Bioinformatic and functional characterisation of *Globodera pallida* effector genes

Peter Thorpe

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

The University of Leeds
School of Biology
Faculty of Biological Sciences

December 2012

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

The findings that form part of chapter 3 of this thesis are going to form part of the *Globodera pallida* genome publication (manuscript in preparation):

Genome sequence and transcriptome of the potato cyst nematode *Globodera pallida*

Catherine Lilley[1], Adam Reid[2], James Cotton[2], Peter Thorpe[1,3], Peter Cock[3], Helen Beasley[2], Vivian Blok[3], Taisei Kikuchi[4], Jason Tsai[1], Martin Hunt[2], Magdalena Zarowiecki[2], Juan Palomares Rius[3], Laura Jones[1], Nancy Holroyd[2], Matthew Berriman[2], John Jones[3], Peter Urwin[1]

1. Center for Plant Sciences, University of Leeds, Leeds. LS2 9JT, United Kingdom
2. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SA, United Kingdom
3. Plant Pathology Program, The James Hutton Institute, Invergowrie, United Kingdom
4. Forestry and Forest Products Research Institute, Tsukuba, Japan

Peter Thorpe: Bioinformatic identification of orthologous and novel effectors from the genome sequence and normalised transcriptome data provided by the Sanger Institute. PFam domain identification in the identified novel effectors. Identification of CAZyme genes and other horizontal gene transfer candidates. Identification and phylogenetic analysis of SPRY and 448 family gene members. Generate an expression profile for all genes of interest from the normalised transcriptome data provided by the Sanger Institute. Annotation of the identified effector genes.

Acknowledgments

I would like to thank Prof. Pete Urwin, Dr John Jones and Dr Catherine Lilley for all the supervision and support received throughout my PhD.

I would like to thank everyone in the nematology laboratory from Leeds and everyone in the plant pathology department at The James Hutton Institute. I would also like to thank my family and friends.

This project was funded by the British Biotechnology Research Council (BBSRC) and Mylnefield Research Services (MRS*)*.

Abstract

Pathogens secrete molecules, termed effectors, to manipulate their host to the benefit of the pathogen. Effectors of plant parasitic nematodes are predicted to have a range of functions such as facilitating invasion, initiation and maintenance of the feeding site, and suppression of host defences. The genome sequence of the potato cyst nematode *Globodera pallida* was analysed to identify putative effectors. They include: 129 effectors similar to those previously identified from cyst nematodes, 53 cell wall modifying enzymes and 117 novel putative effectors. Only four effectors were common between *G. pallida* and the root-knot nematode *Meloidogyne incognita*. These could have a conserved role in plant parasitism. A large SPRY domain containing gene family was identified in *G. pallida*. It has 299 members, of which 30 are predicted to be secreted and therefore categorised as effectors. Phylogenetic analysis showed that the family is hugely expanded and specific to *Globodera* species.

Fifty-four putative effectors were cloned from *G. pallida* cDNA. Transgenic lines of *Arabidopsis thaliana* and *Solanum tuberosum* L. 'Désirée' were produced, to express a range of these effectors and act as tools for functional characterisation. Potato lines that expressed selected effectors were subjected to phenotypic analysis and pathogen susceptibility assays. The largest range of aberrant phenotypes was observed for those plants expressing *GpIA7* and *GpIVG9*. Potato lines expressing *GpIA7* showed altered growth phenotypes and an increased susceptibility to *Phytophthora infestans* CS-12. *GpIVG9*-expressing potato lines showed accelerated growth, distorted leaves and increased susceptibility to nematode invasion.

A more in-depth functional characterisation was conducted on a ubiquitin extension protein effector. The *G. pallida* ubiquitin extension protein suppressed PAMP-triggered immunity and the C-terminal extension was required for this activity. The outcomes from this work and the tools generated for future experimentation will contribute to elucidating the complex interactions between pathogens and their hosts.

Table of Contents

List of figures

List of tables

List of abbreviations

ANOVA	Analysis of variance
DAMPs	Damage associated molecular pattern
DNA	Deoxyribonucleic acid
dpi	Days post infection
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ETI	Effector triggered immunity
ETS	Effector triggered susceptibility
EST	Expressed sequence tag
GFP	Green fluorescent protein
GUS	β-glucuronidase
h	Hours
HR	Hyper-sensitive response
IAA	Indole-3-acetic acid
J2	Juvenile nematode (second stage)
LRR	Leucine rich repeats
LS	Linsmaier and Skoog
mRNA	Messenger RNA
MS	Murashige and Skoog
PAMPs	Pathogen associated molecular patterns
PBS	phosphate buffered saline
PCR	Polymerase chain reaction
PR	Pathogenesis-related protein
PTI	PAMP-triggered immunity
PVDF	Polyvinylidene fluoride
qPCR	Quantitative real-time PCR
R-gene	Resistance gene
RKN	Root-knot nematode
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RPM	Rotations per minute
s	Seconds

SDW	Sterile distilled water
TMHMM	Trans membrane hidden Markov model
Y2H	Yeast-two hybrid

1 Introduction

The phylum Nematoda is probably the largest in the animal kingdom in terms of number of individuals and species number (Williamson and Kumar, 2006). Nematodes occupy a huge range of niches across the world from the polar regions to the tropics and are found in fresh and marine water (Bongers and Ferris, 1999). Nematodes can live as non-pathogenic organisms feeding on bacteria, fungi or dead organic matter. One such nematode, *Caenorhabditis elegans,* was chosen as a model for genetics and developmental biology and was the first multicellular organism to have its genome sequenced (C.elegans sequencing consortium, 1998). In addition to providing a basis for subsequent eukaryotic genome projects, the *C. elegans* project is a valuable resource for a wide range of nematode molecular biology projects, including those on parasitic nematodes. Nematodes can also be parasites of humans and other animals; *Trichuris trichiura* is estimated to infect over 1 billion people worldwide (Stephenson *et al.*, 2000) while over 600 million are thought to suffer from hookworm (*Necator americanus* and *Ancylostoma duodenale*) infections (Bethony *et al.*, 2006). Other nematodes, such as *Trichinella spiralis* (Pozio *et al.*, 1993) and *Haemonchus contortus* (Newton and Meeusen, 2003), infect livestock. Nematodes in the soil can be used as indicators of the below ground food-web and are therefore used widely for pollution monitoring and environmental assessments (Wilson and Kakouli-Duarte, 2009). Nematodes can also be serious economic pathogens of plants.

1.1 Plant-parasitic nematodes (PPN)

Approximately 20% of the 20000 described nematode species are parasites of plants (Oliveira *et al.*, 2007). Plant-parasitic nematodes cause substantial damage to crops worldwide, not just as a result of their feeding, but also due to an increased susceptibility of their host plants to fungal, bacterial and viral infections (Nicol *et al.*, 2011). Some plant-parasitic nematodes can also act as vectors of plant viruses. It is estimated that plant-parasitic nematodes cause losses to worldwide agriculture of around US$125 billion each year (Chitwood, 2003). The most economically important PPNs are the biotrophic sedentary endoparasites in the order Tylenchida, including root-knot nematodes (*Meloidogyne* spp.) and cyst nematodes (*Globodera* and

Heterodera spp.). In the UK, the potato cyst nematodes *Globodera pallida* and *G. rostochiensis* cause damage estimated at £50 million annually (Nicol *et al.*, 2011).

Nematodes have a highly conserved morphology. Classification based on morphological characters is therefore extremely difficult due to the scarcity of informative morphological characters. This, coupled to the absence of a comprehensive fossil record for nematodes, makes reconstructing nematode phylogeny problematic. However, recent studies which use analysis of small subunit ribosomal RNA sequences have allowed a detailed phylogeny of the Nematoda to be produced and this has shown that plant parasitism has evolved independently at least four times in the Phylum (Blaxter *et al.*, 1998; van Megen *et al.*, 2009). The interactions of nematodes with plants and their life cycle strategies are diverse. Plant parasitic nematodes can be migratory or sedentary and can be ectoparasites or endoparasites (Gheysen and Vanholme, 2007).

Ectoparasites remain outside the host for the duration of their life cycle and can be migratory or sedentary. Migratory ectoparasites usually have a broad host range and feed on epidermal root cells of numerous hosts during their lifetime. Their interactions with plants are usually very simple, often limited to grazing on root cells. These nematodes are found in the orders Triplonchida and Dorylaimida (Perry and Moens, 2006) and their feeding habits mean that they are the main nematode vectors of plant viruses (Strange and Scott, 2005). Sedentary ectoparasites remain outside the root throughout their life cycle but may initiate a feeding site from which they feed for some or all of their life (Hofmann and Grundler, 2007).

Endoparasites spend a substantial proportion of their life cycle within their host and can be migratory or sedentary. Migratory endoparasites such as *Radopholus* and *Pratylenchus* move throughout the root system of the plant causing substantial tissue damage and feed on the cortical cells. Some migratory endoparasites can have more complex life cycles. For example, the pine wilt nematode *Bursaphelenchus xylophilus* is a migratory endoparasite which feeds on plant tissues in living trees but also feeds on fungi that colonise dead or dying trees. Its life cycle includes a vector insect, most frequently a beetle of the Genus *Monochamus,* which transports the nematodes to a new host during oviposition (reviewed by Jones *et al.,* 2008). Sedentary endoparasites, such as the root-knot and cyst nematodes, are the most economically damaging

nematodes and are part of the Order Tylenchida (Baldwin *et al.*, 2004). These nematodes often have complex, biotrophic interactions with their hosts. Sedentary endoparasites invade the roots soon after hatching and establish a permanent feeding site. The nematodes undergo a series of moults to the adult stage at the feeding site and females remain sedentary for the rest of their lives (Turner and Stone, 1984). In sexually reproducing species males leave the roots after the moult at the adult stage and locate and fertilise the females.

1.1.1 Sedentary endoparasites

1.1.1.1 Root-knot nematodes (RKN)

Root-knot nematodes (*Meloidogyne* spp.) have a large host range that includes more than 2000 plant species and cause damage to agriculture throughout the world (Roberts, 1995). Second stage juveniles (J2) hatch from eggs, enter the plant root and migrate intercellularly until they reach the vascular cylinder where they induce their feeding sites. Although the feeding sites induced by root knot nematodes are superficially similar to those of cyst nematodes (below), in that they are large, multinucleate and metabolically active, the ontogeny of the two types of feeding site is entirely different. In keeping with this, phylogenetic analysis has shown that biotrophic parasitism of plants has evolved independently in root knot nematodes and cyst nematodes (*e.g.* van Megen *et al.,* 2009). Root knot nematode feeding cells are formed as a result of the nematode inducing repeated cycles of mitosis in the absence of cytokinesis, leading to the formation of multinucleate 'giant cells' (Jones and Goto, 2011).

1.1.1.2 Cyst nematodes

The cyst nematodes include the Genera *Heterodera* and *Globodera* as well as several less well characterised Genera (*e.g. Afenestrata* and *Punctodera*). The name describes the appearance of the survival stage of the nematodes: the cyst is a protective layer formed from the body of the adult female that encloses the eggs within. Cyst nematodes usually have a restricted host range compared to root knot nematodes (den Nijs, 2007). The co-evolution of the nematode is in concert with its host: cyst nematodes hatch as J2s but often only do so in large numbers in response to diffusates from the roots of a plant that they can infect (Perry and Wright, 1998).

1.1.1.2.1 Potato cyst nematodes (PCN)

The potato cyst nematodes *G. rostochiensis* and *G. pallida* are indigenous to South America where they co-evolved with their Solanaceous host plants. It is thought that they were first introduced to Europe with potato in around 1600 with further introductions on potato germplasm brought to Europe after the Irish potato famine in the 1840s. PCN is an important agricultural pest that can reduce potato yields by over 50% (den Nijs, 2007). Nematicides are widely used to control plant parasitic nematodes. However, concerns over the effects of the chemicals on non-target organisms (including humans) have resulted in a decline in the number of nematicides available for use for the past 20 years. In 2007 the only active ingredients still approved for use in the UK were fosthiazate, ethoprophos, oxamyl and 1,3-dichloropropene (Tobin *et al.*, 2008). EU legislation (EC 1107/2009) has further reduced options for growers.

Although viable eggs persist in the soil within cysts for many years, a proportion (approximately 10–20%, depending on soil conditions) dies each year in the absence of a host (Perry and Moens, 2006). Increasing the time between a host crop by the subsequent growing of alternative, un-related crops, known as crop rotation, allows the nematode population to decline (Devine *et al.*, 1999). Crop rotation and/or the use of resistant cultivars have been used to control potato cyst nematodes. For example *H1* confers resistance to *G. rostochiensis* on potato (Janssen *et al.*, 1991) and *Gpa2* controls a small number of *G. pallida* populations (Sacco *et al.,* 2009). Although natural resistance is the most effective means of controlling plant parasitic nematodes there is no major gene resistance available for control of *G. pallida* (Green *et al.*, 2012). It is possible that this situation reflects a larger and more diverse introduction into the EU of *G. pallida* compared to that of *G. rostochiensis*. This is illustrated by the fact that in South America, the centre of origin of PCN, both species display a range of virulence against all characterised resistance sources (Franco and Evans, 1978).

In a survey of England and Wales, PCN was detected in 64% of fields sampled. Of the infected fields, 66% contained only *G. pallida* and 25% had a mixture of *G. rostochiensis* and *G. pallida*. Compared to earlier surveys, this represents a substantial increase in the occurrence of *G. pallida.* This is likely to be due to the repeated use of

cultivars containing the *H1* resistance gene to control *G. rostochiensis*, thus selecting for *G. pallida,* which is not controlled by *H1* (Minnis *et al.*, 2002).

Potato cyst nematodes hatch as J2 in response to diffusates from host plant roots, with the moult from J1 to J2 having taken place inside the egg. The J2 nematode penetrates the plant root and migrates intracellularly through the zone of elongation to a site near the vascular tissue. Once a suitable cell – the initial syncytial cell – is located a feeding site, or syncytium, is initiated, most likely as a result of oesophageal gland secretions that are injected through the stylet into the chosen cell (Tytgat *et al.*, 2004; Williamson and Gleason, 2003; Hussey, 1989). Details of the syncytium and its development are provided below (section 1.2). Both male and female nematodes feed from a syncytial cell throughout the J2 and J3 stages. At this stage sexual dimorphism arises, which is controlled by environmental factors such as nutrient supply rather than genotype. Females remain sedentary whereas males stop feeding and regain motility. The females continue to feed and increase in size (Perry and Moens, 2006). Once the female has reached maturity, and after fertilisation by the male, the body tans to form a protective cyst surrounding 200–500 eggs. These eggs remain dormant until the next host crop is detected. This life cycle can take up to three months to complete (den Nijs, 2007).

1.2 Nematode feeding sites

Sedentary endoparasites are biotrophic pathogens that need to induce a feeding site in order to obtain nutrients from their host. These nematodes need to keep the feeding site alive for several weeks in order to complete their life cycle. The nematodes induce profound cytological modifications that increase metabolic activity (Bleve-Zacheo and Zacheo, 1987), change host gene expression (Szakasits *et al.*, 2009a) and increase transport of nutrients within their host (Grundler and Hofmann, 2011). The changes induced by the nematode are not restricted to the infection site but affect the whole plant as a consequence of changes to intrinsic plant signalling pathways (Hofmann and Grundler, 2007).

Cyst nematodes induce a feeding site known as a syncytium whereas root-knot nematodes induce giant cells. Both giant cells and syncytia act as metabolic sinks that deliver plant resources to the parasitic nematode (Williamson and Gleason, 2003). A syncytium is formed by local cell wall degradation and subsequent fusion of the

protoplasts of hundreds of cells (Perry and Moens, 2006). Nuclei within the syncytia undergo repeated S (synthesis) phases of the cell cycle (in which DNA is synthesised – also known as endoreplication) but without nuclear division (Gheysen and Jones, 2006). Syncytia are highly metabolically active and show a proliferation of cytoplasm, hypertrophy of smooth endoplasmic reticulum, ribosomes, mitochondria, plastids and an enlargement of the nucleus. Host cellular changes are controlled by the changes in gene expression induced by the nematode (Sobczak and Golinowski, 2008).

1.2.1 Gene expression in nematode feeding sites

The profound cellular changes that are induced by nematodes in order to provide the nematode with the nutrients required to complete its development are underpinned by major changes in host gene expression. Several large-scale studies of the changes in plant gene expression that occur in feeding sites have been undertaken using functional genomic approaches such as microarray analysis. Briefly, in a study of the interaction between soybean and soybean cyst nematode (*Heterodera glycines*) 1765 genes showed statistically significant changes in expression patterns (1116 up-regulated and 649 down-regulated) at 2 dpi (Ithal *et al.*, 2007a). Upon infection of *Arabidopsis, H. schachtii* was shown to cause up-regulation of 18.4% (3893) and down-regulation of 15.8% (3338) of the 21138 host genes in the syncytium at 5 and 15 dpi (Szakasits *et al.,* 2009). The substantial differences in the numbers of genes found to be differentially regulated in these two studies could be due to the different pathosystems used, the platforms used for the microarray analysis or could reflect variability in microarray data. In an analysis of the interaction between tomato and the RKNs *M. incognita* and *M. javanica* 24h after infection (Bhattarai *et al.*, 2008), 1497 genes in an incompatible interaction and 750 genes in a compatible reaction were found to be differentially expressed. In a study using *Arabidopsis* and *Meloidogyne* 3373 of the 22089 host genes were differentially expressed at different stages during the parasitic infection (Jammes *et al.*, 2005).

