *H. sapiens*.

5.3.3 – In situ hybridisation

In situ hybridisation (ISH) was carried out to identify the location of gene expression of the five cathepsin L-like genes of *G. rostochiensis*. Gr-cpl-like-1, Gr-cpl-like-3, and Gr-cpl-like-5 display intestinal staining in *G. rostochiensis* J2s (Figure 5.6 B, E, G). Gr-cpl-like-1 and Gr-cpl-like-3 also frequently displayed expression in the genital primordia. Gr-cpl-like-2 and Gr-cpl-like-4 showed strong staining in the intestine of mixed parasitic life stages (Figure 5.6 D, F). *In situ* hybridisation was also carried out on Gp-cpl-like-1 (*G. pallida*) to assess whether the intestinal staining pattern observed in *G. rostochiensis* was likely true for cathepsin L-like peptidases in other species analysed. Gp-cpl-like-1 showed staining in the intestine of *G. pallida* in both J2s (Figure 5.6 C) and mixed parasitic stages (*data not shown, personal communication with C. J. Lilley*). These results confirm that the cathepsin L-like peptidases from *G. rostochiensis* and *G. pallida* are not effectors but are more likely to play a role in either embryogenesis, digestion, or nutritional uptake.



Figure 5.6 - mRNA localisation of *Gr-cpl-like-1* to 5, and *Gp-cpl-like-1* by in situ hybridisation. A – A schematic of *Globodera* nematode anatomy highlighting the location of the intestines (purple) and the genital primordia. B – *Gr-cpl-like-1*, C - *Gp-cpl-like-1*, D - *Gr-cpl-like-2*, E - *Gr-cpl-like-3*, F - *Gr-cpl-like-4*, G - *Gr-cpl-like-5*, H – Pre-parasitic J2 negative control, I – parasitic negative control. Scale bars - 50 μ m

5.3.4 – RNAi silencing of Gr-cpl-like-2

Due to the exclusive post-parasitic expression of Gr-cpl-like-2, transgenic plant lines expressing an RNAi hairpin were created to ascertain whether host induced gene silencing of Gr-cpl-like-2 would influence parasitism. Semi-quantitative reverse transcription (SQRT) PCR was used to determine the expression levels of the RNA fragment in the RNAi lines. One high expressing line was chosen (Line 14) alongside two mid-expressing (Lines 12, 23) and one low expressing (Line 20) line (Figure 5.7 A). This was done to determine whether any phenotype observed upon silencing of Gr-cpl-like-2 was influenced by the level of expression of the silencing RNA. It was hypothesised that silencing of Gr-cpl-like-2 would see less females successfully establishing in the RNAi lines. The mean count of *G. rostochiensis* females present on four transgenic lines (Lines 12, 14, 20 and 23) were compared against a transgenic GFP control line. However, there was no statistically significant differences between the number of females counted on the Gr-cpl-like-2 transgenic lines and the GFP control line meaning that there was no significant change to parasitism observed under these parameters (Figure 5.7 C).

Twenty cysts per cell line were randomly selected and assessed for a change in cyst size (area mm²) between the Gr-cpl-like-2 transgenic lines and the GFP control line. This was done to ascertain if there was a difference in female reproductive ability with a smaller cyst area assuming correlation with a smaller internal egg count. There was no significant difference in area between the Gr-cpl-like-2 transgenic lines L12, 14, L20, and the GFP control line (Figure 5.7 D). There was a significant difference (P – 0.001) in average cyst size between L23 and the GFP control line with the cysts from the Gr-cpl-like-2 line being 0.00836 mm² smaller than the cysts grown on the GFP lines. Cysts from L23 were also significantly different (P – 0.029) from those measured from L14 with L23 cysts being 0.006247 mm² smaller than the average cyst from L14. It should be noted that there was no significant difference in average cyst size between L14 and the GFP control line. Significance testing data can be found in sup. File 7.

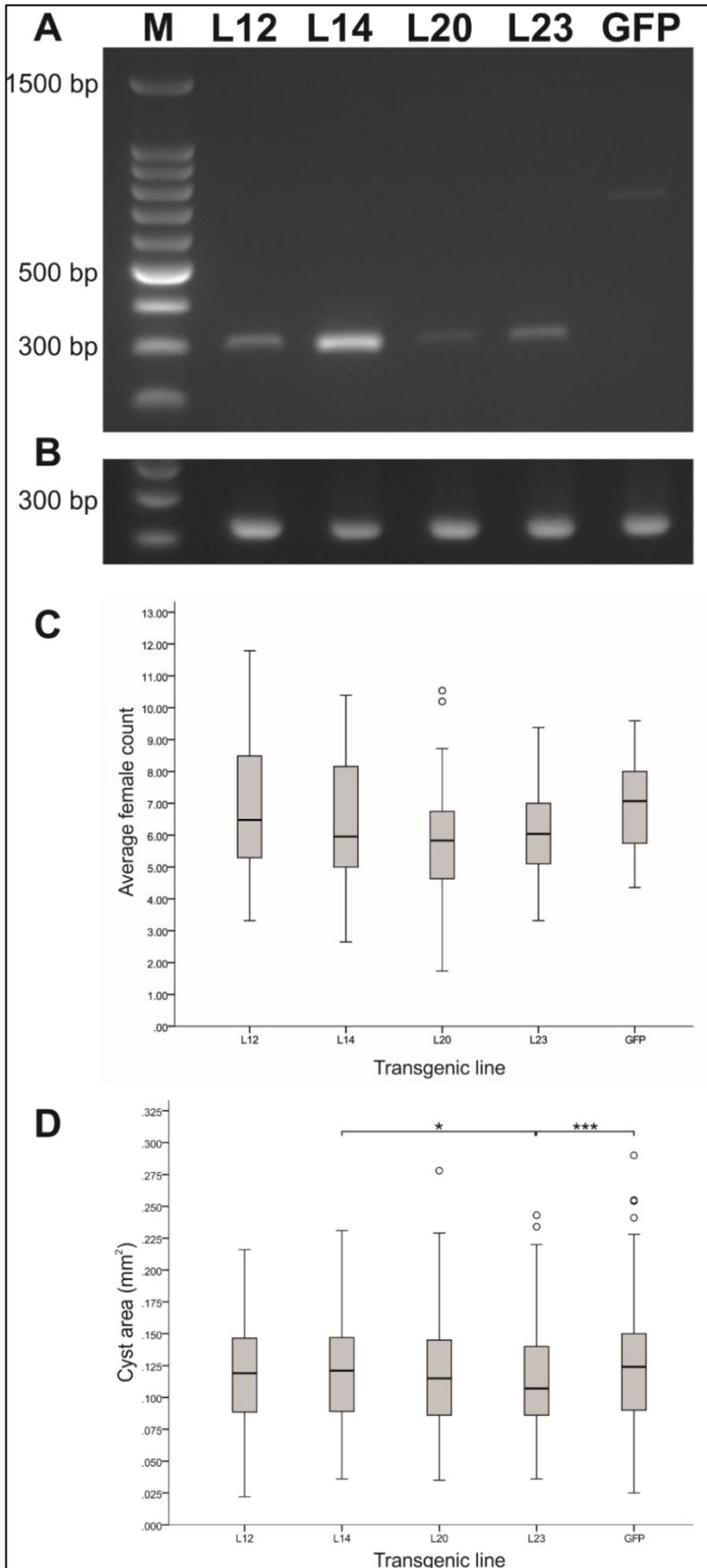


Figure 5.7 – RNA interference of *Gr-cpl-like-2*. A – SQRTPCR indicating the differing levels of construct (*Gr-cpl-like-2* transgenic line or GFP control) expression in transgenic lines after 28 cycles. B – *Elongation factor 1α* (*stEF1α*) control PCR to confirm equal loading of cDNA. C – The average number of females counted on the roots of each transgenic line. No significant difference was observed between the average cyst count on *Gr-cpl-like-2* lines compared with the GFP control line. D – Average area of cysts on transgenic lines. * - $P \leq 0.05$, ** - $P \leq 0.01$, *** - $P \leq 0.001$. M - ladder

5.4 – Discussion

5.4.1 – Identification and phylogenetic analysis of nematode Cathepsin L-like peptidases

Gr-cpl-like-2 from *G. rostochiensis* was initially identified as an effector candidate via the pipeline discussed in chapter 3. Gr-cpl-like-2 was originally hypothesised to play a role in parasitism due to exclusive expression during parasitic life stages, the presence of a signal peptide, and lack of any transmembrane helices. Initial BLAST sequence similarity searches showed Gr-cpl-like-2 was very similar to Hgcp-II from *H. glycines* (Urwin *et al.* 1997). Subsequent BLAST searches lead to the identification of 5 cathepsin L-like genes in each of the following species: *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *H. glycines*. Gp-cpl-like-1, previously known as Gp-cpl-1/AY999065.1, had been previously identified by Lilley *et al.*, (unpublished manuscript). This resulted in the identification of small cathepsin L-like families of peptidases present in multiple PPN species. Each family found across the four species contained five protein members. When a phylogenetic tree was constructed, these separated into five distinct clades, each containing one peptidases from each species (Figure 5.4). For example, Gr-cpl-like-1, Gp-cpl-like-1, Rr-cpl-like-1, and Hgcp-I forming clade 1 etc.

Analysis of the cathepsin L-like peptidases from *G. rostochiensis*, showed that Gr-cpl-like-1, -2, and -3 all contain signal peptides while Gr-cpl-like-4 and -5 do not. This pattern is also conserved across all four nematode species analysed. All proteins in clades 1-3 all have signal peptides which suggests that these peptidases are secreted. Conversely, peptidases found in Clades 4 and 5 across *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *H. glycines* are all predicted to contain a single TMH within the first 100 amino acids of the protein but have no predicted signal peptide. This could suggest differences in function between cathepsin L-like family members. This is supported by gene expression differences across the five clades. Members of clade 2 are highly expressed in post-parasitic stages indicating potential functions in parasitism or reproduction. This contrasts with clade 3 that has high pre-parasitic

expression which suggests a role at an earlier life cycle stage, possibly while the nematode is locating the host plant.

5.4.2 – *In situ* hybridisation

In situ hybridisation has shown that the PPN cathepsin L-like peptidases analysed are primarily expressed in the intestines across different life stages. One possible function is to break down host plant tissue in the gut as the nematode feeds on the host plant, giving them an overall function in digestion and nutritional uptake. As it stands protein degradation for nutritional uptake is not a well-studied area and requires further investigation (Malagon *et al.* 2013). As previously reported in section 5.1.2, cathepsin L peptidases are essential for embryonic development. It was observed that Gr-cpl-like-1 and Gr-cpl-like-3 had expression in the genital primordia. It would follow that Gr-cpl-like-1 and Gr-cpl-like-3 being expressed in the genital primordia as well as the intestines may indicate a reproductive-based role rather than that of digestion and nutritional uptake. The genital primordia is highly transcriptionally active in juvenile nematodes, however, so it cannot be ruled out that this expression may be a false positive. The ISH analysis now means it is unlikely that any of these cathepsin L-like peptidases have roles in parasitism as first hypothesised, but they do have roles that are active alongside those effectors involved in the successful parasitism of the host.

5.4.3 – RNAi silencing of Gr-cpl-like-2

RNAi assays were conducted to ascertain the effect of Gr-cpl-like-2 silencing on parasitism. No significant reduction in the mean count of females was observed in RNAi transgenic lines silencing Gr-cpl-like-2 when compared to those on a GFP transgenic control line. *G. rostochiensis* juveniles were able to successfully establish on the host plant meaning that the silencing of Gr-cpl-like-2 did not affect parasitism or nutritional uptake. As Gr-cpl-like-2 exists as part of a small family of cathepsin L-like peptidases, it is possible that there is an element of functional redundancy or compensation occurring to avoid loss of function. This result has been seen before

in other cathepsin studies. The cathepsin B peptidase SmCB1 (a.k.a SM31) was successfully downregulated by RNAi however there was no observable changes to the survival of the parasitic blood fluke *Schistosoma mansoni* (Skelly *et al.* 2003). It was also proposed here that functional redundancy may play a role due to the discovery of a second cathepsin B peptidase SmCB2 (Caffrey *et al.* 2002).

Out of the four Gr-cpl-like-2 transgenic lines only line 23 (L23) displayed a reduction in cyst size (area mm²) when compared to the GFP control. Out of these four lines it was hypothesised that if any L14 would have a significant cyst size change compared to GFP it would be the line with the highest expression of the Gr-cpl-like-2 RNAi hairpin. L23 was a mid-expressing line so it is currently unclear why this is the only line with a significant cyst size difference. Overall, these results mean that silencing of Gr-cpl-like-2 does not have an observable effect on the female nematodes ability to produce eggs. As discussed in section 5.1.2, inhibition of cpl-1 from *C. elegans* is seen to upregulate serotonin regulated fat loss (Lin *et al.* 2019). The assays in this thesis were not set up to assess changes like this for the RNAi silencing of Gr-cpl-like-2 but it is possible that more complex phenotypes did occur but were not observed or measured.

5.5 – Future work

This chapter has successfully shown the identification and analysis of 5 clades of cathepsin L-like peptidases from the plant-parasitic nematode species *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *H. glycines*. Cloning of the genes and further characterisation of Gr-cpl-like-1 to -5 and Gp-cpl-like-1 were carried out in order to ascertain the location of gene expression in the nematode by *in situ* hybridisation. In order to determine if the same gene expression patterns are followed across the PPN species analysed, further cloning and ISH could be carried out on the cathepsin L-like peptidases identified in *G. pallida*, *R. reniformis*, and *H. glycines*. These may yield interesting results as they are more phylogenetically

divergent from the Globodera species. In addition to this, gene expression profiles for *H. glycines* were not available at the time of analysis so their addition to Figure 5.4 in future would be beneficial to see if they follow the same patterns as the other PPN species in each clade. A multigene approach to RNAi may be taken in the future to assess if functional redundancy and/or compensation of other peptidases were masking RNAi results that silencing of Gr-cpl-like-2 may have displayed. The production of transgenic lines containing multiple RNAi hairpin constructs (the *G. rostochiensis* cathepsin L-like peptidase from each of the five clades for example) may shed more light on the function of these proteins. This may raise the unique issue of multiple different or compounded observable effects if the hypothesis about cathepsin L-like peptidases from different clades having different primary functions is correct.

6. Functional

characterisation of core

effector genes

6.1 – Introduction

Many of the candidate core effectors identified as a result of the analysis described in chapter 3 are predicted or hypothetical proteins. This means there is little to nothing known about their function, their role as effectors, or their host plant targets. Identifying functions of new effectors is of great importance as they may provide future targets for control methods. Effectors also manipulate the plant immune system or, in the case of syncytium-forming nematodes, may target key plant developmental processes. Study of effector functions can therefore help in understanding fundamental questions in plant biology. Understanding the function of novel effector proteins is challenging however, as hypotheses about their potential function are difficult to generate on the basis of the sequence alone. Information about the potential function of such genes can be obtained by identifying the subcellular localisation of the proteins in host plant cells and by identifying host targets of the proteins.

6.1.1 – Subcellular localisation

The process of identifying the subcellular localisation of effectors usually involves using confocal microscopy to visualise fusion proteins, combining the effector protein of interest with a fluorescent tag such as green fluorescent protein (GFP). These fusion proteins can be transiently expressed in living tissues such as plant leaf

material to view their end location, or to track the movement patterns of the protein within or between cells (Runions *et al.* 2007). Analysis of subcellular localisation is a well refined process and has been used to study targeting of effectors from nematode, bacterial, fungal, and oomycete plant pathogens.

A large study was undertaken to identify the subcellular localisation of 52 RXLR effectors from the oomycete pathogen *P. infestans* when transiently expressed in *N. benthamiana* leaf tissue (Wang *et al.* 2019). Out of these 52, 41% were identified as being nucleo-cytoplasmic, 25% were nuclear/nucleolar and 18% were localised to the plasma membrane. The other 16% were either cytoplasmic, cytoskeletal, or localised to other membranes within the cell. Similar, but smaller scale, analyses of effector localisation have also been described for PCN (Jones *et al.* 2009, Thorpe *et al.* 2014). These studies showed that PCN effectors also localise to a diverse range of host structures including the cytoplasm, nucleus/nucleolus, and various other subnuclear bodies, the peroxisome, and peroxisome membranes.

There are several examples of how an understanding of subcellular localisation can be important for protein characterisation. Effector proteins which are targeted to the nucleus of a plant cell are likely to have very different functions from those which target other organelles; therefore, the target location of an effector can give clues as to their potential function. For example, the effector ChEC₂₁ from the hemibiotrophic fungus *Colletotrichum higginsianum* has been shown to localise at the stacks of the Golgi apparatus inside host cells. The Golgi apparatus is an organelle in the cell which is responsible for packaging newly transcribed proteins into vesicles which are then transported to their destination, be that intra- or extracellular. Targeting of the Golgi by effector proteins suggests a function in altering protein transport (Robin *et al.* 2018). Identification of the unusual target location of ChEC₂₁ could lead to streamlining of further investigations into function such as identification of specific host Golgi proteins which may be the effector target.

Effectors have also been shown to target host cell cytoskeletal components such as actin. Actin is necessary for many cellular processes such as cell division and expansion, movement of vesicles and organelles, cell structure and immunity. The diverse and important roles of actin for cell function make it a prime target for pathogens (Porter & Day 2016). One example of this is the HopW1 effector from the bacterial pathogen *Pseudomonas syringae*. HopW1 localises to the actin cytoskeleton and was shown to disrupt this by reducing the density of actin filaments and inhibition of the actin-dependent process of endocytosis (Kang *et al.* 2014).

Analysing the co-localisation of effectors and their host targets can also help inform functional studies. For example, the 30D08 effector (see section 3.1.1.3) has been shown to promote infection and syncytium formation through interaction with a component of the host spliceosome machinery (atSMU2) to alter gene expression levels. 30D08 has been confirmed to localise to the nucleus when inside the host. A nuclear localisation signal (NLS) was identified in the amino acid sequence of this effector which caused a change in localisation from nuclear to cytoplasmic upon site-directed mutagenesis (Verma *et al.* 2018). As gene expression and splicing activity occur within the nucleus, this localisation compliments the previous results known about this effector and its interactions. Co-localisation of the *H. schachtii* 19C07 effector and its host target, an auxin transporter LAX3, at the plasma membrane of host cells has also been demonstrated, providing further evidence in support of the interaction between these two proteins (Lee *et al.* 2011).

6.1.2 – Protein-protein interaction identification methods

Perhaps the best way to understand the likely function of an uncharacterised effector protein is to identify what host proteins they interact with. Yeast two-hybrid analysis is the most commonly used method for identifying host targets of effectors, with co-immunoprecipitation (Co-IP) and/or bimolecular fluorescence complementation (BiFC) used to confirm that the interactions identified in yeast can occur *in planta*.

6.1.2.1 – Yeast two-hybrid

Yeast two-hybrid (Y2H) screening is a genetic method of identifying interactions between proteins. Yeast two-hybrid works by the reconstitution of a functional transcription factor (TF) split into fragments. The fragments are brought together if there is an interaction present between an effector and a host target protein which the fragments are attached to. The TF is commonly split into a binding-domain fragment and an activation domain fragment. In Y2H studies a protein of interest, termed the “bait” protein is tagged with one fragment of the TF. This fusion will be introduced to a library containing many different proteins which potentially interact with the bait. These proteins in the library are all tagged with the other TF fragment and are referred to as “prey” proteins (Figure 6.1 A). If there is an interaction between a bait and prey protein, the transcription factor fragments will reconstitute, allowing them to bind to a promoter and initiate the expression of a reporter gene (Figure 6.1 B). Commonly used reporter genes include those which allow growth on selected media. Many commercial Y2H kits use amino acid production as reporter genes, meaning only those colonies with a positive bait-prey interaction will express the gene to produce a specific amino acid. Production of the amino acid will allow the colony to grow on media excluding that specific amino acid. Examples of amino acids frequently used in this way are histidine and uracil. Another commonly used reporter is the LacZ gene. A positive bait-prey interaction causes the LacZ reporter gene to be expressed, resulting in these colonies turning blue (compared to the white/cream colour of negative samples) when a β -galactosidase assay is carried out. Many Y2H systems will combine multiple reporter genes so growth on selection plates plus a positive additional indicator e.g., a colour change, would give a strong indication of interaction. The gene encoding interacting prey protein from the library can subsequently be recovered and sequenced.

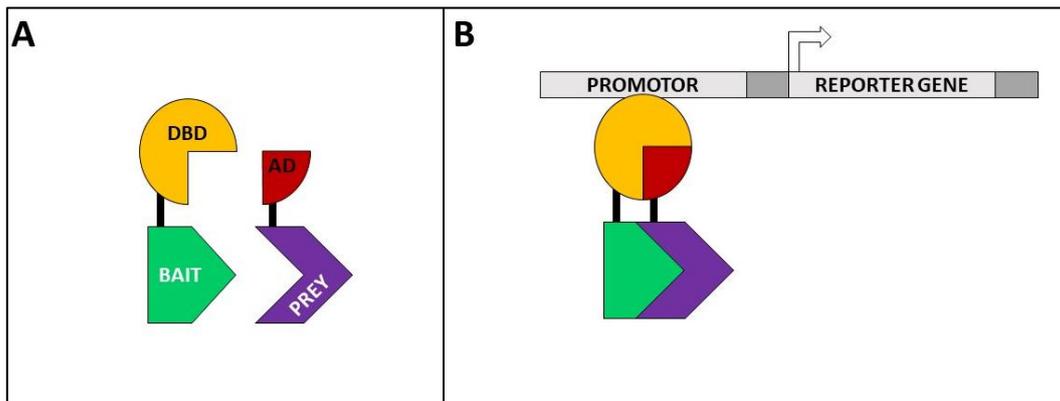


Figure 6.1 – Schematic of yeast two-hybrid reporter gene activation by interacting proteins. A -The bait protein (protein of interest) is fused to the DNA binding domain (DBD) of the transcription factor (TF), while the prey protein from the library is fused to the activation domain (AD) of the TF. B - If there is an interaction between the bait and a prey protein, the proximity causes the TF fragments to reconstitute. The TF can then bind the promoter sequence which causes the expression of the reporter gene.

Yeast two-hybrid is now a common technique for identifying potential targets of effector proteins. Co-immunoprecipitation or BiFC can subsequently be used to validate an interaction discovered from Y2H. A good example of this is the work done on the 30C02 effector from *H. glycines*. As a novel effector it was important to determine the interacting host target proteins to ascertain the role of 30C02 in parasitism. Through Y2H, Hamamouch *et al.* were able to identify a β -1,3-endoglucanase (AT4G16260) as an interacting host protein. This interaction was confirmed using BiFC, where 30C02 and AT4G16260 interacting caused the reconstitution of the two YFP fragments in the cytoplasm of bombarded onion epidermal cells (Hamamouch *et al.* 2012). AT4G16260 is involved in cell expansion and callose formation, so from the interaction studies it is hypothesised that it is targeted by 30C02 to suppress host defences. In keeping with this, overexpression of AT4G16260 also caused a reduction in susceptibility to nematode parasitism. It is important to note that although Y2H has proven to be a successful technique, there are also a number of disadvantages and limitations to its use (Brückner *et al.* 2009). The first being a spatial/locality issue as the reconstitution between the transcription factors must occur in the nucleus. This means that many protein interactions that occur with proteins excluded from the nucleus (e.g., interactions with membrane

bound proteins) are not possible to detect using Y2H. The yeast model itself can also cause certain interactions to not be detected. In many examples yeast is not the native system that these protein interactions are usually found in. This can lead to protein folding issues and steric hinderance that are not seen in the native host and prevent an interaction being detected and producing false negative results. False negatives can also arise from an interaction requiring post-translational modifications which are not present or appropriately functioning in the yeast model. This also applies to those interactions which require environmental factors such as oxidative conditions. False positives can also occur due to non-specific interactions.

6.1.2.2 – Co-immunoprecipitation

Co-immunoprecipitation (Co-IP) uses western blotting to identify interactions between proteins of interest tagged with various antigens, such as GFP, RFP, or epitope tags such as FLAG, HIS, or MYC. During Co-IP assays the effector protein is tagged with one epitope and the putative host interactor protein tagged with a second different epitope. The two fusion proteins are then co-expressed in plant cells. Total protein content is extracted from the cells (referred to as the input sample) and the protein of interest is purified from this mixture using an antibody (raised against the appropriate tag used) bound to agarose or magnetic beads (referred to as the output sample). If the putative interactor is bound to this protein, it will also be purified and can subsequently be detected by western blot using the antibody against the tag fused to the interactor.

6.1.2.3 – Bimolecular fluorescence complementation

Bimolecular fluorescence complementation (BiFC) takes advantage of the fact that yellow fluorescent protein (YFP) can be split into N- and C-terminal portions, neither of which will generate fluorescence alone. When brought into proximity the function of the protein is reconstituted and a yellow fluorescence signal can be detected. In practical terms, one fluorescent protein fragment is fused to the effector protein while the other half is fused to the interacting protein. The two fusion proteins are

then transiently expressed in plant cells and if the effector and putative host target interact, the fragments will reconstitute and produce fluorescence that can then be measured and visualised by microscopy (Kodama & Hu 2012). As both partners in the interaction need to be known in advance, this method is used to validate and confirm an interaction which has already been observed or at least hypothesised due to previous work. Bimolecular fluorescence complementation is not appropriate when “fishing” for unknown interacting proteins.

6.1.3 – Chapter aims

This chapter builds on work from chapter three, taking three uncharacterised effectors GROS_g05682 (20E03), GROS_g02469 (G23G11) and GROS_g02394 (GLAND11) and determining their function through subcellular localisation and protein-protein interaction studies (yeast two-hybrid and Co-IP).

6.2 – Materials and methods

6.2.1 – Cloning

GROS_g05682, GROS_g02469, and GROS_g02394 were transferred from the pDONR201 entry vector into the destination vectors pK7FWG2 (eGFP c-terminal fusion), pK7WGF2 (eGFP n-terminal fusion), and pGRAB_mturq2_GW (mTurquoise2 fusion) using the gateway LR (AttL-AttR) reaction (Invitrogen). Half volume LR reactions were performed following the manufacturer's protocol (Table 6.1). Reactions were incubated at room temperature overnight. The LR reaction was halted by adding 1 μ l of proteinase K before incubating at 37 °C for 10 minutes.

Table 6.1 – Ligation reaction set up

Reagent	Volume (μl)
pDONR201 entry clone	X 50-150ng
pK7FWG2 or pK7WGF2 or pGRAB_mturq2_GW destination vector	0.5 (150ng/ μ l)
LR clonase II enzyme mix	1
EB buffer	Up to 4 μ l

Plasmids were transformed into electrocompetent *E. coli* strain DH5 α cells onto LB agar plates with appropriate antibiotics (Section 2.6, Table 2.5). Colonies obtained were tested for successful insertion by PCR using the primer sets in Table 6.2.

Table 6.2 – Primer sets for GROS q05682, GROS q02469, GROS q02394, and stPRMT1.1 cloning

Gene name	Vector cloned into	Primer name	Sequence
GROS_g05682	pK7WGF2	Cterm_GFP_FOR (Vector specific)	GTGCTCAGGTAGT GGTTGT
		g05682_FOR (Gene specific)	GTTTTAACGAATG ATGGTCCGA
	pK7FWG2	Nterm_GFP_REV (Vector specific)	CGGACACGCTGAA CTTG
		g05682_REV (Gene specific)	GGCCCTATTGCGC TTGA
GROS_g02469	pK7WGF2 and pGRAB_mturq_GW	Cterm_GFP_FOR (Vector specific)	GTGCTCAGGTAGT GGTTGT
		g02469_FOR (Gene specific)	ACCATGGACGCCG GTGGAATGGAT
	pK7FWG2	Nterm_GFP_REV (Vector specific)	CGGACACGCTGAA CTTG
		g02469_REV (Gene specific)	GAAAGCTTGACGG AAAGGCTTAA
GROS_g02394	pK7WGF2	Cterm_GFP_FOR (Vector specific)	GTGCTCAGGTAGT GGTTGT
		g02394_FOR (Gene specific)	ATGGCCAAAGCGT TCAGCAGC
	pK7FWG2	Nterm_GFP_REV (Vector specific)	CGGACACGCTGAA CTTG
		g02394_REV (Gene specific)	TTCGCTGGAAGAG CATTGA

stPRMT1.1	Initial sequencing in pDEST22 vector	pDEST22_preAD_F	TATAACGCGTTTG GAATCACT
		pDEST_R	GTCTCCAATCAAG GTTGTCGGCT
	pCR8_GW_TOPO	M13_FOR	ACTGGCCGTCGTT TTAC
		M13_REV	CAGGAAACAGCTA TGAC
		stPRMT1.1_F	ATGGATTCTGTAA GCAATAATGAGAT TGA
		stPRMT1.1_R	TCATCTCATCCGAT AATACTGAGTTCT
	pGRAB_mturq_GW	Cterm_GFP_FOR (Vector specific)	GTGCTCAGGTAGT GGTTGT
		stPRMT1.1_F	ATGGATTCTGTAA GCAATAATGAGAT TGA
	pK7WGF2 (RFP)	RTL2-M	CAACACATGAGCG AAACCCTATAAGA A
		RFPEndSeq	CTTGATGTCGGTCT TGTAGGC

Positive colonies were grown up in overnight liquid cultures (5 ml LB, with appropriate antibiotics (Section 2.6, Table 2.5)) at 37 °C, with shaking at 200-250 rpm. Plasmid DNA was extracted from cultures using the QIAprep Spin miniprep kit (Qiagen). Plasmids were analysed using Sanger sequencing (The James Hutton Institute). Glycerol stocks were produced from overnight cultures by mixing an equal volume of culture and 60 % v/v glycerol, and was subsequently stored at -80 °C.

When stPRMT1.1 was identified from yeast two-hybrid screens it was initially sequenced using pDEST22_preYAD_F and pDEST_R primers (Table 6.2). Plasmid (pDEST22 containing stPRMT1.1) was miniprep (Section 2.6.3) and transformed into electrocompetent DH5 α strain *E. coli* cells before cloning into the pCR8_GW_TOPO vector (Section 2.5). Success of cloning was confirmed by sequencing with M13 primers and gene specific stPRMT1.1 forward and reverse primers (Table 6.2). stPRMT1.1 was subsequently cloned into the mturquoise2 vector pGRAB_mturq_GW and the RFP vector pk7WGR2 using the LR reaction as described above.

6.2.2 – Transformation into *Agrobacterium tumefaciens*

After the presence of the appropriate gene was confirmed, plasmids were transformed into competent *A. tumefaciens* strains AGL1 (+ pSOUP helper plasmid) or GV3101 by electroporation. Two microlitres of the plasmid was added to 50 μ l of cells in a prechilled cuvette and incubated on ice for 20 minutes. Cells were electroporated using a Micropulser (BioRad) and transferred to a 2 ml Eppendorf, with 1 ml of SOC media and incubated shaking at 28 °C for 2 hours. Transformed cells were plated on LB agar containing rifampicin (50 μ g/ml), gentamicin (25 μ g/ml) and kanamycin (50 μ g/ml) for GV3101 cells or LB agar containing tetracycline (5 μ g/ml) and rifampicin (50 μ g/ml) for AGL1 cells. Plates were incubated at 28 °C for 48 hours. Colonies were tested for successful plasmid transformation by colony PCR using specific gene primers as described in section 2.6.3.

6.2.3 – Agroinfiltration

Five millilitres cultures of LB were inoculated with a single colony and grown overnight at 28 °C. Cultures were pelleted (3500 x g, 10 min) and washed twice in infiltration buffer (IB) (10 mM MES, 1 mM acetosyringone, 10 mM MgCl₂, SDW). The initial OD₆₀₀ was measured before cultures were incubated at room temperature on a shaker for 2-3 hours. Cultures were adjusted to an OD₆₀₀ between 0.02 and 0.2 in fresh IB. A small piercing wound was made to the abaxial surface of the leaf using a

sterile needle. Infiltrations were carried out using a 1 ml syringe (without needle attached) at the previous piercing sites. Infiltrations were carried out on wild type and transgenic *N. benthamiana* lines, actin-RFP (Lifeact-TagRFP – CB174) and nuclei-RFP (mRFP fused to the *N. benthamiana* histone 2B (mRFP-H2B) CB157). *Nicotiana benthamiana* were viewed 24-72 hours post-infiltration.

6.2.4 – Confocal microscopy

All imaging was conducted on a Zeiss 710 confocal laser scanning microscope and viewed/captured using Zen software (ZEN Digital Imaging for Light Microscopy, Zeiss). GFP and RFP tagged samples were sequentially imaged as Z-stacks (optimal sectioning setting). GFP and chlorophyll were excited at 488 nm with emission at 500-533 nm for GFP and 650-690 nm for chlorophyll. RFP was excited using a 561 nm laser and emitted at 592-631 nm. mTurquoise2 was excited at 440 nm with emission at 445-485 nm. Z-stack images were processed as maximum intensity projections.

6.2.5 – The hypersensitive response in *Nicotiana benthamiana*

Overnight cultures of GROS_g05682 and GROS_g02469 in pK7WGF2, GROS_g02394 in pK7FWG2, and eGFP in the pK7WG2 vector in the GV3101 strain of *Agrobacterium* were prepared for infiltrations as per section 6.2.3. All cultures were adjusted to a final OD₆₀₀ of 0.2. Four *N. benthamiana* plants were infiltrated per construct with three leaves per plant selected for infiltration. Each leaf was infiltrated four times (12 infiltration sites per plant, 48 total per construct). Each leaf was monitored and photographed at the same time each afternoon for 7 days post infection (dpi). Individual inoculation sites were determined to be positive for an HR response if over 50 % of the infiltrated area displayed a programmed cell death (PCD) lesion by 7 dpi.

For statistical analysis a positive HR response of 50% or over was given a value of 1. A negative HR response of less than 50% was given a value of 0. A one-way ANOVA with post hoc Tukey test was performed to determine significance ($P < 0.05$) using

SPSS (IBM Corp 2017). Error bars for resulting mean percentage graphs were produced using standard error.

6.2.6 – UV imaging

Images of 7 dpi *N. benthamiana* leaves from the HR response assay (Section 6.2.5) were taken under UV light to better visualise the cell death response. Images were taken using a Canon EOS 70D camera with a EFS 60 mm lens. The camera was focused under white light before switching to UV light for image capture. The aperture was set to F16, international organisation for standardisation (ISO) set to 600, and exposure time set to 5 seconds for all images. UV light was provided by 2x UVP Black-ray 365 nm spotlights.

6.2.7 – Yeast two-hybrid experiments

Yeast two-hybrid experiments were carried out using the ProQuest Two-hybrid system (ThermoFisher).

6.2.7.1 – Yeast cultures

To begin the process of yeast two-hybrid screening, two cultures were set up. 10 ml YPA (40% glucose, 0.2% adenine hemisulphate, in YP media (6 g yeast extract, 12 g bacto-peptone, 600 ml SDW)) was inoculated with the *Saccharomyces cerevisiae* yeast strain MAV203. A second flask of 40 ml YPA was also set up as a control. These cultures were incubated at 28 °C shaking overnight at an angle. The control YPA media should remain clear after incubation to indicate sterility of yeast culture inoculation.

6.2.7.2 – Y2H small scale transformation testing

Forty millilitres of YPA was inoculated with the 10 ml of MAV203 culture and the OD₆₀₀ was adjusted to 0.4. This culture was split into two 20 ml cultures and grown for approximately 3 hours at 28 °C shaking at an angle. Bait and prey plasmids (100 ng

each) were aliquoted into PCR tubes (Table 6.3). Two microlitres of sheared Herring sperm DNA (5 mg/ml) was added to each tube.

Table 6.3 - Y2H plasmid combinations for screening

Tube no.	1st plasmid added	2nd plasmid added
1	pDEST32_g02394	-
2	pDEST32_g02469	-
3	pDEST32_g05682	-
4	pDEST32_g02394	pDEST22 Empty Vector
5	pDEST32_g02469	pDEST22 Empty Vector
6	pDEST32_g05682	pDEST22 Empty Vector
7	Negative control (sperm DNA only)	pDEST22 Empty Vector

Yeast cultures were centrifuged at 2500 rpm for 5 minutes at room temperature. The supernatant was removed and pellets were washed in 40 ml SDW followed by brief agitation using a vortex. Yeast cultures were centrifuged at 2500 rpm for 5 minutes at room temperature. The supernatant was removed and pellets were resuspended in 2 ml 1x LiAc/0.5x TE (10x 1M lithium acetate, 10x TE (100mM Tris-HCl, 10mM EDTA, pH 7.5), SDW) to make the yeast cells competent. Ten microlitres of competent yeast cells were added to each sample tube (Table 6.3) and gently mixed by pipetting (Table 6.3). Seventy microlitres of 1x LiAc/1x TE/40% PEG (PEG-3350) was added to each tube with very gentle mixing by pipetting. All tubes were incubated at 30 °C for 30 minutes. 8.8 µl of DMSO was added to the top of each tube and samples were gently stirred by pipetting. Cells in each sample were heat shocked at 42 °C for 7 minutes. Samples were centrifuged at 1000 g for 1 minute. The supernatant was removed gently leaving ~10 µl. Eighty microlitres of SDW was added to each sample tube. Tubes 1-3 (pDEST32_g02394, pDEST32_g02469 and pDEST32_g05682) were plated on synthetic complete (SC) -L single dropout medium (yeast nitrogen base, yeast synthetic drop out medium supplements without leucine (Sigma), SDW, agar, 40% glucose, pH 5.6). Tubes 4-7 were plated on SC-LT medium

(yeast nitrogen base, yeast synthetic drop out medium supplements without leucine and tryptophan (Sigma), SDW, agar, 40% glucose, pH 5.6). Plates were grown at 28 °C for 3-4 days until large colonies were obtained.

6.2.7.3 – Control vectors

Three control vectors from the proQuest Two-hybrid system were grown in MAV203. Briefly, the control system works using the interaction between Krev1 and RalGDS. When pEXPTM32/Krev1 is expressed with the control labelled A containing the plasmid pEXPTM22/RALGDS-mutant2 there is no detectable interaction and acts as a negative control. Control B contains pEXPTM22/RalGDS-mutant1 which displays a weak interaction with pEXPTM32/Krev1, this acts as a weak positive control. Control C contains pEXPTM22/RalGDS-wildtype and displays a strong interaction with pEXPTM32/Krev1.

6.2.7.4 – Small scale Y2H reporter gene assay

A small scale Y2H reporter gene assay was conducted to check for auto activation between effector (bait) plasmids and the empty pDEST22 (Prey) vector. Three colonies from per plate (One plate per plasmid combination sample listed in Table 6.3) were placed into tubes containing 100 µl SDW each (conducted in 8 well strip tubes). Colonies were resuspended by pipetting. Three times 10 µl of each sample was pipetted onto each of the following plates: 1 x SC-LT plate – all yeast samples present should have grown on this plate, 1 x SC-LTH (yeast nitrogen base, -leucine/-tryptophan/-histidine drop out supplement (Clontech), SDW, agar, 10 mM 3-amino-1,2,4-triazole, 40% glucose, pH 5.6) – for the HIS reporter screen and 1 x SC-LTU (yeast nitrogen base, -leucine/-tryptophan/-uracil drop out supplement (Clontech), SDW, agar, 40% glucose, pH 5.6) – uracil reporter screen. A fourth plate; SC-LT with a 9 x 11cm Hybond N+ nylon membrane on top of media has 3 x 3 µl of each sample was pipetted on top. The plate containing Hybond N+ nylon membrane is used for the β-galactosidase assay. Each plate contains 10 µl of three control samples detailed above: A – no interaction, B – a weak interaction and C – a strong bait/prey

interaction for comparison against effector samples. All plates were incubated at 30 °C for 24 hours before photographing results. See Figure 6.2 for plate set up diagram.

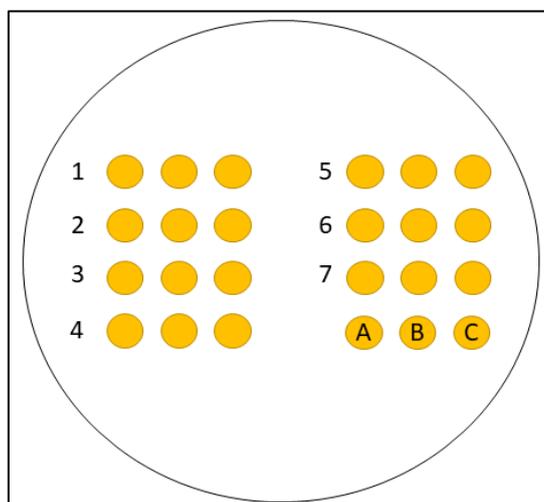


Figure 6.2 – Yeast two-hybrid plate set up. Three individual colonies were taken from each culture plate: 1 - pDEST32_g02394, 2 - pDEST32_g02469, 3 - pDEST32_g05682, 4 - pDEST32_g02394 + pDEST22 empty vector, 5 - pDEST32_g02469 + pDEST22 empty vector, 6 - pDEST32_g05682 + pDEST22 empty vector, 7 – Negative control (Herring sperm DNA) + pDEST22 empty vector. A - pEXPTM32/Krev1 + pEXPTM22/RALGDS-mutant2 negative control, B - pEXPTM32/Krev1 + pEXPTM22/RalGDS-mutant1 weak positive control, C - pEXPTM32/Krev1 + pEXPTM22/RalGDS-wild type strong positive control

6.2.7.5 – β -galactosidase assay

Filter paper was soaked in X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) solution (X-gal, dimethylformamide, β -mercaptoethanol, Z-buffer (Na₂HPO₄ 7H₂O, NaH₂PO₄ H₂O, KCl, MgSO₄ 7H₂O, SDW)). The Hybond N+ nylon membrane with yeast samples grown on it the day before was submerged in liquid nitrogen for 30 seconds to lyse the yeast cell walls. The membrane was placed “yeast side up” on the filter paper and incubated at an angle at 37 °C overnight.

6.2.7.6 – Y2H screen

Ten millilitres of SC-L media (yeast nitrogen base, amino acid minus leucine, SDW, 40% glucose, pH 5.6) was inoculated with a yeast colony from each of the transformations of the pDEST32 baits of interest: pDEST32_g02394, pDEST32_g02469, or pDEST32_g05682. Cultures were incubated at 30 °C with

shaking at an angle. Cultures were diluted to OD₆₀₀ of 0.1 in 150 ml of SC-L media each, then grown for 4-6 hours until an OD₆₀₀ of 0.5 was reached. Each culture was split into 3 x 50 ml falcon tubes which were centrifuged at 2500 rpm for 5 minutes. The supernatant was removed before each pellet was washed in 15 ml SDW and cultures were recombined from the three tubes into one tube per transformation (45 ml in total). Cultures were centrifuged at 2500 rpm for 5 minutes. Supernatant was removed and pellets were resuspended in 750 µl 1 x LiAc/0.5x TE each.

For each effector: pDEST32_g02394, pDEST32_g02469, and pDEST32_g05682 to be tested, two screens were conducted with 15 cold Eppendorf tubes per screen. All tubes (30 in total per effector) were set up containing 2 µl of sheared herring sperm DNA (5 mg/ml) and 1 µg Y2H library DNA (1 µg/µl). Tube 1 (and tube 16 from second screen set) for each effector was a negative control and contained no Y2H library DNA, tubes 2 to 15 were replicates of the screen. Fifty microlitres of yeast culture containing pDEST32_g02394, pDEST32_g02469, or pDEST32_g05682 was added to each of the 15 tubes used for each effector screen. Three hundred microlitres of sterile 1x LiAc/1x TE/40% PEG was added to each tube before being mixed by inverting twice. All 15 tubes per screen were incubated at 30 °C for 30 minutes before 36 µl of dimethyl sulfoxide (DMSO) was added to each. All tubes were incubated at 42 °C for 10 minutes.

Dilutions were produced by removing 10 µl from tubes 2-15 and 17-30 and adding them to 990 µl of SDW per screen for pDEST32_g02394, pDEST32_g02469, and pDEST32_g05682. Secondary dilutions were produced by adding 10 µl from the first dilutions into 90 µl SDW. Ten microlitres of the secondary dilutions were plated onto SC-LT plates. The remaining 90 µl for each dilution was spread on 15 cm SC-LTH plates. Plates were incubated at 30 °C for 5-6 days. Control tubes 1 and 16 for each screen were plated out on SC-LTH. Total colony numbers were counted across the 2 sets of 15 screen plates.

Transformation efficiency for each sample was calculated using the following formula:

$$\text{No. of colonies} \times \text{DF} \times \frac{\text{Transformation vol.}}{0.01 \text{ ml plated}}$$

No. of colonies – Total number of colonies present on plate

DF (dilution factor) = 100

Transformation volume = 5.544 ml

A successful screen should be at least 1×10^6 transformants

6.2.7.7 – Reporter assays for Y2H screen

Colonies were picked into SDW in 8 strip PCR tubes (volume varied depending on colony size ~100 μ l-30 μ l). An autoactivation control (from pDEST32 + EV plate) and the three controls for strength of interaction A, B and C were also picked into SDW. Three microlitres of each were added to the following plates: SC-LT, SC-LTH, SC-LTU and SC-LT plus Hybond N⁺ membrane. Plates were incubated at 30 °C for 24 hours before being photographed. β -Galactosidase assays were conducted as described above (Section 6.2.7.5).