A microarray analysis comparing *Arabidopsis* infected with *H. schachtii* (beet cyst nematode – BCN – compatible) and *H. glycines* (SCN – incompatible) 3 days after infection identified 12 genes that are commonly altered in expression in both compatible and incompatible interactions. 116 genes were identified whose expression patterns changed in the compatible parasitic interaction with BCN (Puthoff *et al.*, 2003). These included plant defence associated genes such as coronatine-induced

proteins, heat shock proteins, thaumatin-like protein and ethylene responsive element binding protein (EREBP). Plant cell-wall modifying genes including proline rich proteins, polygalacturonase and a beta expansin were also up-regulated. Transcription factors and protein kinases involved in signal transduction pathways such as serine-threonine kinases, a calmodulin-related protein and a calcium-dependent protein kinase were down-regulated in infected roots (Puthoff *et al.*, 2003).

Changes in expression patterns of genes that control the cell cycle have been observed in feeding sites induced by cyst nematodes. For example, *Cdc2a* is expressed during cyst nematode infection and an *Arabidopsis thaliana* line containing the *cyc1* promoter linked to *GUS* showed high levels of GUS activity in young syncytia when infected with BCN (beet cyst nematode). This suggests that nematodes can manipulate the expression of cell cycle genes during feeding site induction. It has been shown that there is a similarity in cell cycle related gene expression in areas of lateral root formation and nematode feeding sites (Goverse *et al.*, 2000a). The induction of a nematode feeding site causes long term rearrangements to the cytoskeleton of the plant cell. Consequently, genes encoding cytoskeletal components, such as actin, are highly up-regulated. Tubulin genes are slightly up-regulated in syncytia and highly up-regulated in giant cells (Gheysen and Jones, 2006; de Almeida Engler *et al.*, 2004). The production of PR proteins and toxins suggests that the plant recognises the nematode in a compatible interaction (Bar-Or *et al.*, 2005). *WRKY* genes are thought to repress *PR* genes, including peroxidases which are associated with a hypersensitive response (HR). Several *PR* genes (*PR 1–5*), including a peroxidase, are down-regulated as a result of RKN nematode interaction (Bar-Or *et al.*, 2005). However, only *PR-4* was down-regulated by *H. schachtii* on *Arabidopsis* (Hamamouch *et al.*, 2011).

It is clear that the expression profiles of large numbers of genes are affected during plant–nematode interactions. The largest groups of differentially regulated genes during induction of a syncytium include genes related to metabolism, transcription, signalling, cell-wall related proteins and ribosomal genes, presumably reflecting increased metabolic activity in the feeding site (Puthoff *et al.*, 2003; Szakasits *et al.*, 2009a). The function of many differentially regulated genes is still unknown as many of them belong to complex gene families. Functional studies are required to gain an understanding of their mode of action. There is a vast amount of data produced in

micro-array experiments and further experimental evidence is needed to understand the functional significance of the observed changes.

1.2.2 Changes in plant hormones in the NFS

Nematodes manipulate the levels and distribution of plant hormones in order to induce and protect the feeding site. One of the most important plant hormones, auxin, is involved in many plant developmental processes including lateral root formation, apical dominance and gravitropism (Friml, 2010). There is a lot of evidence showing that auxin plays an important role in plant–nematode interactions. Auxin-insensitive tomato plants (*dgt* mutants) support fewer *G. rostochiensis* than control plants and this is consistent with the significant reduction in the number of BCN that form on the auxin-insensitive *Arabidopsis* mutant *axr2* (Goverse *et al.*, 2000a). Analysis of a GUS-auxin responsive promoter trap line showed that auxin accumulates rapidly in syncytia induced by *H. schachtii* in *A. thaliana* (Goverse *et al.*, 2000a). A reduction in mRNA levels of the auxin down-regulated genes *adr-6, -11* and *-12* has also been observed in syncytia induced by *H. glycines* (Hermsmeier *et al.*, 1998). In addition, auxin responsive genes have been shown to be up-regulated during RKN infection (Bar-Or *et al.*, 2005). PIN1 is an auxin efflux transporter and is involved in the polar movement of auxin (Blakeslee *et al.*, 2005) and a 40% reduction in *H. schachtii* cysts was observed on *pin1 Arabidopsis* mutants. PIN1 may be involved in the delivery of auxin to the feeding site at the early stages of initiation. The auxin transporter genes *PIN3* and *PIN4* have been shown to be highly expressed in the syncytia. PIN3 and PIN4 re-localise to lateral cell membranes during feeding site induction, suggesting that they are involved with radial expansion of the feeding site via lateral transport of auxin. The mutant lines *pin3* and *pin4* do not show compromised feeding site initiation but nematodes developing on these lines do produce much smaller cysts (Grunewald *et al.*, 2009). Recent work has shown how cyst nematodes may manipulate auxin levels and distribution. For example, an effector from *H. schachtii,* Hs19C07, has been shown to interact with an auxin influx transporter LAX3 resulting in an increase in the auxin influx rate in the syncytium (Lee *et al.*, 2011). These lines of evidence suggest that auxin has a critical role in the early stages of feeding site development (Goverse *et al.*, 2000a).

Ethylene is produced by all plant cells and is often associated with stress or wound responses (O'Donnell *et al.*, 1996). It also has a role in fruit ripening by activating

expression of cell wall modifying genes, as mentioned below (Alexander and Grierson, 2002). *Arabidopsis* ethylene production mutants do not support the formation of fully developed feeding sites while ethylene overproducing mutants are hyper-susceptible to BCN. In addition, ethylene insensitive mutants and ethylene signalling mutants (*etr1-1, ein2-1* and *ein3-1*) are less susceptible to nematode infection. Further to this RKN have also been shown to stimulate ethylene responsive genetic pathways (Bar-Or *et al.*, 2005). This highlights the importance of ethylene for successful completion of the nematode life cycle. Syncytial development is associated with extensive cell wall modifications that are brought about by activation of the plant's own cell wall degrading and modifying machinery. Nematodes change expression profiles of many cell wall modifying proteins, including expansins that are thought to disrupt non-covalent bonds between cellulose chains, cellulases that hydrolyse glucose β-1-4 linkages in cellulose, glycosyl hydrolases that hydrolyse non-crystalline cellulose and various enzymes that disrupt the microfibril network including pectate methylesterase, pectate lyase and polygalacturonase (Sobczak *et al.*, 2011). Expression of many of these enzymes is associated with ripening of fruits, a process that is controlled by ethylene (Alexander and Grierson, 2002).

Cytokinins are adenine-like molecules that are thought to be involved in the control of meristematic-cell division through their influence on the cell cycle (Zhang *et al.*, 1996). Biologically active cytokinins are produced and secreted by *H. schachtii* and *Meloidogyne* spp. (De Meutter *et al.*, 2003). In spite of these observations and although it is known that cytokinins can be produced by nematodes as an end point of tryptophan metabolism, the role of cytokinins in plant–nematode interactions has not been studied in any further detail.

Nematodes, like all biotrophic pathogens, need to suppress host defences in order to complete their life cycle (see section 1.417). Plant defences are controlled by the hormones jasmonic acid (JA) and salicylic acid (SA). The JA pathway is activated in response to herbivores while the SA pathway is activated in response to biotrophic pathogens and these two pathways are mutually antagonistic (Kunkel and Brooks, 2002). Many pathogens exploit this by altering the cross-talk between the jasmonic acid (JA) and salicylic acid (SA) pathways. For example, some bacterial plant pathogens produce coronatine, a JA mimic, causing inactivation of the SA defence pathways (Liu *et al.,* 2008). There is evidence that nematodes may also manipulate SA

and/or JA signalling pathways. Root knot nematodes are less successful on *jai* (JA signalling mutant) plants, but are not less successful on a JA biosynthesis mutant *def 1* (Bostock, 2005). This suggests that for a successful nematode–plant interaction, nematodes do not depend on JA biosynthesis but require an intact *Coi-1* signalling pathway. Root-knot nematodes may therefore produce a functional analogue of jasmonic acid that targets the *Coi-1* signalling pathway in order to avoid the activation of plant defences (Bhattarai *et al.*, 2008). Transcript profiling of developing syncytia induced by *H. schachtii* in *Arabidopsis* has shown a local down-regulation of JA biosynthesis genes, which may lead to a local suppression of host defences (Ithal *et al.*, 2007b). An effector from *H. schachtii* (10A06) has been shown to down-regulate SA-responsive genes (see Section 1.3.7) (Hewezi *et al.*, 2010b).

The studies described above have identified some of the changes to the transcriptome and host hormone levels induced in the nematode feeding site. However, the details of how nematode derived signals induce these changes are still incomplete. A better understanding of the nematode factors that are required for the initiation or maintenance of the nematode feeding site will shed light on this fascinating biological process and could provide targets for future pest control measures.

1.3 Nematode effectors

In addition to initiating a feeding site, nematodes need to invade a host plant and suppress host defence signalling pathways for as long as the feeding site is required. Effectors – defined here as any molecule produced by the nematode in order to manipulate the host to the benefit of the nematode – are responsible for each of these processes. Effectors are produced in the gland cells and secreted through the stylet into the plant (Figure 1.1). Identifying and characterising these effectors is a key goal for many research groups. Various approaches have been adopted to identify effectors including bioinformatic analysis of ESTs (Elling *et al.*, 2009a; Roze *et al.*, 2008; Jones *et al.*, 2009b), cDNA-AFLP analysis followed by *in situ* hybridisation to demonstrate expression in the gland cells (Tytgat *et al.*, 2004), proteomic identification of secreted proteins (Bellafiore *et al.*, 2008) and microaspiration of oesophageal gland cell mRNA followed by EST analysis (Huang *et al.*, 2003).

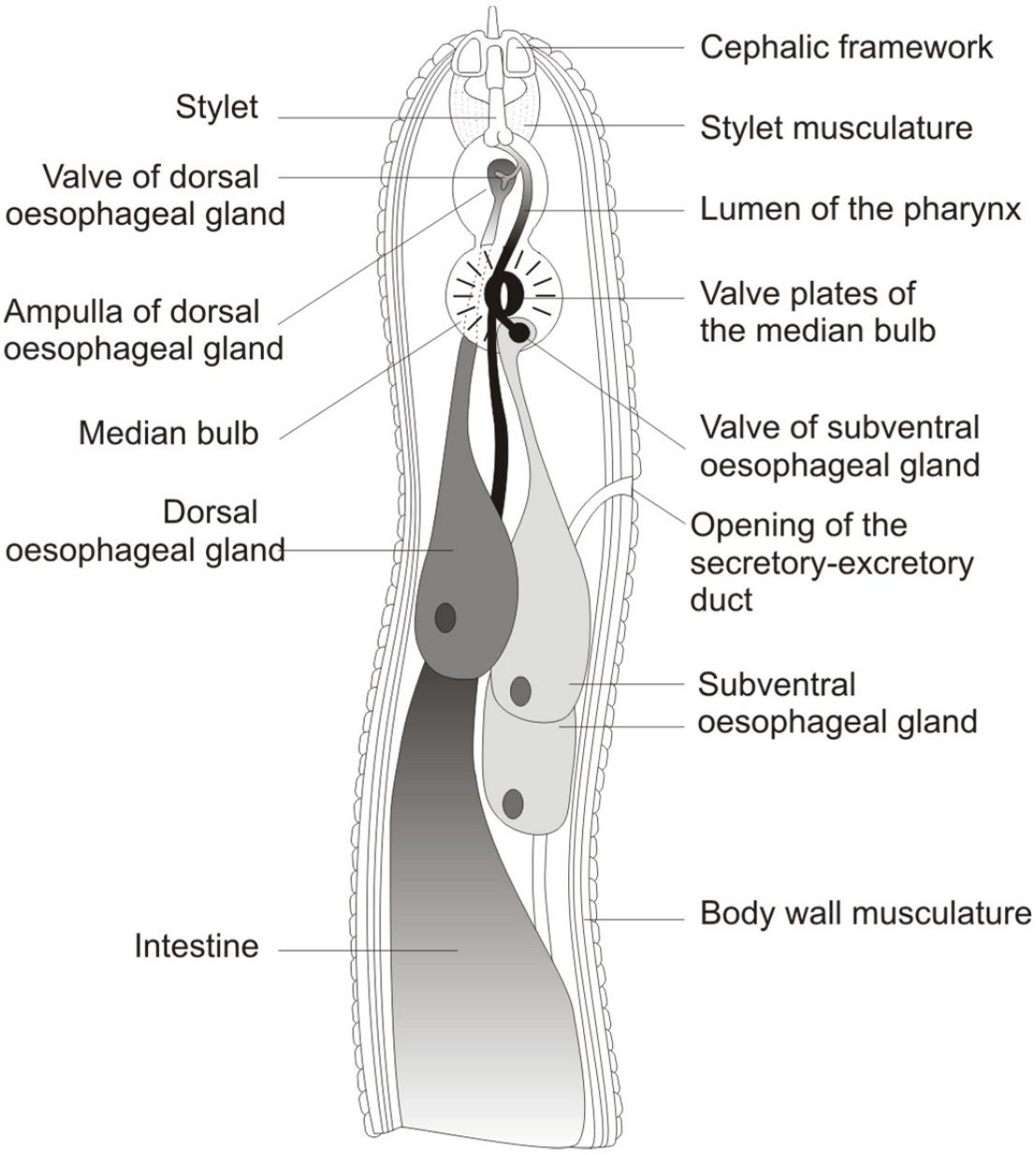


Figure 1.1 Schematic representation of a *Globodera pallida* J2. Effector molecules are synthesised in the dorsal and subventral glands and are then pumped, by the median bulb, out of the stylet into the host cell.

1.3.1 Gland cells

Tylenchid plant parasitic nematodes have two sets of oesophageal gland cells, the dorsal and sub-ventral. After hatching and in the early stages of parasitism (J2) the two sub-ventral gland cells are large and full of secretory granules (Gheysen and Jones, 2006). During later stages of the parasitic cycle the sub-ventral gland cells reduce in size. In addition, antibodies specific for the sub-ventral gland cells show a strong response in early parasitic stages and a much reduced response in adult females (Wyss, 1992). This suggests that the products of the sub-ventral gland cells are involved in the early stages of the parasitic process, including invasion and migration. The dorsal gland cell, by contrast increases in size throughout the life cycle of the nematode (Wyss and Grundler, 1992). Antibodies specific for the dorsal gland cell show a strong response in the adult female and a much reduced response in early parasitic life stages. This suggests that the dorsal gland cell is more important in the later stages of the life cycle and may produce proteins that help maintain the feeding site (Gheysen and Jones, 2006).

1.3.2 Cell wall degrading enzymes

The plant cell wall is the first barrier of defence against pathogen invasion. This obstacle has to be overcome in order for the nematode to successfully invade the host. Nematodes produce a wide range of plant cell wall degrading enzymes and the genes encoding these enzymes are thought to have been acquired via horizontal gene transfer from bacteria or fungi (Haegeman *et al.*, 2011a). Cell wall degrading enzymes facilitate the migration of the nematode through the plant root by softening the plant cell wall (Williamson and Gleason, 2003). The β-1,4-endoglucanases (cellulases) were the first effectors identified from plant nematodes (Smant *et al.*, 1998). Other cell wall degrading enzymes including pectate lyase (Doyle and Lambert, 2002), xylanase (Mitreva-Dautova *et al.*, 2006), polygalacturonase (Jaubert *et al.*, 2002) and, in cyst nematodes, an arabinogalactanase (Vanholme *et al.*, 2009) have subsequently been identified from PPNs. All cell wall modifying enzymes characterised to date are expressed in the subventral gland cells and antibodies against recombinant nematode cellulases have been used to show that they are secreted during migration (De Boer *et al.*, 1999). In addition, other proteins that modify the plant cell wall but that do not have enzymatic activity have been identified. These include expansins (Qin *et al.*, 2004) and cellulose-binding proteins (Ding *et al.*, 1998). Experimental evidence shows

that expansins break the hydrogen bonds between cellulose microfibrils in the cell wall, making the cell wall components more accessible to enzyme activity. The role of the cellulose binding proteins during invasion is not fully determined but one of these proteins has been shown to interact with, and promote the activity of, a host pectin methylesterase (Hewezi *et al.*, 2008). This protein may therefore contribute to the control of the modifications of the syncytial cell wall.

1.3.3 Chorismate mutase

Chorismate mutase is a key control enzyme of the shikimate pathway and is only usually present in bacteria, fungi and plants. However, this enzyme is produced in the subventral oesophageal gland cells of both cyst nematodes and root-knot nematodes (Lambert *et al.*, 1999a). The genes encoding chorismate mutase are thought to have been acquired via horizontal gene transfer from bacteria. It is extremely likely that the chorismate mutase produced by nematodes has a role in manipulation of the host since none of the other components of the pathway in which this enzyme operates are present within the nematode itself.

Chorismate mutase coverts chorismate to prephenate. Chorismate and prephenate are precursors for compounds associated with the auxin signalling pathway, synthesis of aromatic amino acids and production of salicylic acid, phenylpropanoids and a range of secondary metabolites (Jones *et al.*, 2003; Williamson and Gleason, 2003). It has been suggested that chorismate mutase-1 from *M. javanica* (MjCM-1) depletes cytoplasmic chorismate leading to a flux of this compound from the plastid into the cytoplasm. The result of this may be a depletion of IAA within plant tissues, as IAA is synthesised from chorismate in the plastid. Transgenic soybean plants expressing *MjCM-1* have a phenotype (suppressed lateral root and vascular tissue formation) consistent with a deficiency in IAA, that was rescued by the application of exogenous IAA (Doyle and Lambert, 2003). However, chorismate mutase is present in both cyst and root-knot nematodes and, since these nematodes induce entirely different feeding structures, a role for chorismate mutase in induction of the nematode feeding site seems unlikely. Alternative hypotheses have therefore been put forward for the role of chorismate mutase in plant–nematode interactions. Chorismate mutase activity may change levels of flavonoids in the host by increasing the levels of precursors (Gheysen and Fenoll, 2002). Flavonoids are natural inhibitors of auxin transport and it was suggested that an increase in flavonoid levels may allow local manipulation of auxin

levels by nematodes. However, mutants deficient in flavonoid biosynthetic pathways are not resistant to nematodes making it unlikely that this is the role of chorismate mutase (Jones *et al.*, 2007). Although the role of chorismate mutase in the plant–nematode interaction remains unclear, it has more recently been detected in migratory endoparasitic nematodes (Haegeman *et al.*, 2011b), suggesting a potential role in manipulation of host defences. In keeping with this, a chorismate mutase that suppresses host defences has recently been identified from fungi (Djamei *et al.*, 2011).

1.3.4 CLAVATA

A group of peptides similar to CLAVATA3/ESR-related peptides are secreted from the dorsal oesophageal gland cell of cyst nematodes. The proteins encoded by these genes contain one or more copies of a conserved 14 amino acid peptide (the CLE domain) located at the C-terminus of the protein following a variable domain. Plant CLE proteins are believed to be involved in maintenance of the shoot, floral and root meristems, regulation of organ size, apical dominance and control of vascular development. The nematode CLAVATA3/ESR-related peptides are the only reported occurrence of these peptides outside plants (Lu *et al.*, 2009).

Expression of the CLAVATA3/ESR gene *Hg-SYV46* from *H. glycines* in *A. thaliana* produces a *wus*-like phenotype, with a termination of the shoot apical meristem and flowers lacking a central gynoecium (Wang *et al.*, 2005). This phenotype is similar to that observed when over-expressing the endogenous plant peptides *CLV3* and *CLE40* (Fiers *et al.*, 2005). *CLV3* is expressed in the stem cells of shoots and floral meristems and is thought to be involved in controlling the balance of cell proliferation and differentiation through its interaction with CLV1/CLV2 and WUS. Evidence shows that *WUS* is down-regulated by the expression of CLAVATA3/ESR (Muller *et al.*, 2006). WUS is thought to act antagonistically to the CLV pathway by promoting stem cell formation and maintenance (Fletcher, 2002; Leibfried *et al.*, 2005). Where there is an over-expression of CLAVATA3/ESR, this may result in the inability to maintain an adequate number of stem cells in the shoot or floral meristem. Expression of *35S::Hg-SYV46* in a *clv3-1* mutant was able to recover the mutant phenotype. This suggests that the nematode protein has the same function as the plant CLV protein. The nematode protein may mimic the plant functional unit in order to redirect and maintain the differentiation of the root vascular cells into feeding cells (Wang *et al.*, 2005). CLAVATA3/ESR-related peptides have not been

identified in root knot nematodes, although a short peptide with sequence similarity to CLAVATA genes called 16D10 was identified. Transgenic expression of 16D10 could not rescue the *clv3* mutant phenotype and yeast-2-hybrid screens have shown that 16D10 interacts with a *scarecrow* transcription factor whose function is not related to the CLAVATA signalling pathway (Huang *et al.*, 2006b).

1.3.5 Components of the ubiquitin–proteasome pathway

The ubiquitin-proteasome pathway is used by all eukaryotic organisms to tag proteins with ubiquitin (UBI) molecule(s) for further processing, including degradation by the 26S proteasome. Therefore, the abundance of a protein within a cell, and thus its activity, can be altered by ubiquitination (Vierstra, 2009). The UBI-proteasome pathway is discussed in more detail in Chapter 5. Effectors that manipulate components of the ubiquitin-proteasome pathway have been identified from nematodes. Secreted UBI extension proteins have been identified that contain a C-terminal extension coupled to the conserved ubiquitin-like sequence. In *H. glycines* the protein is cleaved and the C-terminal extension is targeted to the nucleolus (Tytgat *et al.*, 2004). Two proteins (S-phase kinase-associated protein (Skp-1) and Ring-H2) that may form functional E3 ligases have also been identified as secreted proteins in ESTs derived from gland cells of cyst nematodes (Gao *et al.*, 2003a). E3 ligases are of key importance in the ubiquitination pathway as they determine the target proteins that are to be degraded. The role of secreted ubiquitin related proteins remains uncertain.

1.3.6 SPRYSECs

A large gene family (SPRYSECs) has been identified in PCN whose members are expressed in the dorsal oesophageal gland in J2s and are upregulated in early parasitic stages of cyst nematodes (Jones *et al.*, 2009a). SPRYSECs, (SECreted SP1a and RYanodine receptor domain) have highly conserved regions that fold into β-strands interspersed with highly variable loops (the SPRY domain). The SPRY domain may act as a hyper-variable binding surface within a stable β-strand scaffold structure. No similar gene family is present in RKN (Rehman *et al.*, 2009a).

One *G. pallida* SPRYSEC (RBP-1) has been identified as the avirulence factor recognised by the NB-LRR protein Gpa2. RBP1 was not recognised by the related NB-LRRs RX or RX2 in *Nicotiana benthamiana* (Sacco *et al.*, 2009b). In addition, a SPRYSEC from *G. rostochiensis* interacts with an “orphan” NB-LRR protein, SW5,

from tomato, although this interaction does not result in a HR. It is possible that SPRYSECs suppress host defences and SPRYSEC proteins have been suggested to act as adapters in multi component E3 ubiquitin ligases (Rehman *et al.*, 2009a).

1.3.7 Other effectors

Many other candidate effectors have been identified from plant parasitic nematodes, often as a result of EST projects coupled to *in situ* hybridisation (*e.g.* Gao *et al.,* 2003). The host targets of some of these proteins have been identified and this has allowed their function to be determined. For example, Hg30C02 from *Heterodera glycines* interacts with a host β-1,3-endoglucanase. Since this is a pathogenesis-related protein the nematode effector may prevent its activity and suppress activation of host defences (Hamamouch *et al.,* 2012). The *H. schachtii* effector HS19C07 interacts with LAX3 and is thought to modulate auxin flow into the cell, as described in section 1.2.2 above (Lee *et al.*, 2011). The 10A06 effector from *H. schachtii* (section 1.2.2) interacts with Spermidine Synthase2, an enzyme involved in polyamine biosynthesis. The interaction of 10A06 with the spermidine synthase increases the abundance of this enzyme which in turn increases polyamine oxidase activity. This then stimulates the activity of anti-oxidants in the syncytia and also disrupts SA defence signalling. Transformed *Arabidopsis* overexpressing 10A06 were more susceptible to nematodes, *Pseudomonas syringae* and *Cucumber mosaic virus* (Hewezi *et al.*, 2010b).

A calreticulin from RKN is secreted both during migration and after giant cell induction and may be involved in calcium sequestration (Vanholme *et al.*, 2004; Caillaud *et al.*, 2008). It has recently been demonstrated that calreticulin suppresses host defences, possibly through its ability to sequester calcium which is required for host defence signalling processes (Jaouannet *et al.*, 2012).

A root-knot nematode gene has been identified that encodes a NODL-like protein. NODL proteins acetylate a polyglucosamine chain as part of the process of production of NOD factors by bacteria. The NOD-L-like gene is thought to have been acquired via horizontal gene transfer from a nitrogen fixing bacterium (Scholl *et al.*, 2003; Bird and Koltai, 2000). Substances secreted by RKN have been shown to elicit responses identical to those induced by nodulation factors from rhizobial bacteria in *Lotus japonicus*. NOD factors and the nematode substances induce subcellular and cytoskeletal reorganisation and, as a consequence, root-hair waviness and branching.

Analysis of nodulation deficient mutants suggests that the nematode substances, like NOD factors, are perceived by NFR1 and NFR5 receptor kinases in the host. The importance of this pathway was confirmed by the observation that *nfr1* and *nfr5* mutants supported fewer RKN than control plants. Nematode NOD-like factors may therefore be involved in the initiation of feeding sites in RKN and they may have gained parts of the symbiont-response pathway to enhance their ability to establish a successful infection (Weerasinghe *et al.*, 2005).

Expressed sequence tag (EST) projects and analysis of the secretome from cyst nematodes and RKN have identified large numbers of secreted proteins produced in the gland cells that have no significant similarity to those in the current databases (Elling *et al.*, 2009a; Jones *et al.*, 2009b). Analysing the function of these proteins represents a significant challenge.

1.4 Plant defence system

Plants lack an adaptive immune system like that of animals and therefore rely on an innate defence system within each cell and subsequent systemic signalling. The plant defence system can be simplified into two main strands – Pathogen Triggered Immunity (PTI) (Zipfel and Robatzek, 2010) and Effector Triggered Immunity (ETI) (Jones and Dangl, 2006). All biotrophic organisms, including nematodes, need to suppress the plant defence system in order to successfully parasitise plants. A detailed understanding of the plant defence system is required before the mode of action of an effector that suppresses host defences can be determined.

The current overview of the plant defence system can be represented by the 'zig zag' model (Figure 1.2). In the first phase pathogen- or microbial-associated molecular patterns (PAMPS or MAMPs) are recognised by plant Pattern Recognition Receptors (PRRs), resulting in activation of the first layer of host defences, PAMP triggered Immunity (PTI). During phase 2, biotrophic pathogens use effectors to suppress PTI leading to effector-triggered susceptibility (ETS). In phase 3 plant resistance proteins, if present, recognise the effectors and produce a hyper-sensitive response (HR) resulting in effector-triggered immunity (ETI). Plant resistance proteins may not always interact with the pathogen effectors directly but may instead monitor host cellular targets of effector action. This is termed the 'guard hypothesis' (Dangl and McDowell, 2006). The HR may be suppressed by other effectors or may be avoided

by the pathogen evolving effectors that are not recognised by the resistance proteins. The plant, in turn, responds by evolving new or modified resistance proteins, leading to a battle between ETI and ETS. There is high selection pressure on effector genes and plant *R* genes. This can be considered as an 'evolutionary arms race' (Jones and Dangl, 2006).

1.4.1 PAMP triggered immunity (PTI)

The first layer of plant defence is provided by a group of extracellular-transmembrane pattern recognition receptors which have evolved to recognise and respond to common classes of slowly evolving MAMPs or PAMPS (Monaghan and Zipfel, 2012). PAMPs are essential for the fitness of the pathogen, are highly conserved, absent from the host and remain unchanged even under strong selection pressure (Boller and Felix, 2009). Examples of PAMPs include FLG22 – a conserved 22 amino acid region of flagellin (Jones and Dangl, 2006), INF 1 – a secreted protein produced in abundance by *P. infestans* (Bos *et al.,* 2010), and chitin – an essential component of the fungal cell wall. No PAMPs have been identified from nematodes, although it has been suggested that chitin present in the stylet may act as a PAMP (Libault *et al.*, 2007). The defence responses activated as a result of PTI include cross-linking of the cell wall, deposition of callose and production of reactive oxygen species (ROS). Signalling pathways that control these responses include an influx of calcium ions, activation of mitogen-activated protein kinases (MAPK), reprogramming of gene expression and systemic signalling to activate defence responses in neighbouring cells (Dodds and Rathjen, 2010). When nematodes are unsuccessful in initiating a feeding site local callose deposition can be observed around the stylet (Hussey *et al.*, 1992) suggesting activation of PTI responses in response to the nematode.

1.4.2 Effector triggered susceptibility (ETS)

All successful biotrophic pathogens have to suppress PTI. Many pathogens have evolved effectors that are secreted into the host apoplast or cytoplasm in order to suppress PTI. For example, during the interaction between *P. infestans* and Solanaceous plants INF1 acts as a PAMP that activates PTI. However, a *P. infestans* effector, AVR3a, has been identified that interacts with and stabilises CMPG1,

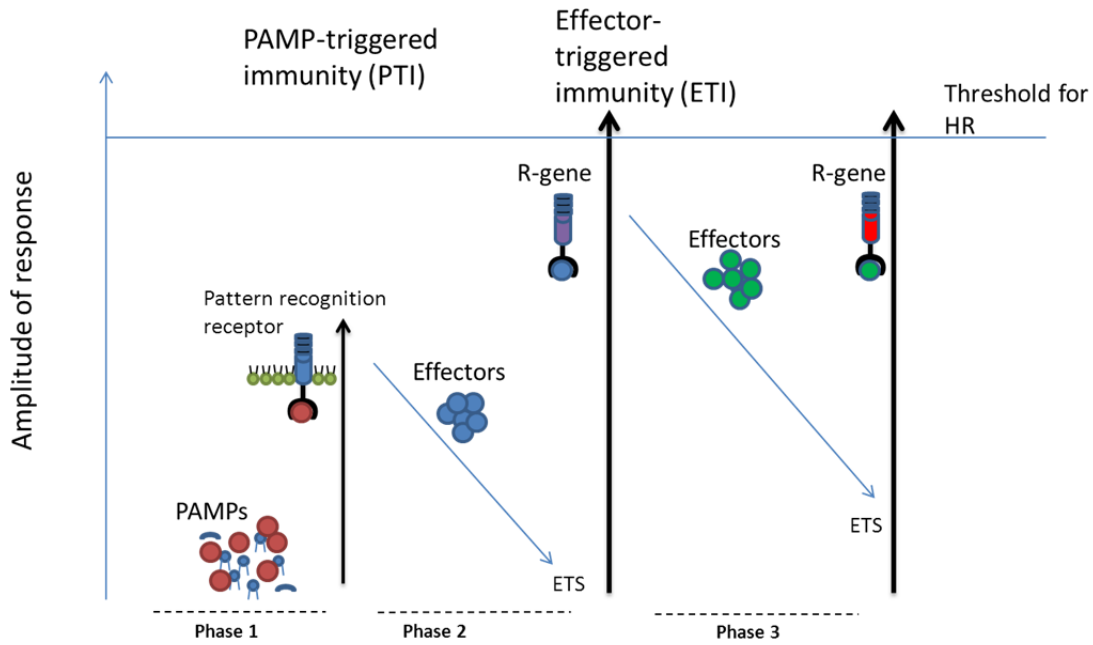


Figure 1.2: The 'zig zag' model representation of the current overview of the plant defence system. Phase 1: Conserved molecules (PAMPS or MAMPs) are recognised by plant transmembrane domain proteins, resulting in PTI. Phase 2: Biotrophic pathogens use effectors to interfere with PTI leading to effector-triggered susceptibility (ETS). Phase 3: Plant R proteins, if present, recognise the effectors or targets of the effectors, termed the 'guard hypothesis' (Dangl and McDowell, 2006) and produce an HR (ETI). The HR, as a result of phase 3, may in turn be suppressed by effectors, and again these effectors may be recognised by plant R proteins, leading to a battle between ETI and ETS (Jones and Dangl, 2006).

which is a component of the downstream signalling pathway in INF1-induced defences. As a result of this interaction, INF1 induced PTI is suppressed. CMPG1 is also a hub for other defence signalling pathways, including those induced by perception of effectors by Cf4, Cf9 and Pto and these defence responses are also suppressed by Avr3a (Gilroy *et al.*, 2011). Other examples of pathogen suppressors of PTI include Ecp6 from *Cladosporium fulvum* which is thought to outcompete the perception of chitin by either sequestering chitin oligosaccharides or interfering with receptors responsible for their perception (de Jonge and Thomma, 2009) and the AvrPtoB effector from *Pseudomonas syringae* that interferes with Flg22 perception by FLS2 by using the host ubiquitin-proteasome pathway in order to degrade the receptor (Xiang *et al.*, 2008). The processes of suppression of PTI by nematodes are less well characterised than those in other pathosystems although two effectors that may be involved in this process, SPRYSEC19 (Postma *et al.*, 2012) and calreticulin (Jaouannet *et al.*, 2012), have recently been identified.

1.4.3 Effector triggered immunity (ETI)

During ETI (effector triggered immunity), plant resistance proteins – most often nucleotide binding leucine rich repeat (NB-LRR) proteins – recognise pathogen effectors, or changes to the targets of the effector (Dangl and McDowell, 2006). Recognition of the presence of the pathogen leads to a strong, localised cell death known as the hypersensitive response (HR) (Dodds and Rathjen, 2010). The hypersensitive response is a 'scorched earth'-like approach that is employed by the plant to leave no colonisable cells for the pathogen, in an attempt to reduce its chance of creating a successful infection (Pritchard and Birch, 2011b). ETI is only effective against pathogens that colonise living tissue (obligate biotrophs or hemi-biotrophs), and is therefore not effective against necrotrophs (Jones and Dangl, 2006).