6.2.7.8 – Analysis of Y2H screen

A selection of interacting strength clones from each screen for pDEST32_g02394, pDEST32_g02469, and pDEST32_g05682 were chosen on the basis of the β -galactosidase assay results. For each clone selected yeast cultures were produced in 4 ml of SC-LT media were mixed by vortex before being incubated overnight at 28 °C. Cultures were centrifuged at 2500 rpm for 2 minutes. The majority of the supernatant was removed, and pellets were transferred to 1.5 ml Eppendorf tubes. Centrifugation at 2500 rpm for 2 minutes was repeated a further two times. Plasmid prep of each clone sample was done using a QIAprep spin miniprep kit (Qiagen). The supernatant was removed, and pellets were resuspended in 150 μ l Qiagen P1 buffer + Zymolyase (13 μ l per 1 ml of P1) and incubated at 37 °C for approximately 30 minutes. The manufacturer's protocol was followed with the following changes

made: 150 µl of P2 buffer and 210 µl of N3 buffer was used. The spin columns (blue) from the miniprep kit were substituted with the spin columns (purple) from the MinElute PCR Purification Kit (Qiagen). Final elution was done in 10-12 µl of elution buffer. DNA from each colony plasmid prep was sent to The James Hutton Institute in house facility for sequencing. Identification of interacting proteins was done using BLAST similarity searches.

6.2.8 – Co-immunoprecipitation

6.2.8.1 – Preparation of sample leaf material

GROS_g02394 was cloned into the GFP vector pK7FWG2 and stPRMT1.1 was cloned into the RFP vector pK7WGR2 as described in section 6.2.1. These shall be referred to as g02394_GFP and stPRMT1.1_RFP here for consistency. The g02394_GFP, stPRMT1.1_RFP constructs alongside the empty GFP and RFP vectors pK7FWG2 and pK7WGR2 were transformed into the *Agrobacterium* strain GV3101 as described in section 6.2.2. Overnight cultures were grown for infiltration into *N. benthamiana* leaves. All combinations for infiltration (Table 6.4) were infiltrated at an OD₆₀₀ of 0.5 each as described in section 6.2.3.

Table 6.4 – Combinations of infiltrated vectors for Co-immunoprecipitation

Combination no.	Vector 1	Vector 2	Predicted Co-IP result
1	pK7FWG2 (GFP empty vector)	RFP-stPRMT1.1	No interaction
2	GROS_g02394-GFP	pK7WGR2 (RFP empty vector)	No interaction
3	GROS_g02394-GFP	RFP-stPRMT1.1	Interaction
4 (- control)	pK7FWG2 (GFP empty vector)	pK7WGR2 (RFP empty vector)	No interaction
5	GROS_g02394-GFP	stPUB17-RFP*	No interaction
6	stKH17-GFP*	RFP-stPRMT1.1	No interaction
7 (+ control)	stKH17-GFP*	stPUB17-RFP*	Interaction

* Positive control vectors were obtained from Dr Hazel McLellan, University of Dundee (McLellan *et al.* 2020).

Four leaf discs were collected per vector combination infiltrated (Table 6.4) using a number 9 cork borer at 2 days post infiltration. Leaf discs were frozen in liquid nitrogen either for immediate use or storage at -80 °C. Leaf discs were ground in liquid nitrogen using sterile micro pestles. Five hundred microlitres of cold extraction buffer (GTEN buffer (10% glycerol, 25mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl), 1x mini-protease inhibitor EDTA-free tablet, 1 M DTT, 10x phenylmethylsulphonyl fluoride (PMSF), 10% Nonidet p40) was added upon grinding before incubation on ice for 30 minutes with occasional mixing by vortex. Samples were centrifuged at 4 °C for 10 minutes at 13,000 rpm. Forty microlitres from each vector combination sample was taken into a fresh Eppendorf with 40 µl of 4x SDS loading buffer. These were labelled as input samples were incubated at 95 °C for 10 minutes before storage at -20 °C. The remaining supernatant for each vector combination sample was transferred into a fresh Eppendorf tube to remove any residual leaf material. Twenty microlitres of either magnetic GFP-trap beads (Chromotek) or magnetic RFP-trap beads (Chromotek) were aliquoted into a fresh Eppendorf tube (one per sample). The trap beads were washed three times in 500 µl of wash buffer (GTEN buffer, 1 x mini-protease inhibitor EDTA-free tablet, PMSF). Between each wash the GFP-trap beads were separated from the buffer using a magnetic rack. Five hundred microlitres of sample supernatant was applied to the GFP-trap beads which were incubated for 1 hour at 4 °C on a rotary mixer. The GFP-trap beads were washed three times in 500 µl of wash buffer. The GFP-trap beads were resuspended in 50 µl 4 X SDS loading buffer and incubated at 95 °C for 10 minutes. GFP-trap beads were separated from sample using a magnetic rack. The output samples were stored at -20 °C or used directly in gel electrophoresis.

6.2.8.2 – SDS-PAGE gels

Gel electrophoresis was used to visualise 10 µl of each input sample adjacent to 10 µl of the corresponding output sample on two polyacrylamide gels (Precast NuPAGE 4-12% Bis-Tris gel, Invitrogen) following the manufacturer's instructions. Samples were electrophoresed alongside the prestained PageRuler Plus protein Standard ladder

(ThermoScientific) at 190 V for approximately 1 hour in 1X MES SDS running buffer ((NuPAGE)ThermoFisher). Duplication of gels allowed for one to be used with GFP antibodies and the other with RFP antibodies during the western blot protocol (Section 6.2.8.4).

6.2.8.3 – SDS page gel wet transfer

For each gel run, a polyvinylidene difluoride (PVDF) membrane was soaked in methanol for 2-3 minutes, washed in SDW and then washed in cold 1X transfer buffer (NuPAGE transfer buffer X20 (Invitrogen), 10% methanol, Deionised H₂O). The gels were removed from the plates and equilibrated for a minimum of 10 minutes in cold 1X transfer buffer. Sponges and filter paper were soaked in cold 1X transfer buffer before stacking in the transfer cassette. In general, the cassette is set up as follows: 2x sponge -> 2x transfer paper -> Gel -> PVDF membrane -> 2x transfer paper -> 2x sponge. More sponges were added at either end to fill any additional space in the cassette if needed. The cassette contents were lightly rolled to remove air bubbles. The transfer cassette was placed into an XCell II blot module (Invitrogen) tank. The tank was filled with cold 1X transfer and placed in a bucket of ice to maintain the cold temperature of the transfer buffer. Proteins were transferred at 90 V for 1-2 hours.

After the transfer the cassette was disassembled, the sponges were rinsed clean in SDW, and the transfer paper was removed. The PVDF membranes were washed three times in SDW and once in methanol. To determine if the transfer of protein was successful the PVDF membranes were stained with Ponceau red stain for 1 minute. Ponceau stain was drained off and the membranes were washed washing three times in SDW and once in methanol before photographing. Membranes can be left hanging dry overnight at this stage or used immediately in western blotting.

6.2.8.4 – Western blot

Membranes were washed in methanol followed by a SDW wash. The membranes were then washed in TBS-T (50mM Tris, 150 mM NaCl, 0.05% Tween 20, pH 7.6) for 1 minute. Membranes were then incubated at room temperature shaking in blocking solution (5% non-fat milk in TBS-T) for 1-2 hours. After blocking the membranes were washed twice in TBS-T for 2 minutes. Each membrane was then incubated at 4 °C overnight in primary antibody solution (15 ml blocking solution, primary antibody). For GFP membrane the primary antibody was rabbit GFP antibody (FL) (Santa-Cruz #sc-8334) and was used at a dilution of 1:5000 μ l. For the RFP membrane the primary antibody used was rat RFP antibody [5F8] (Chromotek) and was used at 1:1500. Membranes were washed three times in TBS-T for 5 minutes. Membranes were incubated in the secondary antibody solution (15 ml blocking solution, secondary antibody) for 1 hour at room temperature. For GFP membrane the secondary antibody was goat anti-rabbit IgG-HRP (Santa-Cruz #sc-2004) and was used at 1:10000. For the RFP membrane the secondary antibody used was goat anti-rat IgG-HRP (Santa-Cruz #sc-2006) and was used at 1:10000. The membranes were washed three times in TBS-T for 5 minutes followed by two washes in 1X TBS for 5 minutes. The membranes were coated in SuperSignal West Pico chemiluminescent substrate (Fisher scientific) as per manufacturers protocol. Images of membranes were taken using a G:BOX gel doc (Syngene) and accompanying GeneSys software.

6.3 – Results

6.3.1 – The hypersensitive response in *Nicotiana benthamiana*

The GROS_g02394, GROS_g02469, and GROS_g05682 effectors were infiltrated into wild type *N. benthamiana* to examine whether any provoked a HR response from the plant. GFP was infiltrated as the control for this experiment as this should not trigger a HR response. The GFP line induced HR in 2.8% (4/144) of replicates. Both GROS_g02394 and GROS_g02469 showed varying, but significantly increased degrees of cell death at infiltration sites across three replicates (Figure 6.3 A, B). GROS_g02394 induced HR in 28.5% (41/144) of the replicates with a P value of 0.00 when compared to the GFP control. GROS_g02469 induced HR in 17.4% (25/144) of replicates with a P value of 0.001 when compared with the GFP control. GROS_g05682 did not induce any significant cell death when compared with the GFP control. GROS_g05682 produced HR in 4.9% (7/144) of replicates and had a P value of 0.948 when compared to the GFP control.

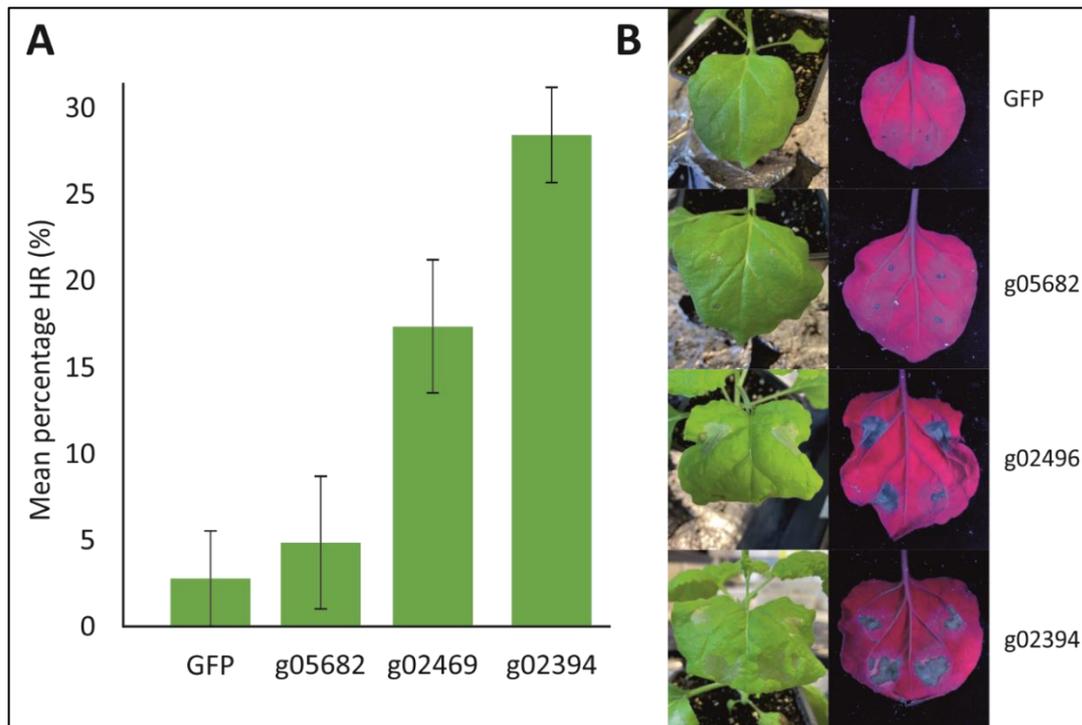


Figure 6.3 – The hypersensitive response of *N. benthamiana* induced by the nematode effectors *GROS_g05682*, *GROS_g02469*, and *GROS_g02394*. A – The mean percentage hypersensitive response observed upon infiltration of leaf tissue with each nematode effector or GFP negative control. B – Photograph of HR response representative of replicates for each effector or control line (adaxial surface, leaf attached to plant) with corresponding UV image (abaxial surface, leaf detached from plant). Error bars were calculated using standard error.

6.3.2 – Subcellular localisation by confocal microscopy

After *GROS_g05682*, *GROS_g02469*, and *GROS_g02394* were confirmed as effectors in chapter three, it was decided to observe where these proteins localise inside the host cell. All three effectors were therefore tagged with GFP in the pK7WGF2 and pK7FWG2 vectors. This produced fusion proteins of each effector tagged with GFP at either the N- or C-terminus. Both configurations were used to confirm that any localisations observed were true and not due to complications (e.g., steric hinderance) caused by application of the tag itself.

6.3.2.1 – GROS_g05682 localisation

When infiltrated into *N. benthamiana*, GROS_g05682 in the pk7WGF2 vector (N-terminal GFP) consistently localised in the cell cytoplasm and nucleus. These infiltrations were repeated at an OD₆₀₀ of 0.2, 0.1, and 0.02 (Figure 6.4 A, B, and C respectively). This result was observed using both wild-type *N. benthamiana* and the transgenic line CB157 which has mRFP fused to histone 2B (mRFP-H2B) resulting in RFP fluorescent nuclei (magenta) (Figure 6.4). GROS_g05682 in the pk7FWG2 vector (C-terminal GFP) did not express consistently or reliably regardless of concentration used or co-infiltration with p19, however very faint fluorescence was observed which suggested this fusion protein also localised to the cytoplasm. The p19 protein comes from the tomato bushy stunt virus and it works to suppress host silencing and therefore increases expression of the co-infiltrated effector protein. Infiltrations of GROS_g05682 in the pk7WGF2 vector at all concentrations were compared to the localisation of free GFP which was observed by agroinfiltration of the pk7WGF2 empty vector at an OD₆₀₀ of 0.02 (Figure 6.4 D). Consistent imaging of free GFP required the addition of p19 at an OD₆₀₀ of 0.1.

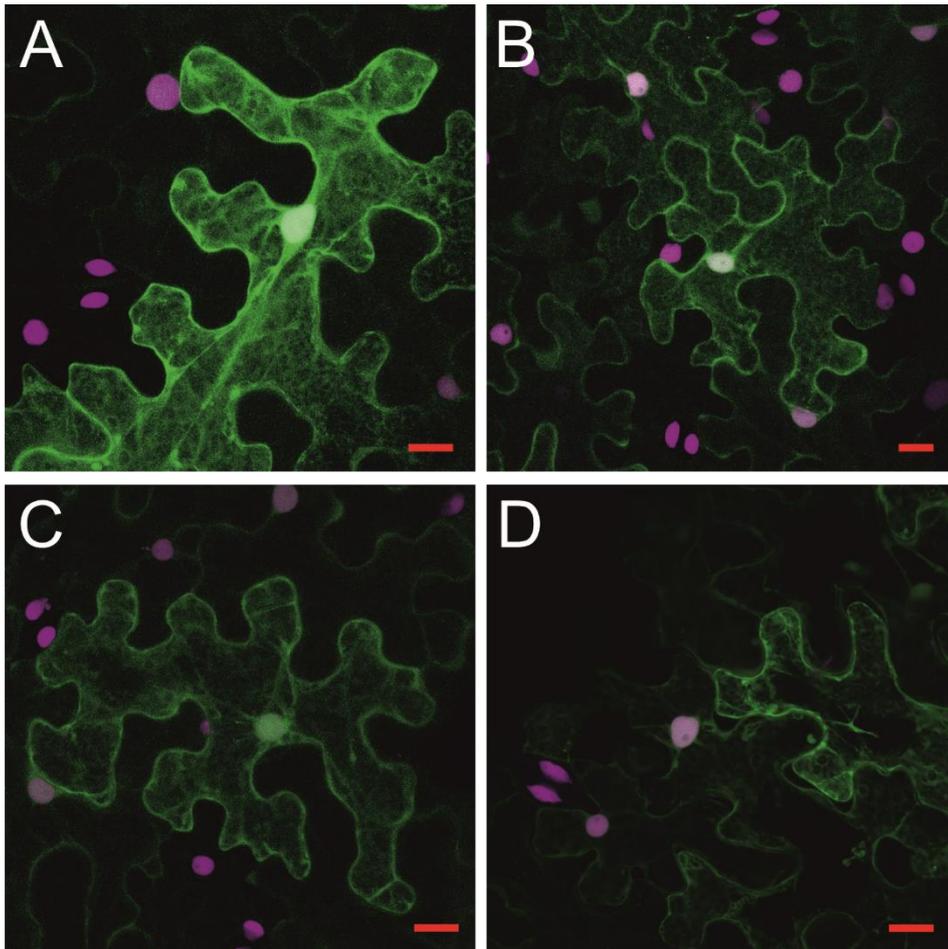


Figure 6.4 - Subcellular localisation of GROS_g05682 in the pk7WGF2 vector transiently expressed in CB157 (RFP-H2B) transgenic *N. benthamiana*. A – Infiltration at an OD_{600} of 0.2, B - Infiltration at an OD_{600} of 0.1, C - Infiltration at an OD_{600} of 0.02. All infiltrations displayed cytoplasmic localisation of this effector. D – Free GFP (pk7WGF2 empty vector) infiltrated at an OD_{600} of 0.02, plus p19 at an OD_{600} of 0.1 displaying the same cytoplasmic localisation. Scale bars – 20 μ m

6.3.2.2 – GROS_g02469 localisation

GROS_g02469 has an interesting localisation pattern that appears disordered and random on first observation. The pattern observed could be loosely described as strings and sphere-like structures distributed across the entirety of the cytoplasm when infiltrated into both wild-type and CB157 (RFP-H2B) transgenic *N. benthamiana*. Initially it was difficult to produce consistent localisations, despite conducting assays to ascertain optimum concentrations for infiltration of each effector. This was solved by co-infiltrating the *N. benthamiana* with the p19 protein. Use of wild-type and CB157 (mRFP-Histone2B) transgenic lines did not provide a

good basis for understanding this localisation, so GROS_g02469 was subsequently infiltrated into the transgenic *N. benthamiana* CB172 line. CB172 expresses a fluorescently tagged endoplasmic reticulum retention signal which results in RFP fluorescent ER. There was no co-localisation observed in this plant line either, therefore GROS_g02469 localising to the ER could be ruled out at this stage. Due to the inconsistent patterns observed by GROS_g02469, it was hypothesised that this effector may be targeting motile bodies e.g., vesicles. To test this theory, short videos (between 30 seconds and 10 minutes) were taken. These videos showed no movement in fluorescence, meaning the movement was either very slow and would require longer periods of video capture, or the target localisation was static.

Observation of the string-like structures led to the hypothesis that these may be parts of the cytoskeleton due to its filamentous components. Using a combination of infiltrating GROS_g02469 into the transgenic *N. benthamiana* CB174 line (RFP-tagged F-actin (LifeAct)) and co-infiltration of GROS_g02469 with the RFP-tagged F-actin (LifeAct) plasmid (obtained from Dr J. Tilsner, referred to as plasmid JT809) it was possible to show that GROS_g02469 localises to the actin filaments of the cytoskeleton. The same localisation of GROS_g02469 to the actin cytoskeleton was observed both when GROS_g02469 is co-infiltrated with JT809 (Figure 6.5 A-C), and infiltration into the transgenic line (Figure 6.5 D-F). Images A-C show there to be less actin filaments overall and there appears to be several “aggregates” (bright dot/sphere shapes) present. When JT809 was infiltrated into wild type *N. benthamiana* individually (control, Figure 6.5 D) more filaments are visible and there are none of these aggregates present. Images F and G of Figure 6.5 contain a co-infiltrated cell as well an un-infiltrated cell in the top right hand corner. The un-infiltrated cell shows the RFP-actin expressed in this transgenic line as it would look natively, which is also the same as image 6.5 H. Using this as a benchmark, it is clear to see that the actin in the cell underneath (bottom right) is disrupted in comparison. There is a marked reduction in visible actin filaments and many aggregates are present. Microtubules were selected as a second cytoskeletal

location as it was hypothesised that there would be no co-localisation of GROS_g02469 with these if the effector was specifically acting upon the host cell actin. For this experiment an mRFP-tagged stCLASP protein was used. This protein has been previously shown to localise to the microtubules (Mei *et al.* 2018). Figure 6.5 I-K shows co-infiltrated GROS_g02469 and mRFP-stCLASP do not co-localise. This is evident in image K as both the green and red fluorescence are individually visible as opposed to the “white” colour observed when GFP and RFP-tagged co-localising proteins are overlaid. In summary GROS_g02469 localises to the actin components of the cytoskeleton and appears to disrupt the native filaments into unorganised aggregates. This may reflect a function in reorganisation of the cytoskeleton during the formation of the syncytium.

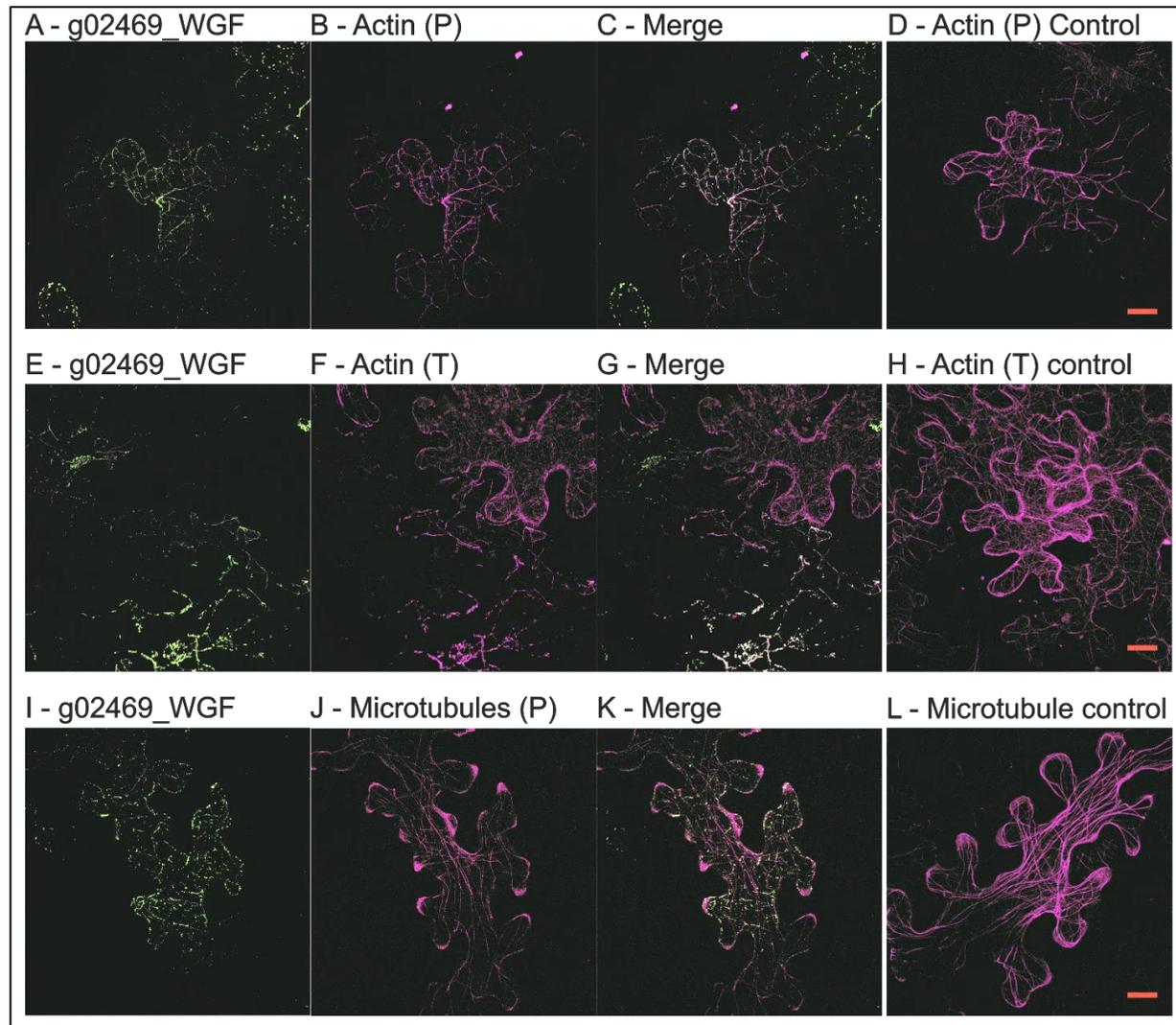


Figure 6.5 – Co-localisation of the GROS_g02469 effector against plant cytoskeleton components. A – GROS_g02469 in *pk7WGF2* (GFP -green) co-infiltrated into WT *N. benthamiana*, B – JT809 vector (RFP-tagged F-actin - magenta) co-infiltrated into WT *N. benthamiana*, C – Merged image of A and B showing co-localisation and disruption of actin by GROS_g02469. D – JT809 individually infiltrated into WT *N. benthamiana* (control), E –GROS_g02469 in *pk7WGF2* co-infiltrated with p19 into the RFP-tagged actin transgenic *N. benthamiana* line CB174. F - RFP-tagged actin transgenic *N. benthamiana* line CB174. G – Merged image of D and E showing co-localisation and disruption of actin by GROS_g02469. H - Transgenic *N. benthamiana* line CB174 with no infiltration (control), I - GROS_g02469 in *pk7WGF2* co-infiltrated into WT *N. benthamiana*, J – mRFP-stCLASP (magenta) which localises at the microtubules co-infiltrated into WT *N. benthamiana*. K – merge of I and J showing independent localisation of GROS_g02469 to the actin and mRFP-stCLASP to the microtubules. L – mRFP-stCLASP individually infiltrated into WT *N. benthamiana* (control). (P) – Infiltrated with plasmid, (T) – Transgenic plant line. Images are increased by 40% brightness/contrast for printing purposes. Scale bars – 20 μ m (all images to same scale).

6.3.2.3 – GROS_g02394 localisation

GROS_g02394 in the pk7WGF2 and pk7FWG2 vectors (encoding N- and C- terminally tagged GFP) was infiltrated at a range of concentrations (OD₆₀₀ of 0.02, 0.1, and 0.2, with and without p19 at an OD₆₀₀ of 0.1) into both WT and the CB157 (RFP-nuclei) transgenic *N. benthamiana* lines. All showed localisation to the actin cytoskeleton (Figure 6.6 A-C). This was evident when compared to the cytoplasmic localisation observed when pk7WGF2 (free GFP) was infiltrated into CB157 transgenic *N. benthamiana* (Figure 6.6 D). Unlike the disruption seen by GROS_g02469 (Section 6.3.2.2), GROS_g02394 does not appear to alter the actin cytoskeleton in any way that can be visualised by fluorescence microscopy (Figure 6.6 E-G). Figure 6.6 G shows the merged image of GROS_g02394 co-infiltrated into the CB174 RFP-tagged actin transgenic line. Although there is clear co-localisation at the actin cytoskeleton, particularly around the cell wall/outer edges of the cytoplasm displayed by the white overlay colour present, there are still some distinctly green and pink areas showing underneath this. This would indicate that there is possibly some of the GFP-tagged effector free in other areas of the cell such as the cytoplasm and the nucleus, that has not bound to the actin. Much like the GROS_g02469 effector, GROS_g02394 was also co-infiltrated to compare with the localisation of the mRFP-stCLASP plasmid. It was determined that there was no co-localisation to the microtubules using this assay (Figure 6.6 I-L).

Since both GROS_g02394 and GROS_g02469 localised to the actin cytoskeleton, co-infiltrations with both effectors were conducted. As both effectors were originally tagged with GFP one was cloned into a secondary fluorescent vector to allow for co-localisation. GROS_g02469 was cloned into pGRAB_mturq2_GW (mTurquoise2 (cyan) fluorescence) which displayed the same “disrupted actin” localisation that the GFP-tagged GROS_g02469 displayed (Figure 6.7 A-C). When infiltrated together GROS_g02394 and GROS_g02469 co-localised to the actin cytoskeleton. The disrupted appearance of the actin caused by GROS_g02469 individually was also observed in these co-infiltrations (Figure 6.7 D-F).

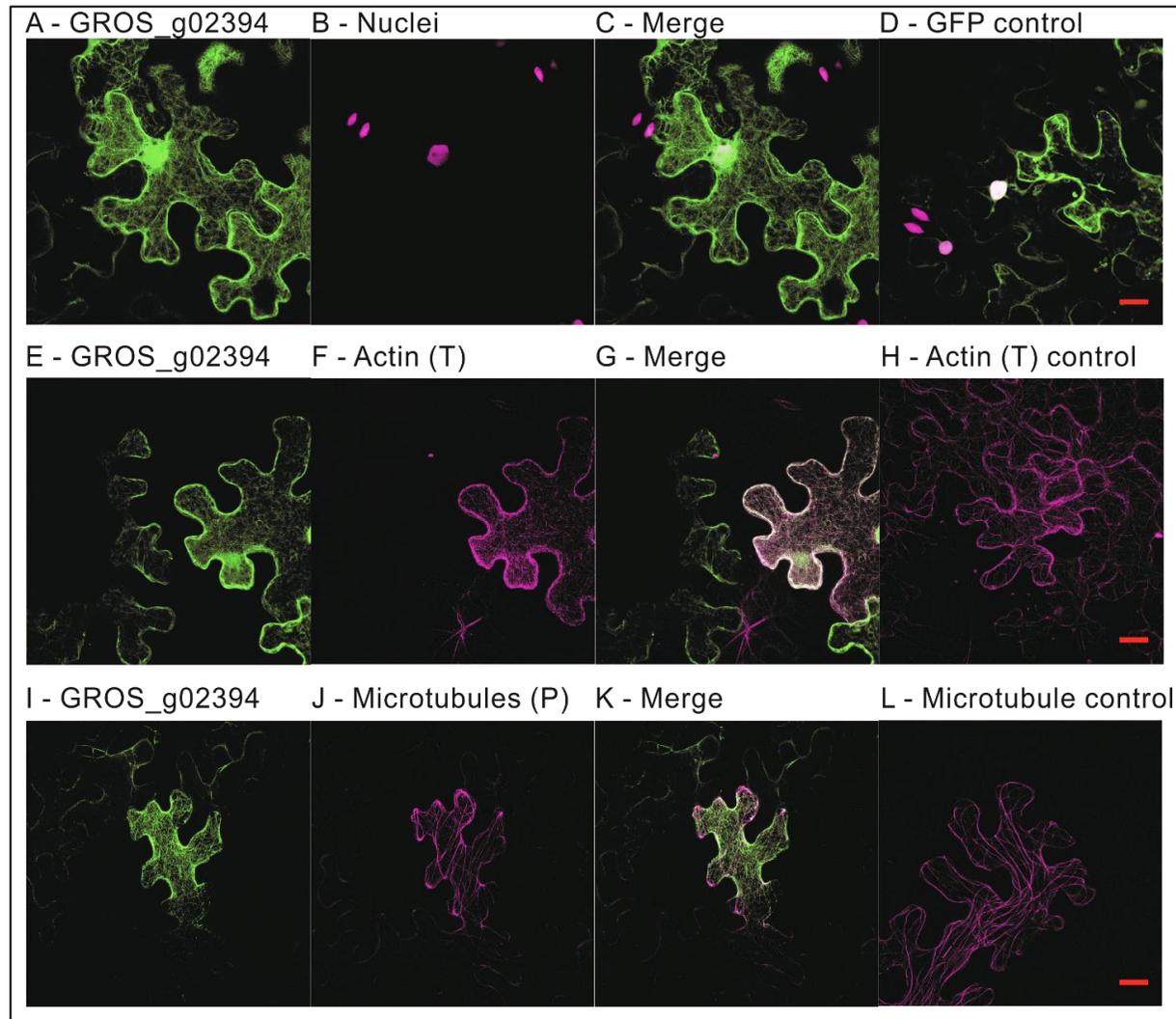


Figure 6.6 - Co-localisation of the GROS_g02394 effector against plant cytoskeleton components. A - GROS_g02394 in *pk7WGF2* co-infiltrated into WT *N. benthamiana*, B - RFP-tagged nuclei (*mRFP-H2B*) in *CB157* transgenic *N. benthamiana*, C - Merged image of A and B. D - Free GFP (*pk7WGF2*) individually infiltrated into *CB157* transgenic *N. benthamiana* (control), E - GROS_g02394 in *pk7WGF2* co-infiltrated with *p19* into the RFP-tagged actin transgenic *N. benthamiana* line *CB174*. F - RFP-tagged actin transgenic *N. benthamiana* line *CB174*. G - Merged image of D and E showing localisation to the actin cytoskeleton by GROS_g02394. H - Transgenic *N. benthamiana* line *CB174* with no infiltration (control), I - GROS_g02394 in *pk7WGF2* co-infiltrated into WT *N. benthamiana*, J - *mRFP-stCLASP* which localises to the microtubules co-infiltrated into WT *N. benthamiana*. K - Merged image of I and J showing independent localisation of GROS_g02469 to the actin and *mRFP-stCLASP* to the microtubules. L - *mRFP-stCLASP* individually infiltrated into WT *N. benthamiana* (control). (P) - infiltrated with plasmid, (T) - transgenic plant line. All images of GROS_g02394 were co-infiltrated with *p19* at an OD_{600} of 0.1. Images are increased by 20/40% brightness/contrast for printing purposes. Scale bars - 20 μm (all images to same scale).

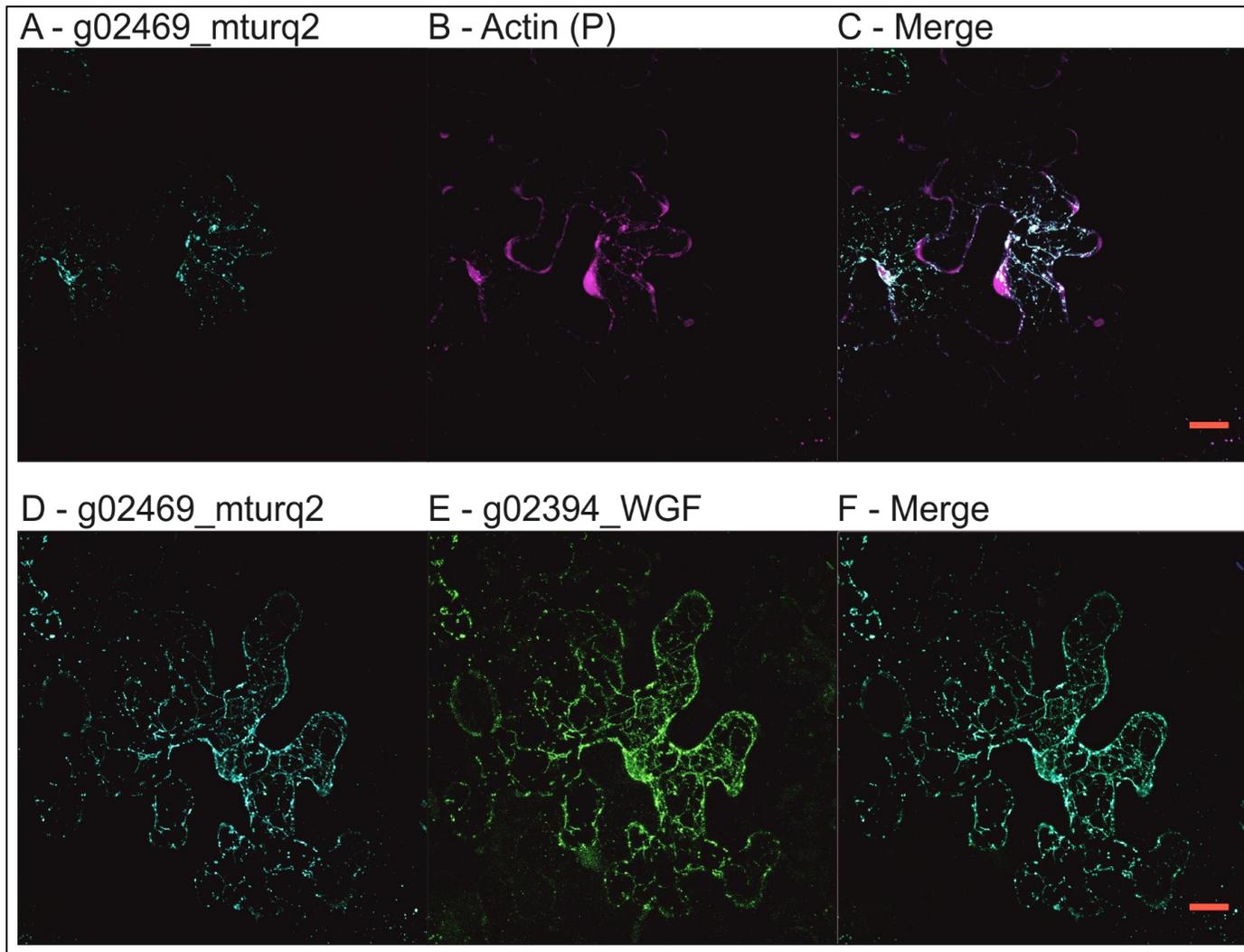


Figure 6.7 - Co-localisation of the GROS_g02394 and GROS_g02469 effectors in WT *N. benthamiana*. A – GROS_g02469 tagged with mturquoise2 fluorescence (Cyan), B – Actin tagged with RFP by infiltration of the JT809 vector (magenta), C – Merged image of A and B displaying co-localisation of the two proteins at the actin cytoskeleton. D – GROS_g02469 tagged with mturquoise2 fluorescence (Cyan), E – GROS_g02394 tagged with GFP fluorescence (pk7WGF2 vector) (Green), F – Merged image of D and E displaying co-localisation of the two proteins. Images are increased by 40% brightness/contrast for printing purposes. Scale bars – 20 μ m (all images to same scale).

6.3.3 – Yeast Two-Hybrid

Yeast two-hybrid assays were used to identify candidate interacting host proteins for each of the three effectors GROS_g05682, GROS_g02394, and GROS_g02469. Unfortunately, despite multiple repeated screens there were no positive interactions identified when using GROS_g05682 or GROS_g02469 as bait.

The Y2H screen using GROS_g02394 as a bait screened 8×10^6 transformants and produced 4 independent positive clones. Sequencing of all 4 yeast colonies identified a fragment (amino acids 53-195) of a single protein. BLAST similarity searches of this sequence returned high identity hits to a predicted arginine N-methyltransferase 1.1-like protein from *Solanum tuberosum* (SeqID: XP_006353524.1). This N-methyltransferase will be referred to as *Solanum tuberosum* Protein Arginine N-methyltransferase 1.1 (stPRMT1.1) hereafter. As initial Y2H screens only identified a section of the methyltransferase, and due to its classification as “predicted” on the NCBI database, the full length stPRMT1.1 gene was cloned from potato cDNA to confirm that the predicted gene exists and is indeed transcribed. The full length (stPRMT1.1_FL) and partial sequence obtained from the Y2H screen (stPRMT1.1_53-195) of stPRMT1.1 are displayed in Figure 6.8. A Y2H assay using this gene and gene fragment shows that colonies co-transformed with either the full length stPRMT1.1 or partial sequence alongside GROS_g02394 grew on both types of selective media (-LTH (-leucine/-tryptophan/-histidine) and -LTU (-leucine/-tryptophan/-uracil)) (Figure 6.9). In this assay both -LTH and -LTU media only promoted growth of colonies containing a positive interaction between GROS_g02394/stPRMT1.1 as these expressed the gene to produce the missing amino acids in each media. Growth on the -LTU plates required a stronger interaction between prey and bait than that on the -LTH plates, which is displayed in both the GROS_g02394/stPRMT1.1_FL and stPRMT1.1_53-195 co-transformations, and by the strong positive control (section 6.2.7.3, proQuest Two-hybrid system control vector C). This was confirmed by positive (colour change to blue) results exhibited during the β -galactosidase assay (Figure 6.9). These are compared to the

GROS_g02394 + pDEST22 empty vector, stPRMT1.1_FL + pDEST32 empty vector and stPRMT1.1_53-195 + pDEST32 empty vector co-transformations. Although there was low level background growth on the -LTH plates for these co-transformations, this is negated by their lack of growth on -LTU and negative β -galactosidase assay results, meaning there is no interaction between these pairings. These data show that the stPRMT1.1 is a candidate interactor for the GROS_g02394 effector.

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
stPRMT1.1_FL	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----													
stPRMT1.1_53-195	HDSVSNNEIENSASYGKKNPSKIKFQYEDDDEQYVQDETVTSSNLEVDLISHCDPPTTAAAVDDSTVGGDKTSADYFDSYSHFGIHEEMLKDYVRTKTYQNVIKNSFLFKDKVYLDVGGAGTILSLFC													
Consensus ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----													
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
stPRMT1.1_FL	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----													
stPRMT1.1_53-195	AKVGAKHYVRIECSHADHAQEIVKLNQFSDVITVIKGVKEEIDLPPVQVDIIISEAHGYFLLYENHLDVLYAROKHLVKDGLVLPDKASLCLTAIEDADYKEDKIEFHWNSVYGFDMSCIRKQAMMEPI													
Consensus ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----													
	261	270	280	290	300	310	320	330	340	350	360	370	380	390
stPRMT1.1_FL	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----													
stPRMT1.1_53-195	VDYDQNGIYVNCQLLKTMDISKMTSGDASFAPFKLIERDDYIHALVAYFDVSEFKCHKLLGFSTGPKSRSTHAKQTVLYLEDYITVCGEAVVGSNTVAPNKNPRVDYIHLKYSYNGKHCRRVSRVQ													
Consensus ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----													
	393	395												
stPRMT1.1_FL	---													
stPRMT1.1_53-195	YYRHR													
Consensus													

Figure 6.8 – stPRMT1.1 amino acid alignment. Full length amino acid sequence of stPRMT1.1 (stPRMT1.1_FL) alongside the partial sequence (stPRMT1.1_53-195) from amino acids 53 to 195 that was originally identified by the Y2H screen.

	-LT	-LTH	-LTU	β -gal
g02394 + stPRMT1.1_FL				
g02394 + stPRMT1.1_53-195				
g02394 + EV (pDEST22)				
stPRMT1.1_FL + EV (pDEST32)				
stPRMT1.1_53-195 + EV (pDEST32)				
(-)				
(+)				
(++)				

Figure 6.9 – Y2H screening with the GROS_g02394 effector. Y2H assay showing interaction between the nematode effector protein GROS_g02394 (bait) and the *S. tuberosum* protein stPRMT1.1 (prey). GROS_g02394 and stPRMT1.1 were grown on the control plate (synthetic complete media (SC) -LT (-Leu/-Trp)), the selective plates -LTH (SC -Leu/-Trp/His) and -LTU (SC -Leu/Trp/Ura). GROS_g02394 and stPRMT1.1 were also grown on -LT + β -galactosidase assay plate. The (+) control should have been blue in the β -galactosidase assay instead of the slight tinge present, however there was issues with loss of function with this culture at the time of this experiment. It was kept as a positive control for growth on the selective media.

When the full-length amino acid sequence for STPRMT1.1 was used as a BLAST query all returned hits were from other plant derived N-methyltransferases containing the words “probable”, “putative”, or “hypothetical”. This is due to the function of PRMT proteins in plants being a relatively unprobed area when compared to research in mammalian species e.g., human PRMTs. Any plant proteins with a sequence similar to functionally characterised PRMTs from other organisms would be annotated this

way. PRMTs are part of a group of proteins which add post-translational modifications (PTM) to other proteins. Well known examples of PTM are phosphorylation and ubiquitination. PRMTs add methyl groups to arginine residues of target proteins. PRMTs have roles in many processes such as cellular differentiation, RNA processing, and DNA repair in mammalian systems. It has also been shown that disrupting these functions can lead to lethal outcomes in mammalian cell systems e.g. cancer development (Ahmad & Cao 2012, Bedford & Richard 2005). Localisation of mammalian PMRTs appears to be quite varied; PRMT8 is targeted to the plasma membrane while PRMT5 is cytoplasmic and PRMT2 is nuclear (Hermann *et al.* 2009).

6.3.4 – Subcellular localisation of GROS_g02394 and the interacting host protein stPRMT1.1

Once the interaction between stPRMT1.1 and GROS_g02394 was identified using Y2H, stPRMT1.1 was fluorescently tagged for use in subcellular localisation studies. stPRMT1.1 was tagged with mTurquoise2 (mturq2) to co-infiltrate and assess the subcellular localisation of the effector and its target simultaneously. The results show that the stPRMT1.1_mturq2 is both cytoplasmic and nuclear when individually infiltrated into the leaves of *N. benthamiana* (Figure 6.10 D, H). It was hypothesised that localisation of stPRMT1.1 would be clustered around areas high in ribosomes - such as the rough endoplasmic reticulum – due to its predicted role in post-translational modification. Due to ribosomes being found free in the cytoplasm, the result is consistent with this original hypothesis. When GROS_g02394 was co-infiltrated with stPRMT1.1 there was no apparent co-localisation between effector and target. GROS_g02394 still localised predominantly to the actin cytoskeleton, while stPRMT1.1 localised the cytoplasm and nucleus (Figure 6.10 A-C, E-G). This result was replicated using both N-terminal and C-terminally tagged GROS_g02394. However, the localisation pattern of stPRMT1.1 when infiltrated individually is distinctly different from the pattern seen when stPRMT1.1 is co-infiltrated with the GROS_g02394 effector. The co-infiltrated stPRMT1.1 shown in Figure 6.10 B and F

does display a cytoplasmic localisation, however it appears to pool at certain points around the cells extremities in a way that suggests it has been shifted towards the cell wall. This is in contrast to stPRMT1.1 being evenly distributed across the cytoplasm and nucleus when individually infiltrated as seen in Figure 6.10 D and H. Given these results, the presence of the GROS_g02394 effector appears to have a spatial effect on the stPRMT1.1 protein. Interestingly GROS_g02394 did appear to co-localise with the actin cytoskeleton at the cell extremities/cell wall in the previous localisation studies conducted (Section 6.3.2.3, Figure 6.6 G). Due to this it was surprising to not see a direct localisation between GROS_g02394 and stPRMT1.1. It is currently unclear where the link lies between the interaction of stPRMT1.1 and GROS_g02394 and the effector localisation to the actin cytoskeleton.