Many resistance genes that recognise the presence of nematodes have been identified. Although some, such as *Mi,* target the developing feeding structure, those against cyst nematodes often show a delayed cell death response targeted at the cells around the syncytium. The initial syncytium is therefore formed but subsequently collapses, or shows restricted development, when the cells surrounding the syncytium degenerate (Sobczak and Golinowski, 2011).

The only known nematode effector/R gene combination that has been identified to date is RBP1 and Gpa2. RBP1 is a SPRYSEC effector (section 1.3.6) of as yet undetermined function. Recognition of RBP1 by Gpa2 is dependent on the presence of a proline residue at position 187, which is located within the SPRY domain. RBP1 is under selection pressure to avoid detection and as a result of this other isoforms of RBP1 that have a serine at position 187 instead of a proline, and that do not elicit a HR, have evolved (Sacco *et al.*, 2009b). It has been suggested that other polymorphisms may also alter the strength of the interaction, and hence the defence response mounted as a result of Gpa2-mediated recognition of RBP1. Six of these polymorphisms are predicted to be located on the extended loop of the SPRY domain (Carpentier *et al.*, 2012). The *Mi* resistance gene which encodes a protein consisting of a Coiled-coil-NB-LRR has been successfully used to control tropical *Meloidogyne* species (Jacquet *et al.*, 2005; Jablonska *et al.*, 2007). Two nematode factors have been identified that may be associated with avirulence against *Mi.* A secreted protein produced in the amphids of *M. incognita* shows polymorphisms that correlate with virulence against *Mi*, although functional evidence showing induction of cell death in the presence of *Mi* and the secreted protein is still lacking (Semblat *et al.*, 2001). A gene called *Cg1*, which is present in avirulent RKN populations but absent from virulent populations, has been identified as a candidate Avr gene for *Mi*. When *Cg1* was silenced in avirulent *M. javanica,* the nematodes were rendered virulent (Gleason *et al.*, 2008). However, the *Cg1* transcript does not encode a secreted protein capable of interacting with a host resistance protein. It is possible that this transcript is involved in regulation of another *M. javanica* gene which may itself be the avirulence factor. Interestingly, recessive mutant *rme-1* tomato plants do not respond to *Mi-1.2* activation by *M. javanica* but the expression of an auto-active form does induce defence responses, suggesting Rme-1 could be a target for a nematode effector (Kaloshian *et al.*, 2011).

Very little is known about how nematodes suppress ETI. However, SPRYSEC19 from *G. rostochiensis* was found to bind to the LRR domain of the R-gene protein product SW5F (Rehman *et al.*, 2009b). SPRYSEC19 was subsequently shown to suppress defence responses from several CC-NB-LRR *R*-gene induced defence pathways, such as SW5B, Gpa2, RX1 and RGH10, but did not suppress defence responses induced by TIR-NB-LRR or extracellular-LRR proteins. SPRYSEC19 does not physically interact

with these proteins and it is therefore thought that it may disrupt *R*-gene induced signalling (Postma *et al.*, 2012).

One of the best studied examples of an effector–R gene interaction is the *P. infestans* effector AVR3, described above. The *P. infestans* PAMP INF1 activates PTI which is suppressed by AVR3a. The R-gene R3a detects one allele of AVR3a, AVR3aKI, resulting in a strong HR. In response, *P. infestans* has evolved a modified version of the AVR3a effector called AVR3aEM which has amino acid changes at positions 80 and 103. AVR3aEM is attenuated in its ability to suppress PTI induced by INF1 perception but does evade R3a recognition (Bos *et al.*, 2009). This is an example of the 'zig zag' model in operation.

1.5 Aims

The aims of this project are:

- Identify effectors from the genome sequence of *G. pallida* using a bioinformatics approach.
- Analyse the evolution of effector gene families in *G. pallida.*
- Determine the function of selected effectors using over-expression in plants and assays for suppression of host defences.

2 Materials and methods

2.1 Biological material

2.1.1 Maintenance and storage of nematodes

Globodera pallida population Lindley was grown on the susceptible potato cultivar Désirée in a glasshouse at the James Hutton Institute in glasshouse conditions at 18–20°C under 16 h/8 h light/ dark. A PCR diagnostic (Pylypenko *et al.*, 2005) was used to confirm the identity of the nematode species. Cysts were stored at 4°C until they were required. All populations had passed through diapause before being used.

2.1.2 Collection of second stage juveniles

Cysts of *G. pallida* from the stock population Lindley E2008 were incubated in sterile distilled water (SDW) at 20°C in the dark for 48h. The cysts were then washed several times in SDW to remove fungal contamination and then incubated in tomato root diffusate for one week at 20°C in the dark. Tomato root diffusate was produced by removing the soil from the roots of 2 tomato plants (cultivar Moneymaker) and standing the plants in 500 ml of SDW for 2 hours. The resulting liquid was filtered and stored at 4°C. Hatched second stage juvenile nematodes were kept at 4°C for a maximum of 1 week before being used for infecting plants.

2.1.3 Infection of plants to obtain parasitic stage nematodes

Nematodes were counted and diluted to a concentration of 250 nematodes ml^{-1}. 10 ml (2500 nematodes) of nematode suspension was applied directly to potato roots 2 weeks after planting in a root trainer (Haxnicks, Oxford, UK). Infected plant root material was harvested at 7, 14, 21 and 28 days post infection and was homogenised using a blender. The resulting homogenate was passed over a series of sieves. The first of these (150 μm) removed larger root pieces and soil while the second (30 μm) collected smaller solid material including the nematodes. Nematodes were purified manually from a resuspension of this material under a binocular microscope. Nematodes were frozen in liquid nitrogen and then stored at -80°C until use.

2.2 *Arabidopsis* culture, sterilisation and growth

Wild type *Arabidopsis thaliana* seeds (Col 0) were sterilised by incubating them in 20% bleach (Domestos, Unilever UK) for 20 mins with rotation. The sterilisation was followed by centrifugation in a microcentrifuge for 30 s at 3000 rpm, the

supernatant was removed and the seeds were resuspended in sterile distilled water in aseptic conditions. This wash procedure was repeated 5 times to remove excess bleach. The sterilised seeds were then incubated overnight in the dark at 4°C to synchronise germination. The sterilised seeds were plated on 0.5x MS10 which consisted of 2.2 gl^{-1} Murashige and Skoog medium (Duchefa Biochemie), 10 gl^{-1} sucrose (Fisher Scientific, Loughborough, UK) adjusted to pH 5.8 using potassium hydroxide (Fisher Scientific). The seeds were grown in a growth chamber at 20°C with 18 hours per day photoperiod.

For glasshouse conditions *Arabidopsis* was grown in compost, sand and loam soil (Sinclair Potting & Growing Medium, East Riding Horticulture) in a ratio of 2:1:1 at 20°C under 16 h/8 h light/ dark.

2.3 Acid fuchsin staining

Acid fuchsin staining was used to stain nematodes in root systems in order to determine infection rates. The infected root system was separated from the stem and upper parts of the plant and weighed in order to allow determination of the infection rate per gram of root due to the variability in size between lines of interest. The root system was then soaked in 1% hypochlorite solution (Fisher Scientific) for 5 mins for potato root, or 2 mins for *Arabidopsis,* followed by three 5-minute washes in water. The root system was then transferred to a beaker of boiling 1X stain solution (10X stock stain solution contains 0.35% w/v acid fuchsin and 25% v/v glacial acid) for 2 mins. Excess stain was removed in water. The resulting stained root system was then stored in acidified glycerol solution until analysis under a microscope. Worms were counted and the life cycle stage of each worm was recorded. Images were obtained using a Leica M165C microscope with a Micropublisher camera controlled by QCapture Pro software.

2.4 Potato cultivation

Solanum tuberosum L. 'Désirée' was used for all experiments. Potato plants were grown in pre-mixed compost (William Sinclair Horticulture Ltd). The compost mix contained the following: Peat 1200 l, sand 100 l, magnesium limestone 2.5 kg, calcium limestone 2.5 kg, osmocote 1.5 kg, celcote 0.5 kg, vermiculite 500 l, sincrocel 3.0 kg. Plants were maintained in glasshouse conditions at 18–20°C under 16 h/8 h light/ dark.

2.5 Molecular biology

2.5.1 Polymerase chain reaction (PCR) and qPCR

Sequences of interest were amplified by PCR. PCR reactions contained 1× *Taq* buffer (Promega, Southampton, UK), 1.5 mM $MgCl_2$, 200 µM dNTPs, 1 µM each primer, and 1 unit of *Taq* DNA polymerase (Promega). Basic PCR cycling conditions consisted of one cycle of denaturing at 94°C for 5 mins followed by 35 cycles of 15 s denaturing at 94°C, 15 s annealing at 54°C and 30 s extension at 72°C. Annealing temperature varied according to the primers being used and extension time was increased where longer sequences were being amplified (1 minute per kilobase). PCR reactions were performed on an ABI Gene Amp 9700 PCR machine (Applied Biosystems, Paisley, UK). Where high fidelity was required a proof reading polymerase (KOD – Novagen) was used in place of *Taq* polymerase.

A Mx3500P qPCR thermo-cycler (Stratagene) controlled by MxPro software was used for all experiments. Each 25 µl qPCR reaction contained: Forward primer and reverse primer at 7.5 µmol l^{-1} each, 1x Bioline (London, UK) SYBR mix without ROX, 5µl of cDNA template. Primer optimisation was performed for all qPCR primers using concentrations of 900 nM, 300 nM, and 50 nM and cDNA dilutions of 1:20, 1:40, and 1:80 to achieve amplification efficiency between 90 and 110%. For all primers used the optimum concentration was 300 nM. qPCR cycling conditions consisted of one cycle of denaturing at 95ºC for 15 mins followed by 40 cycles of 15 s of denaturing at 95°C, 30 s of annealing at 59°C and 30 s of extension at 72°C. For a melting curve analysis, fluorescence data were collected at every 1°C from 59–95°C. Melting curve data confirmed a single product was amplified and no primer dimers were present in the cDNA samples.

2.5.2 Gel electrophoresis

PCR products were analysed by gel electrophoresis in 1% agarose (Invitrogen, Paisley, UK) gels run in 1X TAE buffer (50X stock: 242.2 g Tris, 57.1 ml glacial acetic acid and 18.6 g EDTA disodium salt in 1 litre of water – Fisher Scientific) and stained with SYBR Safe DNA gel stain (Invitrogen). Electrophoresis was typically carried out at 75 V for 25 mins. The gels were imaged using a UVIdoc machine (UVItec Ltd, Cambridge, UK) and images were analysed using the associated software.

2.5.3 DNA purification

PCR products of interest were purified using a QIAquick PCR Purification Kit (Qiagen, Crawley, UK) following the manufacturer's instructions. Occasionally, for example when multiple PCR products were obtained, PCR products were purified from gel fragments excised following agarose gel electrophoresis as described above. In this case a QIAquick gel purification kit (Qiagen) was used following the manufacturer's instructions.

2.5.4 Plant DNA extraction

Total DNA was extracted from plant tissue using a DNeasy Blood and Tissue Kit (QIAgen) following the manufacturer's instructions. Briefly, material was homogenised in extraction buffer and treated with Proteinase K before the DNA was allowed to bind to a DNeasy mini spin column. Bound DNA was washed repeatedly and eluted in TE buffer (10 mM Tris, 0.5 mM EDTA – Fisher Scientific). DNA was quantified using a Nanodrop spectrophotometer ND-1000 (Thermo).

2.5.5 RNA extraction

A maximum of 100 mg of material was frozen using liquid nitrogen and ground to a fine powder, either in a mortal using a mortar and pestle or using a 1.5 ml tube and an RNase and DNA-free plastic pestle. The sample was not allowed to thaw during this stage. Following the protocol from a RNeasy Plant Mini kit (Qiagen) 450 µl of buffer RLT with β-mercaptoethanol was added to the sample. The sample was then allowed to thaw while grinding continued. The resulting mixture was centrifuged through a QIAshredder spin column to remove cell debris. Ethanol was then added to the flow-through to create the required conditions for binding. The lysate was then centrifuged through an RNeasy spin column. Washes using RW1 and RPE buffers were used to clean the membrane. Finally the RNA was eluted off the membrane using 30 µl of RNase-free water. An optional on-column DNase-I digestion within the RNeasy protocol was not used; instead genomic DNA was removed using a separate treatment with DNAase RQ-1 (Promega).

2.6 Cloning

2.6.1 Cloning of PCR products

Purified PCR products were cloned using the pGemT Easy Vector System (Promega) following the manufacturer's instructions. Up to 150 ng of purified PCR product was incubated with 1µl of pGEM T Easy vector in 1X T4 DNA ligase buffer and T4 DNA ligase. Ligations were incubated for 1 hour at room temperature and then transformed into *E. coli* DH5α competent cells by electroporation using a Biorad Micropulser. Transformed cells were plated on LB agar plates (10 g NaCl, 10 g tryptone, 5 g yeast extract and 10 g agar in 1 litre of water) containing 50 µg/ml ampicillin, and 50 µg/ml X-gal. Transformants harbouring recombinant plasmids were identified by blue-white colony selection. Colony PCR was used to identify transformants harbouring the desired clone. Ten bacterial colonies of interest were resuspended separately in 100 µl of SDW (sterile distilled water) and incubated for 10 mins at 90°C. This was then pulse centrifuged to remove cell debris. 1 µl of supernatant was then used per PCR reaction. PCR reactions using vector and gene specific primers were used to confirm the presence and orientation of the gene in the cloning vector.

2.6.2 Cloning into the Gateway System

PCR was used to amplify genes of interest using a proof reading DNA polymerase (KOD Hot Start DNA polymerase). The primers used for this process contained the attB sites to allow BP cloning into a Gateway donor vector. As an alternative, gene sequences were amplified using primers lacking the attB sites and PCR products were cloned into the pCR8 vector using the pCR8/GW/TOPO cloning system (Invitrogen) which allows subsequent LR transfer into the destination vector.

2.6.3 BP reaction

PCR products were purified and cloned into the pDONR221 donor vector using BP Clonase II enzyme (Invitrogen) following the manufacturer's instructions. 15–150 ng of purified PCR product was mixed with 150 ng of pDONR221 vector and 2 µl of BP Clonase II was added. The reaction was left overnight at room temperature and stopped by the addition of 1 µl of Proteinase K. Recombinant plasmid was electroporated into *E. coli* DH10B competent cells and the presence of the expected

inserts was confirmed by colony PCR and analysis of plasmid sequence as described above. All cloned products were sequenced as described in section 2.6.8.

2.6.4 LR cloning

Donor plasmids selected for LR recombination into destination vectors were digested with *Bsp*HI (New England Biolabs, Hitchin, UK) and cleaned using a PCR Purification kit (Qiagen). Digested plasmid was mixed with appropriate destination vector (pMDC32, pGWB5, pGWB6, pMDC43, pMDC83 or pK7WG2) (Karimi *et al.*, 2002a) and LR Clonase II and left overnight at room temperature. Aliquots of the reactions were electroporated into *E. coli* DH10B cells and positive transformants were analysed by colony PCR and sequencing using gene specific primers.

2.6.5 Transformation of *Agrobacterium tumefaciens* GV3101 – heat shock

Heat shock-competent *Agrobacterium tumefaciens* GV3101 cells were prepared by inoculating a 5 ml culture of LB medium containing 50 µg/ml gentamycin (10 g NaCl, 10 g tryptone and 5 g yeast extract in 1 litre of water) and incubating overnight with shaking at 28°C. 2 ml of this culture was added to 50 ml of LB medium in a sterile 250 ml flask and incubated with shaking at 28°C until the $O.D._{600nm}$ (Optical density) had reached between 0.5–1.0 (0.6 was considered optimal). The culture was then chilled on ice before centrifugation at 3000 *g* for 5 mins at 4°C. The supernatant was removed and the pelleted bacteria were washed and resuspended in 1 ml of 20 mM $CaCl_2$. The bacterial cells were centrifuged and washed in 1 ml of 20 mM $CaCl_2$ three more times. After the final wash, the cells were dispensed into pre-chilled tubes in 100 µl aliquots and immediately frozen in liquid nitrogen and stored at -80°C for future use.

To transform these cells, a 100 µl aliquot was thawed on ice and 5 µl (approximately 0.5–1 µg) of purified plasmid was added to the cells and gently mixed. The tube was immediately frozen in liquid nitrogen and thawed in a water bath at 37°C for exactly 5 mins and then 1ml of LB broth was added. This was then incubated with shaking at 28°C for 2 hours. The tube was centrifuged for 30 s at maximum speed in a microcentrifuge and most of the supernatant was removed. The pellet was resuspended in the remaining supernatant (approx. 100 µl) and plated on an LB plate containing appropriate antibiotics.