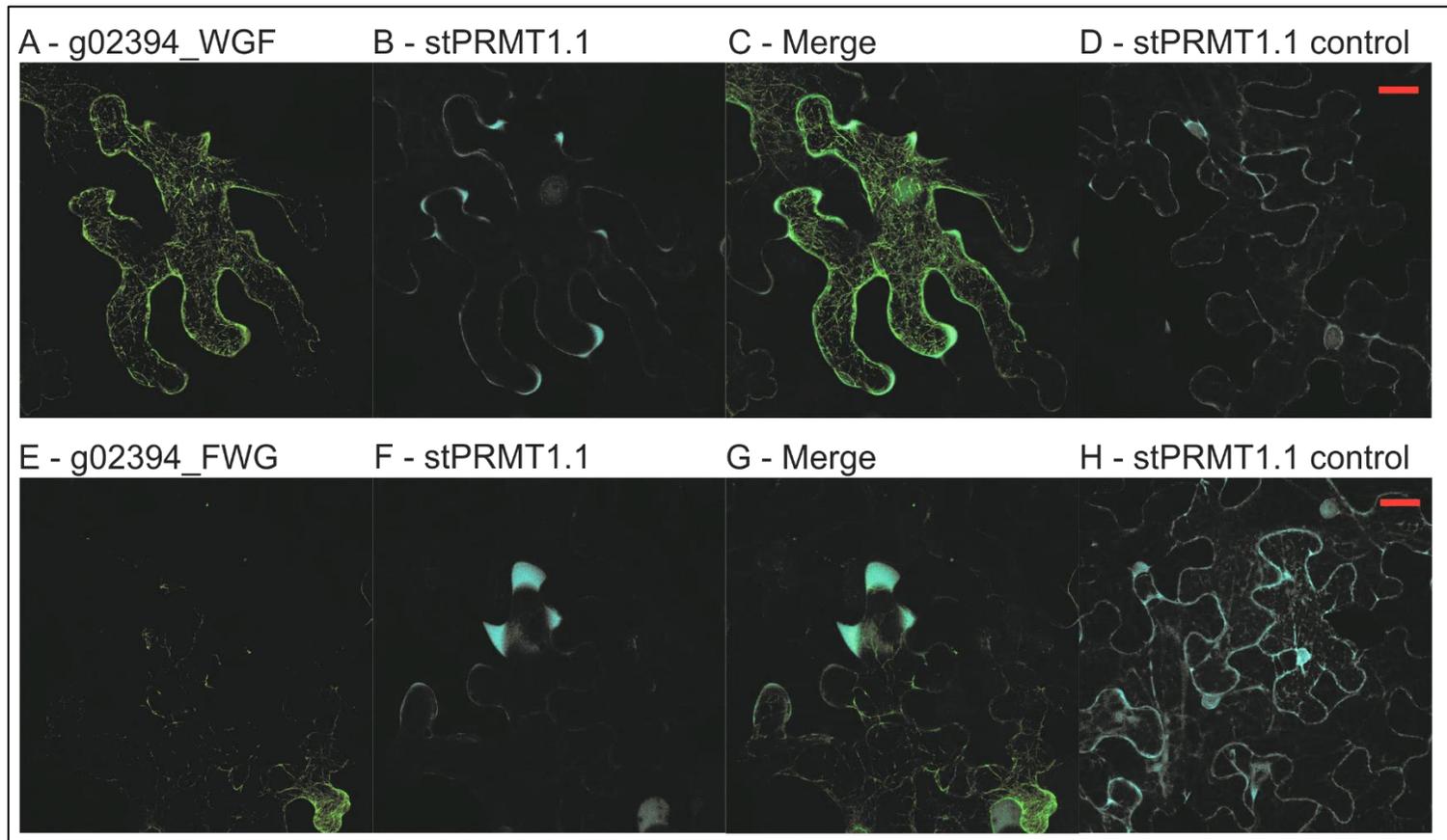


Figure 6.10 - Co-infiltration of *stPRMT1.1* and the *GROS_g02394* effector. A – *GROS_g02394* tagged with GFP (*pk7WGF2* vector) (green) localising to the actin cytoskeleton when co-infiltrated with *stPRMT1.1*, B - *stPRMT1.1* tagged with *mTurquoise2* (cyan) localising to the cytoplasm at the cell extremities and nucleus when co-infiltrated with *GROS_g02394*, C – Merged image of A and B showing lack of co-localisation, D – *stPRMT1.1* infiltrated individually displaying localisation to the cytoplasm and nucleus. E - *GROS_g02394* tagged with GFP (*pk7FWG2* vector) (green) localising to the actin cytoskeleton when co-infiltrated with *stPRMT1.1*, F - *stPRMT1.1* tagged with *mTurquoise2* (cyan) localising to the cytoplasm at the cell extremities and nucleus when co-infiltrated with *GROS_g02394*, G – Merged image of E and F showing lack of co-localisation, H – *stPRMT1.1* infiltrated individually displaying localisation to the cytoplasm and nucleus. Images are increased by 20% brightness/contrast for printing purposes. Scale bars – 20 μ m (all images to same scale).

6.3.5 – Co-immunoprecipitation for confirmation of Y2H results

With interaction between the GROS_g02394 effector and stPRMT1.1 having been identified by yeast two-hybrid; it was important to confirm this interaction outside of the yeast model. GROS_g02394 tagged with GFP (g02394-GFP) and stPRMT1.1 tagged with RFP (stPRMT1.1-RFP) were co-infiltrated into the leaves of *N. benthamiana* to determine if there was an interaction between these proteins in plant cells using co-IP. Other protein pairings co-infiltrated were: Free GFP and stPRMT1.1-RFP (negative control), g02394-GFP and free RFP (negative control), free GFP and free RFP (negative control), and stKH17-GFP and Pub17-RFP (positive control, provided by Dr H. McLellan, University of Dundee (McLellan *et al.* 2020)).

If these proteins interacted in the plant a pull down of g02394 tagged with GFP using GFP trap beads should also pull down stPRMT1.1 tagged with RFP attached to it. This would be confirmed by the presence of bands in the elution sample visible on both blots treated with anti-GFP and anti-RFP antibodies. Preliminary replicates to observe the interaction between g02394 and stPRMT1.1 using magnetic GFP-trap beads were attempted, however no interaction between GROS_g02394 and stPRMT1.1 was observed. On the membrane probed with anti-GFP antibody (referred to as GFP membrane in figure) in Figure 6.11 it is possible to see that each protein pairing has been enriched in the elution from GFP-trap beads (output lanes in Figure 6.11). The bands for both elutions of g02394-GFP + RFP and g02394-GFP + stPRMT1.1-RFP can be observed between 100 and 130 kDa, however they were very faint. These bands were not observed on the anti-RFP antibody treated membrane (RFP membrane in Figure 6.11), suggesting they do not interact. However, there were a number of issues encountered with this assay. The positive control sample using the stKH17-GFP + stPUB17-RFP interaction did not work as expected. This protein pair has been published previously as interacting in both Y2H and co-IP (McLellan *et al.* 2020). No interaction between either the GROS_g02394-GFP + stPRMT1.1-RFP or the positive control pair stKH17-GFP + stPUB17-RFP could be observed when the assay was repeated using magnetic RFP-trap beads. It is currently unclear why no interaction is being observed here but an issue with the antibodies

used cannot be the reason in this case due to the presence of bands on both the anti-GFP and anti-RFP antibody treated membranes. Two bands can be seen on the blot probed with anti-RFP antibody in Figure 6.11; input samples for GFP + stPRMT1.1-RFP and g02394-GFP + stPRMT1.1-RFP. It was expected that bands would be seen in all input samples on this membrane as all samples contain RFP. Further optimisation is required on this assay before a definitive result can be concluded as the current lack of observable interaction between a confirmed interacting control pair places doubts on the lack of interaction observed between our proteins of interest.

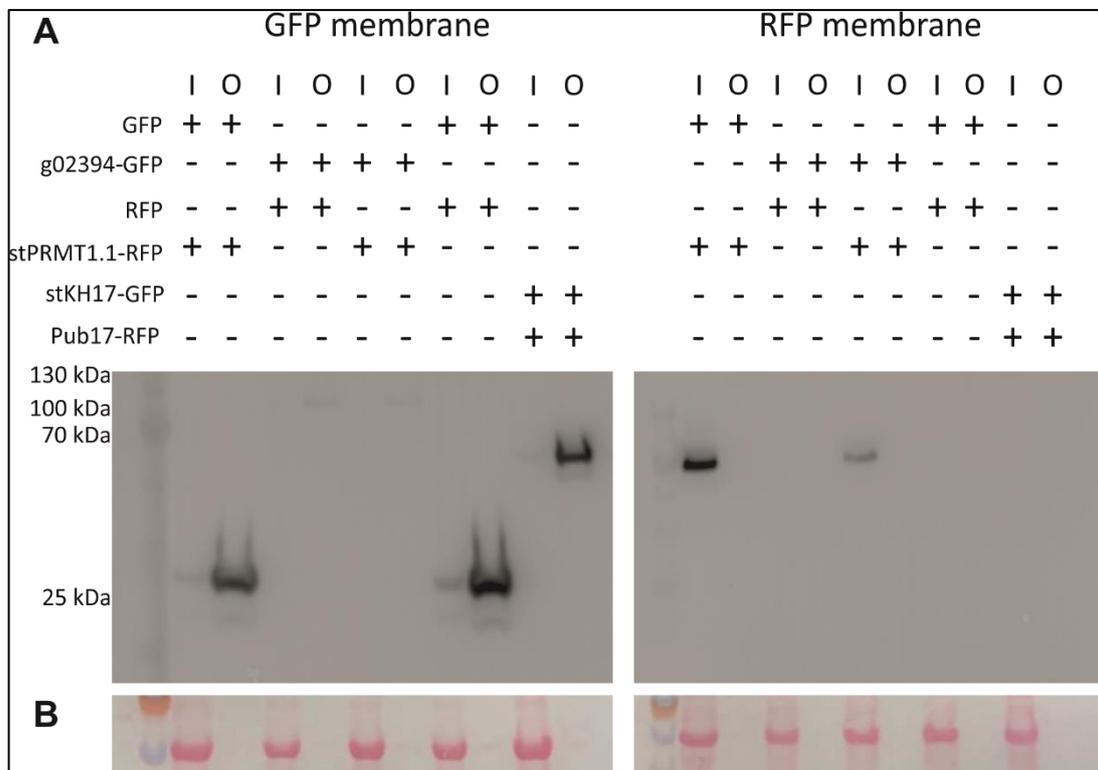


Figure 6.11 – Co-immunoprecipitation of the GROS_g02394 effector and stPRMT1.1 following pull down with GFP-Trap beads. A - Left membrane (GFP membrane) was treated with anti-GFP primary antibody while the right membrane (RFP membrane) was treated with anti-RFP primary antibody. B – Ponceau stain to demonstrate equal protein loading and transfer. Constructs expressed by agroinfiltration are indicated by a +. I – Input sample (total protein extraction from leaf tissue), O – Output sample (protein eluted from Trap beads), kDa - Kilodaltons.

6.4 – Discussion

6.4.1 – The hypersensitive response in *Nicotiana benthamiana*

Both the GROS_g02469 and GROS_g02394 effectors elicited a cell death response when infiltrated into *N. benthamiana* plants. When plants are infected with a pathogen, they can fight against infection using the hypersensitive response (Section 1.4.2). HR is the deliberate killing (programmed cell death (PCD)) of those plant cells which have already been infected by a pathogen and those cells directly neighbouring these. This aims to prevent spread of the pathogen throughout the rest of the plant. The hypersensitive response can be exhibited by both host plants which have evolved/gained R-genes for pathogen recognition or by non-host plants. Usually, an HR response is recognised visually as a relatively small section of dead tissue on an otherwise healthy plant. HR is initiated by host defence genes (R-genes) such as the nucleotide binding (NB) leucine rich repeat (LRR)-related gene family (NLR). Virulence factors e.g., effector proteins are detected and recognised either due to their direct interaction with an R-gene or by their interaction with another host protein. Modifications to host proteins can be detected by R-genes which will then initiate the hypersensitive response (the guard model is discussed in detail in section 1.4.2). The induction of a cell death response due to the presence of these effectors may be due to these proteins being recognised by the plant as part of a defence response. It is also possible that the presence of these proteins in the relatively high concentrations that are introduced by agroinfiltration disrupts normal cell metabolism in a way that causes cell death (rather than programmed cell death). We attempted to avoid this by testing a range of concentrations in these experiments. Alternatively, the effectors may disrupt normal cell metabolism as a result of their (as yet uncharacterised) biochemical function. Future experiments requiring expression of these effectors in plants will need to be performed with the impact of the presence of these proteins on the cell in mind.

The GROS_g05682 effector did not appear to trigger a hypersensitive response at any significantly different rate from the GFP control. It should be noted that

N. benthamiana is not a natural host of PCN so it is possible that this effector protein does not induce any defence response in this plant or the effector is not recognised by the plant in a way that would elicit cell death. HR assays will be more valuable to repeat in future if interacting host proteins were identified and if it were possible to carry out the same assay using a susceptible host species e.g., potato. Co-infiltration studies could be assessed for the initiation of HR by recognition of the interacting protein. Infiltration of effectors on their own will only give half the story.

6.4.2 – Subcellular localisation – Confocal microscopy

6.4.2.1 – GROS_g05682

The localisation of GROS_g05682 was both cytoplasmic and nuclear. This localisation may not provide much information that allows a specific function in terms of syncytial production and maintenance to be determined, but it does allow for the narrowing down of its unlikely functions. For example, it is unlikely that this effector has any functions in the apoplast or cell wall degrading abilities. More generally GROS_g05682 may have an enzymatic function or it could also be involved in altering or disrupting host signalling pathways or metabolic processes throughout the cells.

6.4.2.2 – GROS_g02469

The disruption of actin by GROS_g02469 fits with the overarching understanding that many effectors play a role in syncytial formation and maintenance. As discussed in section 1.2, the nematode uses effectors to alter the native structure of the host cells, modifying them from individual cells into a large multinucleate nutrient sink. This is achieved through plasmodesmata widening and partial cell wall dissolution. It follows that the cytoskeleton of the cell would at least be partially broken down or disrupted and remodelled during this process. It has been previously shown that the actin cytoskeleton is altered from the native state in both syncytia and giant cells (Engler *et al.* 2010). Recently the profilin MiPFN3 effector from the root-knot nematode *M. incognita* was shown to disrupt the actin cytoskeleton of *A. thaliana* cells (protoplasts) (Leelarasamee *et al.* 2018). An alignment of GROS_g02469 with

the protein sequences of MiPFN3 and MiPFN1 showed no sequence similarity despite the apparently similar function (*data not shown*). This supports the initial objective of the thesis; identification of effectors specifically involved in syncytial production and maintenance.

6.4.2.3 – GROS_g02394

Much like the GROS_g02469 effector discussed above (Section 6.4.2.2), it was encouraging to see a second effector that is localising to the actin cytoskeleton as this aligns with the overarching function of syncytial formation and maintenance. It was interesting to see that the interacting host protein stPRMT1.1 doesn't share the same localisation as the effector. Initially this may suggest that the interaction is more transient than first thought, however when co-transformed during the Y2H screens the GROS_g02394-stPRMT1.1 containing colonies showed consistent growth on the -LTU media. Growth on this media indicates this is a strong interaction so is at odds with this hypothesis about it being a transient interaction. Although no clear co-localisation was observed, there did appear to be a shift in the localisation of stPRMT1.1 from throughout the entirety of the cytoplasm to pooling at the cells extremities when co- infiltrated with the GROS_g02394 effector.

6.4.2.4 – Prediction models versus subcellular localisation assays

As previously discussed in chapter 3, three computational models: WoLF PSORT, DeepLoc, and LOCALIZER were used to predict the subcellular localisation of GROS_g02469, GROS_g02394, and GROS_g05682 (Horton *et al.* 2007, Almagro Armenteros *et al.* 2017, Sperschneider *et al.* 2017). Both WoLF PSORT and DeepLoc were used to predict the location of GROS_g02469, GROS_g02394, and GROS_g05682 in the nematode. Due to effector proteins being secreted outside of the nematode it was predicted that these models would return “extracellular” location results. LOCALIZER was used to predict the subcellular location of the effector proteins post-infiltration into the host plant.

Table 6.5 – Summary of predicted vs. confirmed subcellular location of core effectors

Effector	WoLF PSORT prediction		DeepLoc prediction		LOCALIZER prediction	Confirmed localisation in <i>N. benthamiana</i>
	With SP	Without SP	With SP	Without SP		
GROS_g02394 (GLAND11)	Extra	Nucleus cytoplasm	Extra	Mito	Nucleus	Actin cytoskeleton/ nucleus/ possible cytoplasm
GROS_g02469 (G23G11)	Extra	Nucleus	ER	Cytoplasm	Nucleus and/or chloroplast	Actin cytoskeleton
GROS_g05682 (20E03)	Extra	Plasma membrane	Extra	Cytoplasm	Nucleus and/or Mito	Cytoplasm

*Extra – extracellular, Mito – mitochondria, ER – endoplasmic reticulum

The localisation of the three effectors was tested using WoLF PSORT and DeepLoc both including and not including the signal peptide (Table 6.5). The signal peptide targets the protein to the secretory pathway. Both WoLF PSORT and DeepLoc predict that GROS_g02394 and GROS_g05682 effector proteins would be secreted extracellularly when the SP was included. For GROS_g02469, WoLF PSORT predicted extracellular localisation while DeepLoc predicted localisation at the endoplasmic reticulum (ER). A predicted location of “extracellular” supports the assertion that these proteins are secreted outside of the nematode and into the host plant. When the SP was removed and the mature amino acid sequences were used, the localisation for each effector changed. None of the predicted locations given by WoLF PSORT matched those given by DeepLoc. Using the mature amino acid sequences Localiser predicted that all three effectors would be localised to the

nucleus, as well as GROS_g02469 localising to the chloroplast and GROS_g05682 localising to the mitochondria. Through confocal microscopy and fluorescent tagging to observe the localisation of these effectors it has now been shown that only one of these three predictions (GROS_g02394) was correct, partially. This shows the precautionary nature under which prediction/computational models should be used and demonstrates the need for experimental confirmation.

6.4.3 – Yeast two-hybrid

GROS_g02394 interacted with the arginine N-methyltransferase stPRMT1.1 in Y2H assays and stPRMT1.1 is likely to be involved in host post-translational modification. After translation, proteins can go through covalent modifications such as ubiquitination, phosphorylation, and methylation. Post translational modifications (PTM) provide a variety of functions such as being necessary in the formation of the mature protein, altering structure by promoting protein folding, or targeting a protein for degradation after its use, or due to damage and misfolding (Wilkinson 1987, Duan & Walther 2015, Wold 1981). Methylation is the addition of a CH₃ group and can occur on multiple atoms e.g., oxygen (O), sulphur (S), carbon (C) and nitrogen (N). The methylation of nitrogen atoms can occur on the protein amino acid residues arginine, glutamine, asparagine, and histidine. Protein arginine N-methyltransferases (PRMTs) alter protein function by methylation of nitrogen atoms found on the guanidinium group of arginine residues (Krause *et al.* 2007). PRMT proteins can be divided into three types (Type I, II & III) depending on their catalytic activity and product (Cha & Jho 2012, Blanc & phane Richard 2017). PRMTs have variable C- and N-terminal sequences which convey substrate specificity between protein family members.

In mammals there is a small family of PRMTs containing nine protein members (PRMT1-9) (Hermann *et al.* 2009, Yang *et al.* 2015). Of these nine mammalian PRMT, stPRMT1.1 has highest sequence similarity to human PRMT1 (using *H. sapiens* isoform 1 - canonical sequence for comparison). Human PRMT1 is a type I PRMT

which can produce both ω - N^G -monomethylarginine (MMA) and asymmetric ω - N^G,N^G -dimethylarginine (aDMA) using the two terminal guanidino nitrogen atoms (Ahmad & Cao 2012). This is facilitated using the methyl donor S-adenosyl-L-methionine (SAM). PRMT1 appears to have a large suite of substrate proteins which it interacts with, however one of the most notable is histone H4 (Strahl *et al.* 2001). Methylation of H4 can alter chromatin states leading to changes in gene expression. Based on high sequence similarity with PRMT1, it is possible that stPRMT1.1 may also function as a type I PRMT. This is also supported by the fact that human PRMT1 in its active state localises to the cytoplasm and nucleus (done in human embryonic kidney cells), the same localisation displayed by stPRMT1.1 (Herrmann *et al.* 2005).

Protein methylation was first discovered in 1959, however the study of PRMTs has seen few breakthroughs in plants to date (Murn & Shi 2017). PRMTs have been characterised in both *A. thaliana* and *Oryza sativa* (rice). These plant species contain a homolog of each of the 9 mammalian PRMTs as well as a tenth plant-specific member which has been linked to regulation of flowering time (Niu *et al.* 2007). In *A. thaliana* there are two homologs to mammalian PRMT1 – AtPRMT1a and AtPRMT1b (formerly known as AtPRMT11) (Figure 6.12). AtPRMT1a and 1b have been shown to co-localise to both the nucleus and cytoplasm when transiently transfected into onion epidermal cells, complementing the localisation patterns seen by stPRMT1.1 and PRMT1 (Yan *et al.* 2007).

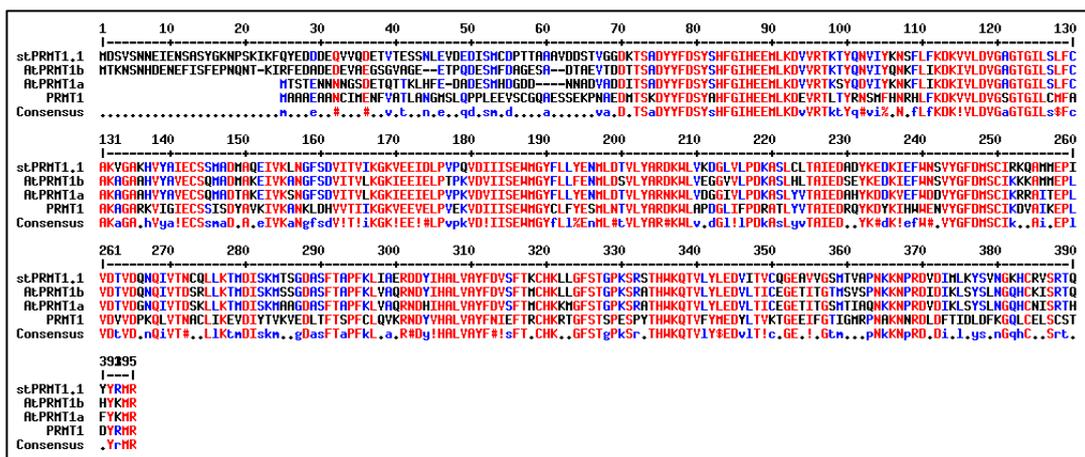


Figure 6.12 - Arginine N-methyltransferase (PRMT) alignment. Sequence conservation between *stPRMT1.1* (*S. tuberosum*), *AtPRMT1a* and *1b* (*A. thaliana*) and *PRMT1* (*H. sapiens*).

AtPRMT1b has been shown to interact with AtMBD7 (methyl-CpG-binding domain 7) (Scebba *et al.* 2007). AtMBD7 methylates DNA sequences which makes the link between DNA methylation seen carried out by mammalian PRMTs and protein methylation that *stPRMT1.1* is likely to carry out (Zemach *et al.* 2008). AtPRMT1b also methylates histone 4B. Histone methylation can alter the transcription levels of certain genes and can alter protein interactions with the histone and chromatin. Given everything that is currently known about the functions of PRMT1, AtPRMT1a, and AtPRMT1b, it is possible that *stPRMT1.1* would function in a similar way. Interacting with post-translational modifying proteins and those with DNA methylation functionality would be beneficial to the invading nematode as it can be used to alter gene expression (either increasing expression of beneficial genes or reducing expression of host defence genes for example), so it is not unlikely that effectors would exist for this purpose.

This still leaves the question of how the actin cytoskeletal localisation of GROS_g02394 fits with this narrative? It is possible that there is a protein associated with the actin cytoskeleton that would be beneficial to the nematode in a methylated state. GROS_g02394 may then interact with *stPRMT1.1* to initiate methylation of this secondary target. There have been studies that show

nuclear actin exists as a component of RNA polymerase II and together with nuclear myosin, plays a significant role in transcription and chromatin structure (Grummt 2006). It should also be noted that actin itself can be post-translationally methylated and arginylated (additional arginine added onto proteins after translation) in order to regulate nuclear proteins (Saha *et al.* 2011). Although this doesn't explain why the GROS_g02394 localises to cytoskeletal actin at present, it does provide future work areas to understand the function behind this interaction going forward.

With regards to the inability to identify interacting proteins with GROS_g05682 or GROS_g02469, this could be due to the library used for the screens. The nematode effector proteins were used as prey to screen against a cDNA library from potatoes which had been infected with *Phytophthora infestans*. *Phytophthora infestans* is an oomycete which is responsible for late/potato blight on species of the solanaceous family of plants. As the sample tissue had been infected with a pathogen, many of the plant's defence genes would be expressed and present in the library. It is assumed that there would be a large overlap in the defence genes upregulated by the plant if it had been infected by a parasitic nematode e.g., *G. rostochiensis* as seen in the *P. infestans* library. Therefore, the decision was taken to use this library as there was no library available for use which had come from nematode infected tissue at the time of this study. As this library consisted mainly of leaf material, it is possible that some target proteins may not have been present in the library used which means artificially negative results may have been observed for GROS_g05682 or GROS_g02469.

As observed with the subcellular localisation results, it was shown that the GROS_g02469 not only binds to the actin filaments of the cytoskeleton, but also appears to disrupt its native state. With this in mind there are other hypotheses that may explain why we failed to identify an interacting host protein:

- 1 - If this effector is having a destructive effect on actin or an actin associated protein then it may not be possible to recover an intact interactor which would give positive

results in the yeast two-hybrid. In general terms if the interacting host target is being broken down then it is less likely to be identified using Y2H as there will be less intact protein to pull out at the end of the assay. The action of the effector may also be lethal to the yeast cells where an essential protein is targeted and degraded, making it very unlikely that an interaction could be detected in this system.

2 – It is possible that GROS_g02469 interacts with an intermediate host protein that in turn interacts with the actin cytoskeleton. This would explain why no actin was detected using the Y2H. If this intermediate interaction is more transient, then it would be difficult to pull this out of a Y2H study as the interaction could be weak and easily broken. GROS_g02469 also does not show sequence similarity with any previously characterised actin binding proteins. With no actin binding domains readily identifiable in GROS_g02469, it is more likely that its localisation with actin is through an interacting protein and not a direct interaction.

3 – It may be possible that the interaction with the actin cytoskeleton is facilitated by another effector that the nematode secretes into the host plant. If true, this second effector would not exist in the Y2H library so it would be impossible to draw it out and identify it as an interacting protein. It may be possible to identify these unknown interacting proteins through a pull down assay followed by mass spectrometry. Much like the protocol for co-IP, the effector would be tagged, for example, with GFP and infiltrated into plant material. Total protein could then be extracted and the effector could be pulled down using GFP beads. Any interacting proteins would be pulled down with the effector. The products from the pull down could then be identified via mass spectrometry.

6.4.4 – Co-immunoprecipitation (Confirmation of Y2H results)

Initial co-IP studies could not confirm an interaction between GROS_g02394-GFP and stPRMT1.1-RFP as observed in the yeast two-hybrid assay. This currently cannot be taken as a definitive negative result, however as it was also not possible to show an interaction between the previously confirmed interacting pair stKH17-GFP and stPUB17-RFP. The co-IP study was the final experiment to be carried out as part of

this thesis so optimisation of the technical issues encountered was not possible due to time constraints. Although a band for g02394-GFP+stPRMT1.1-RFP was not observed on both blots probed with anti-GFP and anti-RFP antibodies to indicate an interaction, this could be due to a lack of adequate protein concentration in the initial samples. It is possible that the initial infiltrations (OD₆₀₀ of 0.5) were at too low a concentration to observe the interaction and future studies would require increased concentrations to observe previously faint or missing bands.

6.5 – Future work

The interaction between the GROS_g02394 effector and stPRMT1.1 requires further study to fully understand the link between the two, and how this interaction benefits the nematode. As described in section 6.4.3, the homologs of stPRMT1.1 in humans and Arabidopsis interact with histone H4 as well as many non-histone proteins in the case of PRMT1. Therefore, it would be interesting to see if stPRMT1.1 also interacts with and methylates histone H4 specifically. To find other (non-histone) interactors with stPRMT1.1 it would be advantageous to use stPRMT1.1 as the prey protein in additional yeast two-hybrid studies. This would potentially allow for the identification of the interacting proteins of stPRMT1.1 itself, giving a more well-rounded understanding of why stPRMT1.1 is targeted by the GROS_g02394 effector. Potentially this could explain the link with the actin localisation seen by the effector as stPRMT1.1 could pull out an actin binding protein as an interactor for example.

As discussed previously the library used to screen for interacting targets of these effectors may not have been the most appropriate. Due to the high cost and time involved in construction of a new library it was not possible to use a cDNA library from potatoes which had been infected with *G. rostochiensis* at the time of analysis. It is possible that interacting targets for both GROS_g05682 or GROS_g02469 may be identified in future if a more appropriate library became available for use. To further this, it may be possible to do a tandem affinity purification (TAP) (or other similar protein “fishing” protocol) using *G. rostochiensis* cell lysis to identify any

proteins our effectors of interest interact with that the nematode also produces e.g., a second effector.

Going forward with the GROS_g02394 and GROS_g02469 effectors, actin binding assays could be performed. These would show definitively if either of the effectors which localise to the actin cytoskeleton directly interact with the actin itself or not. Firstly, an F-actin binding assay could be performed to ascertain binding status. Then a test to see if either effector has actin bundling activity. There are kits currently available to carry out these tests such as those from Cytoskeleton Inc., however these use actin from mammalian sources (rabbit skeletal muscle actin) which may not show the same results as *in vivo* studies as actin sourced from plant material. Changes to this protocol to include actin from a plant source – potato if available, may be necessary.

The subcellular localisation and Y2H assays have helped identify the functions of these effectors, but there is still significant work to be done within this research area. An interacting host protein has been identified for GROS_g02394 and an actin disrupting phenotype was shown for GROS_g02469, however the function of GROS_g05682 remains elusive. RNA interference to knock down these three genes would be beneficial and add to the knowledge base on these effectors. RNAi would show the effect removing these effectors has on parasitism. Possibly there would be less females present and the feeding sites could potentially be compromised with the removal of the actin disrupting effector GROS_g02469. For RNAi assays to be conducted, as GROS_g05682 is expressed highly exclusively at the J2 life stage, this would require the J2 soak method of applying RNAi constructs. Transgenic lines expressing hairpin RNAi constructs targeting GROS_g02469 and GROS_g02394 could be made as these genes are expressed highly at the 14 dpi life stage.

Additional optimisation of co-IP studies is required to confirm the interaction between GROS_g02394 and stPRMT1.1. This interacting status of these proteins is

in limbo currently as it has been demonstrated by Y2H, however confirmation of interaction outside of the yeast model has not been proven during this thesis. When taking the lack of co-localisation observed under confocal microscope there are definitely still numerous questions surrounding this effector and its function in the host plant.

7. General Discussion

This thesis was undertaken in order to further the understanding of how syncytia-forming nematodes interact with their host plants in order to achieve successful parasitism. This has been achieved primarily through identification of new effectors that are present in several syncytium forming species: *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans*. Further to this a subset of these newly identified effectors were functionally characterised. In addition, a family of cathepsin L-like peptidases present in all of these nematodes was characterised.

7.1 – The importance of feeding site production

Plant-parasitic nematodes are highly damaging to agriculture, in part due to the many different host plants that are infected. The most damaging PPN include species which form extended biotrophic interactions with their hosts, most notably those which produce feeding sites. The ability to induce a feeding site has clearly proven to be successful for nematode parasitism, with sedentary endoparasitism and subsequent feeding site production potentially having evolved independently five times in nematodes (Holterman *et al.* 2017).

Nematode feeding sites include syncytia and giant cells. Syncytia are large, multinucleate, nutrient rich feeding sites generated through rounds of protoplast fusion. In some species up to approximately 200 host cells can be incorporated into a syncytium. Syncytia are predominantly produced by cyst nematodes from the *Heterodera* and *Globodera* genera, however other nematode species such as *R. reniformis* and *N. aberrans* also produce these structures. The most extensively studied syncytia are those produced by the PCN species *G. rostochiensis* and *G. pallida*, as well as the cereal, soybean, and beet cyst nematode species *H. avenae*, *H. glycines*, and *H. schachtii* respectively.

Giant cells are produced by the root-knot nematodes *Meloidogyne* genus. Giant cells differ from syncytia as they are not formed through incorporation of neighbouring cells, instead they are formed by enlargement of a small number of cells (approximately 5-10) through reprogramming of host cell development. Giant cells go through multiple rounds of mitosis in the absence of cytokinesis, resulting in large, multinucleate cells that the nematode feeds from. Giant cells also differ from syncytia as they induce the production of a gall in the surrounding root tissue. Syncytia and giant cells are induced as a response to secreted effector proteins, which initiate and sustain these structures.

Feeding site production is not solely contained to the endoparasitic nematodes species, as they are also induced by *Xiphinema* and *Longidorus* spp. The dagger nematode *Xiphinema index* is a migratory ectoparasite which feeds by probing its long stylet through the cell wall into individual cells as it moves along the root surface. It has been observed that upon feeding on root tip cells *Xiphinema index* induces these to become large and multinucleate due to rounds of mitosis without subsequent cytokinesis, as described for RKN. These giant cells are much smaller than those produced by RKN species and are not maintained for sustained use, eventually becoming necrotic due to continual feeding (Bleve-Zacheo & Zacheo 1983, Wyss 2002). Species of *Longidorus* such as *L. elongatus* also produce galling of the root tip after feeding induces both cell hypertrophy and hyperplasia (Griffiths & Robertson 1984, Wyss 2002). The use of feeding sites in successful parasitism strategies of this diverse set of nematode species suggests that nematodes can exploit plant developmental strategies with relative ease, despite the host defences that have evolved against this parasitic strategy.

7.2 – Understanding the syncytium

Although there is some understanding of the processes underlying syncytium formation, currently our knowledge is incomplete. For many species the physical development of the syncytium has been observed from the selection of the initial

syncytial cell to the final expanded syncytium in which many hundreds of cells may be incorporated. Effectors are known to play a large role in syncytium formation and maintenance. As discussed in section 1.2.2, nematodes hijack host processes for their own gain. *Heterodera schachtii* redirect the flow of the plant hormone auxin using host PIN proteins as part of the process underlying syncytium formation (Grunewald *et al.* 2009). In the presence of *H. schachtii* the expression and subcellular localisation of host PIN proteins is altered, leading to the redirection of auxin transport. This results in a build-up of the hormone at the syncytium and a subsequent increase in cell size and growth. It is thought that this process is induced by a yet to be identified effector secreted by the nematode. Other effectors such as 19C07 have also been shown to be involved in altering auxin signalling for syncytia growth which is discussed in section 3.1.1.1.

Another set of effectors that target plant hormone function are the CLE peptides. Plant CLEs, discussed in section 3.1.1.2, have roles in maintenance of both shoot and root apical meristem stem cell populations, alongside initiation of organ formation from stem cell populations. Nematodes have evolved CLE-like effectors in order to hijack these processes, suggesting a role in cell proliferation during syncytia formation (Guo *et al.* 2017). Our understanding of syncytium formation and maintenance has been improved with the understanding of CLE-like effectors amongst other examples, however these are usually investigated on an individual basis using *in vitro*, non-natural conditions. It is not yet known how these effectors act simultaneously in the natural host environment and there are many effectors and functions which are as yet undescribed.

There are a diverse range of nematodes that can make a syncytium, and recent technological developments have provided both genome and transcriptome resources for many of these species. Therefore, through this thesis it was sought to learn something new about critical effectors by identifying those present in phylogenetically diverse species that can make syncytia.

The potato cyst nematodes *G. rostochiensis* and *G. pallida* parasitise members of the Solanaceae family, causing a large economic impact on the potato, tomato, and aubergine industries. PCN species now have a wide distribution globally. They originated in South America, but after being introduced to and spread across Europe, they have spread across the world with seed potato from this region. They also are now a threat in Eastern Africa (Evans *et al.* 1975, Mburu *et al.* 2020). PCN come with the added concern that they can persist in crop land for decades after initial infection due to their hardy cyst life stage.

Rotylenchulus reniformis has a wide host range of over 300 host species. *Rotylenchulus reniformis* was first discovered on cowpea plants but has subsequently been identified as a parasite of hundreds of plant species including corn, cotton, soybean, tobacco, tea, tomato, potato, as well as fruit trees such as passion fruit and peach palms. This nematode species also parasitises many weeds such as Sicklepod (*Senna obtusifolia*), Velvetleaf (*Abutilon theophrasti*), and *Ipomoea sp.*, which allows it to persist and thrive on the fringes of field land even in the absence of a crop host (Molin & Stetina 2016). *R. reniformis* is found in many climates ranging from temperate to tropical, in areas across all continents apart from Antarctica.

Nacobbus aberrans has a similarly broad host range and parasitises host species including potato, pepper, carrot, cucumber, lettuce, beets, and subspecies of *Brassica oleracea* (kale, cauliflower, cabbage etc.). This nematode species also parasitises cactus plants such as the spiny star/pincushion cactus (*Escobaria vivipara*) and members of the *Opuntia* (prickly pear) genus. *Nacobbus aberrans* is distributed throughout both North and South America. It has also been identified in Egypt and may be present but undetected in other warm and sub-tropical climates (Abu-Gharbieh & Al-Azzeh 2004). It has been detected in Europe, for example in the United Kingdom, but was subsequently eradicated (Jeger *et al.* 2018).

It is clear that the diversity and number of species of host plants that can be parasitised by these syncytia-forming nematode species poses a major threat to global agriculture. In 2011 Nicol *et al* calculated that PPN species cause damage valued at ~\$80 billion dollars in global crop losses each year, however this value may have increased in recent years (Nicol *et al.* 2011). In addition, the ability of nematodes to induce syncytia in such a diverse range of plant species strongly suggests that they are targeting a fundamental plant developmental process in doing so. In broad terms, if we learn more about nematode-host interactions this will give a better picture as to how we combat these parasites in the future. Resistance breeding is a large research area for tackling PPN, but breeding cultivars is costly and time consuming. An alternative may be to develop a multifaceted approach against effectors used across many species and which may play critical roles in infection.

7.3 – Effectors

Effectors are secreted by many plant-parasitic nematode species. Genome analysis has shown that a very large number of effectors are produced and that these can exist in large gene families. The overarching role of effectors is suppression of host defences and alteration or sequestering of host processes to the benefit of the nematode. The majority of nematode effectors are produced in the oesophageal glands; either the dorsal gland or subventral glands. There is a general trend in effectors which function at the J2 life stage, and thus have a role in the very early stages of parasitism, are produced in the subventral glands, while effectors which function during later parasitic stages are produced in the dorsal gland, although there are exceptions.

There are many effectors with known, well studied functions such as the plethora of cell wall degrading enzymes used to soften and break through host plant cell walls, but the function of many effectors remains unknown. Many effectors are pioneers although they may contain domains of unknown function (DUF) that have been identified in other proteins but whose function is unknown. There are also

undoubtedly many effectors which are yet to be identified. Understanding the function of these effectors is a challenge and prioritising these proteins for analysis is important. The original aim of this thesis was centred around identification of core effectors. These were defined as effectors present in a wide range of syncytium forming nematodes and were thought likely to play important roles in nematode biology. Identification of core effectors presents the unique opportunity to potentially identify genes/proteins that could be targeted as a method of control in future studies.

7.4 – Identification of core effectors of syncytia-forming nematodes

The first research chapter of this thesis describes the identification of core effector proteins from a diverse range of plant-parasitic, syncytia-forming nematode species (*G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans*). A total of 43 orthologous groups of candidate core effectors were identified. Of these, three *G. rostochiensis* core effectors; GROS_g05682, GROS_g02469, and GROS_g02394 from the 20E03, G23G11 (originally thought to be part of the GLAND15 family), and GLAND11 orthologous groups were chosen for further characterisation.

The pipeline used to identify these effectors contained five key steps. A list of known effectors present in *G. rostochiensis*, derived from previous published work as well as those identified as being associated with the DOG Box motif, were used as a starting point for BLAST sequence similarity searches against genome and transcriptome data available for the syncytia-forming species *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *N. aberrans*. Results of these BLAST searches were filtered so that all candidate genes retained had a predicted signal peptide and lacked transmembrane domains. The latter were removed as this may indicate proteins that are anchored to the cell membrane in some way, therefore making them unlikely to be secreted as effectors. However, it should be noted that at least one effector from *P. infestans*, which interacts with an ER localised NAC (NAM, ATAF and CUC family) transcription factor, contains a predicted transmembrane domain

(McLellan *et al.* 2013). Although this step may exclude some potential effectors, it will remove far more non-effectors and was therefore included in the pipeline. Gene expression data for each candidate was assessed and any gene which was not upregulated during the J2 or parasitic life stages was filtered out. This therefore removed those genes which had constitutive expression across all life stages. The amino acid sequences of candidate genes were then used in BLAST searches to identify any similar genes present in root knot nematodes. Any candidate genes with high similarity (percentage identity) to a gene from RKN were removed as this would indicate they do not function specifically in syncytial feeding site formation and/or maintenance. Candidates were analysed for the presence of DOG box motifs in the upstream untranslated regions as well as the presence of any conserved functional domains. Effectors which were still present after this filtering stage had *in situ* hybridisation carried out on the *G. rostochiensis* homologue of each of these candidate families to assess the location of gene expression.

The identification of core effectors from syncytia-forming nematodes was successful, however there may also be valid and interesting effectors which were filtered out of this pipeline due to various criteria not being met. Firstly, gene families with more variable sequences may have been missed due to the stringent E-value used here ($1e-5$). An example of this is seen with the HYP effector family, where there are sub-families containing small, conserved motifs and tandem repeats which are surrounded by sections of high sequence variability. It is unlikely that all of the genes identified as part of the initial HYP analysis would have been identified using the BLAST criteria applied here. The same stands for any other variable gene families initially missed in the pipeline. This is also illustrated by the fact that genes from six candidate families (pectate lyase, Gro_DOG_0057, Gro_DOG_0067, Gro_DOG_0073, Gro_DOG_0079, and Gro_DOG_0211) from *R. reniformis* were initially missed until the E-value for BLAST searches was lowered from $1e-5$ to $1e-4$. There will also be effectors missed due to the setup of the initial gene set for BLAST analysis. As the initial gene list was only made up of known *G. rostochiensis* effectors there will

undoubtedly have been some that were not included. However, the inclusion of sequences downstream of the DOG box motif would have mitigated against this to some extent. It would have been difficult to identify completely unknown, uncharacterised effectors if a related gene was not present in the initial gene list. Although those effectors which were uncharacterised existed in the initial list, overall, the setup of this pipeline would select for those already known in some capacity, making it hard to identify any truly novel genes. Another option if this work was to be repeated and expanded in future would include data from the other genus of cyst nematodes; *Heterodera*. Genome and transcriptome data for *Heterodera* species was not available at the beginning of this project and therefore were not incorporated into the pipeline. The genome of the soybean cyst nematode *H. glycines* and the transcriptome of the sugarcane cyst nematode *Heterodera sacchari* are now published and could be used as part of this study in future to identify core effector candidates in more syncytia-forming species (Masonbrink *et al.* 2019, Pokhare *et al.* 2020). Further to this, the genome of *Heterodera schachtii* is now available as a preprint on BioRxiv and should be published for use in future studies soon (Siddique *et al.* 2021).

The inclusion of *Nacobbus aberrans* in this study has presented a unique set of hurdles. As this species is more distantly related to the other nematodes analysed it was evident that there was a subset of effectors which were not present in *N. aberrans* that were present in *G. rostochiensis*, *G. pallida*, and *R. reniformis*. Ultimately it was decided not to pursue any of the effectors in this subset, however this may still be a valid starting point for analysis of effectors with conserved function in the formation of syncytia induced by cyst nematodes and *R. reniformis* in the future. Due to *N. aberrans* unique phylogenetic position it is possible that there are fundamental differences in its effector set. Potentially there are syncytial specific effectors present in other cyst nematodes that *N. aberrans* does not contain that may be substituted with other effectors found in RKN species. It is also possible that there are equivalent *N. aberrans* genes for those orthologous genes in this subset

however currently no genome data is available for this species. This means that it is not possible to confirm presence or absence of genes definitively, and it is not possible to check if predicted sequences from the transcriptome have the correct nucleotide and amino acid sequences as well as gene expression attributed.

Immediate future work that could be followed up from the work in this chapter is the reassessment of those genes not selected for functional characterisation in chapter six. As discussed, there were seven *G. rostochiensis* genes from candidate effector families which did not give conclusive *in situ* hybridisation results. This meant they could not be confirmed as effectors based on their mRNA location. If the ISH of these was reattempted with optimisation of the protocol to obtain clearer results, then some of these genes may also be interesting lines of study. One example of these genes is GROS_g05985. It was not possible to clone this gene and therefore no ISH was performed. After looking at the protein alignment available for this orthologous family (Gro_DOG_0149 family alignment, Sup. File 5) there appears to be a lot of sequence variation at both the beginning and end of these genes. This would indicate either misprediction of the stop codon or incorporation of an intron segment due to incorrect intron/exon boundary detection. Using BLAST similarity searches GROS_g05985 is similar to phosphatidylinositol 5-phosphate 4-kinases with the highest similarity being 49% identity with a phosphatidylinositol 5-phosphate 4-kinase type-2 alpha (PIP4K2A) from the tardigrade species *Hypsibius dujardini*. Most of the literature surrounding PIP4KA functions are from animal models surrounding insulin hormone signalling. If in future GROS_g05985 was shown to be a secreted effector it would be interesting to see it has an effect on host hormone signalling pathways.

GROS_g11017 and GROS_g11020 are both classed as pioneers and do not produce BLAST similarity hits to any other known proteins. Since the work to identify candidate core effectors in chapter three was conducted new data sets for the RKN species *Meloidogyne graminicola* and *M. enterolobii* have been released. A paper on

the genome of *M. graminicola* was published in September 2020 alongside nucleotide and amino acid sequences from *M. enterolobii* being uploaded as a direct submission to the NCBI protein database in August 2020 (Phan *et al.* 2020). Repeating the BLASTp search using GROS_g11017 and GROS_g11020 as queries at the time of writing this chapter reveals high percentage identity with “hypothetical protein Mgra_00005074” from *M. graminicola* and “unnamed protein product” *M. enterolobii*. There are other lower confidence hits (percentage ID less than 40%) with “unnamed protein product” from *Bursaphelenchus xylophilus*. This means that if the work in chapter three was to be repeated now GROS_g11017 and GROS_g11020 would no longer meet the criteria as being specific to syncytia-forming nematode species. This highlights the importance of continued genome and transcriptome study efforts as our base knowledge is still incomplete. Regardless of their change in status as candidate core effectors under the criteria of this study, these proteins would be worth further research as there is currently no known function attributed to this protein family.

7.5 – Arabinogalactan endo- β -1,4-galactanases of plant-parasitic nematodes

One core function of all PPN is the requirement to break down the plant cell wall. Arabinogalactan endo- β -1,4-galactanases are cell wall degrading enzymes from the GH53 family which are secreted by plant pathogens in order to loosen the host cell wall during infection. Chapter 4 describes the identification, cloning, and functional testing of three GH53 enzymes GrGAL1, GpGAL1, and RrGAL1 from *G. rostochiensis*, *G. pallida*, and *R. reniformis* respectively. Phylogenetic analysis indicates that these genes were most likely to have been obtained through a horizontal gene transfer event from an ancestral bacterial species. Like other plant cell wall degrading enzymes, arabinogalactan endo- β -1,4-galactanases are not present in the genomes of other animal species. Biochemical analysis confirmed that the nematode GH53 enzymes break down the substrate galactan. This shows that they do function as endo- β -1,4-galactanases. Unfortunately, it was not possible to obtain a high quality source of type I arabinogalactan for further biochemical analysis. However, galactan

makes up the main backbone of arabinogalactan so this result is significant. Obtaining AGI to repeat this experiment would be the top priority as further research from this chapter. This research adds to the many CWDE families already known to be produced by nematodes. In future it may be interesting to assess the speed/efficiency of substrate break down using a combination of different nematode CWDEs together. Few such studies on how effectors work in tandem have been described to date.

7.6 – Cathepsin L-like peptidases of plant-parasitic nematodes

Chapter five describes the identification and characterisation of a cathepsin L-like peptidase family present in *G. rostochiensis*, *G. pallida*, *R. reniformis*, and *H. glycines*. BLAST similarity searches using all identified cathepsin L-like peptidases discussed in this thesis returned no similar genes in *Nacobbus aberrans*. This BLAST search should be repeated on the genome data for *N. aberrans* that becomes available in the future as these are proteins with conserved functions that are very likely to be present in this nematode. The cathepsin L-like peptidase family contains genes in each species that can be divided into five defined clades. Despite high conservation between genes, each of the five clades present different gene expression patterns between pre-parasitic (J2) and post-parasitic (adult female) life stages. Significantly, the expression profiles within each clade are conserved across species, suggesting possible functional conservation as well. *In situ* hybridisation has shown that the five cathepsin L-like peptidases from *G. rostochiensis* (Gr-cpl-like-1 – Gr-cpl-like-5) as well as Gp-cpl-like-1 from *G. pallida* localise to the intestines of the nematode. RNA interference to silence Gr-cpl-like-2 did not lead to any significant changes in the nematodes ability to parasitise the host or any impact on cyst size, indicating there was no observable change in egg size or number. Interpreting these results is not straightforward and is discussed further in section 7.9. Future research in this area would be to determine the gene expression profile of the *H. glycines* cathepsin L-like peptidases (HGCP-I, HGCP-II, and Hgcp-like-III to Hgcp-like-V). This would give an insight into if the cathepsin L peptidases present in *Heterodera* species also follow

the same expression patterns observed in each clade. Further to this it would be good to establish if these peptidases follow the same phylogenetic and gene expression patterns in other more diverse species e.g., root-knot nematode species. A preliminary BLAST search shows that there are putative cathepsin L-like genes in *B. xylophilus* and *Meloidogyne* species such as *M. graminicola*, *M. enterolobii*, and *M. incognita*. The potential for using digestive proteinases for control of plant-parasitic nematodes has been demonstrated through the generation of transgenic plants expressing proteinase inhibitors. Such plants have been shown to provide control of a wide range of nematodes including cyst nematodes (including control in field studies), root knot nematodes, and migratory endoparasites in many different host plants (P. E. Urwin *et al.* 1997, Lilley *et al.* 2004, H. J. Atkinson *et al.* 2004).

7.7 – Functional characterisation of core effector genes

The final research chapter (chapter 6) of this thesis examined some of the confirmed core effectors that were identified in chapter 3 in more detail. This work focused on the functional characterisation of the effectors GROS_g05682, GROS_g02394, and GROS_g02469 through confocal microscopy, yeast two-hybrid analysis, co-immunoprecipitation, and hypersensitive response assays.

The GROS_g05682 subventral gland effector was shown to localise to the cytoplasm when infiltrated as a GFP-tagged fusion protein and it also did not appear to induce an HR response in *N. benthamiana* leaves. It was not possible to identify an interacting host protein of GROS_g05682 through yeast two-hybrid assays. Out of the three effectors analysed in this chapter the function of GROS_g05682 remains the most elusive.

The GROS_g02469 dorsal gland effector only induced a hypersensitive response from *N. benthamiana* in 17.4% (25/144) of infiltration sites tested. This figure was surprisingly low considering the subcellular localisation of this effector. GROS_g02469 was shown to localise to and disrupt the actin cytoskeleton. Upon

infiltration of GROS_g02469 into *N. benthamiana* leaf tissue, the native filaments of actin could be seen to form disordered aggregates. However, this interaction with the actin filaments has not yet been tested using actin binding assays. It was not possible to identify an interacting host protein for GROS_g02469 using Y2H assays. This could be due to it having a destructive enzymatic effect (e.g., hydrolysis) on its host target, meaning it may not be possible to pull out an intact target protein. It is also possible that whatever this effector interacts with is excluded from the nucleus, meaning the interaction would be missed in these assays (discussed in section 6.1.2.1). As previously mentioned, the actin binding assays may shed more light on the function of this effector regardless of the negative Y2H results.

GROS_g02394 showed the highest induction of the hypersensitive response with 28.5% (41/144) of the infiltration sites displaying cell death. Interestingly this effector was also shown to localise to the actin cytoskeleton, but unlike the disruption caused by GROS_g02469, GROS_g02394 appears to leave the actin intact. GROS_g02469 and GROS_g02394 were co-infiltrated into *N. benthamiana* leaves to assess their effects on actin in tandem. The two effectors co-localised to the actin as anticipated and the disruption to the native actin filaments occurring with the individual infiltration of GROS_g02469 was still observed during co-infiltration. This is a point of interest for future study as we now have two effectors specific to syncytia-forming nematodes confirmed to interact with the actin cytoskeleton at the same time, but they appear to have different functions. This would be a good place to start a multi-effector functional study and look at whether the lack of one impacts the function of the other. An interacting host protein, stPRMT1.1, was identified for the GROS_g02394 effector using Y2H assays. stPRMT1.1 is an arginine N-methyltransferase. When co-infiltrated it became clear that GROS_g02394 and stPRMT1.1 do not co-localise, but the presence of GROS_g02394 does appear to cause stPRMT1.1 to have a significant location change. The shift of stPRMT1.1 from being found consistently across the entirety of the cytoplasm to the extremities in the presence of GROS_g02394 is potentially significant but the reason for this shift

is yet to be determined. Unfortunately, due to time constraints the interaction between GROS_g02394 and stPRMT1.1 could not be confirmed outside of the yeast model using co-immunoprecipitation experiments. No conclusions could be made from the preliminary co-IP assays performed due to the lack of interaction detected between the positive control proteins. This suggests the presence of technical issues while carrying out this assay which need to be addressed in future repetitions. To confirm this interaction outside of the yeast model is the first experiment required in future to continue the study of this effector. Further work on the interaction between GROS_g02394 and stPRMT1.1 may prove useful. As stPRMT1.1 is an N-methyltransferase it could be shown to have a significant post-translational modification role which was not elucidated in this work. This could shed further light on how syncytia-forming nematodes can alter host gene expression. Many of the N-methyltransferases from mammalian species e.g., humans which stPRMT1.1 shares high sequence conservation with have been shown to methylate histones. Histone methylation can have multiple downstream effects, one of these being the switch from euchromatin to heterochromatin, resulting in changes in gene expression (Cedar & Bergman 2009).

7.8 – Computational identification and prediction of gene sequences

This thesis work has used multiple computational models to analyse both gene and protein data. This is invaluable in furthering our understanding of the complex nematode species being studied, however it is not without its pitfalls and limitations. As described in chapter 3, it is often the case that the *in silico* processes for identification of genes is not 100% accurate. It is possible for mispredictions of start and/or stop codons to be made, as well as errors in exon-intron boundary assignment to occur. Without manual checking of the predicted genes, it may cause issues with further analysis. For example, accurate primer design and subsequent gene cloning procedures may be made impossible if work is carried out with mispredicted gene sequences. After the identification of the 15 high confidence core effector families in chapter 3, cloning of the *G. rostochiensis* genes from each of

these families was attempted. As previously discussed in section 7.4, from preliminary observation of the Gro_DOG_0149 family alignment, genes in this family have highly variable C-terminal regions (candidate core effector family alignments - Sup. File 5). It is possible that the stop codon of GROS_g05985 is mispredicted in which case the primers used to clone this gene would not be fit for purpose. The same circumstances apply to GROS_09112 as the C-terminal regions of genes identified in the Gro_DOG_0169 family are also highly variable. Manual observation and correction of these sequences would be required before further work could continue with these candidates. Finally, it is unclear why it was not possible to clone GROS_g09671 as there are no immediate anomalies visible in the alignment of this gene with the other genes identified in the Gro_DOG_187 family.

Chapter 4 discussed the presence of GH53 arabinogalactan endo- β -1,4-galactanases from the nematode species *G. rostochiensis*, *G. pallida*, and *R. reniformis*. Both GpGAL1 and RrGAL1 had been mispredicted as multiple smaller sequences which meant the full length sequence of each gene had to be manually pieced back together before cloning could be attempted. If the initial mispredicted nucleotide sequences had been taken at face value and not further investigated then the results in this chapter would not have been possible. Although software employed to produce genome and transcriptome data is constantly improving, computational determination and prediction of genes should still be followed up by manual checks.

Being able to produce pipelines for identification of effectors has only been possible due to advances in bioinformatics. Many of the steps in such pipelines, such as signal peptide and transmembrane domain predictions, sequence similarity searches (BLAST), obtaining RNAseq reads to assess gene expression patterns, and motif identification and searches depend on the availability of robust software. However, taking the outputs of such pipelines at face value can be problematic. Originally the cathepsin L-like proteins discussed in chapter 5 were thought to be effector candidates. This was due to Gr-cpl-like-2 (GROS_g03615) being present in the initial

pipeline for effector identification as it met all of the pipeline criteria e.g., contains a signal peptide, lacks transmembrane domains, gene expression peaking at parasitic life stages etc. Gr-cpl-like-2 has since been shown by *in situ* hybridisation to localise to the intestines of the nematode. As effectors are known to only be produced in the oesophageal gland cells or the amphids it was clear from this result that Gr-cpl-like-2 is not an effector. This mistaken classification would have been carried through if manual testing had not been conducted. In a more general sense this demonstrates the importance of continued confirmation of effectors by experimentation using techniques like ISH despite improvements to the accuracy of computational effector predictions being made frequently.

One of the more significant developments in computational prediction of effectors is the identification of conserved motifs in the upstream promotor regions of these genes. The DOG box was identified upstream of 77% of 101 known dorsal gland effectors in *G. rostochiensis* (Eves-van den Akker & Birch 2016, Eves-Van den Akker *et al.* 2016a). This is a powerful new tool in the ability to identify new potential effectors and its discovery has led to a motif associated with potential effectors being found in *B. xylophilus* (Espada *et al.* 2018). A further motif has also been identified in *Pratylenchus penetrans*, however the genome for this species is incomplete so it cannot be fully verified yet (Vieira *et al.* 2018). The newest motif to be discovered is the Mel-DOG from root-knot nematode *Meloidogyne incognita* (Martine Da Rocha *et al.* 2021). These are steps in the direction of being able to predict effectors with high confidence based on nucleotide/amino acid sequence alone. However, the presence of the motif is not definitive proof that the gene in question is an effector. Any protein that needs to be produced at a high level in the dorsal gland may have the DOG box present and this may include proteins involved in gland cell function. In addition, the presence of a DOG box is not an absolute indicator of dorsal gland expression. For example, GROS_g05682 - the 20E03 effector has four DOG box motifs present in the 500 bp upstream promotor region, but when ISH was carried out GROS_g05682 was expressed in the subventral glands. Although GROS_g05682 has

still been shown to be an effector it is unclear why it would have so many DOG box motifs upstream if it is not produced in the dorsal gland. A theory that may explain this is if the DOG box regulated the timing of effector expression rather than the location where the effector is expressed from. It is already known that in general the effectors produced by the subventral glands are expressed at different life stages to those effectors produced by the dorsal gland. The other two effectors chosen for functional characterisation GROS_g02394 (GLAND11 effector) and GROS_g02469 (G23G11 effector) both have two DOG box motifs in their upstream promotor regions and have since been shown by ISH to localise to the dorsal gland. This further supports the point previously made that although computer technology is improving rapidly and is an indispensable part of effector protein identification, results from these pipelines must be followed up by experimentation. The presence of these upstream promotor motifs indicate the presence of an overarching transcription factor(s) involved in the expression of different, unrelated effector genes. Such transcription factors have yet to be identified, partly because there are limitations to our current understanding of nematode genomes. For example, nothing is known about the folding structure of nematode chromatin and how that impacts these unidentified transcription factors interacting with the promotor motifs (Eves-van den Akker 2021). With gaps in the knowledge on this front it is difficult to determine how regulatory elements play a role in the control of nematode gene expression.

7.9 – Future research avenues

There are multiple experiments that could be conducted to directly further the research in this thesis that has been discussed in sections 7.4 – 7.7. In depth studies of effectors and their interacting host proteins have been indispensable in furthering our understanding of how nematodes, and pathogens more generally, are so successful at infecting their hosts. Pathogen-host interactions are very often studied on a 1-to-1 basis which runs the risk of not giving a full picture of the processes occurring naturally *in vivo*. In future, steps could be made to study effector-host interactions in a wider context. This could be by assessing multiple different effectors

or in terms of larger families, multiple effector family members at once. Much research has been done on the stacking of resistance genes in host plants to make the resistance source robust and more difficult for pathogen to evolve to overcome it. Targeting of effectors could present the same opportunity for control from the opposite direction. It is highly likely that targeting a single essential effector as a control method would not be effective for long as the nematode species would evolve either a new effector or a way to compensate without it. It may be possible to target a group of essential effectors at once which would prevent the nematode from successfully parasitising the plant and would be complex for the nematode to adapt to quickly.

Taking the GH53 arabinogalactan endo- β -1,4-galactanases as an example; the cell wall presents the primary barrier for many nematode species to overcome in order to parasitise the host. This means it could be used as an avenue for control in future. A multigene targeting approach of nematode CWDEs may stop the nematode from being able to successfully penetrate the host plant and therefore reduce parasitism. It is unlikely that large reductions to parasitism would be seen if RNAi was applied to these GH53 genes on their own. This is because there is likely a level of functional redundancy and compensatory function from other CWDEs. It would be interesting to see what effect the tandem silencing of multiple types of CWDEs at the same time would have on parasitism. For example, knockdown of a cellulase, a pectate lyase, and a hemicellulase at the same time may present enough of a hurdle to prevent the nematode from successfully progressing through the cell wall. If proven successful this approach would require the approved use of genetically modified transgenic host plant lines in wider agricultural settings to provide a delivery system for such a method.

In chapter five it was shown that the knockdown of the cathepsin L-like peptidase Gr-cpl-like-2 had no significant impact on the ability of the *G. rostochiensis* to parasitise the host or on the size of cysts produced. The results of this individual

experiment do not necessarily mean that this has no effect on the worm, however. The phylogenetic analysis of these peptidases showed that they are actually part of a larger family which is split into five clades. Each member – Gr-cpl-like-1 to Gr-cpl-like-5 – shows different expression profiles and some have variations e.g., not containing signal peptides. It is possible that there is a level of functional redundancy being displayed here with one of the other members compensating for the loss of Gr-cpl-like-2 due to RNAi. It is possible that a different picture would be painted if all five cathepsin L peptidase members of *G. rostochiensis* had been silenced at the same time. This could also be done with other effector families if there were only a small number of proteins in the family.

As well as there being many effectors with yet undefined functions, there are still many other facets of effector biology and plant-nematode interactions in general that are yet to be elucidated. The understanding of how host plants recognise potential threats in their environments through PAMP-recognition receptor interactions is an example of an area where we have little knowledge in terms of plant-nematode interactions. A class of nematode pheromones called the ascarosides have a member, Ascr18, which has been shown to activate host defence responses which are usually attributed to PAMP recognition, such as MAPK cascades (Manosalva *et al.* 2015). In response *A. thaliana* has been shown to metabolise Ascr18 into two smaller ascaroside molecules through peroxisomal acyl-CoA oxidases. Upon the Ascr18 metabolism these smaller ascarosides reduce parasitism by acting as a repellent to nematodes in proximity (Manohar *et al.* 2020). The accompanying pattern recognition receptor which activates host defences in response to Ascr18 has not been identified yet. On the other hand, a pattern recognition receptor called NILR1 which recognises a nematode PAMP present in “NemaWater”, liquid in which J2 nematodes have been incubated and that contains secreted products of the nematodes, has been identified (Section 1.4.1.1) (Mendy *et al.* 2017). NILR1 functions as a co-receptor with BAK1 and initiates the plant defence response when treated with NemaWater. The specific PAMP which NILR1 recognises

is not currently known but it is not Ascr18. A future avenue for control may be explored in priming host plants with nematode PAMPs in order to raise their innate immune response, making it more difficult for nematodes to initially infect or establish within the host.

A large barrier for the study of plant-nematode interactions in the past has been the inability to reliably transform many plant-parasitic species, meaning that standard practice *in vivo* protocols used in other species such as *C. elegans* have been impossible to carry out. Many PPN species have a long life cycle with varied stages which make it difficult to apply genetic modification methods. The sex of many species including PCN is environmentally determined and the location of females inside of the host plant mean that the germline is inaccessible (Eves-van den Akker *et al.* 2021). Understanding plant-nematode interactions would shift greatly with access to transformation methods as experiments such as knockouts or expression of fluorescently tagged effectors becoming possible. For many years the study of effectors from PPN species has heavily relied on *in situ* hybridisation studies to show where mRNA is expressed, and RNAi to determine the phenotypes produced by gene silencing. Transformation of PPN would allow for improvements to be made in terms of localisation studies and a movement away from heavy reliance on reverse genetics. Current work is being conducted on achieving a reliable transformation method in PPN. There has been some success in the delivery of GFP mRNA into J2s of *H. schachtii* via soaking with liposomes (Kranse *et al.* 2021b). GFP was observed in the soaked worms using confocal microscopy. The transient expression of GFP was visible in the worms for up to 30 hours whereafter fluorescence returned to normal background levels. Although this is short term expression it represents a significant advance in the area of transformation. Additionally, Kranse *et al.* have successfully delivered macromolecules such as a fluorescent, membrane permeable DNA dye (Hoechst) into the male gonads of *H. schachtii* and *M. hapla* via microinjection. This is a prerequisite for transformation using the protocols developed for *C. elegans* and also used in animal parasitic species such as *Strongyloides* (Shao *et al.* 2017).

7.10 – Conclusions

The work in this thesis has allowed the identification of effectors conserved in syncytia-forming nematodes and has begun the process of functional characterisation of a subset of these. It provides a framework for future studies in this area as well as providing a series of potentially critical effectors that may be useful targets for control of plant-parasitic nematodes. In addition, they provide a route to developing our understanding of how syncytium forming nematodes are able to manipulate fundamentally important plant developmental processes as part of their infection biology.

8 – References

- ABAD, P., GOUZY, J., AURY, J.-M., CASTAGNONE-SERENO, P., DANCHIN, E.G.J., DELEURY, E., PERFUS-BARBEOCH, L., et al. 2008. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, **26**, 909–915, 10.1038/nbt.1482.
- ABAWI, G.S. & WIDMER, T.L. 2000. Impact of soil health management practices on soilborne pathogens, nematodes and root diseases of vegetable crops. *Applied Soil Ecology*, **15**, 37–47, 10.1016/S0929-1393(00)00070-6.
- ABU-GHARBIH, W. & AL-AZZEH, T. 2004. A checklist on nematode-plant associations in the Arab countries. *Arab Journal of Plant Protection*, **22**, 1–22.
- AGRICULTURE AND HORTICULTURE DEVELOPMENT BOARD (AHDB). 2021. AHDB applies for Emergency Authorisations after Vydate ban Available at: <https://ahdb.org.uk/news/ahdb-applies-for-emergency-authorisations-after-vydate-ban> [Accessed May 10, 2021].
- AHMAD, A. & CAO, X. 2012. Plant PRMTs Broaden the Scope of Arginine Methylation. *Journal of Genetics and Genomics*, **39**, 195–208, 10.1016/J.JGG.2012.04.001.
- AICH, P. & BISWAS, S. 2018. Highly Conserved Arg Residue of ERFNIN Motif of Pro-Domain is Important for pH-Induced Zymogen Activation Process in Cysteine Cathepsins K and L. *Cell Biochemistry and Biophysics*, **76**, 219–229, 10.1007/S12013-017-0838-X.
- ALI, S., MAGNE, M., CHEN, S., CÔTÉ, O., GERIČ STARE, B., OBRADOVIC, N., JAMSHAD, L., WANG, X., BÉLAIR, G. & MOFFETT, P. 2015. Analysis of Putative Apoplastic Effectors from the Nematode, *Globodera rostochiensis*, and Identification of an Expansin-Like Protein That Can Induce and Suppress Host Defenses. *PLOS ONE*, **10**, e0115042, 10.1371/journal.pone.0115042.
- ALMAGRO ARMENTEROS, J.J., SØNDERBY, C.K., SØNDERBY, S.K., NIELSEN, H. & WINTHER, O. 2017. DeepLoc: prediction of protein subcellular localization using deep learning Hancock, J., ed. *Bioinformatics*, **33**, 3387–3395,

10.1093/bioinformatics/btx431.

- ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. & LIPMAN, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410, 10.1016/S0022-2836(05)80360-2.
- ANDERSSON, J.O. 2005. Lateral gene transfer in eukaryotes Review Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences*, **62**, 1182–1197, 10.1007/s00018-005-4539-z.
- ANDOLFO, G. & ERCOLANO, M.R. 2015. Plant Innate Immunity Multicomponent Model. *Frontiers in Plant Science*, **6**, 987, 10.3389/fpls.2015.00987.
- ANNUNZIATO, A. 2008. DNA Packaging: Nucleosomes and Chromatin. *Nature Education*, **1**, 26.
- ARDÈVOL, A. & ROVIRA, C. 2015. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *Journal of the American Chemical Society*, **137**, 7528–7547, 10.1021/jacs.5b01156.
- ARTIMO, P., JONNALAGEDDA, M., ARNOLD, K., BARATIN, D., CSARDI, G., DE CASTRO, E., VERINE DUVAUD, S., et al. 2012. EXPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, **40**, W597–W603, 10.1093/nar/gks400.
- ATKINSON, H.J., BABBITT, P.C. & SAJID, M. 2009. The global cysteine peptidase landscape in parasites. *Trends in Parasitology*, **25**, 573–581, 10.1016/j.pt.2009.09.006.
- ATKINSON, H. J., GRIMWOOD, S., JOHNSTON, K. & GREEN, J. 2004. Prototype demonstration of transgenic resistance to the nematode *Radopholus similis* conferred on banana by a cystatin. *Transgenic Research*, **13**, 135–142, 10.1023/B:TRAG.0000026070.15253.88.
- ATKINSON, HOWARD J., GRIMWOOD, S., JOHNSTON, K. & GREEN, J. 2004. Prototype Demonstration of Transgenic Resistance to the Nematode *Radopholus similis* Conferred on Banana by a Cystatin. *Transgenic Research*, **13**, 135–142, 10.1023/B:TRAG.0000026070.15253.88.
- AVERY, L. & SHTONDA, B.B. 2003. Food transport in the *C. elegans* pharynx. *The*

- Journal of Experimental Biology*, **206**, 2441–2457, 10.1242/jeb.00433.
- BAIRWA, A., VENKATASALAM, E.P., SUDHA, R., UMAMAHESWARI, R. & SINGH, B.P. 2017. Techniques for characterization and eradication of potato cyst nematode: a review. *Journal of Parasitic Diseases*, **41**, 607–620, 10.1007/s12639-016-0873-3.
- BAKER, M. 2011. Making sense of chromatin states. *Nature Methods*, **8**, 717–722, 10.1038/nmeth.1673.
- BAKKER, E., ACHENBACH, U., BAKKER, J., VAN VLIET, J., PELEMAN, J., SEGERS, B., VAN DER HEIJDEN, S., et al. 2004. A high-resolution map of the H1 locus harbouring resistance to the potato cyst nematode *Globodera rostochiensis*. *Theoretical and Applied Genetics*, **109**, 146–152, 10.1007/s00122-004-1606-z.
- BARRETT, A.J. & RAWLINGS, N.D. 2014. Evolutionary Lines of Cysteine Peptidases. *Biological Chemistry*, **382**, 727–734, 10.1515/bchm.2001.382.5.727.
- BARRETT, A.J. & RAWLINGS, N.D. 1996. Families and clans of cysteine peptidases. *Perspectives in Drug Discovery and Design*, **6**, 1–11, 10.1007/BF02174042.
- BARTLEM, D.G., JONES, M.G.K. & HAMMES, U.Z. 2014. Vascularization and nutrient delivery at root-knot nematode feeding sites in host roots. *Journal of Experimental Botany*, **65**, 1789–1798, 10.1093/jxb/ert415.
- BEDFORD, M.T. & RICHARD, S. 2005. Arginine methylation: An emerging regulator of protein function. *Molecular Cell*, **18**, 263–272, 10.1016/j.molcel.2005.04.003.
- BÉRA-MAILLET, C., ARTHAUD, L., ABAD, P. & ROSSO, M.-N. 2000. Biochemical characterization of MI-ENG1, a family 5 endoglucanase secreted by the root-knot nematode *Meloidogyne incognita*. *European Journal of Biochemistry*, **267**, 3255–3263, 10.1046/j.1432-1327.2000.01356.x.
- BERMAN, H.M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T.N., WEISSIG, H., SHINDYALOV, I.N. & BOURNE, P.E. 2000. The Protein Data Bank. *Nucleic Acids Research*, **28**, 235–242.
- BILGRAMI, A. 2008. *Integrated Management and Biocontrol of Vegetable and Grain Crops Nematodes*. Ciancio, A. & Mukerji, K.G., eds. Springer Netherlands, 3–28 pp, 10.1007/978-1-4020-6063-2.

- BIRCH, P.R.J., REHMANY, A.P., PRITCHARD, L., KAMOUN, S. & BEYNON, J.L. 2006. Trafficking arms: oomycete effectors enter host plant cells. *Trends in Microbiology*, **14**, 8–11, 10.1016/J.TIM.2005.11.007.
- BLANC, R.S. & PHANE RICHARD, S. 2017. Arginine Methylation: The Coming of Age. *Molecular Cell*, **65**, 8–24, 10.1016/j.molcel.2016.11.003.
- BLANCHARD, A., FOUVILLE, D., ESQUIBET, M., MUGNIERY, D. & GRENIER, E. 2007. Sequence Polymorphism of 2 Pioneer Genes Expressed in Phytoparasitic Nematodes Showing Different Host Ranges. *Journal of Heredity*, **98**, 611–619, 10.1093/jhered/esm050.
- BLEVE-ZACHEO, T. & ZACHEO, G. 1983. Early stage of disease in fig roots induced by xiphinema index. *Nematologica*, **11**, 175–187.
- BLOK, V.C. & PHILLIPS, M.S. 2012. Biological characterisation of Globodera pallida from Idaho. *Nematology*, **14**, 817–826.
- BÖGER, M., HEKELAAR, J., VAN LEEUWEN, S.S., DIJKHUIZEN, L. & LAMMERTS VAN BUEREN, A. 2019. Structural and functional characterization of a family GH53 β -1,4-galactanase from Bacteroides thetaiotaomicron that facilitates degradation of prebiotic galactooligosaccharides. *Journal of Structural Biology*, **205**, 1–10, 10.1016/j.jsb.2018.12.002.
- BOHLMANN, H. & SOBCZAK, M. 2014. The plant cell wall in the feeding sites of cyst nematodes. *Frontiers in Plant Science*, **5**, 1–10, 10.3389/fpls.2014.00089.
- BOLGER, A.M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, **30**, 2114–2120, 10.1093/bioinformatics/btu170.
- BOLWELL, G.P., BINDSCHEDLER, L. V., BLEE, K.A., BUTT, V.S., DAVIES, D.R., GARDNER, S.L., GERRISH, C. & MINIBAYEVA, F. 2002. The apoplastic oxidative burst in response to biotic stress in plants: a three-component system. *Journal of Experimental Botany*, **53**, 1367–1376, 10.1093/jxb/53.372.1367.
- BRAITHWAITE, K.L., BARNA, T., SPURWAY, T.D., CHARNOCK, S.J., BLACK, G.W., HUGHES, N., LAKEY, J.H., et al. 1997. Evidence That Galactanase A from Pseudomonas fluorescens Subspecies cellulosa Is a Retaining Family 53 Glycosyl Hydrolase in

- Which E161 and E270 Are the Catalytic Residues. *Biochemistry*, **36**, 15489–15500, 10.1021/bi9712394.
- BRAND, U., FLETCHER, J.C., HOBE, M., MEYEROWITZ, E.M. & SIMON, R. 2000. Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. *Science*, **289**, 617–619, 10.1126/science.289.5479.617.
- BRENNER, S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics*, **77**, 71–94.
- BROWN, D.J.E., ROBERTSON, W.M., NEILSON, R., BEM, E. & ROBINSON, D.J. 1996. Characterization and vector relation of a serologically distinct isolate of tobacco rattle tobnavirus (TRV) transmitted by *Trichodorus similis* in northern Greece. Kluwer Academic Publishers, 1382 pp.
- BRÜCKNER, A., POLGE, C., LENTZE, N., AUERBACH, D. & SCHLATTNER, U. 2009. Yeast two-hybrid, a powerful tool for systems biology. *International Journal of Molecular Sciences*, **10**, 2763–2788, 10.3390/IJMS10062763.
- BURMEISTER, A.R. 2015. Horizontal Gene Transfer. *Evolution, Medicine and Public Health*, **2015**, 193–194, 10.1093/emph/eov018.
- BUTTLE, D.J. & MORT, J.S. 2013. Cysteine Proteases. In *Encyclopedia of Biological Chemistry: Second Edition*. Elsevier Inc., 589–592., 10.1016/B978-0-12-378630-2.00009-8.
- CAFFREY, C.R., GOUPIL, L., REBELLO, K.M., DALTON, J.P. & SMITH, D. 2018. Cysteine proteases as digestive enzymes in parasitic helminths. *PLOS Neglected Tropical Diseases*, **12**, e0005840, 10.1371/JOURNAL.PNTD.0005840.
- CAFFREY, C.R., SALTER, J.P., LUCAS, K.D., KHIEM, D., HSIEH, I., LIM, K.C., RUPPEL, A., MCKERROW, J.H. & SAJID, M. 2002. SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. *Molecular and biochemical parasitology*, **121**, 49–61, 10.1016/S0166-6851(02)00022-1.
- CARAPITO, R., VORWERK, S., JELTSCH, J.-M. & PHALIP, V. 2013. Genome-wide transcriptional responses of *Fusarium graminearum* to plant cell wall substrates. *FEMS Microbiology Letters*, **340**, 129–134, 10.1111/1574-6968.12079.
- CASTILLO, P. 2007. Host-Parasite Relationships in Fall-Sown Sugar Beets Infected by

- the Stem and Bulb Nematode, *Ditylenchus dipsaci*. *Plant Disease*, **91**, 71–79, 10.1094/PD-91-0071.
- CEDAR, H. & BERGMAN, Y. 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics*, **10**, 295–304, 10.1038/nrg2540.
- CHA, B. & JHO, E.H. 2012. Protein arginine methyltransferases (PRMTs) as therapeutic targets. *Expert Opinion on Therapeutic Targets*, **16**, 651–664, 10.1517/14728222.2012.688030.
- CHITWOOD, D.J. 2003. Nematicides. In *Encyclopedia of Agrochemicals*. New York: John Wiley & Sons, 1104–1115.
- CHRISTGAU, S., SANDAL, T., KOFOD, L.V. & DALBØGE, H. 1995. Expression cloning, purification and characterization of a β -1,4-galactanase from *Aspergillus aculeatus*. *Current Genetics*, **27**, 135–141, 10.1007/BF00313427.
- CHRONIS, D., CHEN, S., LU, S., HEWEZI, T., CARPENTER, S.C.D., LORIA, R., BAUM, T.J. & WANG, X. 2013. A ubiquitin carboxyl extension protein secreted from a plant-parasitic nematode *Globodera rostochiensis* is cleaved *in planta* to promote plant parasitism. *The Plant Journal*, **74**, 185–196, 10.1111/tpj.12125.
- CHUNG, T., WANG, D., KIM, C.S., YADEGARI, R. & LARKINS, B.A. 2009. Plant SMU-1 and SMU-2 homologues regulate Pre-mRNA splicing and multiple aspects of development. *Plant Physiology*, **151**, 1498–1512, 10.1104/pp.109.141705.
- CLARK, S.E., RUNNING, M.P. & MEYEROWITZ, E.M. 1995. CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development*, **121**, 2057–2967.
- COLL, N.S., EPPLE, P. & DANGL, J.L. 2011. Programmed cell death in the plant immune system. *Cell Death and Differentiation*, **18**, 1247–1256, 10.1038/cdd.2011.37.
- COOMBS, G.H. & MOTTRAM, J.C. 1997. Parasite proteinases and amino acid metabolism: possibilities for chemotherapeutic exploitation. *Parasitology*, **114**, 61–80, 10.1017/S003118209700111X.
- CORPET, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, **16**, 10881–10890, 10.1093/nar/16.22.10881.

- COSGROVE, D.J. 2017. Microbial Expansins. *Annual Review of Microbiology*, **71**, 479–497, 10.1146/annurev-micro-090816-093315.
- COTTON, J.A., LILLEY, C.J., JONES, L.M., KIKUCHI, T., REID, A.J., THORPE, P., TSAI, I.J., et al. 2014. The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology*, **15**, 1–17, 10.1186/gb-2014-15-3-r43.
- COULOMBE, R., GROCHULSKI, P., SIVARAMAN, J., MENARD, R., MORT, J.S. & CYGLER, M. 1996. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO Journal*, **15**, 5492–5503.
- CROW, B. 2017. Nematodes - How do I know if I have a problem? *Green Section Record*, **55**, 1–6.
- CROW, W.T. 2005. Alternatives to Fenamiphos for Management of Plant-Parasitic Nematodes on Bermudagrass 1. *Nematology*, **37**, 477–482.
- DA ROCHA, MARTINE, BOURNAUD, C., DAZENIÈRE, J., THORPE, P., BAILLY-BECHET, M., PELLEGRIN, C., PÉRÉ, A., et al. 2021. Genome Expression Dynamics Reveal the Parasitism Regulatory Landscape of the Root-Knot Nematode *Meloidogyne incognita* and a Promoter Motif Associated with Effector Genes. *Genes*, **12**, 771, 10.3390/GENES12050771.
- DA ROCHA, M., BOURNAUD, C., DAZENIERE, J., THORPE, P., PELLEGRIN, C., BAILLY-BECHET, M., PERE, A., et al. 2021. Genome expression dynamics reveals parasitism regulatory landscape of the root-knot nematode *Meloidogyne incognita* and a promoter motif associated with effector genes. *BioRxiv*, 10.1101/2021.04.02.438169.
- 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, 10.1073/pnas.1008486107.
- DANGL, J.L. & JONES, J.D.G. 2001. Plant pathogens and integrated defence responses to infection. *Nature*, **411**, 826–833, 10.1038/35081161.

- DAVIES, G. & HENRISSAT, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853–859, 10.1016/S0969-2126(01)00220-9.
- DAVIES, L.J., LILLEY, C.J., KNOX, P.J. & 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, 10.1111/j.1469-8137.2012.04238.x.
- DAVIS, E.L., HAEGEMAN, A. & KIKUCHI, T. 2011. Degradation of the plant cell wall by nematodes. In Jones, J., Gheysen, G. & Fenoll, C., eds. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. 255–273.
- DAVIS, E.L. & MACGUIDWIN, A.E. 2000. Lesion nematode disease. *The Plant Health Instructor*, 10.1094/PHI-I-2000-1030-02.
- DE ALMEIDA ENGLER, J., VAN POUCKE, K., KARIMI, M., DE GROODT, R., GHEYSEN, GREETJE, ENGLER, G. & GHEYSEN, GODELIEVE. 2004. Dynamic cytoskeleton rearrangements in giant cells and syncytia of nematode-infected roots. *The Plant Journal*, **38**, 12–26, 10.1111/j.1365-313X.2004.02019.x.
- DE BOER, J.M., DAVIS, E.L., HUSSEY, R.S., POPEIJUS, H., SMANT, G. & BAUM, T.J. 2002. Cloning of a Putative Pectate Lyase Gene Expressed in the Subventral Esophageal Glands of *Heterodera glycines*. *Journal of Nematology*, **34**, 9–11.
- DE DEUS BARBOSA, A.E.A., DA ROCHA FRAGOSO, R., DE LIMA E SOUZA, D. DOS S., FREIRE, É., DE OLIVEIRA NETO, O.B., VIANA, A.A.B., TOGAWA, R.C., et al. 2009. Differentially expressed genes in cotton plant genotypes infected with *Meloidogyne incognita*. *Plant Science*, **177**, 492–497, 10.1016/j.plantsci.2009.07.013.
- DE JONGE, R. 2012. In silico identification and characterization of effector catalogs. *Methods in Molecular Biology*, **835**, 415–425, 10.1007/978-1-61779-501-5_25.
- DECRAEMER, W. & GERAERT, E. 2006. Ectoparasitic nematodes. In Perry, R.N. & Moens, M., eds. *Plant Nematology*. CABI International, Wallingford, Oxon (CABI), 213–216.
- DION, C., CHAPPUIS, E. & RIPOLL, C. 2016. Does larch arabinogalactan enhance immune function? A review of mechanistic and clinical trials. *Nutrition and Metabolism*, **13**, 28, 10.1186/s12986-016-0086-x.

- DOI, R.H. & KOSUGI, A. 2004. Cellulosomes: Plant-cell-wall-degrading enzyme complexes. *Nature Reviews Microbiology*, **2**, 541–551, 10.1038/nrmicro925.
- DOYLE, E.A. & LAMBERT, K.N. 2002. Cloning and characterization of an esophageal-gland-specific pectate lyase from the root-knot nematode *Meloidogyne javanica*. *Molecular Plant-Microbe Interactions*, **15**, 549–556, 10.1094/MPMI.2002.15.6.549.
- DUAN, G. & WALTHER, D. 2015. The roles of post-translational modifications in the context of protein interaction networks. *PLoS computational biology*, **11**, e1004049, 10.1371/journal.pcbi.1004049.
- DUBREUIL, G., DELEURY, E., MAGLIANO, M., JAOUANNET, M., ABAD, P. & ROSSO, M.N. 2011. Peroxiredoxins from the plant parasitic root-knot nematode, *Meloidogyne incognita*, are required for successful development within the host. *International Journal for Parasitology*, **41**, 385–396, 10.1016/J.IJPARA.2010.10.008.
- DVORÁK, J., MASHIYAMA, S. T., SAJID, M., BRASCHI, S., DELCROIX, M., SCHNEIDER, E.L., MCKERROW, W.H., et al. 2009. SmCL3, a gastrodermal cysteine protease of the human blood fluke *Schistosoma mansoni*. *PLoS neglected tropical diseases*, **3**, 10.1371/JOURNAL.PNTD.0000449.
- DZIDIC, S. & BEDEKOVIC, V. 2003. Horizontal gene transfer-emerging multidrug resistance in hospital bacteria. *Acta Pharmacologica Sinica*, **24**, 519–526.
- ECOSPRAY LTD. 2017. *NEMguard PCN Granules: Safety data sheet*. Norfolk, 1–7 pp.
- EDGAR, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 1792–1797, 10.1093/nar/gkh340.
- ELLENBY, C. 1952. Resistance to the potato root eelworm, *Heterodera rostochiensis* Wollenweber. *Nature*, **170**, 1016, 10.1038/1701016a0.
- 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, 10.1016/j.ijpara.2007.03.012.
- EMI, S. & YAMAMOTO, T. 1972. Purification and Properties of Several Galactanases of *Bacillus subtilis* var. *amylosacchariticus*. *Agricultural and Biological Chemistry*,

- 36**, 1945–1954, 10.1271/bbb1961.36.1945.
- ENGLER, J. DE A., RODIUC, N., SMERTENKO, A. & ABAD, P. 2010. Plant actin cytoskeleton re-modeling by plant parasitic nematodes. *Plant Signaling and Behavior*, **5**, 213–217, 10.4161/psb.5.3.10741.
- EPPO. 2006. Testing of potato varieties to assess resistance to *Globodera rostochiensis* and *Globodera pallida*. *EPPO Bulletin*, **36**, 419–420, 10.1111/j.1365-2338.2006.01032.x.
- ESPADA, M., EVES-VAN DEN AKKER, S., MAIER, T., PARAMASIVAN, V., BAUM, T., MOTA, M. & JONES, J.T. 2018. STATAWAARS: a promoter motif associated with spatial expression in the major effector-producing tissues of the plant-parasitic nematode *Bursaphelenchus xylophilus*. *BMC genomics*, **19**, 10.1186/s12864-018-4908-2.
- EVANS, A. 2020. Biofumigation to suppress pests, weeds and diseases. *Scottish Rural Development Programme - Farm advice service*, 1–4.
- EVANS, K., FRANCO, J. & DE SCURRAH, M.M. 1975. Distribution of species of potato cyst-nematodes in South America. *Nematologica*, **21**, 365–369, 10.1163/187529275X00103.
- EVANS, K. & STONE, A.R. 1977. A Review of the Distribution and Biology of the Potato Cyst-Nematodes *Globodera rostochiensis* and *G. pallida*. *Tropical Pest Management*, **23**, 178–189, 10.1080/09670877709412426.
- EVES-VAN DEN AKKER, LILLEY, C., DANCHIN, E., RANCUREL, C., COCK, P., URWIN, P. & JONES, J. 2014. The transcriptome of *Nacobbus aberrans* reveals insights into the evolution of sedentary endoparasitism in plant-parasitic nematodes. *Genome biology and evolution*, **6**, 2181–2194, 10.1093/gbe/evu171.
- EVES-VAN DEN AKKER, S. 2021. Plant–nematode interactions. *Current Opinion in Plant Biology*, **62**, 102035, 10.1016/J.PBI.2021.102035.
- EVES-VAN DEN AKKER, S. & BIRCH, P.R.J. 2016. Opening the Effector Protein Toolbox for Plant-Parasitic Cyst Nematode Interactions. *Molecular Plant*, **9**, 1451–1453, 10.1016/j.molp.2016.09.008.
- EVES-VAN DEN AKKER, S., LAETSCH, D., THORPE, P., LILLEY, C., DANCHIN, E., DA ROCHA, M.,

- RANCUREL, C., et al. 2016a. The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology*, **17**, 124, 10.1186/s13059-016-0985-1.
- EVES-VAN DEN AKKER, S., LILLEY, C.J., JONES, J.T., URWIN, P.E., TRUDGILL, D.L., MEGEN, H. VAN, ELSEN, S. VAN DEN, et al. 2014. Identification and Characterisation of a Hyper-Variable Apoplastic Effector Gene Family of the Potato Cyst Nematodes Williams, D.L., ed. *PLoS Pathogens*, **10**, e1004391, 10.1371/journal.ppat.1004391.
- EVES-VAN DEN AKKER, S., LILLEY, C.J., YUSUP, H.B., JONES, J.T. & URWIN, P.E. 2016b. Functional C-TERMINALLY ENCODED PEPTIDE (CEP) plant hormone domains evolved de novo in the plant parasite *Rotylenchulus reniformis*. *Molecular plant pathology*, **17**, 1265–1275, 10.1111/mpp.12402.
- EVES-VAN DEN AKKER, S., STOJILKOVIĆ, B. & GHEYSEN, G. 2021. Recent applications of biotechnological approaches to elucidate the biology of plant–nematode interactions. *Current Opinion in Biotechnology*, **70**, 122–130, 10.1016/J.COPBIO.2021.03.008.
- FATT, H. V. & DOUGHERTY, E.C. 1963. Genetic Control of Differential Heat Tolerance in Two Strains of the Nematode *Caenorhabditis elegans*. *Science*, **141**, 266–267, 10.1126/science.141.3577.266.
- FEI, Q., ZHANG, Y., XIA, R. & MEYERS, B.C. 2016. Small RNAs Add Zing to the Zig-Zag-Zig Model of Plant Defenses. *Molecular Plant-Microbe Interactions*, **29**, 165–169, 10.1094/MPMI-09-15-0212-FI.
- FINETTI-SIALER, M. 1990. Histopathological changes induced by *Nacobbus aberrans* in resistant and susceptible potato roots- fdi:30525- Horizon. *Revue de Nématologie*, **13**, 155–160.
- FLOR, H.H. 1971. Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, **9**, 275–296.
- FOOT, M.A. 1978. Temperature Responses of Three Potato-Cyst Nematode Populations From New Zealand. *Nematologica*, **24**, 412–417.
- FOSU-NYARKO, J., NICOL, P., NAZ, F., GILL, R. & JONES, M.G.K. 2016. Analysis of the

- Transcriptome of the Infective Stage of the Beet Cyst Nematode, *H. schachtii*
Castagnone-Sereno, P., ed. *PLOS ONE*, **11**, e0147511,
10.1371/journal.pone.0147511.
- FUKUSHIGE, H. 1991. Propagation of *Bursaphelenchus xylophilus*
(Nematoda:Aphelenchoididae) on Fungi Growing in Pine-Shoot Segments.
Applied Entomology and Zoology, **26**, 371–376, 10.1303/aez.26.371.
- GALDIERO, MARILENA, GALDIERO, MASSIMILIANO, FINAMORE, E., ROSSANO, F., GAMBUZZA, M.,
CATANIA, M.R., TETI, G., MIDIRI, A. & MANCUSO, G. 2004. Haemophilus influenzae
porin induces Toll-like receptor 2-mediated cytokine production in human
monocytes and mouse macrophages. *Infection and Immunity*, **72**, 1204–1209,
10.1128/iai.72.2.1204-1209.2004.
- GAO, B., ALLEN, R., MAIER, T., DAVIS, E.L., BAUM, T.J. & HUSSEY, R.S. 2001. Identification
of Putative Parasitism Genes Expressed in the Esophageal Gland Cells of the
Soybean Cyst Nematode *Heterodera glycines*. *The American Phytopathological
Society*, **14**, 1247–1254, 10.1094/MPMI.2001.14.10.1247.
- GAO, B., 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, 10.1094/MPMI.2003.16.8.720.
- GARNA, H., MABON, N., WATHELET, B. & PAQUOT, M. 2004. New Method for a Two-Step
Hydrolysis and Chromatographic Analysis of Pectin Neutral Sugar Chains.
Journal of agricultural and food chemistry, **52**, 4652–4659, 10.1021/jf049647j.
- GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M.R., APPEL, R.D. &
BAIROCH, A. 2005. Protein Identification and Analysis Tools on the ExPASy
Server. *The Proteomics Protocols Handbook*, 571–607, 10.1385/1-59259-890-
0:571.
- GEBHARDT, C., MUGNIERY, D., RITTER, E., SALAMINI, E. & BONNEL, E. 1993. Identification of
RFLP markers closely linked to the HI gene conferring resistance to *Globodera
rostochiensis* in potato. Springer-Verlag, 541–544 pp.
- GEILFUS, C.M. 2017. The pH of the Apoplast: Dynamic Factor with Functional Impact
Under Stress. *Molecular Plant*, **10**, 1371–1386, 10.1016/j.molp.2017.09.018.