2.6.6 Transformation of *Agrobacterium tumefaciens* GV3101 – electroporation

Electrocompetent *Agrobacterium tumefaciens* GV3101 cells were prepared by inoculating a 5 ml culture of LB medium and incubating overnight with shaking at 28°C. 2 ml of this culture was added to 50 ml of LB medium in a sterile 250 ml flask and incubated with shaking at 28°C until the $O.D._{600nm}$ had reached between 0.5–1.0 (0.6 was considered optimal). The culture was then chilled on ice before centrifugation at 4000 *g* for 15 mins at 4°C. The supernatant was removed and the pelleted bacteria were washed and resuspended in 20 ml of sterile 10% glycerol solution. This wash step was repeated 3 more times to remove salts, with a final re-suspension in 5 ml of 10% glycerol solution. The cells were dispensed into pre-chilled tubes in 100 µl aliquots and immediately frozen in liquid nitrogen and stored at -80°C for future use.

To transform these cells, a 100 µl aliquot was thawed on ice and 5 µl (approximately 0.5–1 µg) of purified plasmid was added to the cells and gently mixed. The competent cell and plasmid mixture was then transferred to a pre-chilled electroporation cuvette. The outer surface was dried using paper towels and the sample was subjected to electroporation using a Biorad Micropulser on setting EC2. Immediately following electroporation 1 ml of LB was added to the electroporation cuvette. This mixture was then transferred to a 2 ml Eppendorf tube which was incubated for 2 hours at 28°C. Cells were plated on LB plates with appropriate antibiotics.

2.6.7 Plasmid purification

Bacteria containing plasmids with inserts of the anticipated size were grown in 3 ml of LB with appropriate antibiotic overnight at 37°C and plasmids were prepared using a GeneJet plasmid preparation kit (Fermentas) following the manufacturer's instructions. Bacteria were lysed by alkaline lysis and precipitated cell debris was removed by centrifugation. DNA was bound to a column, washed and eluted in sterile distilled water. Plasmid yield and quality were checked by agarose gel electrophoresis as described above and using a Nanodrop spectrophotometer.

2.6.8 Sequencing and sequence analysis

All sequence analysis was performed by the JHI sequencing service using an ABI 3730 DNA sequencer. Sequences were first edited in BioEdit and subsequently analysed in more detail using appropriate software packages for the task in hand.

2.7 Western Blotting

2.7.1 Protein Extraction and blotting

1 cm diameter samples were cut from leaves using a cork borer, frozen in liquid nitrogen and ground to a fine powder using a mortar and pestle. A protease inhibitor tablet (Complete Mini EDTA-Free; Roche Diagnostics, West Sussex, UK) was dissolved in 10 ml of 1X PBS (phosphate buffered saline: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4) and DTT was added to a final concentration of 10 mM. 500 µl of this solution was then added to the ground sample. The resulting liquid was centrifuged at 11,700 *g* for 3 mins. The supernatant was retained and subjected to repeated centrifugations until the collected supernatant was clear. An approximate protein concentration was obtained using a Nanodrop spectrophotometer reading at 280 nm. Where necessary, proteins were concentrated using Microcon columns (Millipore, Watford, UK).

Samples were prepared for Western blotting by mixing protein extract with 2 µl of reducing agent and 5 µl of 4X LDS sample buffer (Invitrogen NuPage). These samples were incubated at 75°C for 10 mins. Running buffer was made from a 20X stock of NuPage MOPS SDS running buffer (Invitrogen) supplemented with 200 µl of NuPage antioxidant for the 200 ml loaded into the inner-chamber of the gel tank. Samples and protein standard (Invitrogen Novex Sharp Pre-Stained Protein Standards) were loaded on NuPage 4–12% Bis-Tris Gels, which were run at 200V for 30 mins.

The proteins on the gel were then transferred to nitrocellulose membrane (Hybond – ECL; GE Life Sciences, Bucks., UK) in NuPage Transfer buffer containing 10% methanol (Fisher Scientific) for 2 hours at 30 V. The nitrocellulose membrane was then incubated overnight at 4°C in blocking solution (PBS containing 2% milk

powder and 0.1% Tween-20). The proteins on the membrane were analysed either by colorimetric detection or chemiluminescent detection.

2.7.2 Colorimetric detection

The membrane was incubated in 10 ml of blocking solution containing the primary antibody (*e.g.* Anti-GFP rabbit serum at 1:2000 dilution – Invitrogen) and secondary antibody (*e.g.* Anti-Rabbit IgG Alkaline Phosphatase Conjugate at 1:10 000 dilution; Sigma, Poole, UK) for 2 h with shaking at room temperature. Primary and secondary antibodies varied depending on the experiment. The membrane was then washed three times in PBS containing 0.1% (v/v) Tween-20 for at least 10 mins each time. The membrane was then incubated in detection solution which was composed of 5 ml of detection buffer (0.1 M Tris, 0.05 M $MgCl_2$ and 0.1 M NaCl), 25 µl NBT of 100 mg/ml stock solution (4-Nitro Blue Tetrazolium Chloride; Roche) and 18.5 µl of 50 mg/ml stock solution BCIP (5-Bromo-4-Chloro-3-Indolyl-Phosphate; Roche). The incubation was stopped in SDW when clear banding patterns were visible and the blot was imaged using an Umax Powerlook III scanner.

2.7.3 Chemiluminescent detection

The membrane was incubated in 10 ml of 1X PBS containing 0.1% (v/v) Tween-20 (Sigma) and the appropriate primary antibody for one hour at room temperature with agitation. For all chemiluminescent experiments performed no secondary antibody was required. The blot was then rinsed three times in 1X PBS; 0.1% Tween-20 for 20 mins with agitation. The membrane was then incubated in 1 ml of luminol/enhancer and 1 ml of peroxide solution from the Pico Super Signal West kit (Thermo Scientific) for 5 mins in the dark with constant mixing. The membrane was then drip dried and placed protein side down in an X-ray cassette. An autoradiography film (Fisher Scientific) was placed over the membrane in the cassette. Exposure time was dependent on the quantity of the signal produced from the detection solution, a strong signal would only need 15s exposure while a very weak signal needed an over-night exposure. The films were developed using an Xograph Compact developer.

2.7.4 Ponceau S red staining

Ponceau S acid red staining was performed to confirm the presence of total protein on membranes during the western blot procedure. Ponceau S acid red (Sigma) (0.5%

Ponceau red [w/v] and 1% acetic acid [v/v]) staining is a non-destructive, reversible method to detect proteins on a nitrocellulose or PVDF membrane before detection using antibodies. After transfer of the proteins from the gel to the membrane the membrane was rehydrated in distilled water, if dry, for 1 min. Excess water was removed, then the membrane was incubated in Ponceau S acid red solution at room temperature for a few minutes until bands could be seen. Several washes of water were then used to remove the stain; during this removal of excess stain images were taken.

2.8 Antibiotics

The following antibiotics were used in this work.

Antibiotic	**Working concentration**	**Solvent**
Ampicillin	50 µg/ml	SDW
Kanamycin	50 µg/ml	SDW
Spectinomycin	100 µg/ml	SDW
Gentamycin	50 µg/ml	SDW
Rifampicin	50 µg/ml	methanol
Cefotaxime	100 µg/ml	SDW

2.9 Primers

2.9.1 Primers used to clone effectors from *G. pallida*

Effector (without signal peptide)	**F/R**	**Gene-Specific Sequence part (5'->3')**	**Finnzyme Tm (°C)**	**PCR product size (bp)**
Gp1106_1	F	AAGCCAGCAGACAAAAAGGC	65.04	564
	R	TTCGTCCATATTGGATTTTGG	62.86	
Gp1106_2	F	GCTCTTCTGGACACGGGTC	64.50	594
	R	TTCGTCCATATTGGATTTTGG	62.86	
Gp30G12	F	TCTCCAGTTCATCCTAATGAAGATG	63.80	579
	R	ATTCAATGCTGACGGCACA	64.96	
Gp30G12	F	TCTCCAGTTCATCCTAATGAAGATG	63.80	468

Effector (without signal peptide)	F/R	Gene-Specific Sequence part (5'->3')	Finnzyme Tm (°C)	PCR product size (bp)
	R	ATTCAATGCTGACGGCACA	64.96	
Gp448_2	F	GCCCCGCATCCATGCT	68.16	573
	R	AAAGTCGTCTTCGTCGGCTTC	66.20	
Gp448_3	F	GCACCCAGGTTCCCGT	64.46	513
	R	TTGTTTTGTGTAAGCGCTGTG	63.32	
Gp448_4	F	CTAATGTCCGGCTACATTGTCA	63.25	468
	R	TTGCTTTGTGTAAGCGCTGT	63.09	
Gp448_1	F	GCCCCACAATTCCCGT	64.14	504
	R	GAGCTTGTGCGAGCCG	64.42	
Gp66_P1	F	GACCTCACACTGGACAGCTTG	64.51	309
	R	GCCACAGCATCCGTAACAG	63.92	
Gp747	F	GAATGCTGCTTTGATGATGG	62.77	171
	R	TTTTGAACCGCCTGTGC	63.15	
GpA42	F	TGTGGTGGTGACTGTTTTGG	64.16	132
	R	TTTTCGTCTTATGAGCTTGCTTC	63.03	
GpCLV3	F	ACAAATGAAAAGGATGATAAAGAAGC	62.91	513
	R	ATGGTGAGGGTCGGGC	64.46	
Gpdgl1d	F	TTCAGCTGTGGCGATACTG	62.28	90
	R	GTTGACTCGTTTGCGAGGT	62.52	
Gpdgl1e	F	GATGGAAATAGAAAACCAAAGAAAAC	62.35	75
	R	CCATGTCTTATTGGGAACTTGG	63.73	
GpE9	F	TACCCTCTGAGCTCGTGGAG	64.08	978
	R	ATGACGAGCTTGGCCATTT	64.08	
GpG12H04	F	ATTCCAGATGAAGCCGTTCA	64.06	1839
	R	CGTACGCAATAAATGGTCGAA	64.02	
GpG16H02	F	CAATTACAATCGAAGAGCATCGG	65.91	345
	R	CAAAAGGCGAAAGCACCG	66.22	
GpG20E03	F	ACACCTAACGATAACCCGATTG	63.21	507
	R	AGCACAGAAAGGCGAAAAGA	64.47	
GpG20E03b	F	ACACCTAACAATAATCCGATCATG	62.19	498
	R	TTTTATTTTTACTTGTGACCAAG	62.56	

Effector (without signal peptide)	F/R	Gene-Specific Sequence part (5'->3')	Finnzyme Tm (°C)	PCR product size (bp)
		GC		
GpG7E05	F	TTCCATTGCTGGGATTCGA	66.03	279
	R	ATTTGGTCCGTTGCACAGC	65.60	
GpG8A07	F	GCCACTGACGGGATGACA	65.73	696
	R	TTTCGTTTTGATCACTCGCC	64.53	
GpHg10C02	F	AACGCCAACGCCAAGG	65.61	228
	R	ATGCTTAGGCTTCTTTCCGC	63.99	
GpHgsec12	F	TGTCAGTGCCGATGGG	62.84	456
	R	ACATTGATGGTCAAATTGTTGC	63.03	
GpHgsec3	F	GAGTTTCCTTTGCCGGTCAC	65.31	966
	R	CTCTCTTTTCTGTCCGCAAGG	64.67	
GpHgsec4	F	CAAGATGATGATGACAAAGATGC	62.91	699
	R	GTTCTTGCCAAGCCCAATT	63.46	
GpHgsec6	F	AATGGAAATACCGGCGG	62.84	684
	R	CGATTCGTCGATTCCGA	63.11	
GpHgsec8	F	GATAAGGGCGCGGATGT	63.24	843
	R	AAACTTCCAACTGCCTACCCT	62.33	
GpSCN1120	F	ATTGACTCGTACTTCATTTCCCAC	63.62	198
	R	TAACATGTACACAGCTTTGTTCTCC	62.64	
Gp4D06	F	GCCCCGCATCCATGC	67.47	531
	R	GTTGGCGGCGCTGTATTT	65.52	
GpChorismate_mutase	F	CCAAAATCGCCCGCTC	65.04	828
	R	TTCATTCAGCAGTTTCTTGGC	63.71	
GpIA7	F	CAGGACGCTGCTCCCAT	64.87	153
	R	GCAAAACTTGCAGGTTTTTGG	64.95	
GpIVG9	F	GGGTCGTGTTTGTCTAGTGG	61.65	222
	R	CCAATTTTTATCCATGTCATCAC	61.38	
Gp29D09	F	GCCCCACAATTCCCGT	64.14	504
	R	GAGCTTGTGAGAGCCGGA	64.05	
M13	F	GTAAAACGACGGCCAGT	55	
	R	GTTTTCCCAGTCACGAC	55	

Additional reverse primers were produced for each effector incorporating a HA tag that allowed detection of the resulting fusion protein by western blotting. The sequences of these primers were:

oligoname	**sequence**	F/R
Gp1106-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTATTCGTCCATATTGGATTTTGG	R-HA
Gp30G12-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAATTCAATGCTGACGGCACA	R-HA
Gp448-2-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAAAAGTCGTCTTCGTCGGCTTC	R-HA
Gp448-3-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTATTGTTTTGTGTAAGCGCTGTG	R-HA
Gp448-4-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTATTGCTTTGTGTAAGCGCTGT	R-HA
Gp448-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAGAGCTTGTGCGAGCCG	R-HA
Gp66-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAGCCACAGCATCCGTAACAG	R-HA
Gp747-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTATTTTGAACCGCCTGTGC	R-HA
GpA42-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTATTTTCGTCTTATGAGCTTGCTTC	R-HA
GpCLV3-1-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAATGGTGAGGGTCGGGC	R-HA
Gpdgl1d-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAGTTGACTCGTTTGCGAGGT	R-HA
Gpdgl1e-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTACCATGTCTTATTGGGAACTTGG	R-HA
GpE9-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAATGACGAGCTTGGCCATTT	R-HA
GpG12H04-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTACGTACGCAATAAATGGTCGAA	R-HA
GpG16H02-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTACAAAAGGCGAAAGCACCG	R-HA
GpG20E03-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAAGCACAGAAAGGCGAAAAGA	R-HA
GpG20E03b	TCAGGCATAATCAGGTACATCATAAGGGTATTTTATTTTTACTTG	R-HA

oligoname	sequence	F/R
-HA-R	TGACCAAGGC	
GpG7E05-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAATTTGGTCCGTTGCACAGC	R-HA
GpG8A07-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTATTTCGTTTTGATCACTCGCC	R-HA
GpHg10C02-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAATGCTTAGGCTTCTTTCCGC	R-HA
GpHgsec12-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAACATTGATGGTCAAATTGTTGC	R-HA
GpHgsec3-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTACTCTCTTTTCTGTCCGCAAGG	R-HA
GpHgsec4-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAGTTCTTGCCAAGCCCAATT	R-HA
GpHgsec6-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTACGATTCGTCGATTCCGA	R-HA
GpHgsec8-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAAAACTTCCAACTGCCTACCCT	R-HA
GpSCN1120-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTATAACATGTACACAGCTTTGTTCTCC	R-HA
Gp4D06-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAGTTGGCGGCGCTGTATTT	R-HA
GpCM-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTATTCATTCAGCAGTTTCTTGGC	R-HA
GpIA7-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAGCAAAACTTGCAGGTTTTTGG	R-HA
GpIVG9-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTACCAATTTTTATCCATGTCATCAC	R-HA
Gp29D09-HA-R	TCAGGCATAATCAGGTACATCATAAGGGTAGAGCTTGTGAGAGCCGGA	R-HA

2.9.2 Primers used for qPCR analysis of SPRY domain containing proteins

Gene	Forward	Reverse	Tm (°C)
GPLIN_001312900	TGAACAGCCAATTCCAAAAA	ACCGTTGGTTTAGTCCCAAG	59
GPLIN_001465500	AGAGCTCGAAAAGGGAATGA	CGGCAACTGTATTCTGCTGT	59
GPLIN_000132500	ATGGTTGACACCGAATTGAA	GAAGTCAAACGGAGGGTTTC	59
GPLIN_000133000	ATGTGTTTCGCTCTGTACGC	GCAAGTCCAATGTCAACACC	59
GPLIN_000320000	TTGACGCTGAATTGGAGAAG	TCTTGCTGCAGAGCATTCTT	59
GPLIN_000195600	CCAAACAAATGGCATTGAAC	CAGCCTCGTGACAAAGAAAA	59
GPLIN_000555800	TTGCTTGTCGTTGTTCTTCC	TTCTTTGTGGGATTGTTCCA	59
GPLIN_000203800	AGTTCGAACGTGCAATGAAG	ACCTTCTCTGGGTGGACAAC	59
GPLIN_000696800	GAACAGCCAATCCCAAAAAT	CAATCACTTTTCCCCTTGGT	59
GPLIN_000583000	ATGGCAATTTTGTGATTGGA	GACCCCAAAATCTGCCTTTA	59
GPLIN_000930100	CCAACAAAGAATGTCGGCTA	GTTTTGTGGCAAGTCCAATG	59