- GHEDIN, E., WANG, S., FOSTER, J.M. & SLATKO, B.E. 2004. First sequenced genome of a parasitic nematode. *Trends in Parasitology*, **20**, 151–153, 10.1016/j.pt.2004.01.011.
- GIBLIN-DAVIS, R.M., MAKINSON, J., CENTER, B.J., DAVIES, K.A., PURCELL, M., TAYLOR, G.S., SCHEFFER, S.J., GOOLSBY, J. & CENTER, T.D. 2001. Fergusobia/Fergusonina-induced Shoot Bud Gall Development on *Melaleuca quinquenervia*. *Journal of Nematology*, **33**, 239–247.
- GOETTIG, P. 2016. Effects of glycosylation on the enzymatic activity and mechanisms of proteases. *International Journal of Molecular Sciences*, **17**, 10.3390/IJMS17121969.
- GOVERSE, A., OVERMARS, H., ENGELBERTINK, J., SCHOTS, A., BAKKER, J. & HELDER, J. 2000. Both Induction and Morphogenesis of Cyst Nematode Feeding Cells Are Mediated by Auxin. *Molecular Plant-Microbe Interactions*, **13**, 1121–1129, 10.1094/MPMI.2000.13.10.1121.
- GRIFFITHS, B.S. & ROBERTSON, W.M. 1984. Nuclear changes induced by the nematode *Longidorus elongatus* in root-tips of ryegrass, *Lolium perenne*. *The Histochemical Journal*, **15**, 927–934, 10.1007/BF01011831.
- GROOTAERT, P. & SMALL, R.W. 1983. Observations On the Predation Abilities of Some Soil Dwelling Predatory Nematodes. *Nematologica*, **29**, 109–118.
- GROVES, M.R., COULOMBE, R., JENKINS, J. & CYGLER, M. 1998. Structural basis for specificity of papain-like cysteine protease proregions toward their cognate enzymes. *Proteins: Structure, function, and genetics*, **32**, 504–514.
- GRUDKOWSKA, M. & ZAGDAŃSKA, B. 2004. Multifunctional role of plant cysteine proteinases. *Acta Biochimica Polonica*, **51**, 609–624.
- GRUMMT, I. 2006. Actin and myosin as transcription factors. *Current Opinion in Genetics & Development*, **16**, 191–196.
- GRUNEWALD, W., CANNOOT, B., FRIML, J. & GHEYSEN, G. 2009. Parasitic nematodes modulate PIN-mediated auxin transport to facilitate infection Dangl, J.L., ed. *PLoS Pathogens*, **5**, 3–9, 10.1371/journal.ppat.1000266.
- GUHATHAKURTA, D., SCHRIEFER, L.A., WATERSTON, R.H. & STORMO, G.D. 2004. Novel

- transcription regulatory elements in *Caenorhabditis elegans* muscle genes. *Genome Research*, **14**, 2457–2468, 10.1101/gr.2961104.
- GUO, X., CHRONIS, D., DE LA TORRE, C.M., SMEDA, J., WANG, X. & MITCHUM, M.G. 2015. Enhanced resistance to soybean cyst nematode *Heterodera glycines* in transgenic soybean by silencing putative CLE receptors. *Plant Biotechnology Journal*, **13**, 801–810, 10.1111/pbi.12313.
- GUO, X., WANG, J., GARDNER, M., FUKUDA, H., KONDO, Y., ETCHELLS, J.P., WANG, X. & MITCHUM, M.G. 2017. Identification of cyst nematode B-type CLE peptides and modulation of the vascular stem cell pathway for feeding cell formation. *PLoS Pathogens*, **13**, 10.1371/journal.ppat.1006142.
- HAEGEMAN, A., JONES, J.T. & DANCHIN, E.G.J. 2011a. Horizontal Gene Transfer in Nematodes: A Catalyst for Plant Parasitism? *The American Phytopathological Society*, **24**, 879–887, 10.1094/MPMI-03-11-0055.
- HAEGEMAN, A., JONES, J.T. & DANCHIN, E.G.J. 2011b. Horizontal Gene Transfer in Nematodes: A Catalyst for Plant Parasitism? / 879 *MPMI*, **24**, 879–887, 10.1094/MPMI.
- 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, 10.1093/jxb/ers058.
- HANDOO, Z.A., CARTA, L.K., SKANTAR, A.M. & CHITWOOD, D.J. 2012. Description of *Globodera ellingtonae* n. sp. (Nematoda: Heteroderidae) from Oregon. *Journal of Nematology*, **44**, 40–57.
- HARRISON, B.D. 1968. Reactions of some old and new British potato cultivars to tobacco rattle virus. *European Potato Journal*, **11**, 165–176, 10.1007/BF02364420.
- HASHMI, S., BRITTON, C., LIU, J., GUILIANO, D.B., OKSOV, Y. & LUSTIGMAN, S. 2002. Cathepsin L is essential for embryogenesis and development of *Caenorhabditis elegans*. *Journal of Biological Chemistry*, **277**, 3477–3486,

10.1074/jbc.M106117200.

- HASNA, M.K., INSUNZA, V., LAGERLÖF, J. & RÄMERT, B. 2007. Food attraction and population growth of fungivorous nematodes with different fungi. *Annals of Applied Biology*, **151**, 175–182, 10.1111/j.1744-7348.2007.00163.x.
- HE, Y., ZHOU, J., SHAN, L. & MENG, X. 2018. Plant cell surface receptor-mediated signaling - A common theme amid diversity. *Journal of Cell Science*, **131**, 10.1242/jcs.209353.
- HENRISSAT, B. & BAIROCH, A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical Journal*, **293 (Pt 3)**, 781–788, 10.1042/bj2930781.
- HENRISSAT, B. & BAIROCH, A. 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal*, **316**, 695, 10.1042/BJ3160695.
- HEPHER, A. & ATKINSON, H.J. 1992. Nematode control with proteinase inhibitors.
- HERMANN, F., PABLY, P., ECKERICH, C., BEDFORD, M.T. & FACKELMAYER, F.O. 2009. Human protein arginine methyltransferases in vivo - Distinct properties of eight canonical members of the PRMT family. *Journal of Cell Science*, **122**, 667–677, 10.1242/jcs.039933.
- HERRMANN, F., LEE, J., BEDFORD, M.T. & FACKELMAYER, F.O. 2005. Dynamics of Human Protein Arginine Methyltransferase 1 (PRMT1) in Vivo . *Journal of biological chemistry*, **280**, 38005–38010, 10.1074/jbc.M502458200.
- HEWEZI, T., HOWE, P.J., MAIER, T.R., HUSSEY, R.S., MITCHUM, M.G., DAVIS, E.L. & BAUM, T.J. 2010. Arabidopsis spermidine synthase is targeted by an effector protein of the cyst nematode *Heterodera schachtii*. *Plant Physiology*, **152**, 968–984, 10.1104/pp.109.150557.
- HOANG, D.T., CHERNOMOR, O., VON HAESLER, A., MINH, B.Q. & VINH, L.S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**, 518–522, 10.1093/molbev/msx281.
- HOGENHOUT, S.A., VAN DER HOORN, R.A.L., TERAUCHI, R. & KAMOUN, S. 2009. Emerging Concepts in Effector Biology of Plant-Associated Organisms. *The American Phytopathological Society*, **22**, 115–122, 10.1094/MPMI-22-2-0115.

- HOLTERMAN, M., KAREGAR, A., MOOIJMAN, P., VAN MEGEN, H., VAN DEN ELSEN, S., VERVOORT, M.T.W., QUIST, C.W., et al. 2017. Disparate gain and loss of parasitic abilities among nematode lineages. *PLoS ONE*, 10.1371/journal.pone.0185445.
- HOLTERMAN, M., KARSSSEN, G., VAN DEN ELSEN, S., VAN MEGEN, H., BAKKER, J. & HELDER, J. 2009. Small Subunit rDNA-Based Phylogeny of the Tylenchida Sheds Light on Relationships Among Some High-Impact Plant-Parasitic Nematodes and the Evolution of Plant Feeding. **99**.
- HOLTERMAN, M., VAN DER WURFF, A., VAN DEN ELSEN, S., VAN MEGEN, H., BONGERS, T., HOLOVACHOV, O., BAKKER, J. & HELDER, J. 2006. Phylum-Wide Analysis of SSU rDNA Reveals Deep Phylogenetic Relationships among Nematodes and Accelerated Evolution toward Crown Clades. *Molecular biology and evolution*, **23**, 1792–1800, 10.1093/molbev/msl044.
- HORTON, P., PARK, K.-J., OBAYASHI, T., FUJITA, N., HARADA, H., ADAMS-COLLIER, C.J. & NAKAI, K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, **35**, W585–W587, 10.1093/nar/gkm259.
- HUANG, G., DONG, R., ALLEN, R., DAVIS, E.L., BAUM, T.J. & HUSSEY, R.S. 2006. A Root-Knot Nematode Secretory Peptide Functions as a Ligand for a Plant Transcription Factor. *Molecular Plant-Microbe Interactions*, **19**, 463–470.
- HUANG, G., DONG, R., ALLEN, R., DAVIS, E.L., BAUM, T.J. & HUSSEY, R.S. 2005. Developmental expression and molecular analysis of two *Meloidogyne incognita* pectate lyase genes. *International Journal for Parasitology*, **35**, 685–692, 10.1016/j.ijpara.2005.01.006.
- HUANG, X., XU, C.-L., YANG, S.-H., LI, J.-Y., WANG, H.-L., ZHANG, Z.-X., CHEN, C. & XIE, H. 2019. Life-stage specific transcriptomes of a migratory endoparasitic plant nematode, *Radopholus similis* elucidate a different parasitic and life strategy of plant parasitic nematodes. *Scientific Reports*, **9**, 1–11, 10.1038/s41598-019-42724-7.
- HUSNIK, F. & McCUTCHEON, J.P. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology*, **16**, 67–79, 10.1038/nrmicro.2017.137.

- HUSSEY, R.S. 1989. Disease-Inducing Secretions of Plant-Parasitic Nematodes. *Annual Review of Phytopathology*, **27**, 123–141, 10.1146/annurev.py.27.090189.001011.
- IBM CORP. 2017. IBM SPSS statistics for Windows.
- IYAMA, K., LAM THI BACH TUYET & STONE, B.A. 1994. Covalent cross-links in the cell wall. *Plant Physiology*, **104**, 315–320, 10.1104/pp.104.2.315.
- INSERRA, R.N., DUNN, R.A. & VOVLAS, N. 1994. Host Response of Ornamental Palms to *Rotylenchulus reniformis*, 737–743 pp.
- J. BRYAN, G., MCLEAN, K., PANDE, B., PURVIS, A., A. HACKETT, C., E. BRADSHAW, J. & WAUGH, R. 2004. Genetical dissection of H3-mediated polygenic PCN resistance in a heterozygous autotetraploid potato population. *Molecular Breeding*, **14**, 105–116, 10.1023/B:MOLB.0000037999.13581.9c.
- JAGDALE, G.B., CASEY, M.L., GREWAL, P.S. & LINDQUIST, R.K. 2004. Application rate and timing, potting medium, and host plant effects on the efficacy of *Steinernema feltiae* against the fungus gnat, *Bradysia coprophila*, in floriculture. *Biological Control*, **29**, 296–305, 10.1016/S1049-9644(03)00164-6.
- JAOUANNET, M., MAGLIANO, M., ARGUEL, M.J., GOURGUES, M., EVANGELISTI, E., ABAD, P. & ROSSO, M.N. 2013. The Root-Knot Nematode Calreticulin Mi-CRT Is a Key Effector in Plant Defense Suppression. / *97 MPMI*, **26**, 97–105, 10.1094/MPMI.
- 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, 10.1016/S0014-5793(02)02906-X.
- JEGER, M., BRAGARD, C., CAFFIER, D., CANDRESSE, T., CHATZIVASSILIOU, E., DEHNEN-SCHMUTZ, K., GILIOLI, G., et al. 2018. Pest categorisation of *Nacobbus aberrans*. *EFSA Journal*, **16**, 5249, 10.2903/j.efsa.2018.5249.
- JIA, Y., MCADAMS, S.A., BRYAN, G.T., HERSHEY, H.P. & VALENT, B. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO Journal*, **19**, 4004–4014, 10.1093/emboj/19.15.4004.
- JONES, J.D.G. & DANGL, J.L. 2006. The plant immune system. *Nature*, **444**, 323–329, 10.1038/nature05286.

- JONES, J.T., HAEGEMAN, A., DANCHIN, E.G.J., GAUR, H.S., HELDER, J., JONES, M.G.K., KIKUCHI, T., et al. 2013. Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, **14**, 946–961, 10.1111/mpp.12057.
- JONES, J.T., KUMAR, A., PYLYPENKO, L.A., THIRUGNANASAMBANDAM, A., CASTELLI, L., CHAPMAN, S., COCK, P.J.A., et al. 2009. Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology*, **10**, 815–828, 10.1111/j.1364-3703.2009.00585.x.
- JONES, M.G.K. 1981a. *B. Rotylenchulus reniformis*. In Zuckerman, B.M. & Rohde, R.A., eds. *Plant Parasitic Nematodes*. London: Academic press, 261.
- JONES, M.G.K. 1981b. Host cell responses to endoparasitic nematode attack: structure and function of giant cells and syncytia. *Annals of Applied Biology*, **97**, 353–372, 10.1111/j.1744-7348.1981.tb05122.x.
- JONES, M.G.K. & NORTHCOTE, D.H. 1972. Nematode-induced syncytium—a multinucleate transfer cell. *Journal of Cell Science*, **10**, 789–809.
- JONES, M.G.K. & PAYNE, H.L. 1977. The structure of syncytia induced by the phytoparasitic nematode *Nacobbus aberrans* in tomato roots, and the possible role of plasmodesmata in their nutrition, 299 pp.
- KALYAANAMOORTHY, S., MINH, B.Q., WONG, T.K.F., VON HAESELER, A. & JERMIIN, L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587–589, 10.1038/nmeth.4285.
- KANEHISA, M., GOTO, S., KAWASHIMA, S. & NAKAYA, A. 2002. The KEGG databases at GenomeNet. *Nucleic Acids Research*, **30**, 42–46, 10.1093/nar/30.1.42.
- KANG, Y., JELENSKA, J., CECCHINI, N.M., LI, Y., LEE, M.W., KOVAR, D.R. & GREENBERG, J.T. 2014. HopW1 from *Pseudomonas syringae* Disrupts the Actin Cytoskeleton to Promote Virulence in *Arabidopsis*. He, S., ed. *PLoS Pathogens*, **10**, e1004232, 10.1371/journal.ppat.1004232.
- KANYUKA, K. & RUDD, J.J. 2019. Cell surface immune receptors: the guardians of the plant's extracellular spaces. *Current Opinion in Plant Biology*, **50**, 1–8, 10.1016/j.pbi.2019.02.005.

- KARRER, K.M., PEIFFERT, S.L. & DITOMAS, M.E. 1993. Two distinct gene subfamilies within the family of cysteine protease genes (tetrahymena/propeptide/cathepsin). *Biochemistry*, **90**, 3063–3067.
- KAY, S., HAHN, S., MAROIS, E., HAUSE, G. & BONAS, U. 2007. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science*, **318**, 648–651, 10.1126/science.1144956.
- KELLEY, L.A., MEZULIS, S., YATES, C.M., WASS, M.N. & STERNBERG, M.J.E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, **10**, 845–858, 10.1038/nprot.2015.053.
- KNOCH, E., DILOKPIMOL, A. & GESHI, N. 2014. Arabinogalactan proteins: Focus on carbohydrate active enzymes. *Frontiers in Plant Science*, **5**, 10.3389/fpls.2014.00198.
- KNOETZE, R., SWART, A. & TIEDT, L. 2013. Description of *Globodera capensis* n. sp. (Nematoda: Heteroderidae) from South Africa. *Nematology*, **15**, 233–250.
- KNOETZE, R., SWART, A., WENTZEL, R. & TIEDT, L.R. 2017. Description of *Globodera agulhasensis* n. sp. (Nematoda: Heteroderidae) from South Africa. *Nematology*, **19**, 805–816, 10.1163/15685411-00003087.
- KODAMA, Y. & HU, C.D. 2012. Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives. *BioTechniques*, **53**, 285–298, 10.2144/000113943.
- KRANSE, O., BEASLEY, H., ADAMS, S., PIRES-DA SILVA, A., BELL, C., LILLEY, C.J., URWIN, P.E., et al. 2021a. Toward genetic modification of plant-parasitic nematodes: Delivery of macromolecules to adults and expression of exogenous mRNA in second stage juveniles. *G3: Genes, Genomes, Genetics*, **11**, 10.1093/g3journal/jkaa058.
- KRANSE, O., BEASLEY, H., ADAMS, S., PIRES-DASILVA, A., BELL, C., LILLEY, C.J., URWIN, P.E., et al. 2021b. Toward genetic modification of plant-parasitic nematodes: delivery of macromolecules to adults and expression of exogenous mRNA in second stage juveniles. *G3 Genes, Genomes, Genetics*, **11**, 10.1093/G3JOURNAL/JKAA058.

- KRAUSE, C.D., YANG, Z.-H., KIM, Y.-S., LEE, J.-H., COOK, J.R. & PESTKA, S. 2007. Protein arginine methyltransferases: Evolution and assessment of their pharmacological and therapeutic potential. *Pharmacology & Therapeutics*, **113**, 50–87, 10.1016/j.pharmthera.2006.06.007.
- KROGH, A., LARSSON, B., VON HEIJNE, G. & SONNHAMMER, E.L.. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, **305**, 567–580, 10.1006/jmbi.2000.4315.
- KUDLA, U., MILAC, A., QIN, L., OVERMARS, H., ROZE, E., HOLTERMAN, M., PETRESCU, A., et al. 2007. Structural and functional characterization of a novel, host penetration-related pectate lyase from the potato cyst nematode *Globodera rostochiensis*. *Molecular Plant Pathology*, **8**, 293–305, 10.1111/j.1364-3703.2007.00394.x.
- LANGMEAD, B. & SALZBERG, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359, 10.1038/nmeth.1923.
- LARKIN, M.A., BLACKSHIELDS, G., BROWN, N.P., CHENNA, R., MCGETTIGAN, P.A., MCWILLIAM, H., VALENTIN, F., et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948, 10.1093/bioinformatics/btm404.
- LAWRENCE, K.S., PRICE, A.J., LAWRENCE, G.W., JONES, J.R. & AKRIDGE, J.R. 2008. Weed Hosts For *Rotylenchulus Reniformis* In Cotton Fields Rotated With Corn In The Southeast Of The United States. *Nematropica*, **38**, 13–22.
- LE NOURS, J., RYTTERSGAARD, C., LO LEGGIO, L., ØSTERGAARD, P.R., BORCHERT, T.V., CHRISTENSEN, L.L.H. & LARSEN, S. 2003. Structure of two fungal β -1,4-galactanases: Searching for the basis for temperature and pH optimum. *Protein Science*, **12**, 1195–1204, 10.1110/ps.0300103.
- 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, 10.1104/pp.110.167197.
- LEELARASAMEE, N., ZHANG, L. & GLEASON, C. 2018. The root-knot nematode effector

- MiPFN3 disrupts plant actin filaments and promotes parasitism Kaloshian, I., *ed. PLOS Pathogens*, **14**, e1006947, 10.1371/journal.ppat.1006947.
- LEWIS, K.A., TZILIVAKIS, J., WARNER, D.J. & GREEN, A. 2016. An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment*, **22**, 1050–1064, 10.1080/10807039.2015.1133242.
- LI, H., HANDSAKER, B., WYSOKER, A., FENNEL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & 1000 GENOME PROJECT DATA PROCESSING SUBGROUP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079, 10.1093/bioinformatics/btp352.
- LILLEY, C.J., URWIN, P.E., JOHNSTON, K.A. & ATKINSON, H.J. 2004. Preferential expression of a plant cystatin at nematode feeding sites confers resistance to *Meloidogyne incognita* and *Globodera pallida*. *Plant Biotechnology Journal*, **2**, 3–12, 10.1046/J.1467-7652.2003.00037.X.
- LIN, K., ZHANG, F., ZHOU, S., LIU, W., GAN, J. & PAN, Z. 2007. Stereoisomeric separation and toxicity of the nematicide fosthiazate. *Environmental Toxicology and Chemistry*, **26**, 2339–2344, 10.1897/07-255R.1.
- LIN, Y., BAO, B., YIN, H., WANG, X., FENG, A., ZHAO, L., NIE, X., YANG, N., SHI, G.-P. & LIU, J. 2019. Peripheral cathepsin L inhibition induces fat loss in *C. elegans* and mice through promoting central serotonin synthesis. *BMC Biology*, **17**, 1–19, 10.1186/S12915-019-0719-4.
- LIVAJA, M., ZEIDLER, D., VON RAD, U. & DURNER, J. 2008. Transcriptional responses of *Arabidopsis thaliana* to the bacteria-derived PAMPs harpin and lipopolysaccharide. *Immunobiology*, **213**, 161–171, 10.1016/j.imbio.2007.10.004.
- LODISH, H., BERK, A., ZIPURSKY, S.L., MATSUDAIRA, P., BALTIMORE, D. & DARNELL, J. 2000. Hierarchical Structure of Proteins. In *Molecular Cell Biology*. W. H. Freeman.
- LOMBARD, V., GOLACONDA RAMULU, H., DRULA, E., COUTINHO, P.M. & HENRISSAT, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, **42**, D490–D495, 10.1093/nar/gkt1178.
- LORD, J.S., LAZZERI, L., ATKINSON, H.J. & URWIN, P.E. 2011. Biofumigation for control of

- pale potato cyst nematodes: Activity of brassica leaf extracts and green manures on *Globodera pallida* in vitro and in soil. *Journal of Agricultural and Food Chemistry*, **59**, 7882–7890, 10.1021/JF200925K.
- LOZANO-TORRES, J.L., WILBERS, R.H.P., GAWRONSKI, P., BOSHOVEN, J.C., FINKERS-TOMCZAK, A., CORDEWENER, J.H.G., AMERICA, A.H.P., et al. 2012. Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 10119–10124, 10.1073/pnas.1202867109.
- LOZANO-TORRES, J.L., WILBERS, R.H.P.P., WARMERDAM, S., FINKERS-TOMCZAK, A., DIAZ-GRANADOS, A., VAN SCHAİK, C.C., HELDER, J., et al. 2014. Apoplastic Venom Allergen-like Proteins of Cyst Nematodes Modulate the Activation of Basal Plant Innate Immunity by Cell Surface Receptors Kaloshian, I., ed. *PLoS Pathogens*, **10**, e1004569, 10.1371/journal.ppat.1004569.
- LYNCH, M. & CONERY, J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science*, **290**, 1151–1155, 10.1126/science.290.5494.1151.
- MADDEN, T. 2003. The BLAST Sequence Analysis Tool. *In The NCBI handbook*. National Center for Biotechnology Information (US), 281–296.
- MADHAIYAN, M., POONGUZHALI, S., SARAVANAN, V.S., HARI, K., LEE, K.C. & LEE, J.S. 2013. *Duganella sacchari* sp. nov. and *Duganella radiceis* sp. nov., two novel species isolated from rhizosphere of field-grown sugar cane. *International Journal of Systematic and Evolutionary Microbiology*, **63**, 1126–1131, 10.1099/ij.s.0.040584-0.
- MAIER, T.R., HEWEZI, T., PENG, J. & BAUM, T.J. 2013. Isolation of Whole Esophageal Gland Cells from Plant-Parasitic Nematodes for Transcriptome Analyses and Effector Identification. *Molecular Plant-Microbe Interactions*, **26**, 31–35, 10.1094/MPMI-05-12-0121-FI.
- MALAGON, D., BENITEZ, R., KASNY, M. & J., A.F. 2013. Peptidases in Parasitic Nematodes: A Review. *In Parasites*. 61–102.
- MANOHAR, M., TENJO-CASTANO, F., CHEN, S., ZHANG, Y.K., KUMARI, A., WILLIAMSON, V.M.,

- WANG, X., KLESSIG, D.F. & SCHROEDER, F.C. 2020. Plant metabolism of nematode pheromones mediates plant-nematode interactions. *Nature Communications*, **11**, 1–11, 10.1038/s41467-019-14104-2.
- MANOSALVA, P., MANOHAR, M., VON REUSS, S.H., CHEN, S., KOCH, A., KAPLAN, F., CHOE, A., et al. 2015. Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. *Nature Communications*, **6**, 1–8, 10.1038/ncomms8795.
- MANZANILLA-LÓPEZ, R.H., COSTILLA, M.A., DOUCET, M., FRANCO, J., INSERRA, R.N., LEHMAN, P.S., CID DEL PRADO-VERA, I., SOUZA, R.M. & EVANS, K. 2002. The genus *Nacobbus* thorne & allen, 1944 (Nematoda: Pratylenchidae): systematics, distribution, biology and management resumen. *Nematropica*, **32**, 149–227.
- MARCHLER-BAUER, A., DERBYSHIRE, M.K., GONZALES, N.R., LU, S., CHITSAZ, F., GEER, L.Y., GEER, R.C., et al. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research*, **43**, D222–D226, 10.1093/nar/gku1221.
- MARHAVÝ, P., KURENDA, A., SIDDIQUE, S., DÉNERVAUD TENDON, V., ZHOU, F., HOLBEIN, J., HASAN, M.S., GRUNDLER, F.M., FARMER, E.E. & GELDNER, N. 2019. Single-cell damage elicits regional, nematode-restricting ethylene responses in roots. *The EMBO Journal*, **38**, 10.15252/embj.2018100972.
- MASLER, E.P. & PERRY, R.N. 2018. 3.3.1 Diapause. In Perry, R.N., Moens, M. & Jones, J.T., eds. *Cyst Nematodes*. 20–22.
- MASONBRINK, R., MAIER, T.R., MUPPIRALA, U., SEETHARAM, A.S., LORD, E., JUVALE, P.S., SCHMUTZ, J., et al. 2019. The genome of the soybean cyst nematode (*Heterodera glycines*) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC Genomics*, **20**, 119, 10.1186/s12864-019-5485-8.
- MBURU, H., CORTADA, L., HAUKELAND, S., RONNO, W., NYONGESA, M., KINYUA, Z., BARGUL, J.L. & COYNE, D. 2020. Potato Cyst Nematodes: A New Threat to Potato Production in East Africa. *Frontiers in Plant Science*, **11**, 670, 10.3389/fpls.2020.00670.
- MCCARTHY, D.J., CHEN, Y. & SMYTH, G.K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic*

- Acids Research*, **40**, 4288–4297, 10.1093/nar/gks042.
- MCDERMOTT, J.E., CORRIGAN, A., PETERSON, E., OEHMEN, C., NIEMANN, G., CAMBRONNE, E.D., SHARP, D., ADKINS, J.N., SAMUDRALA, R. & HEFFRON, F. 2011. Computational prediction of type III and IV secreted effectors in gram-negative bacteria. *Infection and Immunity*, **79**, 23–32, 10.1128/IAI.00537-10.
- MCLELLAN, H., BOEVINK, P.C., ARMSTRONG, M.R., PRITCHARD, L., GOMEZ, S., MORALES, J., WHISSON, S.C., BEYNON, J.L. & BIRCH, P.R.J. 2013. An RxLR Effector from *Phytophthora infestans* Prevents Re-localisation of Two Plant NAC Transcription Factors from the Endoplasmic Reticulum to the Nucleus Tyler, B., ed. *PLoS Pathogens*, **9**, e1003670, 10.1371/journal.ppat.1003670.
- MCLELLAN, H., CHEN, K., HE, Q., WU, X., BOEVINK, P., TIAN, Z. & BIRCH, P.R.J. 2020. The Ubiquitin E3 Ligase PUB17 Positively Regulates Immunity by Targeting a Negative Regulator, KH17, for Degradation. *Plant communications*, **1**, 10.1016/J.XPLC.2020.100020.
- MCNICHOLAS, S., POTTERTON, E., WILSON, K.S., NOBLE, M.E.M. & IUCR. 2011. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography*, **67**, 386–394, 10.1107/S0907444911007281.
- MEI, Y., THORPE, P., GUZHA, A., HAEGEMAN, A., BLOK, V.C., MACKENZIE, K., GHEYSEN, G., JONES, J.T. & MANTELIN, S. 2015. Only a small subset of the SPRY domain gene family in *Globodera pallida* is likely to encode effectors, two of which suppress host defences induced by the potato resistance gene *Gpa2*. *Nematology*, **17**, 409–424, 10.1163/15685411-00002875.
- MEI, Y., WRIGHT, K.M., HAEGEMAN, A., BAUTERS, L., DIAZ-GRANADOS, A., GOVERSE, A., GHEYSEN, G., JONES, J.T. & MANTELIN, S. 2018. The *Globodera pallida* SPRYSEC Effector GpSPRY-414-2 That Suppresses Plant Defenses Targets a Regulatory Component of the Dynamic Microtubule Network. *Frontiers in Plant Science*, **9**, 1–16, 10.3389/fpls.2018.01019.
- MELILLO, M.T., LEONETTI, P., LEONE, A., VERONICO, P. & BLEVE-ZACHEO, T. 2011. ROS and NO production in compatible and incompatible tomato-Meloidogyne incognita

- interactions. *European Journal of Plant Pathology*, **130**, 489–502, 10.1007/s10658-011-9768-4.
- MELKMAN, T. & SENGUPTA, P. 2004. The worm's sense of smell: Development of functional diversity in the chemosensory system of *Caenorhabditis elegans*. *Developmental Biology*, **265**, 302–319, 10.1016/j.ydbio.2003.07.005.
- MENDY, B., WANG'OMBE, M.W., RADAKOVIC, Z.S., HOLBEIN, J., ILYAS, M., CHOPRA, D., HOLTON, N., ZIPFEL, C., GRUNDLER, F.M.W. & SIDDIQUE, S. 2017. Arabidopsis leucine-rich repeat receptor-like kinase NILR1 is required for induction of innate immunity to parasitic nematodes Mackey, D., ed. *PLOS Pathogens*, **13**, e1006284, 10.1371/journal.ppat.1006284.
- MITCHUM, M.G., HUSSEY, R.S., BAUM, T.J., WANG, X., ELLING, A.A., WUBBEN, M. & DAVIS, E.L. 2013. Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist*, **199**, 879–894, 10.1111/nph.12323.
- MOLIN, W.T. & STETINA, S.R. 2016. Weed hosts and relative weed and cover crop susceptibility to *Rotylenchulus reniformis* in the Mississippi delta. *Nematropica*, **46**, 121–131.
- MORGAN, A.A. & RUBENSTEIN, E. 2013. Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome Casarini, D.E., ed. *PLoS ONE*, **8**, e53785, 10.1371/journal.pone.0053785.
- MORRISON, M., POPE, P.B., DENMAN, S.E. & MCSWEENEY, C.S. 2009. Plant biomass degradation by gut microbiomes: more of the same or something new? *Current Opinion in Biotechnology*, **20**, 358–363, 10.1016/j.copbio.2009.05.004.
- MOTA, M.M. & EISENBACK, J.D. 1993. Morphology of Second-stage Juveniles and Males of *Globodera tabacum tabacum*, *G. t. virginiae*, and *G. t. solanacearum* (Nemata: Heteroderinae). *Journal of Nematology*, **25**, 27–33.
- MUGNIERY, D. 1978. Vitesse de développement, en fonction de la température, de *Globodera rostochiensis* et *G. pallida* (Nematoda : Heteroderidae). *Revue de Nématologie*, **1**, 3–12.
- MURN, J. & SHI, Y. 2017. The winding path of protein methylation research:

- Milestones and new frontiers. *Nature Reviews Molecular Cell Biology*, **18**, 517–527, 10.1038/nrm.2017.35.
- MUSYOKA, T.M., NJUGUNA, J.N. & BISHOP, Ö.T. 2019. Comparing sequence and structure of falcipains and human homologs at prodomain and catalytic active site for malarial peptide based inhibitor design. *Malaria Journal*, **18**, 1–21, 10.1186/S12936-019-2790-2.
- NEHER, D.A. 2010. Ecology of Plant and Free-Living Nematodes in Natural and Agricultural Soil. *Annual Review of Phytopathology*, **48**, 371–394, 10.1146/annurev-phyto-073009-114439.
- NGUYEN, L.T., SCHMIDT, H.A., VON HAESLER, A. & MINH, B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274, 10.1093/molbev/msu300.
- NICOL, J.M., TURNER, S.J., COYNE, D.L., NIJS, L. DEN, HOCKLAND, S. & MAAFI, Z.T. 2011. Current Nematode Threats to World Agriculture. In *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Dordrecht: Springer Netherlands, 21–43., 10.1007/978-94-007-0434-3_2.
- NIU, L., LU, F., PEI, Y., LIU, C. & CAO, X. 2007. Regulation of flowering time by the protein arginine methyltransferase AtPRMT10. *EMBO Reports*, **8**, 1190–1195, 10.1038/sj.embor.7401111.
- NOON, J.B., HEWEZI, T., MAIER, T.R., SIMMONS, C., WEI, J.-Z., WU, G., LLACA, V., et al. 2015. Eighteen New Candidate Effectors of the Phytonematode *Heterodera glycines* Produced Specifically in the Secretory Esophageal Gland Cells During Parasitism. *Phytopathology*, **105**, 1362–1372, 10.1094/PHYTO-02-15-0049-R.
- OCHOA-VILLARREAL, M., AISPURO-HERNANDEZ, E., VARGAS-ARISPURO, I. & NGEL, M. 2012. Plant Cell Wall Polymers: Function, Structure and Biological Activity of Their Derivatives. In *Polymerization*. InTech., 10.5772/46094.
- OKA, Y., SHAPIRA, N. & FINE, P. 2007. Control of root-knot nematodes in organic farming systems by organic amendments and soil solarization. *Crop Protection*, **26**, 1556–1565, 10.1016/j.cropro.2007.01.003.
- OKADA, K., UEDA, J., KOMAKI, M.K., BELL, C.J. & SHIMURAA, Y.C. 1991. Requirement of

- the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation. *The Plant Cell*, **3**, 677–684, 10.1105/tpc.3.7.677.
- OPPERMAN, C.H., BIRD, D.M., WILLIAMSON, V.M., ROKHSAR, D.S., BURKE, M., COHN, J., CROMER, J., et al. 2008. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 14802–14807, 10.1073/pnas.0805946105.
- OTTEN, H., MICHALAK, M., MIKKELSEN, J.D. & LARSEN, S. 2013. The binding of zinc ions to *Emericella nidulans* endo- β -1,4-galactanase is essential for crystal formation. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, **69**, 850–854, 10.1107/S1744309113019714.
- PALOMARES-RIUS, J.E., HIROOKA, Y., TSAI, I.J., MASUYA, H., HINO, A., KANZAKI, N., JONES, J.T. & KIKUCHI, T. 2014. Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and fungi. *BMC Evolutionary Biology*, **14**, 69, 10.1186/1471-2148-14-69.
- PAPOLU, P.K., DUTTA, T.K., TYAGI, N., URWIN, P.E., LILLEY, C.J. & RAO, U. 2016. Expression of a Cystatin Transgene in Eggplant Provides Resistance to Root-knot Nematode, *Meloidogyne incognita*. *Frontiers in Plant Science*, **0**, 1122, 10.3389/fpls.2016.01122.
- PARK, C.-J., CADDELL, D.F. & RONALD, P.C. 2012. Protein phosphorylation in plant immunity: insights into the regulation of pattern recognition receptor-mediated signaling. *Frontiers in Plant Science*, **3**, 177, 10.3389/fpls.2012.00177.
- PATEL, N., HAMAMOUCHE, N., LI, C., HUSSEY, R., MITCHUM, M., BAUM, T., WANG, X. & DAVIS, E.L. 2008. Similarity and functional analyses of expressed parasitism genes in *Heterodera schachtii* and *Heterodera glycines*. *Journal of Nematology*, **40**, 299–310.
- PATTATHIL, S., HAHN, M.G., DALE, B.E. & CHUNDAWAT, S.P.S. 2015. Insights into plant cell wall structure, architecture, and integrity using glycome profiling of native and AFEX™-pretreated biomass. *Journal of Experimental Botany*, **66**, 4279–

4294, 10.1093/jxb/erv107.

- PELHAM, H.R.B. 1990. The retention signal for soluble proteins of the endoplasmic reticulum. *Trends in Biochemical Sciences*, **15**, 483–486, 10.1016/0968-0004(90)90303-S.
- PENG, H., CUI, J., LONG, H., HUANG, W., KONG, L., LIU, S., HE, W., HU, X. & PENG, D. 2016. Novel Pectate Lyase Genes of Heterodera glycines Play Key Roles in the Early Stage of Parasitism Jones, J., ed. *PLOS ONE*, **11**, e0149959, 10.1371/journal.pone.0149959.
- PÉREZ, S., RODRÍGUEZ-CARVAJAL, M.A. & DOCO, T. 2003. A complex plant cell wall polysaccharide: Rhamnogalacturonan II. A structure in quest of a function. *Biochimie*, **85**, 109–121, 10.1016/S0300-9084(03)00053-1.
- PERRY, R.N. & BEANE, J. 1982. The effect of brief exposures to potato root difisate on the hatching of Globodera rostochiensis. *Revue de Nématologie*, **5**, 221–224.
- 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, 10.1038/nmeth.1701.
- PEYMEN, K., WATTEYNE, J., FROONINCKX, L., SCHOOF, L. & BEETS, I. 2014. The FMRFamide-like peptide family in nematodes. *Frontiers in Endocrinology*, **5**, 90, 10.3389/fendo.2014.00090.
- PHAN, N.T., ORJUELA, J., DANCHIN, E.G.J., KLOPP, C., PERFUS-BARBEOCH, L., KOZLOWSKI, D.K., KOUTSOVOULOS, G.D., et al. 2020. Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*). *Ecology and Evolution*, **10**, 11006–11021, 10.1002/ECE3.6680.
- PHUKAN, U.J., JEENA, G.S. & SHUKLA, R.K. 2016. WRKY Transcription Factors: Molecular Regulation and Stress Responses in Plants. *Frontiers in Plant Science*, **7**, 760, 10.3389/fpls.2016.00760.
- POK MAN NGOU, B., AHN, H.-K., DING, P. & JONES, J.D. 2020. Mutual Potentiation of Plant Immunity by Cell-surface and Intracellular Receptors. *BioRxiv*, 1–29, 10.1101/2020.04.10.034173.
- POKHARE, S.S., THORPE, P., HEDLEY, P., MORRIS, J., HABASH, S.S., ELASHRY, A., EVES-VAN DEN

- AKKER, S., GRUNDLER, F.M.W. & JONES, J.T. 2020. Signatures of adaptation to a monocot host in the plant-parasitic cyst nematode *Heterodera sacchari*. *The Plant Journal*, tpj.14910, 10.1111/tpj.14910.
- POPEIJUS, H., OVERMARS, H., JONES, J., BLOK, V., GOVERSE, A., HELDER, J., SCHOTS, A., BAKKER, J. & SMANT, G. 2000. Degradation of plant cell walls by a nematode. *Nature*, **406**, 36–37, 10.1038/35017641.
- PORTER, K. & DAY, B. 2016. From filaments to function: The role of the plant actin cytoskeleton in pathogen perception, signaling and immunity. *Journal of Integrative Plant Biology*, **58**, 299–311, 10.1111/jipb.12445.
- POSTMA, W.J., SLOOTWEG, E.J., REHMAN, S., FINKERS-TOMCZAK, A., TYTGAT, T.O.G., VAN GELDEREN, K., LOZANO-TORRES, J.L., et al. 2012. The effector SPRYSEC-19 of *Globodera rostochiensis* suppresses CC-NB-LRR-mediated disease resistance in plants. *Plant Physiology*, **160**, 944–954, 10.1104/pp.112.200188.
- POTNIS, N., TIMILSINA, S., STRAYER, A., SHANTHARAJ, D., BARAK, J.D., PARET, M.L., VALLAD, G.E. & JONES, J.B. 2015. Bacterial spot of tomato and pepper: Diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Molecular Plant Pathology*, **16**, 907–920, 10.1111/mpp.12244.
- POTTER, S.M., JOHNSON, B.A., HENSCHEN, A., ASWAD, D.W. & GUZZETTA, A.W. 1992. The type II isoform of bovine brain protein L-isoaspartyl methyltransferase has an endoplasmic reticulum retention signal (...RDEL) at its C-terminus. *Biochemistry*, **31**, 6339–6347, 10.1021/bi00142a025.
- PRITCHARD, L. & BIRCH, P.R.J. 2014. The zigzag model of plant-microbe interactions: is it time to move on? *Molecular Plant Pathology*, **15**, 865–870, 10.1111/mpp.12210.
- QIN, L., KUDLA, U., ROZE, E.H.A., GOVERSE, A., POPEIJUS, H., NIEUWLAND, J., OVERMARS, H., et al. 2004. A nematode expansin acting on plants. *Nature*, **427**, 30–30, 10.1038/427030a.
- QIN, L., OVERMARS, H., HELDER, J., POPEIJUS, H., ROUPPE VAN DER VOORT, J., 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, 10.1094/MPMI.2000.13.8.830.
- QUINLAN, A.R. & HALL, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842, 10.1093/bioinformatics/btq033.
- QUISPE-HUAMANQUISPE, D.G., GHEYSEN, G. & KREUZE, J.F. 2017. Horizontal gene transfer contributes to plant evolution: The case of agrobacterium T-DNAs. *Frontiers in Plant Science*, **8**, 10.3389/fpls.2017.02015.
- QUIST, C.W., SMANT, G. & HELDER, J. 2015. Evolution of Plant Parasitism in the Phylum Nematoda. *Annual Review of Phytopathology*, **53**, 289–310, 10.1146/annurev-phyto-080614-120057.
- RANCUREL, C., LEGRAND, L. & DANCHIN, E.G.J. 2017. Alieness: Rapid Detection of Candidate Horizontal Gene Transfers across the Tree of Life. *Genes*, **8**, 10.3390/genes8100248.
- REBOIS, R. V, MADDEN, P.A. & JOE ELDRIDGE, B. 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., VOSSEN, J., et al. 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, 10.1094/MPMI-22-3-0330.
- REHMANY, A.P., GORDON, A., ROSE, L.E., ALLEN, R.L., ARMSTRONG, M.R., WHISSON, S.C., KAMOUN, S., TYLER, B.M., BIRCH, P.R.J. & BEYNON, J.L. 2005. Differential Recognition of Highly Divergent Downy Mildew Avirulence Gene Alleles by RPP1 Resistance Genes from Two Arabidopsis Lines. *The Plant Cell*, **17**, 1839–1850, 10.1105/tpc.105.031807.
- RIDLEY, B.L., O'NEILL, M.A. & MOHNEN, D. 2001. Pectins: Structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, **57**, 929–967,

10.1016/S0031-9422(01)00113-3.

- ROBERT-SEILANIANZ, A., GRANT, M. & JONES, J.D.G. 2011. Hormone Crosstalk in Plant Disease and Defense: More Than Just JASMONATE-SALICYLATE Antagonism. *Annual Review of Phytopathology*, **49**, 317–343, 10.1146/annurev-phyto-073009-114447.
- ROBERTSON, L., ROBERTSON, W.M., SOBCZAK, M., HELDER, J., TETAUD, E., ARIYANAYAGAM, M.R., FERGUSON, M.A.J., 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, 10.1016/S0166-6851(00)00295-4.
- ROBIN, G.P., KLEEMANN, J., NEUMANN, U., CABRE, L., DALLERY, J.F., LAPALU, N. & O'CONNELL, R.J. 2018. Subcellular localization screening of *Colletotrichum higginsianum* effector candidates identifies fungal proteins targeted to plant peroxisomes, golgi bodies, and microtubules. *Frontiers in Plant Science*, **9**, 562, 10.3389/fpls.2018.00562.
- ROBINSON, M.D. & OSHLACK, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, **11**, R25, 10.1186/gb-2010-11-3-r25.
- ROBINSON, M.P., ATKINSON, H.J. & PERRY, R.N. 1987. The influence of temperature on the hatching, activity and lipid utilization of second stage juveniles of the potato cyst nematodes *Globodera rostochiensis* and *G. pallida*. *Revue de Nématologie*, **10**, 349–354.
- ROSENTHAL, P., SIJWALI, P., SINGH, A. & SHENAI, B. 2005. Cysteine Proteases of Malaria Parasites: Targets for Chemotherapy. *Current Pharmaceutical Design*, **8**, 1659–1672, 10.2174/1381612023394197.
- ROSSI, M., GOGGIN, F.L., MILLIGAN, S.B., KALOSHIAN, I., ULLMAN, D.E. & WILLIAMSON, V.M. 1998. The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. *Proceedings of the National Academy of Sciences*, **95**, 9750–9754.
- ROSSO, M.N., DUBRANA, M.P., CIMBOLINI, N., JAUBERT, S. & ABAD, P. 2005. Application of

- RNA interference to root-knot nematode genes encoding esophageal gland proteins. *Molecular Plant-Microbe Interactions*, **18**, 615–620, 10.1094/MPMI-18-0615.
- ROUPPE VAN DER VOORT, J., VAN DER VOSSEN, E., BAKKER, E., OVERMARS, H., VAN ZANDVOORT, P., HUTTEN, R., KLEIN LANKHORST, R. & BAKKER, J. 2000. Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. *Theoretical and Applied Genetics*, **101**, 1122–1130, 10.1007/s001220051588.
- RUNIONS, J., HAWES, C. & KURUP, S. 2007. Fluorescent Protein Fusions for Protein Localization in Plants. *Methods in Molecular Biology*, **390**, 239–255.
- RUTTER, J., WINGE, D.R. & SCHIFFMAN, J.D. 2010. Succinate dehydrogenase - Assembly, regulation and role in human disease. *Mitochondrion*, **10**, 393–401, 10.1016/j.mito.2010.03.001.
- RYDER, L.S. & TALBOT, N.J. 2015. Regulation of appressorium development in pathogenic fungi. *Current Opinion in Plant Biology*, **26**, 8–13, 10.1016/j.pbi.2015.05.013.
- RYTTERSGAARD, C., LE NOURS, J., LO LEGGIO, L., JØRGENSEN, C.T., CHRISTENSEN, L.L.H., BJØRNVAD, M. & LARSEN, S. 2004. The structure of endo- β -1,4-galactanase from *Bacillus licheniformis* in complex with two oligosaccharide products. *Journal of Molecular Biology*, **341**, 107–117, 10.1016/j.jmb.2004.05.017.
- RYTTERSGAARD, C., LO LEGGIO, L., COUTINHO, P.M., HENRISSAT, B. & LARSEN, S. 2002. *Aspergillus aculeatus* β -1,4-galactanase: Substrate recognition and relations to other glycoside hydrolases in clan GH-A. *Biochemistry*, **41**, 15135–15143, 10.1021/bi026238c.
- SACCO, M.A., KOROPACKA, K., GRENIER, E., JAUBERT, M.J., BLANCHARD, A., GOVERSE, A., SMANT, G. & MOFFETT, P. 2009. The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death. *PLoS Pathogens*, **5**, 1000564, 10.1371/journal.ppat.1000564.
- SAHA, S., WONG, C.C.L., XU, T., NAMGOONG, S., ZEBROSKI, H., YATES, J.R. & KASHINA, A. 2011. Arginylation and methylation double up to regulate nuclear proteins and

- nuclear architecture in vivo. *Chemistry and Biology*, **18**, 1369–1378, 10.1016/j.chembiol.2011.08.019.
- SAMPEDRO, J. & COSGROVE, D.J. 2005. The expansin superfamily. *Genome Biology*, **6**, 242, 10.1186/gb-2005-6-12-242.
- SATO, K., KADOTA, Y. & SHIRASU, K. 2019. Plant Immune Responses to Parasitic Nematodes. *Frontiers in Plant Science*, **10**, 10.3389/fpls.2019.01165.
- SCEBBA, F., DE BASTIANI, M., BERNACCHIA, G., ANDREUCCI, A., GALLI, A. & PITTO, L. 2007. PRMT11: a new Arabidopsis MBD7 protein partner with arginine methyltransferase activity. *The Plant Journal*, **52**, 210–222, 10.1111/j.1365-313X.2007.03238.x.
- SCHEEPMAKER, J.W.A., GEELS, F.P., SMITS, P.H. & GRIENSVEN, L.J.L.D. VAN. 1997. Control of the mushroom pests *Lycoriella auripila* (Diptera: Sciaridae) and *Megaselia halterata* (Diptera: Phoridae) by *Steinernema feltiae* (Nematoda: Steinernematidae) in field experiments. *Annals of Applied Biology*, **131**, 359–368, 10.1111/j.1744-7348.1997.tb05165.x.
- SCHNEIDER, C.A., RASBAND, W.S. & ELICEIRI, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, **9**, 671–675, 10.1038/nmeth.2089.
- SCHOOFF, H., LENHARD, M., HAECKER, A., MAYER, K.F.X., JÜ, G. & LAUX, T. 2000. The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. *Cell*, **100**, 635–644.
- SEIBOTH, B. & METZ, B. 2011. Fungal arabinan and L-arabinose metabolism. *Applied Microbiology and Biotechnology*, **89**, 1665–1673, 10.1007/s00253-010-3071-8.
- SEITNER, D., UHSE, S., GALLEI, M. & DJAMEI, A. 2018. The core effector Cce1 is required for early infection of maize by *Ustilago maydis*. *Molecular Plant Pathology*, **19**, 2277–2287, 10.1111/mpp.12698.
- SHANKS, C.M., RICE, J.H., ZUBO, Y., SCHALLER, G.E., HEWEZI, T. & KIEBER, J.J. 2016. The Role of Cytokinin During Infection of Arabidopsis thaliana by the Cyst Nematode *Heterodera schachtii*. *The American Phytopathological Society*, **29**, 10.1094/MPMI-07-15-0156-R.
- SHAO, H., LI, X. & LOK, J.B. 2017. Heritable genetic transformation of *Strongyloides*

- stercoralis by microinjection of plasmid DNA constructs into the male germline. *International Journal for Parasitology*, **47**, 511–515, 10.1016/J.IJPARA.2017.04.003.
- SHINGLES, J., LILLEY, C.J., ATKINSON, H.J. & URWIN, P.E. 2007. Meloidogyne incognita: Molecular and biochemical characterisation of a cathepsin L cysteine proteinase and the effect on parasitism following RNAi. *Experimental Parasitology*, **115**, 114–120, 10.1016/j.exppara.2006.07.008.
- SHOWMAKER, K.C., SANDERS, W.S., DEN AKKER, S.E. VAN, MARTIN, B.E., PLATT, R.N., STOKES, J. V., HSU, C.Y., BARTLETT, B.D., PETERSON, D.G. & WUBBEN, M.J. 2019. A genomic resource for the sedentary semi-endoparasitic reniform nematode, *Rotylenchulus reniformis* Linford & Oliveira. *Journal of Nematology*, **51**, 10.21307/jofnem-2019-013.
- SHOWMAKER, K.C., SANDERS, W.S., EVES-VAN DEN AKKER, S., MARTIN, B.E., CALLAHAN, F.E., PETERSON, D.G. & WUBBEN, M.J. 2018. Five life stage-specific transcriptome assemblies for the reniform nematode, *Rotylenchulus reniformis* Linford and Oliveira. *Journal of Nematology*, **50**, 529–530, 10.21307/jofnem-2018-053.
- SHTONDA, B.B. & AVERY, L. 2006. Dietary choice behavior in *Caenorhabditis elegans*. *The Journal of Experimental Biology*, **209**, 89–102, 10.1242/jeb.01955.
- SIDDIQUE, S. & GRUNDLER, F.M. 2018. Parasitic nematodes manipulate plant development to establish feeding sites. *Current Opinion in Microbiology*, **46**, 102–108, 10.1016/j.mib.2018.09.004.
- SIDDIQUE, S., RADA KOVIC, Z.S., HILTL, C., PELLEGRIN, C., BAUM, T.J., BEASLEY, H., CHITAMBO, O., et al. 2021. The genome and lifestage-specific transcriptomes of a plant-parasitic nematode and its host reveal susceptibility genes involved in trans-kingdom synthesis of vitamin B5. *BioRxiv*, 2021.10.01.462558, 10.1101/2021.10.01.462558.
- SIQUEIRA-NETO, J.L., DEBNATH, A., MCCALL, L.I., BERNATCHEZ, J.A., NDAO, M., REED, S.L. & ROSENTHAL, P.J. 2018. Cysteine proteases in protozoan parasites. *PLoS Neglected Tropical Diseases*, **12**, 10.1371/journal.pntd.0006512.
- SKELLY, P.J., DA'DARA, A. & HARN, D.A. 2003. Suppression of cathepsin B expression in

- Schistosoma mansoni by RNA interference. *International Journal for Parasitology*, **33**, 363–369, 10.1016/S0020-7519(03)00030-4.
- SMANT, G., STOKKERMANS, J.P.W.G., YAN, Y., DE BOER, J.M., BAUM, T.J., WANG, X., HUSSEY, R.S., et al. 1998. Endogenous cellulases in animals: Isolation of β -1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences*, **95**, 4906–4911, 10.1073/pnas.95.9.4906.
- SMANT, GEERT, STOKKERMANS, J.P.W.G., YAN, Y., DE BOER, J.M., BAUM, T.J., WANG, X., HUSSEY, R.S., et al. 1998. Endogenous cellulases in animals: Isolation of β -1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 4906–4911, 10.1073/pnas.95.9.4906.
- SOBCZAK, M., AVROVA, A., JUPOWICZ, J., PHILLIPS, M.S., ERNST, K. & KUMAR, A. 2005. Characterization of Susceptibility and Resistance Responses to Potato Cyst Nematode (*Globodera* spp.) Infection of Tomato Lines in the Absence and Presence of the Broad-Spectrum Nematode Resistance Hero Gene. *Molecular Plant-Microbe Interactions MPMI*, **18**, 158–168, 10.1094/MPMI.
- 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*. Springer Science, 61–82.
- SONG, J., WIN, J., TIAN, M., SCHORNACK, S., KASCHANI, F., ILYAS, M., VAN DER HOORN, R.A.L. & KAMOUN, S. 2009. Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3.
- SPERSCHNEIDER, J., CATANZARITI, A.-M., DEBOER, K., PETRE, B., GARDINER, D.M., SINGH, K.B., DODDS, P.N. & TAYLOR, J.M. 2017. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports*, **7**, 44598, 10.1038/srep44598.
- SPICUGLIA, S. & VANHILLE, L. 2012. Chromatin signatures of active enhancers. *Nucleus*, **3**, 126–131, 10.4161/nucl.19232.
- STAPLETON, J.J. & DEVAY, J.E. 1986. Soil solarization: a non-chemical approach for

- management of plant pathogens and pests. *Crop Protection*, **5**, 190–198, 10.1016/0261-2194(86)90101-8.
- STARR, J.L. 1991. *Rotylenchulus reniformis* on Greenhouse-grown Foliage Plants: Host Range and Sources of Inoculum. *Supplement to Journal of Nematology*, **23**, 634–638.
- STRACHAN, S.M., ARMSTRONG, M.R., KAUR, A., WRIGHT, K.M., YIN LIM, T., BAKER, K., JONES, J., BRYAN, G., BLOK, V. & HEIN, I. 2019. Mapping the H2 resistance effective against *Globodera pallida* pathotype Pa1 in tetraploid potato. *Theoretical and Applied Genetics*, **132**, 1283–1294, 10.1007/s00122-019-03278-4.
- STRAHL, B.D., BRIGGS, S.D., BRAME, C.J., CALDWELL, J.A., KOH, S.S., MA, H., COOK, R.G., et al. 2001. Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1. *Current Biology*, **11**, 996–1000, 10.1016/S0960-9822(01)00294-9.
- SULLIVAN, M.J., INSERRA, R.N., FRANCO, J., MORENO-LEHEUDÉ, I. & GRECO, N. 2007. Potato cyst nematodes: plant host status and their regulatory impact. *Nematropica*, **37**, 193–201.
- SWARUP, R. & PÉRET, B. 2012. AUX/LAX family of auxin influx carriers-An overview. *Frontiers in Plant Science*, **3**, 10.3389/fpls.2012.00225.
- SYRACUSE, A.J., JOHNSON, C.S., EISENBACK, J.D., NESSLER, C.L. & SMITH, E.P. 2004. Intraspecific Variability within *Globodera tabacum solanacearum* Using Random Amplified Polymorphic DNA. *Journal of Nematology*, **36**, 443–439.
- TEN HAVE, A., MULDER, W., VISSER, J. & VAN KAN, J.A.L. 1998. The Endopolygalacturonase Gene *Bcpg1* Is Required for Full Virulence of *Botrytis cinerea*, 1009–1016 pp.
- THE C. ELEGANS SEQUENCING CONSORTIUM. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, **282**, 2012–2018, 10.1126/science.282.5396.2012.
- THERMO SCIENTIFIC. 2012. *T123-Technical bulletin - Interpretation of Nucleic Acid 260/280 Ratios*. Delaware, 1 pp.
- THORPE, P., MANTELIN, S., COCK, P.J., BLOK, V.C., COKE, M.C., EVES-VAN DEN AKKER, S.,

- GUZEEVA, E., et al. 2014. Genomic characterisation of the effector complement of the potato cyst nematode *Globodera pallida*. *BMC genomics*, **15**.
- TIWARI, S.P. & KHARE, M.N. 2003. White tip caused by *Aphelenchoides besseyi*, an important seed borne disease of rice. In Trivedi, P., ed. *Advances in nematology*. Scientific Publishers, 103–114.
- TOGASHI, K. & HOSHINO, S. 2001. Distribution pattern and mortality of the white tip nematode, *Aphelenchoides besseyi* (Nematoda: Aphelenchoididae), among rice seeds. *Nematology*, **3**, 17–24.
- TOTH, I.K., PRITCHARD, L. & BIRCH, P.R.J. 2006. Comparative Genomics Reveals What Makes An Enterobacterial Plant Pathogen. *Annual Review of Phytopathology*, **44**, 305–336, 10.1146/annurev.phyto.44.070505.143444.
- TRANG, A. & KHANDHAR, P.B. 2019. *Physiology, Acetylcholinesterase*. StatPearls Publishing.
- TRENHOLM, L.E., LICKFELDT, D.W. & CROW, W.T. 2005. Use of 1,3-dichloropropene to reduce irrigation requirements of sting nematode-infested bermudagrass. *HortScience*, **40**, 1543–1548, 10.21273/hortsci.40.5.1543.
- TURNER, S.J. & ROWE, J.A. 2006. Cyst nematodes. In *Plant Nematology*. CABI International, Wallingford, Oxon (CABI), 91–122.
- UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. 2000. Ethylene Dibromide (Dibromoethane).
- URWIN, LILLEY, C.J., MCPHERSON, M.J. & ATKINSON, H.J. 1997. Characterization of two cDNAs encoding cysteine proteinases from the soybean cyst nematode *Heterodera glycines*. *Parasitology*, **114**, 605–613.
- URWIN, P.E., ATKINSON, H.J., WALLER, D.A. & MCPHERSON, M.J. 1995. Engineered oryzacystatin-I expressed in transgenic hairy roots confers resistance to *Globodera pallida*. *The Plant Journal*, **8**, 121–131, 10.1046/J.1365-313X.1995.08010121.X.
- URWIN, P. E., LILLEY, C.J., MCPHERSON, M.J. & ATKINSON, H.J. 1997. Resistance to both cyst and root-knot nematodes conferred by transgenic *Arabidopsis* expressing a modified plant cystatin. *The Plant Journal*, **12**, 455–461, 10.1046/J.1365-

313X.1997.12020455.X.

- VAN DEN BERG, E., PALOMARES-RIUS, J.E., VOVLAS, N., TIEDT, L.R., CASTILLO, P. & SUBBOTIN, S.A. 2016. Morphological and molecular characterisation of one new and several known species of the reniform nematode, *Rotylenchulus* Linford & Oliveira, 1940 (Hoplolaimidae: Rotylenchulinae), and a phylogeny of the genus. *Nematology*, **18**, 67–107, 10.1163/15685411-00002945.
- VAN DER HOORN, R.A.L. & KAMOUN, S. 2008. From Guard to Decoy: a new model for perception of plant pathogen effectors. *The Plant Cell*, **20**, 2009–2017, 10.1105/tpc.108.060194.
- VAN MEGEN, H., VAN DEN ELSSEN, S., HOLTERMAN, M., KARSEN, 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, 10.1163/156854109X456862.
- VANHOLME, B., HAEGEMAN, A., JACOB, J., CANNOOT, B. & GHEYSEN, G. 2009. Arabinogalactan endo-1,4- β -galactosidase: a putative plant cell wall-degrading enzyme of plant-parasitic nematodes. *Nematology*, **11**, 739–747, 10.1163/156854109X404599.
- VANHOLME, B., VAN THUYNE, W., VANHOUTEGHEM, K., DE MEUTTER, J., CANNOOT, B. & GHEYSEN, G. 2007. Molecular characterization and functional importance of pectate lyase secreted by the cyst nematode *Heterodera schachtii*. *Molecular Plant Pathology*, **8**, 267–278, 10.1111/j.1364-3703.2007.00392.x.
- VARYPATAKIS, K., JONES, J.T. & BLOK, V.C. 2019. Susceptibility of potato varieties to populations of *Globodera pallida* selected for increased virulence. 995–998.
- VERMA, A., LEE, C., MORRIS, S., ODU, F., KENNING, C., RIZZO, N., SPOLLEN, W.G., et al. 2018. The novel cyst nematode effector protein 30D08 targets host nuclear functions to alter gene expression in feeding sites. *New Phytologist*, **219**, 697–713, 10.1111/nph.15179.
- VERMA, S., DIXIT, R. & PANDEY, K.C. 2016. Cysteine proteases: Modes of activation and future prospects as pharmacological targets. *Frontiers in Pharmacology*, **7**, 10.3389/fphar.2016.00107.

- VERNET, T., BERTI, P.J., DE MONTIGNY, C., MUSIL, R., TESSIER, D.C., MENARD, R., MAGNY, M.C., STORER, A.C. & THOMAS, D.Y. 1995. Processing of the Papain Precursor: The ionization state of a conserved amino acid motif within the pro region participates in the regulation of intramolecular processing. *Journal of Biological Chemistry*, **270**, 10838–10846, 10.1074/JBC.270.18.10838.
- VIEIRA, P., MAIER, T.R., EVES-VAN DEN AKKER, S., HOWE, D.K., ZASADA, I., BAUM, T.J., EISENBACK, J.D. & KAMO, K. 2018. Identification of candidate effector genes of *Pratylenchus penetrans*. *Molecular Plant Pathology*, **19**, 1887–1907, 10.1111/MPP.12666.
- VOIGT, C.A. 2014. Callose-mediated resistance to pathogenic intruders in plant defense-related papillae. *Frontiers in Plant Science*, **5**, 10.3389/fpls.2014.00168.
- VORWERK, S., SOMERVILLE, S. & SOMERVILLE, C. 2004. The role of plant cell wall polysaccharide composition in disease resistance. *Trends in Plant Science*, **9**, 203–209, 10.1016/j.tplants.2004.02.005.
- VOS, P., SIMONS, G., JESSE, T., WIJBRANDI, J., HEINEN, L., HOGERS, R., FRIJTERS, A., et al. 1998. The tomato Mi-1 gene confers resistance to both root-knot nematodes and potato aphids. *Nature Biotechnology*, **16**, 1365–1369, 10.1038/4350.
- WANG, J., DHROSO, A., LIU, X., BAUM, T.J., HUSSEY, R.S., DAVIS, E.L., WANG, X., KORKIN, D. & MITCHUM, M.G. 2021. Phytonematode peptide effectors exploit a host post-translational trafficking mechanism to the ER using a novel translocation signal. *New Phytologist*, **229**, 563–574, 10.1111/nph.16765.
- 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, 10.1111/j.1469-8137.2010.03300.x.
- WANG, S., MCLELLAN, H., BUKHAROVA, T., HE, Q., MURPHY, F., SHI, J., SUN, S., et al. 2019. *Phytophthora infestans* RXLR effectors act in concert at diverse subcellular locations to enhance host colonization. *Journal of Experimental Botany*, **70**,

- 343–356, 10.1093/jxb/ery360.
- WANG, X., MITCHUM, M.G., GAO, B., LI, C., 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, 10.1111/J.1364-3703.2005.00270.X.
- WANG, X., XUE, B., DAI, J., QIN, X., LIU, L., CHI, Y., JONES, J.T. & LI, H. 2018. A novel *Meloidogyne incognita* chorismate mutase effector suppresses plant immunity by manipulating the salicylic acid pathway and functions mainly during the early stages of nematode parasitism. *Plant Pathology*, **67**, 1436–1448, 10.1111/ppa.12841.
- WAR, A.R., PAULRAJ, M.G., WAR, M.Y. & IGNACIMUTHU, S. 2011. Role of salicylic acid in induction of plant defense system in chickpea (*cicer arietinum* L.). *Plant Signaling and Behavior*, **6**, 1787–1792, 10.4161/psb.6.11.17685.
- WARDLOW, L.R., PIGGOTT, S. & GOLDSWORTHY, R. 2001. Foliar application of *Steinernema feltiae* for the control of flower thrips. *Mededelingen (Rijksuniversiteit te Gent. Fakulteit van de Landbouwkundige en Toegepaste Biologische Wetenschappen)*, **66**, 285–291.
- 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, 10.1093/bioinformatics/btp033.
- WAWRA, S., TRUSCH, F., MATENA, A., APOSTOLAKIS, K., LINNE, U., ZHUKOV, I., STANEK, J., et al. 2017. The RxLR Motif of the Host Targeting Effector AVR3a of *Phytophthora infestans* Is Cleaved before Secretion. *The Plant Cell*, **29**, 1184–1195, 10.1105/tpc.16.00552.
- WEFERS, D., TYL, C.E. & BUNZEL, M. 2014. Novel arabinan and galactan oligosaccharides from dicotyledonous plants. *Frontiers in Chemistry*, **2**, 10.3389/fchem.2014.00100.
- WHISSON, S.C., BOEVINK, P.C., MOLELEKI, L., AVROVA, A.O., MORALES, J.G., GILROY, E.M., ARMSTRONG, M.R., et al. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*, **450**, 115–118,

10.1038/nature06203.

- WHITFORD, R., FERNANDEZ, A., DE GROODT, R., ORTEGA, E. & HILSON, P. 2008. Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 18625–18630, 10.1073/pnas.0809395105.
- WILKINSON, K.D. 1987. Protein ubiquitination: a regulatory post-translational modification. *Anti-cancer drug design*, **2**, 211–229.
- WILLIAMSON, V.M. & HUSSEY, R.S. 1996. Nematode Pathogenesis and Resistance in Plants. *The Plant Cell*, **8**, 1735–1745, 10.1105/tpc.8.10.1735.
- WISE, R.J., BARR, P.J., WONG, P.A., KIEFER, M.C., BRAKE, A.J. & KAUFMAN, R.J. 1990. Expression of a human proprotein processing enzyme: correct cleavage of the von Willebrand factor precursor at a paired basic amino acid site. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 9378, 10.1073/PNAS.87.23.9378.
- WOLD, F. 1981. In Vivo Chemical Modification of Proteins (Post-Translational Modification). *Annual Review of Biochemistry*, **50**, 783–814, 10.1146/annurev.bi.50.070181.004031.
- WOLF, A.J. & UNDERHILL, D.M. 2018. Peptidoglycan recognition by the innate immune system. *Nature Reviews Immunology*, **18**, 243–254, 10.1038/nri.2017.136.
- WORLD HEALTH ORGANIZATION. 2017. Guideline: Preventive chemotherapy to control soil-transmitted helminth infections in at-risk population groups. 1–74.
- WUBBEN, M.J., GANJI, S. & CALLAHAN, F.E. 2010. Identification and molecular characterization of a β -1,4-endoglucanase gene (Rr-eng-1) from *Rotylenchulus reniformis*. *Journal of Nematology*, **42**, 342–351.
- WYSS, U. 2002. Feeding behaviour of plant-parasitic nematodes. In Lee, D.L., ed. *The Biology of Nematodes*. Taylor and Francis Inc., 233–260.
- XUE, Q., WU, X.-Q., ZHANG, W.-J., DENG, L.-N., WU, M.-M., XUE, Q., WU, X.-Q., ZHANG, W.-J., DENG, L.-N. & WU, M.-M. 2019. Cathepsin L-like Cysteine Proteinase Genes Are Associated with the Development and Pathogenicity of Pine Wood Nematode, *Bursaphelenchus xylophilus*. *International Journal of Molecular*

- Sciences*, **20**, 215, 10.3390/ijms20010215.
- YADAV, J.K. & PRAKASH, V. 2011. Stabilization of α -Amylase, the Key Enzyme in Carbohydrates Properties Alterations, at Low pH. *International Journal of Food Properties*, **14**, 1182–1196, 10.1080/10942911003592795.
- YAMAGUCHI, Y.L., ISHIDA, T. & SAWA, S. 2016. CLE peptides and their signaling pathways in plant development. *Journal of Experimental Botany*, **67**, 4813–4826, 10.1093/jxb/erw208.
- YAN, D., ZHANG, Y., NIU, L., YUAN, Y. & CAO, X. 2007. Identification and characterization of two closely related histone H4 arginine 3 methyltransferases in *Arabidopsis thaliana*. *Biochemical Journal*, **408**, 113–121, 10.1042/BJ20070786.
- YANG, Y. & ANDERSON, C.T. 2020. Biosynthesis, Localisation, and Function of Pectins in Plants. In Kontogiorgos, V., ed. *Pectin: Technological and Physiological Properties*. Springer, 2–16.
- YANG, Y., HADJIKYRIACOU, A., XIA, Z., GAYATRI, S., KIM, D., ZURITA-LOPEZ, C., KELLY, R., et al. 2015. PRMT9 is a Type II methyltransferase that methylates the splicing factor SAP145. *Nature Communications*, 10.1038/ncomms7428.
- YEATES, G.W., BONGERS, T., DE GOEDE, R.G.M., FRECKMAN, D.W. & GEORGIEVA, S.S. 1993. Feeding Habits in Soil Nematode Families and Genera - An Outline for Soil Ecologists. *Journal of Nematology*, **25**, 315–331.
- ZEMACH, A., GASPAN, O. & GRAFI, G. 2008. The three methyl-CpG-binding domains of AtMBD7 control its subnuclear localization and mobility. *Journal of Biological Chemistry*, **283**, 8406–8411, 10.1074/jbc.M706221200.
- ZHANG, L., LILLEY, C.J., IMREN, M., PAUL KNOX, J. & URWIN, P.E. 2017. The complex cell wall composition of syncytia induced by plant parasitic cyst nematodes reflects both function and host plant. *Frontiers in Plant Science*, **8**, 1–12, 10.3389/fpls.2017.01087.
- ZIPFEL, C. 2014. Plant pattern-recognition receptors. *Trends in Immunology*, **35**, 345–351, 10.1016/J.IT.2014.05.004.
- ZIPFEL, C., KUNZE, G., CHINCHILLA, D., CANIARD, A. & JONES, J.D.G. 2006. Perception of

the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. *Cell*, **125**, 749–760.

ZUNKE, U. 1990. Observations on the Invasion and Endoparasitic Behavior of the Root Lesion Nematode *Pratylenchus penetrans*. *Journal of Nematology*, **22**, 309–320.

9. Appendix

Appendix containing all supplementary figures referenced throughout this thesis.

9.1 – Supplementary file 1

Excel spreadsheet “GROS_Initial_list_for_BLAST” can be found on accompanying data CD.

9.2 – Supplementary file 2

Python script “extract_blast_data_using_list” can be found on accompanying data CD.

9.3 – Supplementary file 3

Python script “extract_seq_from_fasta_keep_order_of_list” can be found on accompanying data CD.

9.4 – Supplementary file 4

Python script

“extract_seq_from_fasta_keep_order_of_list_modded_for_Reniformis_Trinity_kmer2_cd9
0_nt_or_aa” can be found on accompanying data CD.

Gro DOG 0116 family alignment

GROS_g01949	1	-----M I G P A L V F L F L-----F A T N I R E D D D I P M D G P - D-----P E V D E H I R Q H F Q G H M D A F S S G M K E G L A E I O K E S D E-----M A A R G E K D P G N L A D L I K P E N V A K F V S N W L P E R T R V L L P K	104
GROS_g10784	1	M R C P K I N T N V P V I L M T I I---G L L T C - T I S G G K K K E S G G - D G A L N D L L G G L D L D G M L S G V Q E L L G K H A G G S A D G K K G--D G L E K L K E M A A N L D P K M L D I L G K M D T E K I S G T I Q N M V K P E N I W E M L P A H V K E M L P E	130
g12045.t1	1	-----M G I S I I F L F V-----L L T N - L V I D G Q D T K A G G - N G E Q D D-----P K A E E M I R E Q L K G H M D A F V S G M K D G M E E I I K E A G E-----	67
GPLIN_000806500	1	M K C R V Y E V S A R V H S L R F I L V G L L T C - T I S G G K Q K E S G G Q D G A L N G L L G G L D F D R M L S G V Q E L L G K H A G G S A D G K K G--N G L D K L K E M A A N L D P K M L D I L G K M D T E K I T E T I Q N M V K P E N I W E M L P A H V K E M L P E	134
GPLIN_000344300	1	M R C P K I N T I V P V I L M T I G-----L L T C - T I S G G K K K E G G E - N G A L N D I L G G L D L E G M L S G V Q E L L G K H A G G S A D G K K G--N G L D K L K E M A A N L D P K M L D I L G K M D T E K I S E T I Q N M V K P E N I W E M L P A H V K E M L P E	129
cds.Nab_28487_c0_seq1_m.28420	1	-----M S L F P I S F S L F L-----V P L F C P L L L F G A O N E G S G E G E K D F O E M L G G I-----L S G M Q Q M V G Q - F E K T A G N G K G G N L D G L K G L-----M N N F D A D K M G Q L L Q N F G S P D M L A D M L P E D W K E L I P E	107
GROS_g01949	105	E W L---O P L I K A L T T E S G A I A E L Y A G A D R I G S A D E F W R E L R K K A P T T A E L L E N - A M G O V T R S W D F N R M L--N E E S K Q Y V E K V I S G A E W F V R Q V A G O F L R L S N E A T L T M N L F E A W P L L N T V L S K A M S N R M L L E V	235
GROS_g10784	131	G W F L I G K F I G N V L S K E W D K I M E L L D G V D K L G S M E E L R E E L K K K A P I L S D L L E N D W W G N M E R W O K F E G N L--G E E A K Q Y T E N V K A T G W E F L S R K V F A P A L R L S P E A M N L K O S L G N T L P A L K P L L E K A A T N K K L K E W	265
g12045.t1	68	-----G K I E K F K T-----	75
GPLIN_000806500	135	G W F L I G K F V G N V L S K E W D K I Y E L L D G V D T L G S L E E L R E E L K K K A P I L A D L L E N D W W G N M K R W E K F E A N L--G E E A K Q Y T E N V K A T G W E F L S R K V F A E A L R L S P E A M N L K O S L G N T L P A L K P L L E K A A T N K K L R E W	269
GPLIN_000344300	130	G W F V I G K F V G N V L S K E W D K I M E L L D G V D K L G S M E E L R E E L K K K A P I L A D L L E N D W W G N M E R W E K F E A N L--G A E A K Q Y T E N I K A T G W E F L S R K V F A P A L R L S P E A M N L K O S L G N T L P A L K P L L E K A A T N K K L K E W	264
cds.Nab_28487_c0_seq1_m.28420	108	E W R---E L L G E V I A T E M D K L P G W W A A F T G M E S V D G F F E V Y R K K A P V L S E L L Y E - W A V I M K R W A T F E K T L O T D G E A L S Y V M L K A T G G K W I L T K G V H P F O G L S A K A K E T L R A L Q K A F P V L D V L L G K V Y V G E V L T K W	240
GROS_g01949	236	A G-----E Q Y K K A E E Q M R A K E Q E K S A T K E E F N E S M M A G M K E M F N M N L G E F D D I M G M D E E-----E E G E K E S D G K E E K D D E E T V A E Q K E G K E V E A G A E T H-----	328
GROS_g10784	266	A E-----A W M M A E L E A M F S K V O L A-----T D G L K E K A G G G N A-----	299
g12045.t1	270	A E-----A W M M A E L E A M F S K V O L A-----T D G L K E K A G G G N A-----	303
GPLIN_000806500	265	A D-----L O M M A E L E A M F S K V O L A-----T D G L K E K A D G G G N A-----	298
GPLIN_000344300	241	L A G L A G I K L P E G A E P E K H A K E G K M K K E K A Q O D M K E E E E G D E D E E E E Y G D M E G E E E E D E L E E Y L E D E E A T T D A E T A T E G G G S P M M K G I P I I R T S L S N R R S S K D S L S S F V S R T S R V A G A I A S I A S P P A S R R	377
cds.Nab_28487_c0_seq1_m.28420	329	-----G E Q E E G Q C S K S D E Q C E A K E S T E K A E E E L E K N V-----	362
GROS_g01949	300	-----G R T E D E-----	305
GROS_g10784	304	-----G R T Q D E-----	309
g12045.t1	299	-----G R T E Y E-----	304
GPLIN_000806500	378	L Q L E G R K R T I L G D E P V P R L D L R G P D E A E E L C S Q Y G A A E M S G H R R Q H L S L G R R S S F T T T T A A I T R A P S G G N Q I A P E L S D L V I Y M Q A V K F G F V V T N T T E L P V A S G T G Q R D T P G T S G G D S G L A S G E L R P C S N S F A G I S	514
GPLIN_000344300			
cds.Nab_28487_c0_seq1_m.28420	515	R M R S H A P S V L S S Y S A N V P S A A V V Q Q Q Q A G O F A L Q Q F R R P K S S S H V A A Q Q L E S G R I L O P V S A Q L S S G S D K A D S A A G T P P S G O P Q H N S L S V A H Q G N A A G S L P P P G T T S R S S P K S H A S C Y Q V S S L T E A S A R K L C R K	651
GROS_g01949			
GROS_g10784			
g12045.t1			
GPLIN_000806500			
GPLIN_000344300	652	H A O K C I S Y T R N H I M R T Y P G G M R I D S N S F P Q R L L E F W Q S G L Q M V A L N F Q T A D V T L A V N T A M F E Q S G S G Y M L K P R V L W D S S H P L F C R F N P L A K D L S S R P A L I L Q L T V I S G Q Y V A P G S F A A S P F L E I E V I G I Q A D C V	788
cds.Nab_28487_c0_seq1_m.28420			
GROS_g01949			
GROS_g10784			
g12045.t1			
GPLIN_000806500			
GPLIN_000344300	789	K E K S K T V G R N G V N P W W N H G C T F R I L F A E L A F L R I A V C D G A A N G R V I S Q R V P V R C L R P G Y R H L P L R T A S N O P L E Q S S L F I R S R F E Q E E F I S L H D E D L L P H S A P Y S A Q H S P H S A C S A S H P C P A N P F N P H C G G P P L A F E	925
cds.Nab_28487_c0_seq1_m.28420			
GROS_g01949	363	-----E E T T R D E L-----	370
GROS_g10784	306	-----E G T K E E L-----	313
g12045.t1			
GPLIN_000806500	310	-----E G T K E E L-----	317
GPLIN_000344300	305	-----E G T T E E L-----	312
cds.Nab_28487_c0_seq1_m.28420	926	P E L T Y Q V L K L D A E A N V K L S I L R R Q I F I I R V T G L F S D E T P T I V H A E S S T V R N V I A Q A L A N A G K P A E S P D D Y I L S E E T A P L V N S P S P S I G H L L A T H S A A A E A G G P L C A G D G G G P S T G D Q Q Q G Q Q	1051