2.9.3 Primers used for analysis of transgenic plants and semi-quantitative RT-PCR

Effector (without signal peptide)	F/R	Gene-Specific Sequence part (5'->3')	Tm (°C)	PCR product size (bp)
Gp1106_1	F	ACCATCCAGTGCTCCAGTTGT	55	225
	R	ACATCCAGGTCCATTAAGAAC	55	
Gp448_1	F	GTTAAAGCCATGTTCGAGTTGG	55	216
	R	AATCACATTGTTCGGGTGTGG	55	
Gp448_2	F	AGAATTGTGGCCGAAATCGACG	55	226
	R	TGGTATGTTCCGTGGCAAGTG	55	
Gp448_3	F	TTGGAGGCGATGAACAACTGC	55	227
	R	AACGCGTTGCAAATGTTGTCG	55	
Gp448_4	F	GTCAACTTCACGAACTCGGTGG	55	230
	R	CCTGTCGCTTTTGCGGCCAAAT	55	
Gp66_P1	F	TCACACTGGACAGCTTGATGC	55	259
	R	TAACAGGTTCCGTCACCGTAG	55	
Gp747	F	GAATGCTGCTTTGATGATGG	55	153
	R	TGAACCGCCTGTGCCTGTGC	55	
GpA42	F	TGTGGTGGTGACTGTTTTGG	55	121
	R	TCGTCTTATGAGCTTGCTTC	55	
Gpdgl1e	F	TTCAGCTGTGGCGATACTGG	55	81
	R	GTTGACTCGTTTGCGAGGT	55	

Effector (without signal peptide)	F/R	Gene-Specific Sequence part (5'->3')	Tm (°C)	PCR product size (bp)
GpE9	F	CACGATTTCAACATACTTACT	55	215
	R	TTCCTGCTCGTTCGGCTTGAG	55	
GpG16H02	F	TATCCGAGTCCTTCACTACTG	55	227
	R	AAGATGATCATCCAGTCCAAG	55	
GpG20E03	F	ACACCTAACGATAACCCGAT	55	225
	R	ACATCCAGGTCCATTAAGAAC	55	
GpG20E03	F	ACCGCAATACAGACGATGATGG	55	238
	R	GCTGTCGGGAGTTTGTCACAC	55	
GpG7E05	F	ACGCAATAATTCTGCTCAACG	55	224
	R	GCAATATGAACAGTGAATTGG	55	
GpG8A07	F	TCCTATTGCATTTTCCTCTCG	55	226
	R	ACAGCTCCTCCTCGTGTTTGC	55	
GpHg10C02	F	CAAGGCCGAAGCTGAAGCC	55	208
	R	ATGCTTAGGCTTCTTTCCGC	55	
GpHgsec4	F	AAGGAGCACAAAGAGCCTGC	55	211
	R	AAATGTTCTCGAAGATGGACG	55	
GpSCN1120	F	ATTGACTCGTACTTCATTTCC	55	185
	R	ATGTACACAGCTTTGTTCTCC	55	
GpE9	F	CACGATTTCAACATACTTACT	55	215
	R	TTCCTGCTCGTTCGGCTTGAG	55	
Gp4D06	F	CATGTGGTTGCTCTAATGACC	55	240
	R	ATTGTGCGAGCACCCATCTTC	55	
GpIVG9	F	TCGTGTTTGTCTAGTGGCACTG	55	203
	R	TTATCCATGTCATCACTGGCG	55	
GpIA7	F	AGGACGCTGCTCCCATCACC	55	134
	R	GCAGGTTTTTGGGCACATCG	55	
18S	F	TCGGCTTGCTCTGATGATTC	55	
	R	CCGACCAATGCACACCAAAG	55	
p35S	F	AAGGAAGTTCATTTCATTTGGAGAGGA	63	
t35S	R	CAACACATGAGCGAAACCCTATAAGAA	63	

2.9.4 Primers used in functional analysis of a *Globodera pallida* effector similar to ubiquitin extension proteins (Chapter 5)

GpUBI-EP	F/R	Gene-Specific Sequence part (5'->3')	Tm (°C)	PCR product size (bp)
UBI_F	F	GACACTGACCGGCAAAAC	55	288
UBI_R	R	GGTATCAGCCGCCCCGGA	55	
UBIpDONR221F	F	GGGGACCACTTTGTACAAGAAAGAAGACACTGACCGGCAAAAC	65	313
UBIpDONR221R	R	GGGGACCACTTTGTACAAGAAAGTGGTATCAGCCGCCCCGGA	65	
UBItruncpDONR221R	R	GGACCACTTTGTACAAGAAAGCTGGGTATTGGTTCTCGCACTCGTTGG	65	277
EF1αF	F	AACATCTCTGTGAAGGACATTCG	59	196
EF1αR	R	TCTCCTTAAGTTCGGCGAATTTGC	59	
EIF4αF	F	CGAAACAGGACCAACAAATG	59	94
EIF4αR	R	GTTCAGATCAGCTCCCCAAT	59	
UBI_Rev_all_qPCR	R	CATTGGTTCTCGCACTCGTTGGG	59	100 and 97
UBI_For_qPCR	F	ACAGCTCGAAGATGGCCGCA	59	
UBI_R_WT_SP_qPCR	R	TCTCGCACTCGTTGGGTCCATGT	59	100

3 Identification and characterisation of the *G. pallida* effectorome

3.1 Introduction

Until recently, sequencing a genome was a hugely expensive procedure and was restricted to model organisms used in wide areas of biology, or species of major economic or pathogenic importance to man. Fortunately for the field of nematology, the free living nematode *Caenorhabditis elegans* was chosen as a model organism for genetics and developmental biology and was subsequently the first multi-cellular organism to have its genome sequenced (C.elegans sequencing consortium, 1998). Following on from this, the genome of the closely related *C. briggsae* was sequenced in order to facilitate comparative genomics studies (Stein *et al.*, 2003). However, recent advances in sequencing technology, such as 454 (Rothberg and Leamon, 2008) and Illumina sequencing (Bennett, 2004), have dramatically reduced the costs of sequencing while massively increasing the data output (Table 3.1), resulting in an exponential growth in sequence data (Mardis, 2008). Genomics has now started to become applied to the study of plant-parasitic nematodes. Expressed sequence tags (ESTs), single pass sequence reads from cDNA libraries, have been generated from a wide range of plant-parasitic nematodes over the last 10-15 years (e.g. Popeilus *et al.,* 2000). More recently the first genome sequences for plant parasitic nematodes have been generated. These include two root-knot nematodes, *M. incognita* (Abad *et al.*, 2008) and *M. hapla* (Opperman *et al*., 2008), as well as the pine wilt nematode *Bursaphelenchus xylophilus* (Kikuchi *et al.*, 2011). In addition, sequencing projects are currently in progress for the cyst nematodes *G. pallida*, *G. rostochiensis* and *Heterodera glycines* as well as migratory endoparasitic nematodes such as *Pratylenchus coffeae* and *Radopholus similis* (C. Opperman pers. comm). By contrast, genome sequences have been available for other plant pathogens, particularly bacterial pathogens for many years. The availability of genome sequences has had a major impact on the study of each of these organisms.

Sequencer	454 GS FLX	Illumina HiSeq 2000	SOLiDv4	Sanger 3730xl
Read Length	700bp	100bp paired end	50+50bp	400 - 800bp
Output per run	0.7Gb	600Gb	10Gb	1.9~84Kb
Time per run	24h	3~10 days	7-14 days	20 min - 3h
Instrument cost	$500 000	$690 000	$495 000	$95 000
Cost per run	$7000	$6000	$15000/100Gb	$4 per 800bp reaction
Advantage	Read length and fast	High throughput	Accuracy	High quality long reads
Disadvantage	Error rate. High cost. Low throughput	Short read assembly	short read assembly	High cost. Low throughput

Table 3.1: Comparison of features for 454, Illumina, SOLiD and Sanger sequencing platform, detailing their advantages and disadvantages (Liu *et al.*, 2012).

3.1.1 *C. elegans*

The genome of *C. elegans* is 97Mb and encodes 19,099 genes (*C.elegans* sequencing consortium, 1998). A variety of tools have been developed for characterisation of gene function in *C. elegans* including RNAi, transformation and the generation of mutants via transposon insertion (Tabara *et al.*, 1999). Large panels of the generated mutants have been screened for a range of phenotypic characters. The ability to transform *C. elegans* has underpinned many functional studies reviewed in Jones *et al.,* (2011).. Such information has provided insights into a wide range of biological processes including ageing, regulation of fat deposition, RNAi, nervous system function and development, metabolism and detoxification. This information has proved to be widely applicable to distantly related organisms, including humans, and has also underpinned many studies on parasitic nematodes.

3.1.2 Genomes of plant parasitic nematodes

Although several plant nematode genomes are in the process of being sequenced, the only published plant parasitic nematode genomes are for *M. incognita* (Abad *et al.*, 2008), *M. hapla* (Opperman *et al*., 2008) and *B. xylophilus* (Kikuchi *et al.*, 2011). These genome sequences provide a basis for comparative genomics of PPN, although no major analyses in this area have been published at the time of writing. The *M. hapla* genome is one of the smallest metazoan genomes characterised to date. This nematode has a genome of 54Mbp with relatively low amounts of repetitive regions (~12%) and a high AT content (Opperman *et al*., 2008). *M. hapla* reproduces sexually and has a much narrower host range in comparison to *M. incognita.* The genome size of *M. incognita* is 86 Mbp (Abad *et al.*, 2008), but is partially duplicated. By contrast, *H. glycines* has a predicted genome size of 92.5 Mbp (Opperman and Bird, 1998) and the predicted genome size for *G. pallida* is over 100 Mbp. Analysis of the *M. incognita* genome has provided, amongst other useful information, an insight into the process and consequences of reproducing by asexual mitosis. For example, triplicated genome regions have been identified that are diverging away from each other resulting in pseudo alleles that show high divergence at the protein level. It has been suggested that this method of introducing genetic variation could account for the ability to successfully parasitize on a wide host range allowing rapid adaptation to environmental and geographic locations (Abad and McCarter, 2011).

Parasitic nematode genomes investigated to date encode fewer genes than those of non-parasitic nematodes. It is thought that the host environment, selection pressure as a result of interactions with the host and host nutrient supply enables this observed loss of non-essential genes. For example, *M. incognita* contains fewer genes encoding a variety of detoxifying enzymes, possibly because living within a host provides protection from biotic and abiotic stresses encountered by non-parasitic nematodes (Abad *et al.,* 2008). In a comparison using the predicted genes from *M. incognita*, *B. malayi* and *C. elegans* only 3533 common orthologous genes were identified. This relatively low number of orthologous genes may be explained by their ancient divergent history (Abad *et al.*, 2008). However, some conserved pathways were identified such as innate immunity signalling, some sex determination genes, dauer formation and RNAi machinery (Abad and McCarter, 2011).

Plant parasitism is thought to have evolved several times within Nematoda (Blaxter *et al.*, 1998), reducing the probability that conserved genes are present between different plant parasitic nematodes that underlie their interactions with the host. In support of this argument, sequence similarity searches using 31 *M. incognita* putative effector genes against *H. glycines* putative parasitism genes revealed few common effectors (Gao *et al.*, 2003a). This suggests that these nematodes have evolved different effectors to achieve successful parasitism.

In order to characterise the function of genes within *C. elegans,* RNA interference (RNAi) has been used extensively (Kamath *et al.*, 2003). There are 2,958 genes that give a lethal RNAi phenotype in *C. elegans* and 1,083 of these have orthologues in *M. incognita* (Abad *et al.*, 2008). An orthologue of a gene that has a lethal RNAi phenotype in *C. elegans* was used as a target for plant delivered dsRNA to induce RNAi in *M. incognita*. The target was a splicing factor and silencing the gene encoding for this protein resulted in a reduction in *M. incognita* gall formation in tobacco (Yadav *et al.*, 2006). An alternative strategy may be to target effector gene(s) that are specific to plant parasitic nematodes, or to a species of interest. This may allow development of highly targeted control measures. For example, the 16D10 effector, common to all *Meloidogyne* species studied to date, was used as a target for plant delivered dsRNA to induce RNAi. Several species of *Meloidogyne* showed reduced infectivity as a result of the RNAi induced down-regulation of this effector

(Huang *et al.*, 2006a). Therefore the identification of homologues of *C. elegans* genes that have a lethal RNAi phenotype could be used as targets for alternative control measures, or the identification of effectors that could be used as a highly directed control measure.

3.1.2.1 Mining genome sequences for effectors

The availability of a genome sequence allows bioinformatics to be used to identify genes of interest such as effectors. Effectors, defined here as any nematode protein secreted into the host that manipulates the host to the benefit of the nematode, can be identified using various strategies. Simple BLAST searches can be used to identify orthologues of known effectors from other species. Secreted proteins, some of which may also be effectors, can be identified by analysing the predicted proteins in a genome sequence for the presence of a signal peptide and the absence of a transmembrane domain (Jones *et al.*, 2009b).

3.1.3 Effector identification from transcriptomes of plant parasitic nematodes

There are many examples of important plant parasitic nematodes for which a genome sequence is not currently available. For these nematodes, the sequencing of RNA transcripts has proved to be an extremely useful approach for identifying genes of interest. These have, in the past, primarily taken the form of relatively small scale expressed-sequence tag (EST) analyses, using Sanger sequencing of cDNA libraries (*e.g.* Popeijus *et al.,* 2000). EST analyses have been performed on many PPN and the sequences used for identification of putative effectors. Examples include *H. glycines* (Elling *et al.*, 2009a; Gao *et al.*, 2003a), *G. pallida* and *G. rostochiensis* (Popeijus *et al.*, 2000), *Meloidogyne chitwoodi* (Roze *et al.*, 2008), and the pine wood nematodes *Bursaphelenchus xylophilus* and *B. mucronatus* (Kikuchi *et al.*, 2007).

More recent studies have used next generation sequencing, particularly 454 FLX analysis, because the long reads generated in this technique are well suited to *de novo* transcriptome analysis. For example, EST analysis of *P. coffeae* identified 49 sequences with similarity to cell wall modifying proteins and orthologues of 15 known effectors from other PPN. Interestingly this includes chorismate mutase which was previously only thought to occur in sedentary nematodes. The EST analysis also provided a list of genes encoding proteins that have a signal peptide and no transmembrane domain which could encode putative effectors (Haegeman *et al.*,

2011b). An EST analysis of *Pratylenchus thornei* revealed 12 genes similar to known parasitism genes including cell wall modifying enzymes and putative effectors of unknown function that could be subjects of future research (Nicol *et al.*, 2012). The advancement and reduction in cost of sequencing technology will allow more transcriptome projects to be performed and will also allow quantitative analysis of gene expression via RNAseq data. The availability of ESTs/transcriptome information can also provide useful data for the training of gene prediction software (Jones *et al.*, 2009b).

3.1.4 Expression patterns of effectors

Understanding the expression profiles of effectors can provide an insight into their functional roles. For example, effectors that are important during migration such as cell wall degrading enzymes, are highly expressed at J2 and in males as these are the life stages that migrate through the root. Similarly, it can be argued that effectors that suppress PTI are likely to be expressed at J2 and early parasitic stages as nematodes need to suppress host defences as a feeding site is established. Effectors are synthesised in the oesophageal gland cells, which comprise one dorsal and two subventral glands. The subventral glands are more metabolically active in early parasitic stages and are full of secretory products at the J2 life stage. The subventral gland cells subsequently become less active throughout the parasitic interaction. In contrast, the dorsal gland is less metabolically active in early stages and becomes more active in later parasitic stages (Gheysen and Jones, 2006). Therefore knowledge of the expression profile of an effector could provide information about where it is expressed and the stage of the life cycle at which it is important.

3.1.5 Horizontal gene transfer

Analysis of the genomes and transcriptomes of many plant parasitic nematodes has revealed that these organisms have a large number of genes that are likely to have been acquired from bacteria and fungi as a result of horizontal gene transfer (HGT - reviewed by Haegeman *etl al,.* 2011). HGT can be defined as inter-species asexual movement of genetic material (Haegeman *et al.*, 2011a) (see section 1.3.2). The most well defined examples are provided by the cell wall degrading enzymes. Although it is widespread in plant parasitic nematodes, few occurrences of HGT have been documented in other eukaryotic organisms. Using genome sequence data it has been

possible to determine the full extent of HGT within some nematode genomes. For example, 61 enzymes that metabolise carbohydrates (CAZymes) acquired via HGT were identified in *M. incognita* (Abad *et al.*, 2008) and 33 such genes were identified in *M. hapla* (Opperman et al., 2008). In many cases these genes have undergone duplication following acquisition and are present in gene families.

Phylogenetic analysis of CAZymes has shown that there is some common ancestry between Genera of PPN in terms of genes acquired by HGT (Danchin *et al.*, 2010). For example GH5 cellulases are present in root-knot nematodes, Pratylenchidae, cyst nematodes, *Radopholus* and *Aphelenchus*. This suggests that there was a common ancestor of extant Clade 12 nematodes that acquired GH5 cellulases. In contrast to this, GH28 polygalacturonases have only been identified in RKN (Haegeman *et al.*, 2011a). Therefore there may have been multiple occurrences of HGT throughout evolution of plant nematodes. Further evidence for multiple independent horizontal gene transfer events comes from the analysis of the *B. xylophilus* genome. Glycosidase hydrolase family (GH45) genes, thought to have originated from fungi, have been identified in *B. xylophilus* but have not been identified in any other plant parasitic nematode. Furthermore no GH5, GH30, GH43 or GH28 domain containing genes were identified in the *B. xylophilus* genome, whereas expansin and pectate lyase genes have been identified in *B. xylophilus* and other plant parasitic nematodes. *Bursaphelenchus* species are mainly fungal feeders and it is interesting to observe that the *B. xylophilus* genome not only encodes plant cell wall modifying enzymes but also enzymes that metabolise the fungal cell wall, including six β-1,3 endoglucanases (GH16). Genes encoding GH16 proteins that degrade β-1,3-glucan, a core component of the fungal cell wall, are thought to have been acquired from bacteria (Kikuchi *et al.*, 2011). The presence/absence of certain HGT candidates detailed above suggests *B. xylophilus* acquired such genes in several independent transfer events.