G23G11 family alignment

GROS_g02469	1	MFRQVPLLENNMIFLLLFLLNKAHADAGGDDVYITLGGASNLKVSADKLEARFLADAGIKRREVSVAAGAPFAAAAPASRSNRRVSTRTTITTTITTELS--RKYVTFVEDVDVLEFETGIIWPK134
GROS_g02470	1	MTLRLLIIL--FISPFILHIANGV--IKRGLTYSAGVABM--RKYVRELLAKDEVRASDRRAERVOYERTVYVYDDDEFFIRRFSTTSKHEIYVEFEFGIOWPK103
g951.t1	1	NMRIRALLSPGLGLLLLLLDLIGRVRA--DIADYLSLQISRPRLKTSR--HLTGGSEEDDDEDLRLRGPATR--L--KTEMRRRNVTLVE--EYGRKIVFGIOWPK107
g952.t1	1	----MFVLPRLIILLMANLAHFQVRAFDSSRSRFRWNTLSPKTKLKS--KRW-HKLDHSDSSSEVSTKLGKK--V--RKMVHHEGDKVTEVREGGSRRTLOKVVFTGVIOWPK112
g25264.t1	1	----MFVLPRLIILLMANLAHFQVRAFDSSRSRFRWNTLSPKTKLKS--V-HKLDHSDSSSEVSTKLGKK--V--RKMVHHEGDKVTEVREGGSRRTLOKVVFTGVIOWPK104
g25265.t1	1	MMEIRVLSPLGLGLVLLLLLDLIGRVCA--DIADYLSLQISRPRLKTSR--YLTGVRGSDGDEDDPPRLRGPATR--L--KTEMRRRNVTLVE--DDDGRKIWFGLWPK107
g33471.t1	1	----MMLIGPLGVLLVLLLDLIGRVCA--DWADYLSRRTLRTRLLKTSR--EIESDDEDEVPMLRGRQR--L--KTEMRRRNVTLVE--DEDYDVKIWFGLWPK99
SPLIN_000763000	1	MNBLIIL--FISPFILHIANGV--IKRGLTYSAGVABM--RKYVRELLAKDEVRASDRRAERVOYERTVYVYDDDEFFIRRFSTTSRKEIYVEFEFGIOWPK103
cds.Nab_57241_c0_seq1_m.50052	1	MDCCFLRLPCLLLLLFVVFVFGHGE--LQKLFASAFNRFGGS--IKKSPFPAPPPPKKALACPHTKTL--EEVTKRKRIRKHVR--KKAHFAPAEEKTLTLGLIOWPK106
cds.Nab_57240_c0_seq1_m.50041	1	MDCCFLRLPCLLLLLFVVFVFGHGE--LQKLFASAFNRFGGS--IKKSPFPAPPPPKKALACPHTKTL--EEVTKRKRIRKHVR--KKAHFAPAEEKTLTLGLIOWPK106
cds.Nab_57239_c0_seq1_m.50031	1	MDCCFLRLPCLLLLLFVVFVFGHGE--LQKLFASAFNRFGGS--IKKSPFPAPPPPKKALACPHTKTL--EEVTKRKRIRKHVR--KKAHFAPAEEKTLTLGLIOWPK106
cds.Nab_57238_c0_seq1_m.50020	1	MDCCFLRLPCLLLLLFVVFVFGHGE--LQKLFASAFNRFGGS--IKKSPFPAPPPPKKALACPHTKTL--EEVTKRKRIRKHVR--KKAHFAPAEEKTLTLGLIOWPK106
cds.Nab_57237_c0_seq1_m.50010	1	MDCCFLRLPCLLLLLFVVFVFGHGE--LQKLFASAFNRFGGS--IKKSPFPAPPPPKKALACPHTKTL--EEVTKRKRIRKHVR--KKAHFAPAEEKTLTLGLIOWPK106
cds.Nab_28879_c0_seq1_m.29300	1	MDCCFLRLPCLLLLLFVVFVFGHGE--LQKLFASAFNRFGGS--IKKSPFPAPPPPKKALACPHTKTL--EEVTKRKRIRKHVR--KKAHFAPAEEKTLTLGLIOWPK106
GROS_g02469	135	GVYMDVDFMSDEEKVTRRSRKKVLDKEDGKFAVKIHFISKK--DSDVLIHFGGKELHYVLDLQVGRMDVIFDLDN-RTVEIYVPRKVEDEKIKLEEDK--244
GROS_g02470	104	FVLYLDIKYFEMNDYKVKGM-AAPAMLDKESQSFVWIKIITDSESEGI--EMLMHMHFG-RK1SHLYLKENNLDIYIHWLE-KMKMIVYENNERAEKLEK--211
g951.t1	108	DKGVYHLDLDFEMTKDEHKVQR-AYPESLDGDDGYFANFTV-SGER--FVYIYIHFQYKDD-PHYH-ERWRVKDADVOIFDLT-TMEVSSARLRRLDVLKLEERDKK--218
g952.t1	113	AEVDYVDFVFRMSDDKKVRGS-AKPSDITGKDFEEMNFTEYDVEEDTSDSDSESPPP-RRKSDIEVDFKIHIFGHEK-PHYE-LDIAELQADVIVYDFDK-KSIFYIVAPRKVDKLEELERDK--243
g25264.t1	105	AEVDYVDFVFRMSDDKKVRGS-AKPSDITGKDFEEMNFTEYDVEEDTSDSDSESPPP-RRRKNSDIEVDFKIHIFGHEK-PHYE-LDIAELQADVIVYDFDK-KSIFYIVAPRKVDKLEELERDK--236
g25265.t1	108	DKVDHLDLDFAMTKDQRVKQR-ADPMSFETDSDSFEADFTV-HGEGK--FRVYIYIHFQYKDD-PKHIRKVEYDQDVHIFDLS-TMKVYVRNRLKALKLEERDKK--218
g33471.t1	100	DKGVYHLDLDFEMTKDQRVKQR-ADPMSFETDSDSFEADFTV-HGEGK--FRVYIYIHFQYKDD-PKHIRKVEYDQDVHIFDLS-TMKVYVRNRLKALKLEERDKK--218
SPLIN_000763000	104	FVLYLDIKYFEMNDYKVKGM-AAPAMLDKESQSFVWIKIITDSESEGV--EMLMHMHFG-RK1SHLYLKENNLDIYIHWLE-KMKMIVYENNERAEKLEK--211
cds.Nab_57241_c0_seq1_m.50052	107	NNYSIRLNIAGNEK--IAPVETKTPDGLFEAVVSLDKTAE--PKNVELTISGGVEMWDEE-FFLEEVADIFVYIDLEOPEEWMDFPSLSEALEEMEGRH--206
cds.Nab_57240_c0_seq1_m.50041	107	NNYSIRLNIAGNEK--IAPVETKTPDGLFEAVVSLDKTAE--PKNVELTISGGVEMWDEE-FFLEEVADIFVYIDLEOPEEWMDFPSLSEALEEMEGRH--206
cds.Nab_57239_c0_seq1_m.50031	107	NNYSIRLNIAGNEK--IAPVETKTPDGLFEAVVSLDKTAE--PKNVELTISGGVEMWDEE-FFLEEVADIFVYIDLEOPEEWMDFPSLSEALEEMEGRH--206
cds.Nab_57238_c0_seq1_m.50020	107	NNYSIRLNIAGNEK--IAPVETKTPDGLFEAVVSLDKTAE--PKNVELTISGGVEMWDEE-FFLEEVADIFVYIDLEOPEEWMDFPSLSEALEEMEGRH--206
cds.Nab_57237_c0_seq1_m.50010	107	NNYSIRLNIAGNEK--IAPVETKTPDGLFEAVVSLDKTAE--PKNVELTISGGVEMWDEE-FFLEEVADIFVYIDLEOPEEWMDFPSLSEALEEMEGRH--206
cds.Nab_28879_c0_seq1_m.29300	107	NNYSIRLNIAGNEK--IAPVETKTPDGLFEAVVSLDKTAE--PKNVELTISGGVEMWDEE-FFLEEVADIFVYIDLEOPEEWMDFPSLSEALEEMEGRH--206
GROS_g02469	245	PVKLEPKBOTMFAAGLWPPV--ADKTYLVSFRDFDRRQAVKRNHFDPDTMVEENAKGMDFDLYVIT--ERKESLYWAAANAK-SARKLEVILRARDLDTTKALDMSI--SKAEFRORLE--367
GROS_g02470	212	PLVKDILIK--EYEFNLHIWPEG--YVOYFLVEGKFEHED--KRVIGSATHAKATRRDGGVFELEHDSAEEDSYMHIF--FD--DVC--294
g951.t1	219	PAEAKKAEKES-EVAFGLWPPV--DKYEMHVDKDFELDRGSRVVKHRTHADPKAEFDGGMKAKIDWE--VKFDALHIWVKSTKSGVKKFVMSNDPDP-NYKVALDIFKFAKLEVEFPKKE--347
g952.t1	244	PK--TKIFSGIWPAAK--YNTFTLMSPEMFDLDDDERVKRNRSHPKVKAETDRDDHFEVELEE--EMPAKLHLWINGQDTAKSVLLEPKED--KKYAFDMRK--SASOEV--349
g25264.t1	237	PK--TKIFSGIWPAAK--YKFTLMSPEMFDLDDDERVKRNRSHPKVKAETDRDDHFEVELEE--EMPAKLHLWINGQDTAKSVLLEPKED--KKYAFDMRK--SASOEV--342
g25265.t1	219	PAEAKKAEKES-EVAFGLWPPV--DKYEMHVDKDFELDRGSRVVKHRTHADPKAEFDGGMKAKIDWE--VKFDALHIWVKSTKSGVKKFVMSNDPDP-NYKVALDIFKFAKLEVEFPKKE--347
g33471.t1	212	PLVNDIYIK--EYEFNLHIWPEG--YVOYFLVEGKFEHED--KRVIGSATHAKATRRDGGVFELEHDSAEEDSYMHIFDOKTQVH--VERVDRB--SOFVLEATIYSTKSEEGFPLR--320
SPLIN_000763000	207	PAEOKV--KRLFRGIWVVGAGGGPYRMFVATGGEFMKKEEGKWEKGDHPPEVAAFHENOTFEAEMULL--EHSOLRVAVORKGHATAHYFTVEGNFD-KFYALDIRT--SEOTRFNNLEE--326
cds.Nab_57240_c0_seq1_m.50041	207	PAEOKV--KRLFRGIWVVGAGGGPYRMFVATGGEFMKKEEGKWEKGDHPPEVAAFHENOTFEAEMULL--EHSOLRVAVORKGHATAHYFTVEGNFD-KFYALDIRT--SEOTRFNNLEE--326
cds.Nab_57239_c0_seq1_m.50031	207	PAEOKV--KRLFRGIWVVGAGGGPYRMFVATGGEFMKKEEGKWEKGDHPPEVAAFHENOTFEAEMULL--EHSOLRVAVORKGHATAHYFTVEGNFD-KFYALDIRT--SEOTRFNNLEE--326
cds.Nab_57238_c0_seq1_m.50020	207	PAEOKV--KRLFRGIWVVGAGGGPYRMFVATGGEFMKKEEGKWEKGDHPPEVAAFHENOTFEAEMULL--EHSOLRVAVORKGHATAHYFTVEGNFD-KFYALDIRT--SEOTRFNNLEE--326
cds.Nab_57237_c0_seq1_m.50010	207	PAEOKV--KRLFRGIWVVGAGGGPYRMFVATGGEFMKKEEGKWEKGDHPPEVAAFHENOTFEAEMULL--EHSOLRVAVORKGHATAHYFTVEGNFD-KFYALDIRT--SEOTRFNNLEE--326
cds.Nab_28879_c0_seq1_m.29300	207	PAEOKV--KRLFRGIWVVGAGGGPYRMFVATGGEFMKKEEGKWEKGDHPPEVAAFHENOTFEAEMULL--EHSOLRVAVORKGHATAHYFTVEGNFD-KFYALDIRT--SEOTRFNNLEE--326
GROS_g02469	368	VLKLLADGMINNEQLSDFPKKVEQKAEAKHA--SDDMLNSIKAMVDDLGKALKKEGEGKALDDELEEVGGLVYVGRKRLDQDDIKDMHKKLREI--EEVIGAEEDVLYKK--489
GROS_g02470	348	LEELMAGMDIKOIKKBEKKAEAAKKEAEAKKTEAKANGGGGGGNAELQKNLQKMLDELEDGKKEENAKLKVFEKELEDVGRKLDAMDKRLDAILKEIKSMNNMGEDIKDAVKKAIKAEEDKLYKK--486
g952.t1	350	AEDVRKKGW--GKRS--364
g25264.t1	343	AEDVRKKGW--GKRS--364
g25265.t1	348	LEELMAGMDIKOIKN--GG--GNAELQKNLQKMLDELEDGKKEENAKLKVFEKELEDVGRKLDAMDKRLDAILKEIKSMNNMGEDIKDAVKKAIKAEEDKLYKK--457
g33471.t1	330	VYMYIYDQIMHKQRVYDD--WEVHIIDIRNKKDMLAVYVNIWIKLDDLNDRDNRERRFRFVGLDSDID--404
SPLIN_000763000	327	LEELKKGMDIIRTVKR--QKAAFPFPPPA--SEHELLTAIRAMVAAKGAELEAMVEEHLRTDLERELKQVGRPMEAIAOGMEOLKGLSALGOLEKGL--VOAVASPPAALADK--443
cds.Nab_57241_c0_seq1_m.50052	327	LEELKKGMDIIRTVKR--QKAAFPFPPPA--SEHELLTAIRAMVAAKGAELEAMVEEHLRTDLERELKQVGRPMEAIAOGMEOLKGLSALGOLEKGL--VOAVASPPAALADK--443
cds.Nab_57240_c0_seq1_m.50041	327	LEELKKGMDIIRTVKR--QKAAFPFPPPA--SEHELLTAIRAMVAAKGAELEAMVEEHLRTDLERELKQVGRPMEAIAOGMEOLKGLSALGOLEKGL--VOAVASPPAALADK--443
cds.Nab_57239_c0_seq1_m.50031	327	LEELKKGMDIIRTVKR--QKAAFPFPPPA--SEHELLTAIRAMVAAKGAELEAMVEEHLRTDLERELKQVGRPMEAIAOGMEOLKGLSALGOLEKGL--VOAVASPPAALADK--443
cds.Nab_57238_c0_seq1_m.50020	327	LEELKKGMDIIRTVKR--QKAAFPFPPPA--SEHELLTAIRAMVAAKGAELEAMVEEHLRTDLERELKQVGRPMEAIAOGMEOLKGLSALGOLEKGL--VOAVASPPAALADK--443
cds.Nab_57237_c0_seq1_m.50010	327	LEELKKGMDIIRTVKR--QKAAFPFPPPA--SEHELLTAIRAMVAAKGAELEAMVEEHLRTDLERELKQVGRPMEAIAOGMEOLKGLSALGOLEKGL--VOAVASPPAALADK--443
cds.Nab_28879_c0_seq1_m.29300	327	LEELKKGMDIIRTVKR--QKAAFPFPPPA--SEHELLTAIRAMVAAKGAELEAMVEEHLRTDLERELKQVGRPMEAIAOGMEOLKGLSALGOLEKGL--VOAVASPPAALADK--443
GROS_g02469	490	LIGLIDGKA--498
GROS_g02470	487	VIGRFEDK--494
g951.t1	-----	-----
g952.t1	-----	-----
g25264.t1	-----	-----
g25265.t1	458	VVGRFEDK--465
g33471.t1	-----	-----
SPLIN_000763000	-----	-----
cds.Nab_57241_c0_seq1_m.50052	444	LVLAWA--449
cds.Nab_57240_c0_seq1_m.50041	444	LVLAWA--449
cds.Nab_57239_c0_seq1_m.50031	444	LVLAWA--449
cds.Nab_57238_c0_seq1_m.50020	444	LVLAWA--449
cds.Nab_57237_c0_seq1_m.50010	444	LVLAWA--449
cds.Nab_28879_c0_seq1_m.29300	444	LVLAWA--449

Gro DOG 0043 family alignment continued

GROS_g09987	657	NGI I IYCQTRADCE I I AQL I NKI GG I P TDVYHA GLSD I KREEVQQK WMS GEV P VV V A T I A F G M S V D K A T V R F V V H M P P Q N M A N Y Y Q E S G R A G R D G H R S F A R V Y Y S R D Y H G M M R Y K L A S S L K Q V K T K N V P K Q L V E Q R K	794	
GROS_g09990				
GROS_g05703				
g27892.t1				
g14819.t1				
g3594.t1				
GPLIN_001140700				
GPLIN_001140900				
GPLIN_000476700				
GPLIN_000919800				
cds.Nab_21641_c0_seq1_m.16185				
GROS_g09987	137		MAQ-	139
GROS_g09990	795	KE I O R G F D K M V E Y L E G T G C R H A L L G E Y F D D R L D A C A S H C D A C K T P Q R V R D S L K C M E L Q O S P A K R A K D A A E G G D A S L Y G G G R K G Q Q R E A E Y Y Q S M A T D D A R G D E T H G E R E E R L R V R V F L Q E L K K R R R A G E A S S P N F S	932	
GROS_g05703	135		I K R N D D K M E Q Q A E N W W A N M R T V D N R S G K F F F L S K N K -	171
g27892.t1	148		V E K -	150
g14819.t1				
g3594.t1				
GPLIN_001140700	147		MAQ-	149
GPLIN_001140900	143		MTQ-	145
GPLIN_000476700	143		MTQ-	145
GPLIN_000919800	146		MAQ-	148
cds.Nab_21641_c0_seq1_m.16185	162		G T D Y E D E N A H G E L L L L L -	178
GROS_g09987				
GROS_g09990	933	S A T N L L M A Q Q Q K V E E G K H G D P S L N P E N O G N F P F L R S T D T A N L S K I P G L T L R R R D Q N R S K L F E E L R M N O S S A V V N F P C T E H K M L T T E A L H A M G G L E E Q I F I S A K I G N I Y A N R I G A K I F E I R K M T S R G E V Y P R F E E S E Q	1070	
GROS_g05703				
g27892.t1				
g14819.t1				
g3594.t1				
GPLIN_001140700				
GPLIN_001140900				
GPLIN_000476700				
GPLIN_000919800				
cds.Nab_21641_c0_seq1_m.16185				
GROS_g09987	---			
GROS_g09990	1071	S N E		1073
GROS_g05703	---			
g27892.t1	---			
g14819.t1	---			
g3594.t1	---			
GPLIN_001140700	---			
GPLIN_001140900	---			
GPLIN_000476700	---			
GPLIN_000919800	---			
cds.Nab_21641_c0_seq1_m.16185	---			

Gro DOG 0203 & Gro DOG 0017 family alignment

GROS_g11017	1	MTFVNLCLISLTVLITMASIT	---	DAI	GFPS	GGGG	---	GC	---	SG	GGGGGGG	CCS	GG	---	GGGG	---	GGGG	---	GC	60	
GROS_g11020	1	MSNFMNCLIALIILVTFVSMI	---	DS	GMFM	---	QSC	---	CG	---	GGGG	---	GG	---	GGGG	---	CCG	---	CG	47	
g14106.t1	1	MCRFLALLFALL	---	---	---	---	---	---	---	---	GGGG	---	GG	---	GGGG	---	CCG	---	CG	14	
g19010.t1	1	MRFRLSVITLTVLITLMSI	---	KAM	GFJK	---	QSC	---	CG	---	GGGG	---	GC	---	GGGG	---	GGG	---	CG	51	
g30766.t1	1	MLSLSILLLLAIVPLMAQNK	---	GRA	VNPLFL	GGP	INP	QOL	---	---	SLRS	GES	IEA	FRNG	---	PS	NFN	FR	NS	77	
g19878.t1	1	MNIICFMSILTLILCCVNL	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	27	
g29037.t1	1	MLSLSILLLLAIVPLMAQNON	---	GRA	VNPLFL	GGP	INP	QOL	---	---	LOO	SLRS	GES	IEA	FRNG	---	PS	NFN	FR	85	
GPLIN_000183000	1	MSNCLNCLIVLIILVTFVSMI	---	DAI	GMFM	---	QSC	---	CG	---	GGGG	---	GG	---	GGGG	---	CCG	---	CG	56	
GPLIN_000182800	1	MSLMLNCLIALIILVTFVSMI	---	EAI	GMFM	---	QSC	---	CG	---	GGGG	---	GG	---	GGGG	---	CCG	---	CG	47	
GPLIN_000182900	1	MSKFMNCLIALIILMTFVSMI	---	DS	GMFM	---	QSC	---	CG	---	GGGG	---	GG	---	GGGG	---	CCG	---	CG	47	
GPLIN_001020100	1	MSKFMNCLIALIILVAFVSMI	---	EAI	GMFM	---	QSC	---	CG	---	GGGG	---	GG	---	GGGG	---	CCG	---	CG	47	
GPLIN_000183200	1	MQKVIITASLVVLLALSMI	---	NAI	GFPS	GGGG	---	CC	---	---	SGGG	---	GG	---	GGGG	---	CCG	---	CG	52	
GPLIN_000183400	1	MNQNVLFVSLFAMVYMLALII	---	DAI	GLPO	---	QNS	CC	QOQ	QSC	GGGG	GGG	---	GG	---	GLGG	---	---	---	60	
GPLIN_000183300	1	MCRSFFALIILILLVIM	---	HO	TNG	GLPO	---	QOQ	CC	QOQ	CC	GGGG	GGG	---	GGGG	---	CCG	---	CG	64	
GPLIN_001396600	1	MLFSKCSVSNFLMVLPLMFLIYVAMLSNSVSAI	---	GMPO	---	KME	CC	PA	MS	CC	---	AA	---	---	---	---	---	---	---	114	
GPLIN_000225800	1	MHFCLLGGOLLCCGLMLLNVE	---	---	---	---	---	---	---	---	GGG	---	CG	---	GGGG	---	GGGGGG	---	CG	39	
GPLIN_000968000	1	MTYFKCFILIVLTFNFI	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	59	
GPLIN_001043700	1	MLASSTIFVILFVITSC	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	38	
cds.Nab_23010_c0_seq1_m.18132	1	MACFLSNLLLLCLLLALVFL	---	LLQ	VNS	FGMP	PA	PA	FA	---	---	---	---	---	---	---	---	---	---	52	
cds.Nab_25727_c0_seq1_m.22943	1	MSSSILLALCLLVAFIALS	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	52	
GROS_g11017	61	RRKREAE	---	AV	OPHY	---	---	---	---	---	CP	KAW	EL	EE	IR	---	ED	AT	AS	116	
GROS_g11020	48	RKKREAAOD	---	AV	OPHY	---	---	---	---	---	CP	TEW	QI	LD	EN	IR	---	DD	AI	GS	105
g14106.t1	15	---	---	AV	OPHL	---	---	---	---	---	CP	TEW	RYL	ME	EA	MA	---	RD	AV	102	
g19010.t1	52	RRKRETETA	---	AV	OPHY	---	---	---	---	---	CP	TEW	KI	LD	GA	LP	---	ND	AL	95	
g30766.t1	78	RRKQPPPNQNF	---	DI	RR	AG	OP	TQ	HK	TA	AV	ON	PO	EE	EN	---	CP	TEW	KI	152	
g19878.t1	28	HSSETP	---	SL	FEV	---	---	---	---	---	CA	PEW	KL	IS	LP	---	DD	AS	100		
g29037.t1	86	RRKQPPPNQNF	---	DI	RR	AG	OP	TQ	OK	SA	IV	AN	PO	EE	EN	---	CP	TEW	KI	80	
GPLIN_000183000	57	RKKRETGAAL	---	AV	OPHY	---	---	---	---	---	CP	TEW	KI	LD	EN	IR	---	DD	AI	115	
GPLIN_000182800	48	RKKREAAOD	---	AV	OPHY	---	---	---	---	---	CP	TEW	QI	LD	EN	IR	---	DD	AI	105	
GPLIN_000182900	48	RKKREAAOD	---	AV	OPHY	---	---	---	---	---	CP	TEW	QI	LD	EN	IR	---	DD	AI	105	
GPLIN_001020100	48	RKKREAAOD	---	AV	OPHY	---	---	---	---	---	CP	TEW	QI	LD	EN	IR	---	DD	AI	105	
GPLIN_000183200	53	RRKRTG	---	AV	OPHY	---	---	---	---	---	CP	TEW	KL	LE	EN	IR	---	DD	AI	105	
GPLIN_000183400	61	RRRRBAKELGVSAALCPHF	---	---	---	---	---	---	---	---	CP	TEW	QI	LD	EN	IR	---	DD	AI	105	
GPLIN_000183300	115	RRKRRGS	---	AIG	AOV	---	---	---	---	---	CP	SEW	EL	LE	EN	IR	---	DD	AI	105	
GPLIN_000713100	65	RRKRE	---	AV	OPHI	---	---	---	---	---	CP	TEW	RL	ME	EA	MA	---	RD	AV	119	
GPLIN_001396600	60	RRKRE	---	AV	OPHY	---	---	---	---	---	CP	TEW	RE	IK	IS	LP	---	DD	AI	118	
GPLIN_000225800	40	EDYESDVRNTHFI	---	TS	AV	PH	TS	SY	RR	KMK	V	---	---	---	---	---	---	---	---	117	
GPLIN_000968000	40	EDYENDVRNTHFI	---	TS	AV	PH	TS	SY	RR	KMK	V	---	---	---	---	---	---	---	---	117	
GPLIN_001043700	99	GGPMNQGFIALQOP	---	OT	VF	---	---	---	---	---	---	---	---	---	---	---	---	---	---	217	
cds.Nab_23010_c0_seq1_m.18132	53	RRRRE	---	AV	OPHI	---	---	---	---	---	CP	TEW	RL	ME	EA	MA	---	RD	AV	108	
cds.Nab_25727_c0_seq1_m.22943	53	RRKRE	---	AV	OPHL	---	---	---	---	---	CP	TEW	RE	EL	MS	IR	---	DD	AI	106	
GROS_g11017	117	FLVTC	---	AG	---	EA	KN	GA	---	---	---	---	---	---	---	---	---	---	---	164	
GROS_g11020	106	FLVTC	---	SAD	---	EA	KN	GA	---	---	---	---	---	---	---	---	---	---	---	154	
g14106.t1	63	FLVNGAAKR	---	GES	AL	AV	GEK	---	---	---	AP	EA	AV	---	VT	NI	RF	AS	GG	123	
g19010.t1	96	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	97	
g30766.t1	153	FLMVLG	---	---	---	DE	EKA	---	---	---	FG	KT	IK	LE	---	KG	---	OV	VE	222	
g19878.t1	86	FLMVLG	---	---	---	DE	EKA	---	---	---	FG	KT	IK	LE	---	KG	---	OV	VE	222	
GPLIN_000183000	116	FLMVLG	---	---	---	DE	EKA	---	---	---	FG	KT	IK	LE	---	KG	---	OV	VE	228	
GPLIN_000182800	106	FLVTC	---	SAD	---	EA	KN	GA	---	---	---	---	---	---	---	---	---	---	---	164	
GPLIN_000182900	106	FLVTC	---	SAD	---	EA	KN	GA	---	---	---	---	---	---	---	---	---	---	---	154	
GPLIN_001020100	106	FLVTC	---	SAD	---	EA	KN	GA	---	---	---	---	---	---	---	---	---	---	---	154	
GPLIN_000183200	110	FLVTA	---	---	---	DE	EKA	---	---	---	FG	KT	IK	LE	---	KG	---	OV	VE	156	
GPLIN_000183400	124	FLVTA	---	DAN	---	ED	KG	---	---	---	VA	RV	HF	---	SG	GG	---	OV	VE	172	
GPLIN_000183300	117	FLVTA	---	TAND	ATV	---	EA	AD	---	---	AA	KA	---	---	---	---	---	---	---	223	
GPLIN_000713100	120	FLVMGA	---	LAGE	IKR	---	AD	SK	---	---	LP	NE	HL	---	---	---	---	---	---	174	
GPLIN_001396600	119	FLVIGMLDEKHVEEKG	---	---	---	SD	VA	---	---	---	ME	LL	HM	---	---	---	---	---	---	177	
GPLIN_000225800	118	FLVMG	---	---	---	LA	NO	---	---	---	IA	KK	---	---	---	---	---	---	---	162	
GPLIN_000968000	118	FLVMG	---	---	---	LA	NO	---	---	---	IA	KK	---	---	---	---	---	---	---	162	
GPLIN_001043700	218	FLVAG	---	---	---	PA	DR	---	---	---	FA	OK	---	---	---	---	---	---	---	262	
cds.Nab_23010_c0_seq1_m.18132	109	FLVVG	---	LAS	---	EL	TE	---	---	---	VP	NO	---	---	---	---	---	---	---	159	
cds.Nab_25727_c0_seq1_m.22943	107	FLVNGANGORASGEKE	---	ST	AL	AV	---	---	---	---	EO	KL	---	---	---	---	---	---	---	173	

Gro DOG 0199 family alignment

<i>GROS_g07013</i>	1	MIFQTL-LTIAMFLADFINGDIDNAVKGKKCAKDDVVNKLTVYEDGAEAECGPLPCGATA-STCSE-----DMSTCKTPSEVFSGMKWAQNGSLLRCSSIRATNKVYIGTDLVTLGSYYTGGTVPE	123
<i>g30503.t1</i>	1	MMFLCA-LISSLLLIANCV-GALONAVKGKRCANQVVNKLTVYEDGAEAECGPLPCGTTSYDTCSD-----DLSSCKAPTEVFSGMKWAHNGSLLRCCLQIKDKVYIGTDLVTLGSYYTGGIVPE	123
<i>GPLIN_001036900</i>	1	MIFQPL-LIITAIFFLLANFINGDIDNAVKGKKCAKDDVVNKLTVYEDGAEAECGPLPCGATA-STCSEVQIFFQFKFKEIQDMSTCKTPSEVFSGMKWAQNGSLLRCSSIRATNKVYIGTDLVTLGSYYTGGTVPE	137
<i>cds.Nab_17897_c0_seq1_m.12768</i>	1	MSFLHLALLSFLLLFGSV-QSLRNAVKGKKCS TNQVVNKLTVYEDGAEAECGPLPCGGTG--VCAE-----DLSSCKAASEIFSGMKWAHNGSLLRCALHAANKVYIGTDLVTLGSYYTGGVVPE	122
<i>GROS_g07013</i>	124	DLMSPGGGLI EYDFVANARTEGGGVRVWVYRVVCG-GQKIASNSA-----GGQGHAPPEASEINDGGSCHKFSDGQNVPAEAEINAAAGPEEAESEADQNPLOYPSPSRRTMAGY-----	235
<i>g30503.t1</i>	124	DRMSQGGGLI EYDFVANARTEGGGVRVWVYRAVCGAGHQISAOQQQVAKQQQVAAGPQPAAVETGSETDGQPEH-----VQQIKAPEEAESEADQNPLOYPSPSRRTMAGY-----	237
<i>GPLIN_001036900</i>	138	DLMSPGGGLI EYDFVANARTEGGGVRVWVYRVVCG-GQKIASNSVQ-----GGQGHAPPEASEINDGQTHCHKFSDGQNVPAEAEINAAADPTPEAESEADQNPLOYPSPSRRTMAGY-----	250
<i>cds.Nab_17897_c0_seq1_m.12768</i>	123	KMMG-GGGTI EYDFVANARTEGGGVRVWVYRIVCG-G-----GGQKAGVGSKGDQQLAE-----	176

Gro DOG 0149 family alignment

GROS_g05985	1	-----M I I L L S T I V A L-----	11
GROS_g13738	1	-----I N T M I I L L S T I V A L-----	14
GROS_g09111			
GROS_g09869	1	MKQKQKHKS VTCQKKLDERKKKWLKMKMKQKPAAKKHSSEEMVVIPTNLE	50
GROS_g03837	1	-----M F Y F X Y L F F S L K M-----	13
GROS_g12421	1	-----M L G S T V P I S L R I L P L F H I L I R I-----	23
g21345.t1	1	-----M L L I I F I L Y L F S A V-----	23
g21349.t1	1	-----M T L F L L V L T G S F L-----	98
g17738.t1	1	-----M H I N C I D Q L F Y I W I C L-----	24
GPLIN_001355500	1	-----M I S L P K A L I S A F I L K F I I Q R S V A D P K N Q L G F A Y T S V S N A G N Y G D D V D G D I L Q F I V I G D I G G E E Y E T N S T V Y E S T V A D S T V V K G T V K A S T T V N G T V K A S T M V K H R P T Q T K A V A D A I D E L S K E K G K V N F L L N T G D N	135
cds.Nab_59091_c0_seq1_m.56063	1	-----M R H S I L L L L L F C V F G-----	38
GROS_g05985	12	-----L G I A N G V G N N P R-----	49
GROS_g13738	15	-----L G I A N G V G N N P L-----	53
GROS_g09111	1		
GROS_g09869	51	-----F G A E E E E E S E L Y E Q T P F I S A S L G T D N L M R L F Q I Q F C-----	114
GROS_g03837	14	-----D G V A G K I G H R R I-----	23
GROS_g12421	24		
g21345.t1	99	Y A E L L K P F D E F I K F R Y P V T V E A T K P E G V L K S S A R G D L D G G G F V G V I D G D O O G E K K A T L R F Y A A G N V Y V E N K V D H A V K V E K E L H E-----	211
g21349.t1	25		
g17738.t1	24	-----I D A A G C N G H F D V D V P N Y N N F C K N L N-----	76
GPLIN_001355500	136	F Y E N G V T F A D V I P R F T E N F E K V Y K H Q S L N V P W Y T I A G N N D A E S T D G I R P Q I E M L E G K W T F P A Q Q Y V V Y H T F S K V E D O F G N R V O E R R S V R I M I D T T V L C G T A S T N A K K H F T W L R K L E A K R R R V N F L F V G H P L H L	274
cds.Nab_59091_c0_seq1_m.56063	39	-----E M V K E G G G E N D E E-----	78
GROS_g05985	50	E V L A T-----	88
GROS_g13738	54	Q V L T P-----	99
GROS_g09111	24	P O L V D-----	62
GROS_g09869	115	N A F C F-----	162
GROS_g03837	51	N S I-----	81
GROS_g12421	48	H A L S N-----	59
g21345.t1	212	S A I V E-----	258
g21349.t1	52	E L L E K-----	89
g17738.t1	77	C S L E T-----	109
GPLIN_001355500	275	D K E R G Y P C A A R V H A L M R Y Y Q V A Y F A G H T H S L Q Y A D L S E A G O T N L G I D S I D T K F I V S G A G S R M E N A I L E S R F P K A Y L K F M K F V Y P T N V G I V S K V K G E T T H S R G A F D G G A F V G V E I N A K E M L A K L N F Y A A G M F Y N K G D T K	413
cds.Nab_59091_c0_seq1_m.56063	79	E K L D N F W L L K E K W A E Q K A E K K Q K I E K T K E Q K S D E D K E A K K F A N K K E L I E H G L-----	156
GROS_g05985	89	--S W T D P E G A H E F-----	166
GROS_g13738	100	--L V W T V E G S D H S-----	177
GROS_g09111	63		90
GROS_g09869	163	--E P T S K S V N N G T F-----	245
GROS_g03837	82	--S Q V T P S H A Y G D F-----	138
GROS_g12421	60	--Q K L I G T G T R Q F-----	337
g21345.t1	259	--F T W T D D E N D Q K-----	112
g21349.t1	90	--P I W T D Q-----	119
g17738.t1	110	--S A S V R H R F P I F-----	232
GPLIN_001355500	414	K P I E K W H K F E G E I Q M K P R R I D K T D E N V D I C N E F L E L G P Q L V D D D H F K M D D T E R I Q S I L G T F F Y A G E A F D A L R E K L G-----	522
cds.Nab_59091_c0_seq1_m.56063	157	--E W S A K D S S-----	232
GROS_g05985	167	M E Q V L D K-----	270
GROS_g13738	178	M Q V L D K-----	270
GROS_g09111	91		149
GROS_g09869	246	F D T L F E R T R S Y D D G P I D F Y V G A W V Y Y E Y L R S Q I L G K K H L L P K F Y A L F D L E N I Q S S E K-----	353
GROS_g03837	160	L R K L L P G-----	240
GROS_g12421	139		162
g21345.t1	338	M E Q T L G H-----	424
g21349.t1	113	I E E T L G H-----	199
g17738.t1	190	L A R F L A D-----	272
GPLIN_001355500	523	L M K T L F N-----	608
cds.Nab_59091_c0_seq1_m.56063	233	L E Q L V L G-----	333

Gro DOG 0149 family alignment continued

GROS_g05985	271	D F M G I A R G	-----DKVYH A F F	K N G I E L D V A K	T O L A G T L H N D A E Y L S K	I E L N D Y S L L L G V R F L D	-----K A P T K N D N N C	-----	I H A T Q R D	-----C R R R	-----P D A K Y E E	L S H L C	362					
GROS_g13738	280	D F M G I N R D	-----G R L I H A F F	K N G I E L D V K K Y K O L A	G T L Q N D A Q Y L S E	I E L N D Y S L L L G V H F L D	-----K A P T K N D N N C	-----	I H A T Q R Q	-----C M R L	-----P D A K Y E E	L S H L C	371					
GROS_g09111	141	N F F G G T K E	-----L L P D K M E A M F	K N G I L L D K A M H K N I	T E R L K D F E I L A K N N V D W S I	I F G V H F V S	-----N E E H K I S D E S P V D	-----	G I L P Y F N G R C I Q	-----C R T H	-----P D A S E F L D	D A T L R	244					
GROS_g09869	354	D F E R Y N P E	-----D K D Q T M L E G L F	R H G I C I E A K H H E Q L A K T	L E N D T L F L K S H N I	I D Y S L L L G V H F I D	-----N E K L P A E M R G E	-----	L F G V L R G T Q D N	-----C N H E	-----F R N K Y E N	D K L L R	454					
GROS_g03837	241	D F	-----N O E F	P D G I L L L D A Q V Y E T L M D V I K	P D Q L V L E S F K I M D Y S L L M G V H N I D	-----	-----G D E G G O T M S C N	-----	S E Q L H C F L F P D D E G G V R A R	-----	-----S T K G E R L I	329						
GROS_g12421	163	D L Y D E T E K K M S E A N R A M L K E K R E	G C F F R N G I L L E O D V Y T K F I G I L S R O A L F L K E S	G R T D Y S L L L G L T F E E L S E R O M E L L Q K N N A N	-----	-----	-----	-----	I L V A S C E D	-----G S R D	-----R Q T Y S N V I	E N F V	272					
GROS_g21345.t1	425	H Y F G W Q D Y	-----G M E A M F	P E G I V L D Q R D H O M L T E K L K R O F T V L A E	I N V D W S I I F A V H F E D	-----	-----O G R D I E M A D E N	-----	A R S P Y I R A H C N Q	-----C R S H	-----	508						
GROS_g21349.t1	200	H F F G W Q D Y	-----D G M C A M F	P E G I V L D K Q D H Q R L T E K L K R D F E I L S K	D V S D W S I I F A V N F E D K K E Q N D A T I	I D D E N A T	-----	-----	S S Y I R A H C N K	-----C R S H	-----P N S S E E T	293						
GROS_g17738.t1	273	D F	-----C H I F G R S G I L L D R S I V R S L M V A I R M D G H F L E R R K	I M D Y S L L L G V H F N L E	-----	-----	-----	-----	E G K S A E E Q S A E R A M R L Q E M P E G I R Y W Y F L K I N A L Y G P L N I	-----L P A W	-----	N S R G G K L L	374					
GROS_gPLIN_001355500	609	N F F G G T K E	-----L L P D K M E A M F	K N S I L L D K A T H K N I A E R L K D F E I L A K N D V N D W S I	I F G V H F V G	-----N E K H K I S V G E S S V D	-----	-----	C I L P Y F R G R C N Q	-----C N H	-----P D A S E F L L N	N A T L R	713					
GROS_gcds.Nab_59091_c0_seq1_m.56063	334	N F L G T E E N M L F A T E R F P K E L V M E A L F	P E G I L L S P O O H K E V M G K L S R D I E L L S S K G I N D Y S L L L A A R F V N	-----	-----	-----	-----	-----	S G G T N P P I V N G I E G I P A M C N R	-----	-----C R P S Y S T G D G A Q R R D	N A H L L	454					
GROS_g05985	363	L Y M A I V D I I T P S S E K V O D Y F M N R M K Q	R K A E F A N I E S M D Q F	-----N G F O L E T I S P I P A P K Y V G R F L G T V L G C I F R	-----P T F S P G N K Q	-----	-----	-----	I M L N G L S O G I C H S F G E	-----	-----K R D T V F A K K N A I S D K K E A K O W D Y E	484						
GROS_g13738	372	L Y M A V D L L T P S T K E A Q E H F M M L K K H R P G E F D N I K T M D A F	-----N G F H L E T V S P I P A P K Y V G R F L G T V L G C I F R	-----P T F S P G N K Q	-----	-----	-----	-----	I M L N G L S O G I C H S F G E	-----	-----E R D T F E K K K K	-----E L S N K W D N K	489					
GROS_g09111	245	L Y I G I V D F T P F S E E R Q T R Y I E G L Q K F I K N V N S N H L M D Y V K	-----	-----A G S V W E T F G A I P P E M Y G H R M L S F S M Y C A F R W Q N D Q T V D A P	-----	-----	-----	-----	E A I R Q A L I Q A F G D S Y G K Q L E K S E E S M L A M L D	-----	-----	D K S K D W K K E L O R G	371					
GROS_g09869	455	L S I G I I D I L Q T Y D	-----F F R R C E N A Y K S S K S L F I Q S E G L F P P S T R R K Y R P N S V K K P N V Y R K F L D A I L N V L V R	-----	-----	-----	-----	-----	K D C G A T E T V D G D E Q D E H Y S A E E D E S E E	-----	-----	551						
GROS_g03837	330	I Y L G I I D I L Q S Y R	-----L F K K L E H M K S V L H D G	-----	-----	-----	-----	-----	S I S Y T N P G F Y A M F E	-----	-----	O T F L S K V F R	-----R T S V D N I Q P	-----	V R H P S S K F R S F V H S Y I A M K Q T P M R Q Q Q R P R I S	-----	E H E E Q R V E D T M D	435
GROS_g12421	273	I P G I I D I L A O F M R K K I D L V N R V V V K H E T E K F L K E S G R T D	Y S L L L G L T F E E L S E R O M E L L Q K N N A N I L V A S C E D C S R D R G Q T Y S N	-----	-----	-----	-----	-----	V I E N F V I P G I I D I L A O F M R K K I D L V N R V V V K H E T E K K A K S A I A P E	-----	-----	405						
GROS_g21345.t1	509	-----	-----	-----	-----	-----	-----	-----	N S E L R N N Q P V H I Y M G I V	-----	-----	D F L M P Y S E	-----	Q R K A Y M K N L K E K	548			
GROS_g21349.t1	294	-----	-----	-----	-----	-----	-----	-----	F G P I P P T M S E R M L A F F M G C A M R W P N A A S V S	-----	-----	E D A R R T M I G G F C D A Y G K	-----	P V D G E E M R A K T E	-----	A K G Y S L L E E E R	365	
GROS_g17738.t1	375	V Y I G I I D I L Q T Y G	-----	Y R K R L E H L F K W L L F Q	-----	-----	-----	-----	R N F S V A K P R Y Y A R F F H N H M A C S I F R	-----	-----	D T F L	-----	432				
GROS_gPLIN_001355500	714	L Y I G I V D F T P F S E A R K E R Y M E G L K K F I K N V N S N H L M D Y V K	-----	-----A G S V W E T F G A I P P E M Y G H R M L S F S M Y C A F R W Q K E E A L D A P	-----	-----	-----	-----	E A I R Q A L I Q A F G D S Y G K N L D K S E E S I L A M L D	-----	-----	D K S K D W K N E L L O R D	840					
GROS_gcds.Nab_59091_c0_seq1_m.56063	455	V F V S L I D M I T P F D E E R E K L Y I R Q L K R F G I D L G T D M G K N L	-----	-----S D Q M E N V S P I P A E I K R E L G V V S G C I F R	-----	-----	-----	-----	Q N F S A Y A P T R O W I	-----	-----	A V R Q G L V Q T I G M E F G K S R K V K E T L M W S L D	-----	P V R R D V S	577			
GROS_g05985	485	A F O H K F D G V	-----	G L Y V F E O F K E E M D	-----	-----	-----	-----	-----	-----	-----	506						
GROS_g13738	490	A F O H K F D G V	-----	A L Y V F E O F K E E M D	-----	K R K M S W D L L S L S Q T P S K F I N E P I G K E Q F V Q Y M	-----	-----	-----	-----	-----	545						
GROS_g09111	372	N Y L F K H H T V	-----	T I F G V P L F K R L M E	-----	K W E K Y F W N F L N Q L A E N K L Q E N G T F V H E N R A F V V R K L S K	-----	-----	-----	-----	-----	432						
GROS_g09869	552	L F I F R A K K M E E S D Q Q T D Q S T T L I S P S K K K K H R O L L S K Q I K Q D I V E D L K K L D L M N V S S I	-----	-----	-----	-----	-----	-----	-----	-----	-----	609						
GROS_g03837	436	S V I L Q Y Q	-----	Q O L G G P L S O R G T S	-----	E R R S Y Y A K M T N I G P S G E	-----	-----	-----	-----	-----	472						
GROS_g12421	406	D F Y F R F M	-----	C A V A T K V F R R T I N	-----	-----	-----	-----	-----	-----	-----	425						
GROS_g21345.t1	549	M K V D R G A E L	-----	M N Y V K G V V L D N L E	-----	S K I	-----	-----	-----	-----	-----	573						
GROS_g21349.t1	366	N V Q D D Q E R	-----	T L Y G V L F P E L M E	-----	E K H H K H W T L M N V L A E D	-----	-----	-----	-----	-----	403						
GROS_g17738.t1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----					
GROS_gPLIN_001355500	841	T Y L F R N H T V	-----	T I F G V P L F K R L M E	-----	K W E K Y F W N F L N Q L A E D L K K E K E T F V O K N R A F V V R K L S K	-----	-----	-----	-----	-----	901						
GROS_gcds.Nab_59091_c0_seq1_m.56063	578	V A V D K K Y T L	-----	S S F G F G L F K A L M L D L T V K T P S W E L I S L A N G K L L E E E Y E H E Q G O K Q K E S I G T D E I V L T S E D K E F R I K S L K K E Q L E Y F R A N L F I Y Y S D Q L I K P A G H S T P R Y Y A I F O L K D N A Y N F L M K K G M L E K A K K E G I	-----	-----	-----	-----	-----	-----	-----	714						
GROS_g05985	507	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	519						
GROS_g13738	546	-----	-----	E A I S A N D G I E K A G K K Y G N V G K V G N R	-----	-----	-----	-----	K D E N A R F L O R Y Y A I L S Y E Q	-----	-----	581						
GROS_g09111	433	-----	-----	N D V K N V K T E L G C A R O G O P K Y Y Y A L L D L O K N G T K D A H E F Y A V T S D K K A G N G K	-----	-----	-----	-----	-----	-----	-----	485						
GROS_g09869	610	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	641						
GROS_g03837	473	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	502						
GROS_g12421	426	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	448						
GROS_g21345.t1	574	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	596						
GROS_g21349.t1	404	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	433						
GROS_g17738.t1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----						
GROS_gPLIN_001355500	902	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	954						
GROS_gcds.Nab_59091_c0_seq1_m.56063	715	E D E N E V L S K I Q O M G G K P Y F Y V I P A K D D D A S A G E E N K P K T T G E E N T R K R K D H G G O A E G H C H A D K K R S A N A N I G K S D M N R	-----	-----	-----	-----	-----	-----	-----	-----	-----	820						

GLAND11 family alignment

GROS_g02394	1	MKSSLVSIIFITVFLLSVVAEFA-----LSAKAFSGGLANLVAKAFVNSVNTWDS--TFIGIGNDGENRYKINLDSLDKLNKFRVPMFAGGLEKLLKSRVDFV-----KENIGTKGSEVLMFSLML	119
g22583.t1	1	-----MNL LAP IFLF-LSAGFG-----SAYVIKPLNNGELANLVAKAFVNTDPCFKY--NIFGLGSRGDLRYQIDIDKVRVTKCKFRMPMP IGGGLEKLLTAKKIPTA-----KNDGARGSERAVANDKLV	115
g26304.t1	1	-----MNL LAP IFLF-FSSGFG-----SAYVIKPLNNGELANLVAKAFVNTDPCFKY--NIFGLGSRGDLRYQIDIDKVRVTKCKFRMPMP IGGGLEKLLTAKKIPTA-----KNDGARGSERAVANDKLV	115
g24562.t1	1	-----MNL LAP IFLF-LSAGFG-----SAYVIKPLNNGELANLVAKAFVNTDPCFKY--NIFGLGSRGDLRYQIDIDKVRVTKCKFRMPMP IGGGLEKLLTAKKIPTA-----KNDGARGSERAVANDKLV	115
GPLIN_000714100	1	MKSSLVSIIFITVFLLSVVAEFA-----SSAKISPFSGGLANLVAKAFVNSVNTWDS--TFIGVGYKDGEDRYKINLDSLDRLNLSFRVPMFAGGLEKLLRSQRVDFV-----KENIGTKGNERVLAFLYKLM	121
cds.Nab_59725_c0_seq1_m.58226	1	MSAQMCSFSIALLA IALLASLAQTQFEGNPLAAASEAHMSNEKLANQVAKAVLESINCWSAASKSIF--VQMDNKYRIDVGO LKAPCTFRVPMPTKRGLKLMGERRVNSA-----RDNDAAKGS TRSATKEAII	129
cds.Nab_59727_c0_seq1_m.58236	1	MSAQMCSFSIALLA IALLASLAQTQFEGNPLAAASEAHMSNEKLANQVAKAVLESINCWSAASKSIF--VQMDNKYRIDVGO LKAPCTFRVPMPTKRGLKLMRERRVNSA-----RDNDAAKGS TRSATKEAII	129
cds.Nab_59726_c0_seq1_m.58231	1	MSAQMCSFSIALLA IALLASLAQTQFEGNPLAAASEAHMSNEKLANQVAKAVLESINCWSAASKSIF--VQMDNKYRIDVGO LKAPCTFRVPMPTKRGLKLMRERRVNSAWNDNDAATRENDARGFTSSVTKEAII	136
GROS_g02394	120	ELMEKLG N--VLVTDKTLRGIIDVYEKKADEAAKRAAAMPKKVDPAQATARRTW-----VAKNDGSDSKGKD----SPTRERIVTIT-KQVWME--DADTPRWRSLTFSRKNEKOKSWLESDSEDESPTE	241
g22583.t1	116	KLLEEIGTEQLVVKDKTFRGKIDVYKTEVEAKKAGPKQKEDGPKTANMEDARRKR-----SKRVKFDIDSDSESPF-----PRSRWKRITTYSRFVWRHYSTSKPGVRRTEKVVRERTVTVEETSDSDSESPRRR	245
g26304.t1	116	KLLEEIGTEQLVVKDKTFRGKIDVYKTEVEAKKAGPKQKEDGPKTANMEDARRKR-----SKRVKFDIDSDSESPF-----PRSRWKRITTYSRFVWRHYSTSKPGVRRTEKVVRERTVTVEETSDSDSESPRRR	245
g24562.t1	116	KLLEEIGTEQLVVKDKTFRGKIDVYKTEVEAKKAGPKQKEDGPKL-----SPDRSRAREN-SY-----GGDGHFLGR-----NEETPGSPRQS	143
GPLIN_000714100	122	ELMEKLG N--VSVTDKTLRGIIDVYEKKADEAAKRAAAMPKKVDPAQATARRTW-----VAKNDGSDSKGSDKGDSPTRHIVTIT-KQVWME--DADADTPRWRSLTFSRKNDKOKSWLESDSEDESPTE	250
cds.Nab_59725_c0_seq1_m.58226	130	VLLVTVGF IIVVSDDI LDTIRRRRVDKEARTAAA-----NQISNAKTFGAPHPTVKKPHATPAAVAPPKPKPFAFNWRTA-QWPPAPATFAGFAPPLKLSAL-----RKKSAFSPKTPPAPAA	250
cds.Nab_59727_c0_seq1_m.58236	130	VLLVTVGF IIVVSDDI LDTIRRRRVDKEARTAAA-----NQISNAKTFGAPHPTVKKPHATPAAVAPPKPKPFAFNWRTA-QWPPAPATFAGFAPPLKLSAL-----RKKSAFSPKTPPAPAA	250
cds.Nab_59726_c0_seq1_m.58231	137	LLESFDSFYVWVNDHILDTIRRRRVDKEARTAAA-----NQISNAKTFGAPHPTVKKPHATPAAVAPPKPKPFAFNWRTA-QWPPAPATFAGFAPPLKLSAL-----RKKSAFSPKTPPAPAA	267
GROS_g02394	242	RSRRRSVQI LDNDSDEARPGARRSPTFERSASFNVRWTSRTT-----TTTSW--VDEDNVERRKVFGEVSVVSPPE-----ETRQRVTEVERRTENVLI LKLDFMDSSEFLDKVFN	350
g22583.t1	246	RASRSRRRYSSSSDS-----SNHRWIQNR-----RTGNWSKVYVDESSDRSKWKVGEKSPS-----KSRKAKAVKEDPSKDSIAVKLLTVEAGAVLEKFMG	338
g26304.t1	144	-----SNHRWIQNR-----RTGNWSKVYVDESSDRSKWKVGEKSPS-----KSRKAKAVKEDPSKDSIAVKLLTVEAGAVLEKFMG	216
g24562.t1	195	AKAPLQXILPTIGGSR I-----DALAI GAKW-----W--LMTVRESSDRSKWKVGEKSPS-----KSRKAKAVKEDPSKDSIAVKLLTVEAGAVLEKFMG	281
GPLIN_000714100	251	RSRRRSVQI LDNDSDE--PGARRSPTFERSASFNVRWTSRTT-----TTTSW--VDEENERRKVKFGEVSVVSPPEGGGRRTETRQRVTEVERRTENVLI LKLDFASEFLDKVFN	365
cds.Nab_59725_c0_seq1_m.58226	251	SADLLKRLKMTVEEQKK-OGSVGFQSKPPAPAAPAALFTSKSTHFAPPAAPAAFTFSASRVPSQKRKW--NILPTNPFGRVVEQKASGRPTAAAEHNRRSTWARRGVTF AAGAAKGFPIVNLNDGVHLLDLIA	386
cds.Nab_59727_c0_seq1_m.58236	261	SADLLKRLKMTVEEQKK-OGSVGFQSKPPAPAAPAALFTSKSTHFAPPAAPAAFTFSASRVPSQKRKW--NILPTNPFGRVVEQKASGRPTAAAEHNRRSTWARRGVTF AAGAAKGFPIVNLNDGVHLLDLIA	396
cds.Nab_59726_c0_seq1_m.58231	268	SADLLKRLKMTVEEQKK-OGSVGFQSKPPAPAAPAALFTSKSTHFAPPAAPAAFTFSASRVPSQKRKW--NILPTNPFGRVVEQKASGRPTAAAEHNRRSTWARRGVTF AAGAAKGFPIVNLNDGVHLLDLIA	403
GROS_g02394	351	YFGSLGAAR-----KDEGGVORMCFALKNLDPKDKLNVAVLG-TPSVANFFGFEFKKAL-NDVGLNNOAELRLVWKSNSCAFDLALIFALAVTIDLLSKLMAE-----RMVPTTEELDDSTELVNRLVKLFD	470
g22583.t1	339	YFSKFIAKK-----GKDKGDKLCPVLKFEFDSKHAQLKTIID-VSFVAAAFKEYKKTAL-KSIDLYQEEQIRLAPKSVTCAYDLSLIYVAAQFGKHLLEKIKAD-----RMDTGKELTESKQELVSI	457
g26304.t1	217	YFSKFIAKK-----GKDKGDKLCPVLKFEFDSKHAQLKTIID-VSFVAAAFKEYKKTAL-KSIDLYQEEQIRLAPKSVTCAYDLSLIYVAAQFGKHLLEKIKAD-----RMDTGKELTESKQELVSI	457
g24562.t1	282	YFSKFIAKK-----GKDKGDKLCPVLKFEFDSKHAQLKTIID-VSFVAAAFKEYKKTAL-KSIDL-----VTCAYDLSLIYVAAQFGKHLLEKIKAD-----RMDTGKELTESKQELVSI	388
GPLIN_000714100	366	YFASLNAAR-----RDEGGVORMCFALKNLDPKDKLNVAVLG-TPSVANFFGFEFKKAL-NDVGLTNOAELRLVWKSNSCAFDLALIFALAVTIDLLSKLMAE-----RMVPTTEELDDSTELVNRLVKLFD	485
cds.Nab_59725_c0_seq1_m.58226	387	YFAGLKAFNRRLFQQQNRTE SFNLRGRVLYHSDGHGLEA I LKRVNDL TNAFGCKEMIGTEGLTRSNVDPLEFENCAYDLSLIY ILSGLGGLVNSI VARIQGTGRN I LIDLIGKIDDILVGIHKELVRF	525
cds.Nab_59727_c0_seq1_m.58236	397	YFAGLKAFNRRLFQQQNRTE SFNLRGRVLYHSDGHGLEA I LKRVNDL TNAFGCKEMIGTEGLTRSNVDPLEFENCAYDLSLIY ILSGLGGLVNSI VARIQGTGRN I LIDLIGKIDDILVGIHKELVRF	535
cds.Nab_59726_c0_seq1_m.58231	404	YFAGLKAFNRRLFQQQNRTE SFNLRGRVLYHSDGHGLEA I LKRVNDL TNAFGCKEMIGTEGLTRSNVDPLEFENCAYDLSLIY ILSGLGGLVNSI VARIQGTGRN I LIDLIGKIDDILVGIHKELVRF	542
GROS_g02394	471	KFLDGLKVKGAERSTEVDR--RVEAATEIEEDLK-----TLEKVR--LT-----DATVTDATGKVMALQRVARDSDQTEPRSLRTLQIWDALMKKDTHWQIRKAMTNLEKQADSEGSTNGLTCMET	586
g22583.t1	458	GMEAEMLK-KGKPNESTRDKALTEAQEVIREQAE-----VLKEHRMKPN-----EEVVGDSMGEVLHALLRRVANHKDSAHOSTLKELGHWYHGLVDDGFWL I KKSVSGLKCGPGAD---GLK KCLNG	572
g26304.t1	336	GMEAEMLK-KGKPNESTRDKALTEAQEVIREQAE-----VLKEHRMKPN-----EEVVGDSMGEVLHALLRRVANHKDSAHOSTLKELGHWYHGLVDDGFWL I KKSVSGLKCGPGAD---GLK KCLNG	450
g24562.t1	389	GMEAEMLK-KGKPNESTRDKALTEAQEVIREQAE-----VLKEHRMKPN-----EEVVGDSMGEVLHALLRRVANHKDSAHOSTLKELGHWYHGLVDDGFWL I KKSVSGLKCGPGAD---GLK KCLNG	503
GPLIN_000714100	486	KFLDGLKVKGAERSTEVDR--RVEAATEIEEDLN-----TLEKVR--LT-----DATVTDAMGKVMALQRVARDSDQTEPRSLRTLQIWDALMKKDTHWQIRKAMSLSLEKQADTEGNTNGLTCMET	601
cds.Nab_59725_c0_seq1_m.58226	526	NPAALVLPFGGPKPNRND--TEKIANIVSRERLRWKEENIADWTVGVVVEVQHALS FYAEFAS IKKELGELMQLKVVDA SHPPAPADSP LINSHN-----WHIRGALS KLTI CGSRN	638
cds.Nab_59727_c0_seq1_m.58236	536	NPAALVLPFGGPKPNRND--TEKIANIVSRERLRWKEENIADWTVGVVVEVQHALS FYAEFAS IKKELGELMQLKVVDA SHPPAPADSP LINSHN-----WHIRGALS KLTI CGSRN	648
cds.Nab_59726_c0_seq1_m.58231	543	NPAALVLPFGGPKPNRND--TEKIANIVSRERLRWKEENIADWTVGVVVEVQHALS FYAEFAS IKKELGELMQLKVVDA SHPPAPADSP LINSHN-----WHIRGALS KLTI CGSRN	655
GROS_g02394	587	VMGTVGK Y I DDVDDWFKTNKPI DMEDANWLVDIQFLIRWKS I	628
g22583.t1	573	AKGKLAEVAADI EDWNGSLKQI SPDDAKWLSAIK FELTWSDI	614
g26304.t1	451	AKGKLAEVAADI EDWNGSLKQI SPDDAKWLSAIK FELTWSDI	492
g24562.t1	504	AKGKLAEVAADI EDWNGSLKQI SPDDAKWLSAIK FELTWSDI	545
GPLIN_000714100	602	VMGGIGK Y I DDVDEWFKAHRPI DMEDANWLVDIQFLIRWKS I	643
cds.Nab_59725_c0_seq1_m.58226	639	-----WLT-----	641
cds.Nab_59727_c0_seq1_m.58236	649	-----WLT-----	651
cds.Nab_59726_c0_seq1_m.58231	656	-----WLT-----	658

Gro DOG 0015 family alignment

GROS_g04556	1	-----MFRTRIVLALFSL-AFLVAMCTAMERRNHS-----	28
GROS_g10456	1	-----MDNOROFANFCYLLLI-I-IIASICA-----	25
GROS_g05089	1	MAIRHFASFLLASLEFCSILHFSGPDEL-----LSPFDVTKPRFWYPTTSKTSXXXXXXXXXXXX-XXXXXXXXXXTRPPPKKQDNDMNCFTWAQNDSCQDSEVLRRLCRRSCMCQNYTV	122
g35960.t1	1	-----MRTATVLLLVTVSLLLLLAG-----	22
g20182.t1	1	-----MGPPLATVLLLVTVS-LLLP-SMC-----	22
g30627.t1	1	-----MYSSNGHRRDGIQRRCPLQLLFLV--LQLVQV-----DTMNCYTWAEDASCEQDPMDEKMCRRSCQLCTNYTV	74
g16213.t1	1	-----MIGFSPFLRFVLA--LVPFCLDTILHFASKIDMQCEIVSLLGNTLLSSETDDQCYGQ-----	58
GPLIN_001083100	1	-----MDNRROFANFYLLLI-I-IIASICA-----	25
GPLIN_000258300	1	MAIRHFASFLLASLEFCSILHFSGPDEL-----LPPFDVTKPRFWYPTTTSKTSPTTTTITTTIT-T-TTTTITMTPRPPPKKQDNDMNCFTWAQNDSCQDSEVLRRLCRRSCMCQNYTV	122
GPLIN_001268900	1	-----MDNRROFANFYLLLI-I-IIASICA-----	25
cds.Nab_23179_c0_seq1_m.18402	1	MLLQRCFLFLRLLFFASLLLLFPABVSAFSKTTTELAVTEVPVKSTAECPKIPRTWS-RSTTRRTTTTITTTITTTIT-T-STTTTATQTFRPRRQKCRDLDMNCYSWAKEPERCELDSEMLKYCRLSRCRNG-NFTV	136
GROS_g04556	29	-----DYDQSRVPSNLR-PISWLLGIWRSEHDGKAMFPTIPKFTFGEELISIALSD-SRMSGP-PAFNYTAFAWSA-----HDDG-----FVTVEEGPVSSSH	113
GROS_g10456	26	-----ISTNAELNQNKK-KLSWIVGKARSEFSGKVFVPSIPTMTFGEELVVAEAPLARTAGV-QFLNWSARAWSH-----STKDFHDEWGYITVESIGN-ATLMTAGNGFTTYEVGEVSK	136
GROS_g05089	123	DVQDNEVEESNYNHRLRFAYDRLRRPRLHLH-RLAFLIGRWKSDFGGKADFPTIPRFITYGETLDFSLN--THLK-Q-PVLNYSAFAWDN--SDLTELHSEHGFLLAGPSSSEVALNTYMSNGFVTVERGEERDHT	251
g35960.t1	23	-----NTELSQLNQ-KLSWIVGKARSEFSGKVFVPSIPTMTFGEELIYVAEAPMAMTSGV-QFLNWSARAWSH-----STKDFHDEWGYITVETIGN-ATLMTAGNGFTTYEVGEVRFK	130
g20182.t1	23	-----NTELSQLNQ-KLSWIVGKARSEFSGKVFVPSIPTMTFGEELIYVAEAPMAMTSGV-QFLNWSARAWSH-----STKDFHDEWGYITVETIGN-ATLMTAGNGFTTYEVGEVRFK	130
g30627.t1	75	EVREVEDE-----RYDLRRVPSLH-RLAFLIGRWKSDFGGKADFPTIPRFITYGETLDFSLN--TRLK-V-PVLNYSAFAWDN--GDLSELHSEHGFLLAGPSSSEVALNTYMSNGFVTVERGEERDHT	190
g16213.t1	59	-----RFNVSRMTHHSSAALSGFVGIWRSQNLGRAFFP-----IGLSGMRKRKRFGLTLERAVAWGMN--QTSSELHSEFGYIVVSRRGK-LALNTYMSNGFITVEEGRDKS	160
GPLIN_001083100	26	-----TSTNAELNQNKK-KLSWIVGKARSEFSGKVFVPSIPTMTFGEELVVAEAPLARTAGV-QFLNWSARAWSH-----STKDFHDEWGYITVESIGN-ATLMTAGNGFTTYEVGEVSK	136
GPLIN_000258300	123	DVQDNEVEE-----MYDLRRVPSLH-RLAFLIGRWKSDFGGKADFPTIPRFITYGETLDFSLN--THLK-Q-PVLNYSAFAWDN--SDLTELHSEHGFLLAGPSSSEVALNTYMSNGFVTVERGEERDHT	240
GPLIN_001268900	26	-----TSTNAELNQNKK-KLSWIVGKARSEFSGKVFVPSIPTMTFGEELVVAEAPLARTAGV-QFLNWSARAWSH-----STKDFHDEWGYITVESIGN-ATLMTAGNGFTTYEVGEVSK	136
cds.Nab_23179_c0_seq1_m.18402	137	EVV--EVEE-----KYDLRKVPPNLH-KLAFLIGRWKSDFGGKADFPTIPRFITYGEELDFSVS--SRMN-V-PVLNYSAFAWDNSESTLGNTELHSEHGFLLAGPSSSLVALNTYMSNGFVTVEEGEEKNS	257
GROS_g04556	114	IRFKLVDIG-RIS-----FSRDLPVHDLV-----EWTLMQ-HNHFEARLD	152
GROS_g10456	137	MVLTLLKDIG-RIS-----FSRDLVEDLRE-----TFIKHQ-DTYMEQVLE	175
GROS_g05089	252	IRFELDRIQ-RIN-----FSRDLVRRRQF-----	275
g35960.t1	131	MVLTLLKDIG-RIS-----FSRDLVEDLRE-----TFIKHQ-DTYLEQILE	169
g20182.t1	131	MVLTLLKDIG-RIS-----FSRDLVEDLRE-----TFIKHQ-DTYLEQILE	169
g30627.t1	191	IRFELDRIQ-RIS-----LSKELPVRRF-----	212
g16213.t1	161	ARLRRLRSVGIIDRND-----FAHDFRVRTIVR-----EWTLLERGORMESRLS	202
GPLIN_001083100	137	MVLTLLKDIG-RISFSRDLVEDLRRFTIKHDDTYMEQVLEMRTATHPKDFHDEWGYITVESIGNATLMTAGNGFTTYEVGEVSKMVLTKDIGRISFSRDLVEDLRE-----TFIKHQ-DTYMEQVLE	262
GPLIN_000258300	241	IRFELDRIQ-RIN-----FSRDLVRRVIF-----	264
GPLIN_001268900	137	MVLTLLKDIG-RIS-----FSRDLVEDVGVVELHGFDRDCEKACVTDGGGELEDPD	186
cds.Nab_23179_c0_seq1_m.18402	258	IRFELQRIQ-RIN-----FSRDLVRRRIF-----EWTLLN-ETHLEARLQ	296
GROS_g04556	153	-----METLTH--GMOEHTMNIYTRVL-----	172
GROS_g10456	176	-----MRTATHPKVGYMEHTRVITYKIS-----	198
GROS_g05089	276	-----MKQALQL-----	283
g35960.t1	170	-----MRTATHPKRSQYLEHTRVITYKIS-----	192
g20182.t1	170	-----MRTATHPKRSQYLEHTRVITYKIS-----	192
g30627.t1	-----	-----	-----
g16213.t1	203	-----METDPSIGLREHTIAIYKEYP-----	224
GPLIN_001083100	263	-----MRTATHPKVGYMEHTRVITYKIT-----	285
GPLIN_000258300	265	-----F-----	265
GPLIN_001268900	187	GSAGSDKKPIRSFFPR--SADDAILSLVCNRTGIFELKK	223
cds.Nab_23179_c0_seq1_m.18402	297	-----MATATH--RMLGHTHIYKRVHPL-----	318

Gro DOG 0169 family alignment continued

GROS_g09112	
GROS_g09110	
GROS_g13625	428 ESFLSGLDLVAQNGLGKSSSL	447
GROS_g13626	
GROS_g07878	480 RDSSGNIDRDE	490
GROS_g05684	
GROS_g02733	
GROS_g01721	
GROS_g13400	
g33126.t1	
g21345.t1	
g34102.t1	
GPLIN_001355500	499 ESFLSGLDLVSLNSLGGKSSKYELDTLMLKTLGNVYDHLFKMGEKSFLSRIYLVFRINTEDEGNHFI LMNNVFENAKDVPALLTFDVKGVFSHORAI VKVTEQDRLNDVVLKWNFFGGTKELLGDKMEAMFKNSILLDKATHK638	
GPLIN_000962700	
cds.Nab_29527_c0_seq1_m.30990	
cds.Nab_29366_c0_seq1_m.30587	
cds.Nab_58262_c0_seq1_m.53270	
GROS_g09112	
GROS_g09110	
GROS_g13625	
GROS_g13626	
GROS_g07878	
GROS_g05684	
GROS_g02733	
GROS_g01721	
GROS_g13400	
g33126.t1	
g21345.t1	
g34102.t1	
GPLIN_001355500	639 NIAERLKKDFEILAKNDVNDWSIIFGVHFGVGEKHKISVGESSVDCILGFYFRGRCNQCRIHPCDASPELLNNOTLRLYIGIVDFFTRFSEARKERYMEGLKFKIKNVNSHLMYVKAQSWETFGAIPPEMYGHRMLAFS778	
GPLIN_000962700	
cds.Nab_29527_c0_seq1_m.30990	
cds.Nab_29366_c0_seq1_m.30587	
cds.Nab_58262_c0_seq1_m.53270	
GROS_g09112	
GROS_g09110	
GROS_g13625	448	
GROS_g13626	
GROS_g07878	
GROS_g05684	
GROS_g02733	
GROS_g01721	
GROS_g13400	
g33126.t1	
g21345.t1	
g34102.t1	
GPLIN_001355500	779 MVCAFRWQKEEALDAEAIIRQALIQAFCDYSGLKNDLKEEESILAMLDDKSKDWKNELORDYLFKNHTVTIFGVPLFKRLMEKMEKYFWFNLNQLAEDELKKEKETFGQKNRAFVVRKLSKNDVKNVKTLEGCARCG918	
GPLIN_000962700	
cds.Nab_29527_c0_seq1_m.30990	
cds.Nab_29366_c0_seq1_m.30587	
cds.Nab_58262_c0_seq1_m.53270	
GROS_g09112	
GROS_g09110	
GROS_g13625	481 ESSRIYLVFRINTEKNGTKDAHEFYAV-TDGGKTENRK	516
GROS_g13626	
GROS_g07878	539 AAFREKTDTRDGGVIRIYYEEFLAMVFLKMMAMMEAR-	574
GROS_g05684	
GROS_g02733	
GROS_g01721	
GROS_g13400	
g33126.t1	
g21345.t1	
g34102.t1	
GPLIN_001355500	919 CGKDYYYAL-LDLQKNGTKDAHEFYAVISNGKTEGK	954
GPLIN_000962700	
cds.Nab_29527_c0_seq1_m.30990	
cds.Nab_29366_c0_seq1_m.30587	
cds.Nab_58262_c0_seq1_m.53270	

Gro DOG 0187 family alignment

GROS_g09671	1	-----MLTFAS SPVL -----	LFLWGNVLADLSQ-----	-----P	-----DKELVLL DS -----	LWRHGDR SPT -----	STFK TD -----	IY Q DFW Q -----	GW GLSR -----	67		
GROS_g03808	1	-----MIFLNLV F -----	LIFLSDNFSIA-----	-----	-----SDELLLV QA -----	LWRHGDR SPT -----	GTFR SD -----	PN Q EDAW Q -----	GW GLSR -----	62		
GROS_g13673	1	MFGALSKNA IEL ST Q FLF -----	GILGAILIQNAK Q W T OS PS R T T S PP AN -----	-----	-----KDELV Y HL-----	WRHGDR TRV -----	F I F ND -----	PN Q EW Q -----	GF GL IT -----	92		
GROS_g02007	1	-----METV H RO-----	-----	-----	-----RSEMI R KK-----	-----	-----	-----	-----	22		
GROS_g01693	1	-----MRR TF IR Q F Y V -----	LFLAD C FAL V G V I F L I NA F K V OHVSM HN -----	-----	-----ASELL F PA H KA E M I E GN L S D L V S A I S R L K E V N L L N -----	-----	-----	-----	-----	94		
GROS_g09164	1	-----MNK I IQ C CL F -----	LLF Y AN Q -----	-----	-----AEL C Y Q II-----	IWRHGDR AF L E -----	KA Y R E D -----	L N D E N N S R -----	GW GL IT -----	61		
GROS_g12317	1	MNYE G K R V I V I V I L L M S L L -----	LILLLL N F S I F N N -----	-----	-----OR L L Y Q A-----	IWRHGDR AF G K -----	LP Y G D -----	PH D E K Y R -----	GW S Q L T N -----	89		
GROS_g05767	1	-----MOK K P O G S I Y I Y -----	LCLL F N V I R L H -----	-----	-----L L T K F V H V -----	LWRHGVR T P A -----	IE M A K E R -----	EW Q -----	GL EL IT -----	75		
GROS_g02879	1	-----M S L P S L F R L -----	SLL T S L F V S I A V G -----	-----	-----SS K L H I G M -----	FFRHGDR TP T F -----	L N E N E -----	N I S S F G T E L A F D L E M T -----	69			
g34866.t1	1	-----MSD K Y F S L L L F L I S L S V S L C F R W A R T R Q V L A K R V V C T D S R F N V I E L M D S D A L E M T D D V M S L O D E L I F V O I -----	-----	-----	-----IWRHGDR AF -----	-----	-----	-----	-----	GF EL IT -----	98	
g29945.t1	1	-----MRF L H I L L F F -----	GLFA A I S C -----	-----	-----IE K L R Y Q A-----	IWRHGDR AP K E -----	K P Y R A D -----	L H E H K W R -----	GW G L T K -----	60		
g598.t1	1	-----MRF L H I L L F F -----	GLFA A I Y C -----	-----	-----IE K L O Y Q A-----	IWRHGDR AP K E -----	K P Y R A D -----	L H D E N K W R -----	GW G L T K -----	60		
g24286.t1	1	-----ML F L L L L -----	LALHL N V O S O A O -----	-----	-----D O P E P P L V A -----	KREL V M A O A-----	WRHGDR AP S K -----	LP Y P G D -----	Y D E Q F W R -----	GW S Q L T N -----	74	
g16589.t1	1	-----ML F L L L L -----	LALHL N V O S O A R -----	-----	-----A D O P E P P R V A -----	KREL V M A O A-----	WRHGDR AP S K -----	LP Y P G D -----	Y D E Q F W R -----	GW S Q L T N W W K K I W 81 -----	76	
g17037.t1	1	-----M L L L L L -----	LAI F W S L A S L T V -----	-----	-----R E V O R S R T E -----	LD T L K F V H V -----	WRHGDR TP S -----	K M R S D T V N T L E K W K A K F G L G L T A -----	76			
GPLIN_001406900	1	-----M I F L N V L L F -----	LIFL S D O K F S I A -----	-----	-----S D E L L L V Q A-----	LWRHGDR S P T -----	GTFR S D -----	PN Q EDAW Q -----	GW GL SR -----	63		
GPLIN_001195200	1	-----M S A A S I D G I L L V V -----	VVVV L L T P A L L G O -----	-----	-----K L A N D K A D -----	LD T L K F V H A -----	LWRHGDR TP S -----	K M I R S D O N T L D K W K A K F G L G L T D -----	80			
GPLIN_000618300	1	-----M S A A S I D G I L L F -----	VVVV L L T P A L L G O -----	-----	-----K L A N D K A D -----	LD T L K F V H A -----	LWRHGDR TP S -----	K M I R S D O N T L D K W K A K F G L G L T D -----	79			
GPLIN_000508200	1	-----M S A A S I D G I L L F -----	VVVV L L T P A L L G O -----	-----	-----M L E S D R A D -----	LD T L K F V H A -----	LWRHGDR TP T I -----	K M I R S D O N T L D K W T A K F G L G L T S -----	76			
cds.Nab_27332_c0_seq1_m.25942	1	-----M L V L L L L F L P S L -----	L C R S S H A A I V Q O S A R T A O S A V T S P P M -----	-----	-----S A B D O A M G G E A D E L L F V O I -----	IWRHGDR TP I -----	Y T T R S D -----	PN Q EDAW Q -----	GF EL IT -----	98		
cds.Nab_24620_c0_seq1_m.20957	1	-----M P I S W M I L V F F -----	L L L S S A A I L L V G -----	-----	-----G O P O E L A D D -----	OP L V L V Q V -----	AN Y R R D -----	PH G E D E W A F -----	GW GL IT -----	77		
GROS_g09671	68	-----R G M -----	S O V V L G K K R Y V E L K -----	-----	-----Y V G G S N S H -----	EL Y V R A T D V R T L I S A I -----	-----	-----	-----	115		
GROS_g03808	63	-----K G M -----	A D H V L G S K L A R Y I D E L K -----	-----	-----F V S E R Y L N K -----	E I Y L R S D V N R T L T S A I -----	-----	-----	-----	110		
GROS_g13673	93	-----R G M -----	A D H V L G H F L K R Y I D E L K -----	-----	-----L L S O N S K -----	E V Y I S D V N R T L O S A T -----	-----	-----	-----	140		
GROS_g02007	23	-----K G M -----	A D H V L G S K L A R Y I D E L K -----	-----	-----F V S A R L N K -----	E I Y V R S D M N R T L T S A I -----	-----	-----	-----	70		
GROS_g01693	92	-----D G A L K E A K N V D S L A K M A D K K T L A D T K -----	L C O V P T S K D N L T K T K F V T E A G N O Y L I R G S P A T D D A E V V K R M E A G A I L V A I T V L E N L A L S W S G D S G Y G I A N N F H D T R I T G S S G E S A L G I 219 -----	-----	-----	-----	-----	-----	-----	-----	-----	219
GROS_g09164	62	-----E G M -----	K O M E E L G O F L E R E Y N A N -----	-----	-----F L A G O F D K K -----	A T I T R S D V D R A L E S A O -----	-----	-----	-----	108		
GROS_g12317	90	-----L G M -----	A D M R E L G O F L R O Y I M E A T -----	-----	-----F L G A S N R D -----	D V F I O S G D R A L V S A O -----	-----	-----	-----	137		
GROS_g05767	76	-----K G M -----	G L L F R L G Q W I R R Y D O G -----	-----	-----F L S E P E N N Y -----	E L Y V R S D F N R T L M S A O -----	-----	-----	-----	121		
GROS_g02879	70	-----K G M -----	W E F V L G Q K V S L L G O -----	-----	-----F L G E S R S R -----	E I C A L S G R O N R T L I A S A O -----	-----	-----	-----	114		
g34866.t1	99	-----R G M -----	A D O L H G K F L K R Y I D L E -----	-----	-----L L S P H S A K -----	E V M V I S D T N R T L I S A T -----	-----	-----	-----	146		
g29945.t1	61	-----K G M -----	R E M K E L G O F L D R Y V V N A S -----	-----	-----F V D G I N N K K -----	E V O I H S D V D R A L E S A L -----	-----	-----	-----	108		
g598.t1	61	-----K G M -----	R E M K E L G O F L D R Y V V N A T -----	-----	-----F V D G I N N K K -----	E V O I H S D V D R A L E S A L -----	-----	-----	-----	108		
g24286.t1	75	-----L G M -----	A D M R E L G O F L N R Y L O E -----	-----	-----L L S A K D O D -----	E V I L S N D R A L T S A O -----	-----	-----	-----	107		
g16589.t1	82	-----N F I Q L G M -----	A D M R E L G O F L N R Y L O E -----	-----	-----L L S A K D O D -----	E V I L S N D R A L T S A O -----	-----	-----	-----	135		
g17037.t1	77	-----D G A -----	R D H I F O R M L H R Y I D E L K -----	-----	-----G O O Y G H G H P P -----	F L A D H F S D -----	-----	-----	-----	131		
GPLIN_001406900	64	-----K G M -----	A D H V L G S K L A R Y I D E L K -----	-----	-----F V S E R Y L N K -----	E I Y L R S D V N R T L T S A I -----	-----	-----	-----	111		
GPLIN_001195200	81	-----D G A -----	R D H E N L G R L L H R Y D O -----	-----	-----F L S C I F T P E -----	E F Y V R S D V N R T L I A S A O -----	-----	-----	-----	125		
GPLIN_000618300	60	-----D G A -----	R D H E N L G R L L H R Y D O -----	-----	-----F L S C I F T P E -----	E F Y V R S D V N R T L I A S A O -----	-----	-----	-----	124		
GPLIN_000508200	77	-----D G A -----	R D H E N L G R L L H R Y D O -----	-----	-----F L S C I F T P</							

Gro DOG 0201 family alignment

GROCS_g03615	1	MFSTVLSFLAYS IYEMVY I L L V L R L I Y V L E I C H Y V D V V F Y R L E I I D D T S W T F T S E I I Y T V A T F L I Y F F N Y I Y I F H I A I A L N R C T S W A V Y K G A F W I R A C H I I I F A F A V S F L F T A H R L F A V A G L V L D S D D T K 138	
GROCS_g03660			
GROCS_g06676			
GROCS_g11484			
GROCS_g02883			
GROCS_g09263			
GROCS_g06777			
GROCS_g10895			
GROCS_g11524			
GROCS_g12051			
GROCS_g13246			
GROCS_g07942			
GROCS_g05547			
GROCS_g08721			
p18199#1			
p25539#1			
p14349#1			
p2797#1			
p19341#1			
p23306#1			
p17659#1			
p12778#1			
p32695#1			
p30861#1			
p21071#1			
p811#1			
p19307#1			
p19694#1			
p30861#2			
SPLIM_001009100			
SPLIM_001068400			
SPLIM_001099500			
SPLIM_000983400			
SPLIM_000402300			
cds.Nab_32144_c0_seq1_m.35269			
cds.Nab_37164_c0_seq1_m.36703			
cds.Nab_59225_c0_seq1_m.50546			
cds.Nab_11237_c0_seq1_m.9474			
cds.Nab_19927_c0_seq1_m.11543			
cds.Nab_28809_c0_seq1_m.29088			
cds.Nab_23144_c0_seq1_m.18337			
cds.Nab_21101_c0_seq1_m.15529			
cds.Nab_25601_c0_seq1_m.22724			
cds.Nab_19861_c0_seq1_m.14190			
cds.Nab_28125_c0_seq1_m.27586			
GROCS_g03615	1		
GROCS_g03660			
GROCS_g06676			
GROCS_g11484			
GROCS_g02883			
GROCS_g09263			
GROCS_g06777	139	1	1
GROCS_g10895			
GROCS_g11524			
GROCS_g12051			
GROCS_g13246			
GROCS_g07942			
GROCS_g05547			
GROCS_g08721			
p18199#1			
p25539#1			
p14349#1			
p2797#1			
p19341#1			
p23306#1			
p17659#1			
p12778#1			
p32695#1			
p30861#1			
p21071#1			
p811#1			
p19307#1			
p19694#1			
p30861#2			
SPLIM_001009100			
SPLIM_001068400			
SPLIM_001099500			
SPLIM_000983400			
SPLIM_000402300			
cds.Nab_32144_c0_seq1_m.35269			
cds.Nab_37164_c0_seq1_m.36703			
cds.Nab_59225_c0_seq1_m.50546			
cds.Nab_11237_c0_seq1_m.9474			
cds.Nab_19927_c0_seq1_m.11543			
cds.Nab_28809_c0_seq1_m.29088			
cds.Nab_23144_c0_seq1_m.18337			
cds.Nab_21101_c0_seq1_m.15529			
cds.Nab_25601_c0_seq1_m.22724			
cds.Nab_19861_c0_seq1_m.14190			
cds.Nab_28125_c0_seq1_m.27586			
GROCS_g03615	20		
GROCS_g03660	19		
GROCS_g06676	56		
GROCS_g11484	24		
GROCS_g02883	20		
GROCS_g09263			
GROCS_g06777	277		
GROCS_g10895			
GROCS_g11524			
GROCS_g12051			
GROCS_g13246			
GROCS_g07942			
GROCS_g05547			
GROCS_g08721	69		
p18199#1	9		
p25539#1	19		
p14349#1	19		
p2797#1	65		
p19341#1	65		
p23306#1	19		
p17659#1	19		
p12778#1	19		
p32695#1	19		
p30861#1	19		
p21071#1	19		
p811#1	19		
p19307#1	22		
p19694#1	19		
p30861#2			
SPLIM_001009100			
SPLIM_001068400	54		
SPLIM_001099500	19		
SPLIM_000983400	19		
SPLIM_000402300	19		
cds.Nab_32144_c0_seq1_m.35269	16		
cds.Nab_37164_c0_seq1_m.36703	16		
cds.Nab_59225_c0_seq1_m.50546	16		
cds.Nab_11237_c0_seq1_m.9474	16		
cds.Nab_19927_c0_seq1_m.11543	20		
cds.Nab_28809_c0_seq1_m.29088	23		
cds.Nab_23144_c0_seq1_m.18337	19		
cds.Nab_21101_c0_seq1_m.15529	62		
cds.Nab_25601_c0_seq1_m.22724	25		
cds.Nab_19861_c0_seq1_m.14190	19		
cds.Nab_28125_c0_seq1_m.27586	22		

Gro DOG 0200 family alignment

GROS_g04903	1	MTLALP	LLLF	FAGI	AF	IQLFM	23																																																																																																																									
GROS_g08379	1	MTLALP	LLLF	FVGI	AF	IQLFM	23																																																																																																																									
GROS_g08377	1	MTLALP	LLLF	FVGL	FL	18																																																																																																																									
GROS_g08380	1	MTLALP	LLLF	FAGI	AF	FL	19																																																																																																																									
GROS_g08646	1	MTSPS	LLLF	FVCM	AF	FYLEL	21																																																																																																																									
GROS_g08378	1	MNSSNASTSEL	GA	V	IRAN	FAP	LCVLF	GLK	LFVLA	IG	LL	S	ITP	IY	LL	RSNC	NLLI	ALNF	LNTAL	QL	EA	L	SATV	VLS	GO	I	Q	I	R	L	N	C	F	L	Q	I	V	Y	I	G	A	T	M	Q	W	A	I	C	F	A	I	A	L	D	R	L	V	G	L	F	F	F	Y	S	N	F	F	K	R	R	I	V	L	V	T	I	139																																																					
GROS_g13812	1	MTHK	V	W	L	S	I	F	V	M	L	A	T	VL	20																																																																																																																
g1257.t1	1	MTHK	V	W	L	S	I	F	V	V	L	A	T	FL	20																																																																																																																
g18097.t1	1	MLNS	L	S	S	S	S	V	H	L	L	L	V	L	F	L	M	I	L	25																																																																																																												
g14551.t1	1	MLNS	L	S	S	S	S	V	H	L	L	L	V	L	F	V	I	L	24																																																																																																													
g11381.t1	1	MTLALP	LLLF	FAS	I	ASM	FL	21																																																																																																																							
GPLIN_000913700	1	MTLALP	LLLF	FVGL	FL	18																																																																																																																									
GPLIN_000913800	1	MTLALP	LLLF	FVGL	FL	18																																																																																																																									
GPLIN_000913500	1	NK	L	A	L	P	M	L	S	F	A	G	I	A	F	M	O	L	F	23																																																																																																												
GPLIN_000562200	1	MTLALP	LLLF	FVSL	FL	18																																																																																																																									
GPLIN_000158600	1	MTLALP	MLFS	FAGI	AF	MOLFL	23																																																																																																																									
GPLIN_000913300	1	MALALP	LLLF	FAGI	AF	IQLFM	23																																																																																																																									
GPLIN_000913400	1	MTLALP	LLLF	FAGI	AF	IQLFM	23																																																																																																																									
GPLIN_000310800	1	MTLALP	LLLF	FVGL	FL	18																																																																																																																									
GPLIN_000922300	1	MTSPS	CM	A	F	F	Y	L	F	15																																																																																																																					
cds.Nab_26040_c0_seq1_m.23527	1	N	P	L	L	L	L	Y	A	K	L	F	A	L	A	V	L	A	L	I	F	S	O	N	B	I	A	N	F	E	K	C	L	E	P	G	V	M	T	S	C	T	Y	N	I	S	T	P	S	A	S	K	G	56																																																																										
GROS_g04903	140	ASV	C	A	F	T	L	Y	K	A	F	L	S	Y	T	F	S	M	A	S	E	K	L	V	I	C	G	T	D	I	S	O	N	S	F	D	Y	M	O	N	L	A	I	N	V	A	T	A	S	C	Y	I	F	L	W	A	L	R	C	K	L	N	D	H	D	C	S	R	R	I	F	R	S	V	S	A	I	M	L	L	E	V	L	G	M	L	T	N	L	L	R	L	I	L	F	L	L	K	N	L	T	D	L	T	L	Y	S	N	S	A	I	N	C	M	A	K	L	A	G	A	V	L	R	E	T	A	T	N	R	278
GROS_g08379	24	GL	T	D	G	A	V	S	O	N	L	E	G	V	C	S	D	P	E	N	E	G	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	E	L	T	K	C	N	I	P	G	W	A	E	S	Y	K	G	T	86																																																																				
GROS_g08377	19	GL	T	D	G	A	V	S	O	N	L	E	G	V	C	S	D	P	E	N	E	G	T	V	C	D	C	K	S	C	T	F	K	N	C	K	N	V	T	S	O	A	K	N	I	P	G	W	A	A	N	V	G	S	86																																																																							
GROS_g08380	20	G	M	S	D	G	A	V	T	C	N	L	D	G	V	C	R	D	R	T	S	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	E	L	T	K	C	P	I	P	G	W	A	K	Y	K	G	T	83																																																																		
GROS_g08646	22	G	M	S	E	G	A	V	I	C	G	L	D	G	N	C	R	D	R	T	S	A	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	V	Y	A	R	C	N	I	P	G	W	A	R	N	Y	K	G	N	84																																																																
GROS_g08378	22	G	L	R	E	A	V	T	C	G	L	D	G	N	C	R	D	R	T	S	S	K	N	K	G	T	V	C	I	F	K	N	C	K	N	V	T	S	P	A	K	C	E	I	P	G	W	D	K	S	Y	V	G	N	78																																																																							
GROS_g08649	1	M	S	E	G	A	V	I	C	G	L	D	G	N	C	R	D	R	T	S	A	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	V	Y	A	R	C	N	I	P	G	W	A	R	N	Y	K	G	N	64																																																																	
GROS_g13812	279	N	S	G	W	R	G	I	G	K	I	M	E	I	C	L	S	G	N	F	V	L	A	L	F	A	E	S	T	V	I	V	D	F	G	K	L	S	F	V	E	K	L	R	A	D	H	K	R	O	G	L	S	E	A	V	T	C	G	L	D	G	V	C	R	D	S	K	N	K	G	T	V	C	393																																																					
GROS_g13812	1	G	L	D	G	A	V	T	C	G	L	D	G	V	C	R	D	R	T	S	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	P	A	K	C	E	I	P	G	W	A	K	Y	K	G	T	65																																																																				
g1257.t1	21	R	H	T	D	A	L	I	H	E	R	N	V	C	D	S	D	P	V	T	F	T	K	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	E	L	T	K	C	P	I	P	G	W	A	K	Y	K	G	T	83																																																																							
g18097.t1	21	R	H	T	D	A	L	I	H	E	R	N	V	C	D	S	D	P	V	T	F	T	K	F	K	S	C	T	F	K	N	C	K	N	V	T	S	P	A	K	C	E	I	P	G	W	A	R	S	D	N	83																																																																										
g14551.t1	26	H	S	V	S	I	V	R	K	B	H	G	H	G	S	D	P	V	T	F	T	K	P	G	A	L	I	C	S	G	Y	K	C	H	L	A	N	P	R	K	H	S	I	P	G	W	S	D	N	93																																																																												
g11381.t1	25	H	S	V	S	I	V	R	K	B	H	G	H	G	S	D	P	V	T	F	T	K	P	G	A	L	I	C	S	G	Y	K	C	H	L	A	N	P	R	K	H	S	I	P	G	W	S	D	N	91																																																																												
GPLIN_000913700	22	S	M	S	E	G	A	V	T	C	N	L	D	G	V	C	R	D	R	T	S	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	P	A	K	C	E	I	P	G	W	A	K	Y	K	G	T	86																																																																			
GPLIN_000913800	19	G	M	S	D	G	A	V	T	C	N	L	D	G	V	C	R	D	R	T	S	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	E	L	T	K	C	P	I	P	G	W	A	K	Y	K	G	T	83																																																																		
GPLIN_000913500	19	G	M	S	D	G	A	V	T	C	N	L	D	G	V	C	R	D	R	T	S	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	P	A	K	C	E	I	P	G	W	A	R	N	Y	K	G	N	88																																																																		
GPLIN_000562200	19	S	M	S	E	G	A	V	I	C	G	L	D	G	N	C	R	D	R	T	S	A	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	V	H	A	C	N	H	P	G	W	A	K	Y	K	G	N	83																																																																		
GPLIN_000158600	19	G	L	D	G	A	V	T	C	N	L	D	G	V	C	R	D	R	T	S	A	S	D	G	S	T	V	C	I	A	V	A	P	S	K	N	C	K	N	V	T	S	P	A	K	C	E	I	P	G	W	A	R	N	Y	K	G	N	74																																																																			
GPLIN_000913300	24	GL	T	D	G	A	V	S	O	N	L	E	G	V	C	S	D	P	E	N	E	G	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	P	A	K	C	E	I	P	G	W	A	K	Y	K	G	N	86																																																																						
GPLIN_000913400	24	GL	T	E	A	V	S	O	N	L	E	G	V	C	S	D	S	E	T	E	G	T	E	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	P	A	K	C	E	I	P	G	W	A	R	N	Y	K	G	N	86																																																																						
GPLIN_000310800	19	G	M	S	D	G	A	V	T	C	N	L	D	G	V	C	R	D	R	T	S	S	D	G	S	T	V	C	N	C	K	S	C	T	F	K	N	C	K	N	V	T	S	E	L	T	K	C	P	I	P	G	W	A	K	Y	K	G	T	83																																																																		
GPLIN_000922300	16	G	L	S	K	A	A	V	T	C	G	L	D	G	V	C	R	D	R	T	S	S	K	N	K	G	T	G	W	D	K	S	Y	V	G	N	50																																																																																									
cds.Nab_26040_c0_seq1_m.23527	57	R	K	R	S	D	V	S	G	A	N	G	V	E	A	Q																																																																																																																

9.7 – Supplementary file 7

Significance testing values for cathepsin L peptidase RNAi screening

Dependent Variable: SQRT of female count

Tukey HSD

(I) Line	(J) Line	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
GFP	Line 12	.10620	.46532	.999	-1.1789	1.3913
	Line 14	.33889	.45799	.947	-.9259	1.6037
	Line 20	1.02144	.46155	.181	-.2532	2.2961
	Line 23	.73892	.46532	.507	-.5462	2.0240
Line 12	GFP	-.10620	.46532	.999	-1.3913	1.1789
	Line 14	.23268	.45799	.986	-1.0321	1.4975
	Line 20	.91524	.46155	.279	-.3594	2.1899
	Line 23	.63272	.46532	.654	-.6524	1.9178
Line 14	GFP	-.33889	.45799	.947	-1.6037	.9259
	Line 12	-.23268	.45799	.986	-1.4975	1.0321
	Line 20	.68256	.45416	.562	-.5717	1.9368
	Line 23	.40003	.45799	.906	-.8648	1.6649
Line 20	GFP	-1.02144	.46155	.181	-2.2961	.2532
	Line 12	-.91524	.46155	.279	-2.1899	.3594
	Line 14	-.68256	.45416	.562	-1.9368	.5717
	Line 23	-.28252	.46155	.973	-1.5572	.9921
Line 23	GFP	-.73892	.46532	.507	-2.0240	.5462
	Line 12	-.63272	.46532	.654	-1.9178	.6524
	Line 14	-.40003	.45799	.906	-1.6649	.8648
	Line 20	.28252	.46155	.973	-.9921	1.5572

Dependent Variable: Area

Tukey HSD

(I) Line	(J) Line	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
GFP	Line12	.003486	.002085	.451	-.00220	.00918
	Line14	.002113	.002094	.851	-.00360	.00783
	Line20	.004814	.002140	.162	-.00103	.01066
	Line23	.008360*	.002126	.001	.00256	.01416
Line12	GFP	-.003486	.002085	.451	-.00918	.00220
	Line14	-.001373	.002100	.966	-.00710	.00436
	Line20	.001328	.002146	.972	-.00453	.00719
	Line23	.004873	.002132	.150	-.00095	.01069
Line14	GFP	-.002113	.002094	.851	-.00783	.00360
	Line12	.001373	.002100	.966	-.00436	.00710
	Line20	.002701	.002155	.720	-.00318	.00858
	Line23	.006247*	.002140	.029	.00041	.01209
Line20	GFP	-.004814	.002140	.162	-.01066	.00103
	Line12	-.001328	.002146	.972	-.00719	.00453
	Line14	-.002701	.002155	.720	-.00858	.00318
	Line23	.003545	.002186	.483	-.00242	.00951
Line23	GFP	-.008360*	.002126	.001	-.01416	-.00256
	Line12	-.004873	.002132	.150	-.01069	.00095
	Line14	-.006247*	.002140	.029	-.01209	-.00041
	Line20	-.003545	.002186	.483	-.00951	.00242

*. The mean difference is significant at the 0.05 level.