Other examples of HGT within plant parasitic nematodes have been identified. Chorismate mutase, polyglutamate synthase, cyanate lyase, and several components of the vitamin B6 synthetic pathway are all present in various plant parasitic nematodes and are thought to have been acquired via HGT reviewed in Haegeman *et al.,* (2011).

3.2 Aims

The aims of this part of the project were:

- Identify the full complement of *G. pallida* effector genes from the genome sequence of this nematode, including orthologues of effectors from other plant parasitic nematodes and novel candidate effectors.
- Identify genes encoding plant cell wall degrading and modifying proteins and other *G. pallida* sequences potentially acquired by HGT.
- Use RNAseq data to determine the expression profiles of effectors, CAZymes and genes acquired by HGT.
- Analyse the phylogeny and evolution of effectors present as large gene families in *G. pallida.*
- Identify any putative promoter regions that may be involved in the regulation of effector expression.

3.3 Materials and methods

3.3.1 Effectors

The *G. pallida* predicted protein set version 1.0 (16th May 2012) was used for identification of effectors. This protein set is available at ftp://ftp.sanger.ac.uk/pub/pathogens/Globodera/pallida/ and was used for the detailed analysis of the *G. pallida* genome.

Two approaches were used to identify effectors. First, *G. pallida* orthologues of previously characterised effectors were identified by BLAST searching. In a second approach, novel effectors were identified using a bioinformatic approach which collected all secreted proteins up-regulated in hatched J2s (as compared to unhatched J2s) or in early parasitic stage nematodes as compared to hatched J2s.

3.3.1.1 Identification *G. pallida* orthologues of previously characterised effectors

A list of known effectors from other plant parasitic nematodes was collated using data from SCN gland cell ESTs (Gao *et al.*, 2003a; Wang *et al.*, 2001), microarray analysis (De Boer *et al.*, 2002), effectors identified from cDNA-AFLP analysis on *G. rostochiensis* (Qin *et al.*, 2000), *G. rostochiensis* and *G. pallida* ESTs (Popeijus *et al.*, 2000) and effectors identified from *M. incognita* (Abad *et al.*, 2008; Huang *et al.*, 2003). The list also included effectors that had previously been identified from *G. pallida* (Blanchard *et al.*, 2007; Jones *et al.*, 2009b). In addition, a list of *G. rostochiensis* effectors was provided by Dr G. Smant (Wageningen University). The collated effector list was subjected to a local, command line BLAST (Altschul *et al.*, 1997) against the *G. pallida* genome sequence. This search used an E-value threshold of 10^{-5} with low complexity filtering turned off. The BLAST parameters used were: BLASTP –db database.fasta –query sequences.fas –evalue 0.00005 –seg no -num_threads 2 –out outfile.txt.

RNAseq reads mapping to the regions of the genome in which each of the identified effectors were analysed visually using Gbrowse. This allowed a check of the accuracy of the gene prediction on the basis of RNAseq coverage to be made.

3.3.1.2 Analysis of effector expression profiles

The expression profiles of putative effectors identified by BLAST searching with effectors from other plant nematodes were analysed using the normalised RNAseq data. MBClusterseq (http://cran.r-project.org/web/packages/MBCluster.Seq/index.html) was used to separate the effectors into clusters that show similar expression profiles. Inspection of the results of this analysis revealed that some clusters showed very similar profiles and genes in such clusters were subsequently merged.

3.3.1.3 Identification of novel candidate effectors

The predicted *G. pallida* protein set was first analysed using a standard secretory protein identification protocol. Proteins that had a predicted signal peptide and no transmembrane domain were identified using SignalP 3.0 (Dyrl Bendtsen *et al.*, 2004) followed by TMHMM (Krogh *et al.*, 2001), based on the methodology used in Jones *et al.* (2009b). Output from this analysis was collated using custom made Python 2.6 and Biopython (Cock *et al.*, 2009) scripts signalPoutput_checker.py", signalPoutput_checker_count_yes_in_column.py" (Appendix 3). Expression profiles of the genes that passed these filters were then analysed using DESeq (http://bioconductor.org/packages/release/bioc/html/DESeq.html) (Anders and Huber, 2010) in order to identify genes that were significantly more highly expressed at J2 compared to eggs or at 7 dpi compared to J2. Genes that passed the secretion and expression profiling filters were taken forward as putative effectors.

These sequences were then BLAST searched against the NR database and those that obviously had functions unrelated to parasitism (*e.g.* collagens, digestive proteinases) but which came through this screen were manually removed. In some cases the results of this BLAST searching provided functional information about the novel putative effectors. The putative effector list was analysed for any known domains using PFam rules defined in ftp://ftp.sanger.ac.uk/pub/databases/Pfam/current_release/Pfam-A.hmm.gz (July 2012), using a command line tool called HMMER (see section 3.3.4.1).

3.3.1.4 Effector annotation

The gene name, location, scaffold number and annotation based on the BLAST hit were added to the annotation database for all identified effectors. The putative effectors were also subjected to local BLAST (tBLASTn) against the *G. pallida*

genome sequence in order to identify any regions of the genome that could contain genes similar to the effectors but that had not been predicted by the gene finding software. All annotation data was sent to the genome annotation team and was incorporated into the final released annotation.

3.3.2 Phylogenetic analysis

Some of the *G. pallida* effectors were found to be present in substantial gene families. The phylogenetic relationships of these sequences and similar sequences in other species were examined.

3.3.2.1 Generating alignments

MUSCLE (Edgar, 2004) was used to generate alignments of sequences of interest . A second revision of the alignment was always performed using the "refine" command. Alignments were visualised using Jalview (Waterhouse *et al.*, 2009).

3.3.2.2 Phylogenetic trees

Phylogenetic trees were drawn using TOPALi V2 (Milne *et al.*, 2009) using the maximum likelihood algorithm PhyML (Guindon *et al.*, 2010) and the substitution model WAG, with GAMMA option and 100 bootstraps.

3.3.3 Identification of genes potentially acquired via HGT

In order to ensure that all potential cell wall degrading and modifying enzymes were identified a combination of approaches that included BLAST searching with cell wall modifying enzymes from other nematodes. CAZyme and InterProScan analysis was used.

3.3.3.1 CAZymes

The CAZymes Analysis Toolkit (CAT) (Park *et al.*, 2010) was used to identify putative carbohydrate active enzymes (CAZymes) using a predefined CAZyme database on the *G. pallida* predicted protein set V1.0. Putative CAZymes were manually annotated using a combination of BLASTP Vs. NR database, NCBI's Conserved Domain Database service (Marchler-Bauer *et al.*, 2011) and InterProScan (Quevillon *et al.*, 2005) to determine to presence of the catalytic domains. Genes of interest were identified by parsing the CAT output files using: CAZyme_finder.py", cayzes_finder_all_together001.py", cayzes_finder_all_together001_domain_info.py" and common_lists_or_not.py" (Appendix 3).

3.3.3.2 Expansins and CBM finding

CAZyme analysis does not identify expansins or CBM proteins with no catalytic domains. Therefore databases of known expansins and CBM genes were generated composed of sequences from other plant-parasitic nematodes. BLASTP was used to identify putative expansins and CBM proteins for manual annotation. Sequences identified in this way were subjected to expression analysis as described above.

3.3.3.3 Other genes acquired by HGT

BLAST searching was used to determine whether the *G. pallida* genome contained homologues of other genes acquired by HGT that have been identified in other plant-parasitic nematodes. For this, a BLASTP search against the *G. pallida* genome was performed using Chorismate Mutase, cyanate lyase and vitamin B6 biosynthetic protein sequences, using the parameters described in section 3.3.1.1 above. Cyanate lyase, along with BLAST searching, was also identified using HMMER (see section 3.3.4.1) using PFam definitions defined in http://pfam.sanger.ac.uk/family/PF02560#tabview=tab6.

3.3.4 SPRY domain proteins

Preliminary analysis suggested that the *G. pallida* genome harbours a substantial family of proteins containing SPRY domains and that some of these are effectors. This family was analysed in some detail.

3.3.4.1 Identification of proteins containing a SPRY domain

Domain definitions were downloaded from http://pfam.sanger.ac.uk/ and, using the HMMER command line tool (Finn *et al.*, 2011), proteins that contained a SPRY domain as defined by the program's internal gathering threshold were identified. The HMMER parameters used for this analysis were: Hmmsearch --cut_ga --domtblout filename.out definition infile.fasta.

The full length *G. pallida* SPRY domain proteins were so diverged that it was not possible to generate a phylogenetic tree of good quality using the whole protein. Therefore a script (get_SPRY_region_i_want_from_fasta_withHMMRoutput.py" Appendix 3) was designed that extracted the SPRY domain alone from each full length protein. This SPRY region was then aligned and refined using MUSCLE (see section 3.3.2.1 above) and subjected to phylogenetic analysis.

A further alignment was then generated by aligning the SPRY domains back on to the PFam SPRY definition file using HMMERalign. The HMMERalign parameters used for this alignment were: hmmalign definition sequences.fasta >outfile.sto.

Phylogenetic analysis of protein sequences is limited in the potential amount of information available to the analysis software; this is because an amino acid can be encoded in several different ways at the nucleotide level. Therefore phylogenetic analysis of sequences of interest will contain more information if the analysis was conducted using the original DNA sequences, which will include base change information that may not be seen in a protein alignment. Alignments at the protein level are more consistent and easier to produce, therefore once a protein alignment has been made, the original DNA coding region needs to be mapped back on to this alignment, preserving the alignment. To do this, the DNA sequence for the domain of interest was obtained using "Gpal_get_nucleotide_SPRY_region_i_want_ from_fasta006.py" (Appendix 3). The nucleotide sequence was then mapped back on to the aligned protein sequence using a back translation Python program made by Peter Cock (JHI) "align_back_trans.py" (Appendix 3).

3.3.4.1.1 Random occurrence of SPRY domains

In order to check that the predicted SPRY domains were not simply generated by chance, a Python script ("shuffle_genes.py"– Appendix 3) was written which randomly shuffled the predicted protein set from *G. pallida*, preserving the length and amino acid composition of each predicted protein. 30 random shuffles of the predicted protein set were generated using this script and analysed for the presence of SPRY domains as described above.

3.3.4.2 SEEDS

Phylogenetic analysis of the SPRY domains was also conducted using a set of aligned SPRY domain containing proteins from PFam termed SEEDs (http://pfam.sanger.ac.uk/family/SPRY#tabview=tab3).

3.3.4.3 Signal peptide analysis of regions upstream of SPRY domain proteins

Many of the proteins identified as having a SPRY domain did not have a predicted signal peptide, despite previous work showing some effectors have SPRY domains. Predicting the appropriate N-terminus of a protein is particularly challenging for gene finding software and it is therefore possible that one explanation for the absence

of signal peptides is problems with the initial gene models. To determine whether any of the SPRY domain containing proteins from *G. pallida* could have been missing the true N-terminal sequence, a Python script "SPRY_get_upstream_regions_directions002.py" (Appendix 3) was written. This script collected a user-determined number of base pairs upstream of each of the SPRY domain proteins not predicted to have a signal peptide. The resulting nucleotide sequences were translated in same direction that the SPRY genes were predicted using "Get_open_reading_frames version 0.0.1", a Python script written by Peter Cock (JHI) available as a Galaxy tool on the Galaxy Tool Shed (Goecks *et al.*, 2010). The resulting predicted amino acid sequences were analysed using SignalP 3.0. The output from SignalP 3.0 was parsed using "signalPoutput_checker_count_SECRETED_in_column.py" (Appendix 3) to return sequences that contained predicted signal peptides.

3.3.4.4 Functional categorisation of SPRY domain proteins on the basis of RNAseq

The expression profiles of the SPRY domain proteins were analysed using RNAseq data and genome data from gpal.201201.Aug_hints.NT.fa. Expression profiles of the genes were clustered into groups using MBClusterseq (http://cran.r-project.org/web/packages/MBCluster.Seq/index.html). Additional more stringent filters removed genes with near zero expression, potentially due to mis-mapped reads (details below).

To get the RNAseq data for a list of genes from the normalised_gene_expression_database, the Python script "get_expression_for_ genes.py" was used (Appendix 3) and the output was reformatted using "expression_get_reformat001.py" (Appendix 3).

The numbers of RNAseq reads mapping to each gene was averaged across the replicates and any expression that was considered as less than "threshold" was set to 0. For SPRY genes the threshold was set to 2 and for the other effectors the threshold was set to 1; all values below these numbers were replaced with 0. All numbers were returned as integers instead of floats to more easily gain an insight into the expression. This was done using "remove_low_expression_and_average.py" (Appendix 3). To add annotation to the expression database produced by

"remove_low_expression_and_average.py" (Appendix 3), a script was written to parse through cluster genes names and the effector annotation file "gene_names_to _annotation_names001.py" (Appendix 3).

Each SPRY domain protein was placed into one of the following expression profiles: J2 specific, J2 and parasitic, constitutive expression, later parasitic and not expressed. At least 2 sequences from each category were chosen for confirmation of expression profiles by qPCR.

3.3.4.5 qPCR

To design primers that were specific to the gene of interest, and that would not amplify off target genes, the output of primer3 was analysed using a command line tool called Emboss Primer Search (Rice *et al.*, 2000). A 20% mis-match threshold was used. The Primersearch parameters used were primersearch –seqall(fasta database) –infile(primerfile) –outfile(name). qPCR was performed as described in 2.5.1.

3.3.5 Identification of putative promoter elements

Lists of known dorsal and subventral gland genes were generated (Appendix 3). To identify putative promoter sequences a script was written ("Dorsal_get_upstream_regions_directions001.py. Subventral_get_upstream_regions _directions001.py" Appendix 3) that returns a user-defined number of nucleotides upstream of the gene of interest, taking into account the gene directions. The resulting upstream region was then analysed for motifs using MEMES (Bailey *et al.*, 1994).

3.4 Results

3.4.1 *Globodera pallida* orthologues of previously identified effectors

The *G. pallida* genome was BLAST searched with a list of known effectors from the cyst nematodes *G. rostochiensis* and *H. glycines* and the root knot nematode *M. incognita.* This analysis revealed 129 putative *G. pallida* orthologues of effectors from these species out of a starting list of 137 putative effectors (Table 3.2). In addition, this analysis revealed a substantial family of *G. pallida* proteins that included a SPRY domain. These SPRY domain sequences are considered in more detail in 3.4.2 below. Some of the *G. pallida* effectors were present in large gene families. Examples included DGL1, 747, CLAVATA (CLE), 448 and 1106 as well as the SPRYSECs. This appears to be different to RKN, where effectors (other than cell wall degrading enzymes) do not exist as large gene families (John Jones pers. comm.).

Comparisons of the *G. pallida* putative effector list and the *M. incognita* genome showed that, with the exception of the cell wall modifying proteins and chorismate mutase, there are very few effectors common to both species. Just three *G. pallida* effectors had good matches in *M. incognita.* GPLIN_000604400 and GPLIN_000555600 are similar to the *M. incognita* effector AY135365 (Huang *et al.,* 2003) (Figure 3.1) and a similar gene is also present in some migratory endoparasitic nematodes (A. Haegeman, pers. comm.). GPLIN_001475500 is similar to RKN gland cell protein 28 (Figure 3.1). Neither of these sequences have good matches in non-plant parasitic nematodes, or in any nematodes outside Clade 12. Some effectors on the list used for BLAST searching were not present in *G. pallida*. For example, 19 of the 52 *H. glycines* effectors used for searching had no matches in the *G. pallida* genome (Table 3.3). There were three *G. rostochiensis* effectors not identified in the *G. pallida* predicted gene models. The "missing" sequences were those from the "66" gene family, A42 and one effector from the 747 family. Effectors similar to the 66 effector family were identified in previous assemblies of the genome sequence, and one sequence was successfully cloned from *G. pallida* cDNA. A42 is present in the genome sequence and has been cloned from cDNA, however the gene has not been called by the software. Members of the 747 family were present in the *G.*

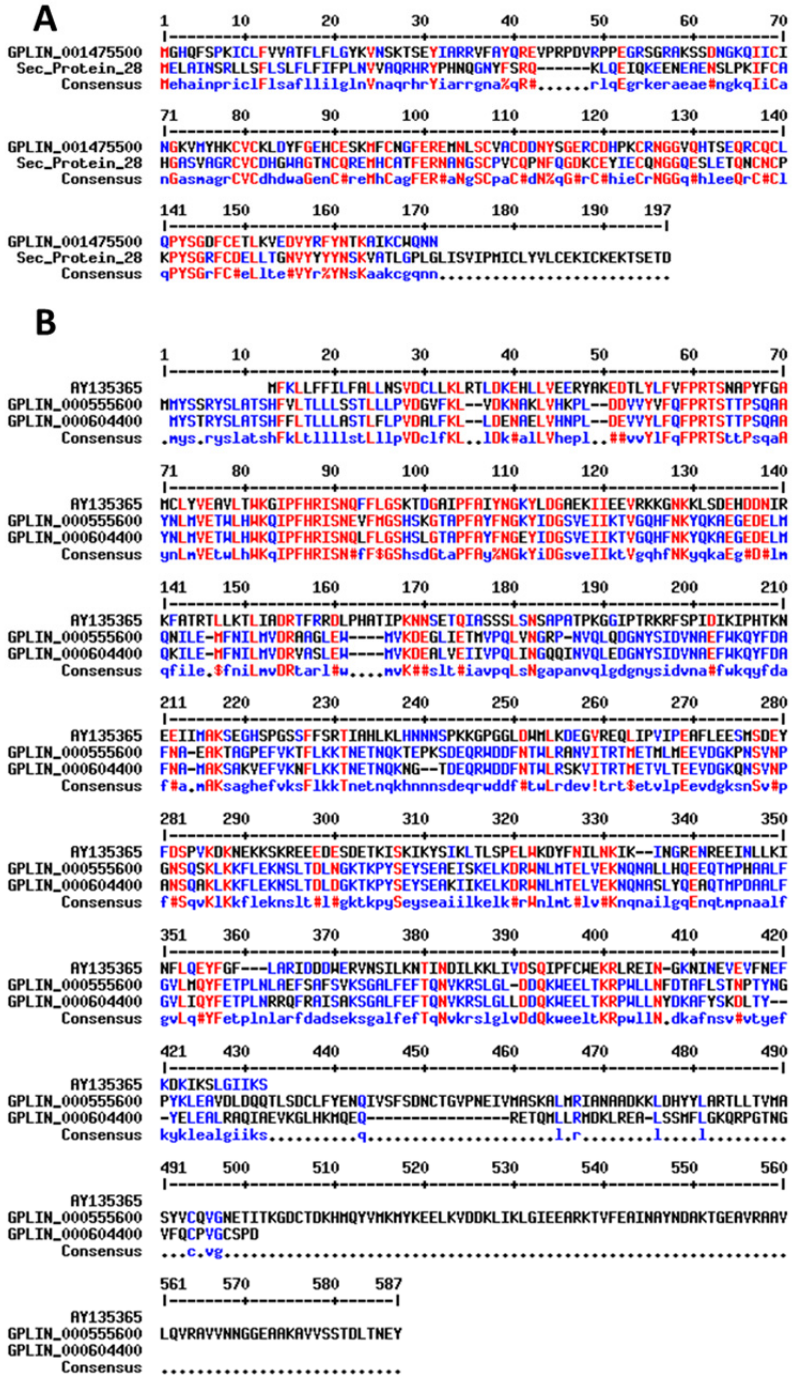


Figure 3.1 : Alignment of putative orthologous effectors from *M. incognita* and *G. pallida.* (A): Alignment of *M. incognita* gland cell protein 28 and putative orthologous effector from *G. pallida.* (B): An alignment of *M. incognita* effector AY135365 and the two *G. pallida* putative orthologues. Conserved amino acids are indicated by upper case red letters, non-conserved amino acids are indicated by lowercase blue letters in the consensus lines.

pallida predicted gene models; however one variant from *G. rostochiensis* was not similar enough to produce any significant similarity search result.

Effectors were identified that are specific to *G. pallida.* These are not present in EST datasets from other nematodes, including the closely related *G. rostochiensis.* These *G. pallida* specific effectors are IA7 and IVG9 (Blanchard *et al.*, 2007) which have 2 (with 5 putative additional genes that have not been predicted in the current annotation) and 5 family members in the *G. pallida* genome, respectively. Some of the *G. pallida* effector gene families are considered in more detail in the sections below.

3.4.1.1 Cell wall modifying enzymes

The plant cell wall is the first barrier of defence against pathogens. Therefore this obstacle has to be overcome in order for the nematode to successfully invade the host. Nematodes have a wide range of plant cell wall modifying proteins which are thought to have been acquired via horizontal gene transfer, from bacteria or fungi (Haegeman *et al.*, 2011a). Many of these proteins are cell wall degrading enzymes that can be identified using CAZyme analysis. A CAZyme search of the predicted *G. pallida* protein set revealed 16 GH5 β-1,4 endoglucanases, 7 PL3 Pectate lyases, 1 GH43 Arabinase and 2 GH53 arabinogalactan endo-1,4-β-galactosidases. All of these enzymes have been previously identified in other plant-parasitic nematodes, although the only other known occurrence of an arabinogalactan endo-1,4-β-galactosidase is from the closely related cyst nematode *H. glycines* (Vanholme *et al.*, 2009) and the migratory nematode *P. coffeae* (Haegeman *et al.*, 2011b). In addition, 9 putative expansin genes were identified, 2 of which contain a CBM domain as well as the expansin domain. Six proteins consisting only of a carbohydrate-binding module (CBM2) were also identified. The full complement of cell wall modifying proteins present in *G. pallida* as compared to other nematodes is summarised in Table 3.4.

.

G. pallida gene	Effector names
GPLIN_000591100	*G. pallida* IVG9 effector
GPLIN_001541500	Paralogue of IVG9 effector
GPLIN_000293500	Paralogue of IVG9 effector
GPLIN_001098200	Possible paralogue of IVG9 effector
GPLIN_001110200	Possible paralogue of IVG9 effector
GPLIN_000638300	*G. pallida* IA7 effector
GPLIN_000740500	Paralogue of IA7 effector
GPLIN_000359000	Member of 1106 effector gene family
GPLIN_000235400	Member of 1106 effector gene family
GPLIN_000793000	Member of 1106 effector gene family
GPLIN_000119200	Member of 1106 effector gene family
GPLIN_000314000	Member of 1106 effector gene family
GPLIN_000768400	Member of 1106 effector gene family
GPLIN_000850500	Member of 1106 effector gene family
GPLIN_001613000	Member of 1106 effector gene family
GPLIN_000684200	Member of 1106 effector gene family
GPLIN_001295300	Member of 1106 effector gene family
GPLIN_000683800	Member of 1106 effector gene family
GPLIN_001043600	Member of 747 effector gene family
GPLIN_000812600	Member of 747 effector gene family
GPLIN_000931100	Member of 747 effector gene family
GPLIN_000376700	Chorismate mutase effector
GPLIN_000666500	Chorismate mutase effector
GPLIN_000594000	C52 effector protein-like
GPLIN_000697600	Member of CLE effector protein family, 4 CLE repeats
GPLIN_001090600	Member of CLE effector protein family, one CLE motif
GPLIN_001090500	Member of CLE effector protein family
GPLIN_000950900	Member of CLE effector protein family
GPLIN_000950800	Member of CLE effector protein family, one CLE motif
GPLIN_000201400	Similar to *G. rostochiensis* E9 effector protein
GPLIN_000057600	Similar to *G. rostochiensis* E9 effector protein
GPLIN_000760900	Similar to *G. rostochiensis* E9 effector protein
GPLIN_000187800	Similar to *G. rostochiensis* E9 effector protein
GPLIN_000854400	*G. pallida* homologue of *H. glycines* G16H02 effector
GPLIN_000780600	*G. pallida* homologue of *H. glycines* effector G19C07
GPLIN_001203000	*G. pallida* homologue of *H. glycines* effector 10C02
GPLIN_000668700	*G. pallida* homologue of *H. glycines* effectors 25A01 and 30G12
GPLIN_000015300	*G. pallida* homologue of *H. glycines* effector G7E05
GPLIN_000167300	Possible orthologue of *H. glycines* G10A06 effector; similarity to E3 Ligases
GPLIN_000785400	Possible orthologue of *H. glycines* G10A06 effector; similarity to E3 Ligases
GPLIN_000393900	Large protein includes sequence similar to *H. glycines* effector scn1120
GPLIN_001559100	Similar to *H. glycines* secretory protein 11 putative effector. Similar to transthyretin-like proteins

G. pallida gene	Effector names
GPLIN_000178900	Similar to *H. glycines* secretory protein 11 putative effector. Similar to transthyretin-like proteins
GPLIN_000869800	Similar to *H. glycines* secretory protein 11 putative effector. Similar to transthyretin-like proteins
GPLIN_000738800	Similar to *H. glycines* secretory protein 11 putative effector. Similar to transthyretin-like proteins
GPLIN_000870000	Similar to *H. glycines* secretory protein 11 putative effector. Similar to transthyretin-like proteins
GPLIN_000169700	Similar to *H. glycines* secretory protein 12 putative effector. Similar to metalloprotease inhibitor
GPLIN_000621200	Similar to *H. glycines* secretory protein 8 putative effector
GPLIN_001317500	Member of the d gl-1 effector family
GPLIN_000901900	Member of the d gl-1 effector family
GPLIN_000901700	Member of the d gl-1 effector family
GPLIN_000325200	Member of the d gl-1 effector family
GPLIN_001199500	Member of the d gl-1 effector family
GPLIN_000207700	Member of the d gl-1 effector family
GPLIN_000442900	Contains *G. pallida* orthologue of *H. glycines* G8A07 effector
Not annotated	Member of the *G. pallida* A42 effector family
Not annotated	Member of the *G. pallida* A42 effector family
GPLIN_000604400	Similar to *M. incognita* effector AY135365, J2 specific
GPLIN_000555600	Similar to *M. incognita* effector AY135365, J2 specific
GPLIN_001416500	Similar to *H. glycines* effector G19B10
GPLIN_000370900	Similar to *H. glycines* effector G19B10
GPLIN_000996800	Similar to *H. glycines* effector G12H04
GPLIN_000926600	Similar to *H. glycines* G20E03 effector
GPLIN_000962200	Similar to *H. glycines* G20E03 effector
GPLIN_000662500	Similar to *H. glycines* G20E03 effector
GPLIN_000977100	Similar to *H. glycines* G20E03 effector
GPLIN_000668700	Similar to *H. glycines* 30G12 effector
GPLIN_000638800	Similar to *H. glycines* 30G12 effector
GPLIN_000637900	Similar to *H. glycines* 30G12 effector
GPLIN_000668600	Similar to *H. glycines* 30G12 effector
GPLIN_001339200	Similar to *H. glycines* 30G12 effector
GPLIN_000120300	Similar to *H. glycines* 30G12 effector
GPLIN_000667500	Similar to *H. glycines* G4G05 and 30G12 effectors
GPLIN_000574800	Similar to *H. glycines* effector gland cell secretory protein 3. Contains thioredoxin-like domain
GPLIN_000990400	Similar to *H. glycines* effector gland cell secretory protein 3. Contains thioredoxin-like domain
GPLIN_001205000	Similar to *H. glycines* effector gland cell secretory protein 3. Contains thioredoxin-like domain
GPLIN_000248100	Similar to *H. glycines* effector G16A01
GPLIN_000933000	Similar to *H. glycines* effector G17G01
GPLIN_001526900	Similar to *H. glycines* effector G17G01
GPLIN_000297600	Similar to *H. glycines* effector G17G01
GPLIN_000167700	GpUBI-EP effector similar to Ubiquitin extension proteins

G. pallida gene	Effector names
GPLIN_000642100	Effector similar to Ubiquitin extension proteins
GPLIN_001038900	Similar to *H. glycines* G18H08 effector
GPLIN_000060800	Member of 448 effector gene family
GPLIN_001471200	Member of 448 effector gene family
GPLIN_001038900	Member of 448 effector gene family
GPLIN_000388900	Member of 448 effector gene family
GPLIN_001255700	Member of 448 effector gene family
GPLIN_000203300	Member of 448 effector gene family
GPLIN_000481100	Member of 448 effector gene family
GPLIN_000796500	Member of 448 effector gene family
GPLIN_000912100	Member of 448 effector gene family
GPLIN_000969800	Member of 448 effector gene family
GPLIN_000970000	Member of 448 effector gene family
GPLIN_001606400	Member of 448 effector gene family
GPLIN_001221800	Member of 448 effector gene family
GPLIN_001596100	Member of 448 effector gene family
GPLIN_000950100	Member of 448 effector gene family
GPLIN_000243800	Member of 448 effector gene family
GPLIN_001390400	Member of 448 effector gene family
GPLIN_000243700	Member of 448 effector gene family
GPLIN_000950600	Member of 448 effector gene family
GPLIN_001221900	Member of 448 effector gene family
GPLIN_000860700	Member of 448 effector gene family
GPLIN_001162100	Member of 448 effector gene family
GPLIN_000970100	Member of 448 effector gene family
GPLIN_001030900	Member of 448 effector gene family
GPLIN_000803500	Member of 448 effector gene family
GPLIN_000792900	Member of 448 effector gene family
GPLIN_001337800	Member of 448 effector gene family
GPLIN_001358800	Member of 448 effector gene family
GPLIN_000969900	Member of 448 effector gene family
GPLIN_000072400	Member of 448 effector gene family
GPLIN_001456900	Member of 448 effector gene family
GPLIN_000407400	Member of 448 effector gene family
GPLIN_001431400	Member of 448 effector gene family
GPLIN_001443600	Member of 448 effector gene family
GPLIN_000126500	Member of 448 effector gene family
GPLIN_000308900	Member of 448 effector gene family
GPLIN_000309000	Member of 448 effector gene family
GPLIN_001390500	Member of 448 effector gene family
GPLIN_001582700	Member of 448 effector gene family
GPLIN_001384700	Putative effector similar to *H. glycines* esophageal gland cell protein Hgg-20. Contains Kinase domain
GPLIN_000349200	Putative effector similar to *H. avenae* gland cell protein and *H. glycines*

G. pallida gene	Effector names
	effector Hgg 20
GPLIN_001475500	Similar to RKN effector (gland cell protein 28). Similar to other nematode proteins
GPLIN_000763000	Similar to *H. glycines* effector G23G11
GPLIN_000872800	Similar to *H. glycines* effector 33A09
GPLIN_000188200	Putative effector similar to *H. avenae* gland cell protein
GPLIN_000107400	Putative effector similar to *H. glycines* Hgg17 effector

Table 3.2: *G. pallida* putative effectors identified by sequence similarity, and their orthologous genes

Effectors not present in *G. pallida* predicted proteins	**First identified in**	**Present in current genome sequence**
AF345801_1	*H. glycines*	No
Hgg-25	*H. glycines*	No
G16A01	*H. glycines*	No
AF273728_1 gland cell secretory protein 1	*H. glycines*	No
AF273733_1 gland cell secretory protein 6	*H. glycines*	No
Gland cell secretory protein 10	*H. glycines*	No
Gland cell secretory protein 9	*H. glycines*	No
Gland cell secretory protein 5	*H. glycines*	No
Gland cell secretory protein 2	*H. glycines*	No
Hgg-26	*H. glycines*	No
G30C02	*H. glycines*	No
G34B08	*H. glycines*	No
G23G12	*H. glycines*	No
G21E12	*H. glycines*	No
G30D08	*H. glycines*	No
G28B03	*H. glycines*	No
G8H07	*H. glycines*	No
G18H08	*H. glycines*	No
G17G06	*H. glycines*	No
AF345800_1 SCN secretory protein	*H. glycines*	No
66 family	*G. rostochiensis*	No – present in previous assemblies
747_22_2	*G. rostochiensis*	Yes
A42	*G. rostochiensis*	Yes

Table 3.3: Summary of putative effectors not present in the *G. pallida* predicted gene models. Some are present in the genome but are not called as genes while others were present in previous genome assemblies.

Species	GH5	GH45	GH30	GH43	GH28	GH53	PL3	CBM	Expansin	Total
G. pallida	15	0	0	1	0	2	7	6	9	53
M. incognita	21	0	6	2	2	0	30	9	20	90
M. hapla	6	0	1	2	2	0	22	2	6	41
B. xylophilus	0	11	0	0	0	0	15	0	8	34
C. elegans	0	0	0	0	0	0	0	0	0	0
P. pacificus	6	0	0	0	0	0	0	0	0	6
B. malayi	0	0	0	0	0	0	0	0	0	0

Table 3.4: The number of genes that are found in cell wall modifying classes for a number of different nematode species.

The expression profiles of the cell wall modifying enzymes were examined. Pectate lyase gene family members are highly up-regulated at J2 and are not expressed at any other life stage. This may reflect a functional role in invasion and migration (Figure 3.2). Other enzymes involved in migration, including expansins (Figure 3.3) and cellulases (Figure 3.4), were up-regulated at J2 and male life stages. CBMs are thought to have two functional roles. Some are thought to be involved in migration while in *H. glycines* one CBM protein has been shown to interact with a host pectin methylesterase which is involved in the regulation of cell growth and expansion. This CBM is thought to be involved in syncytium expansion (Hewezi *et al.*, 2008). The CBMs present in *G. pallida* showed expression profiles reflecting these two functional roles; one CBM gene (GPLIN_000536400) is up-regulated at J2 while another two CBM genes are up-regulated at parasitic stages suggesting they could be involved in syncytium development (Figure 3.5). The two GH53 arabinogalactan endo-1,4-β-galactosidases show different expression profiles; GPLIN143000 is up-regulated in parasitic stages while 142900 is expressed at very low levels, with some up-regulation in males (Figure 3.6). The single GH43 (glycoside hydrolase) present (Figure 3.7) is expressed at extremely low levels in all life stages examined.

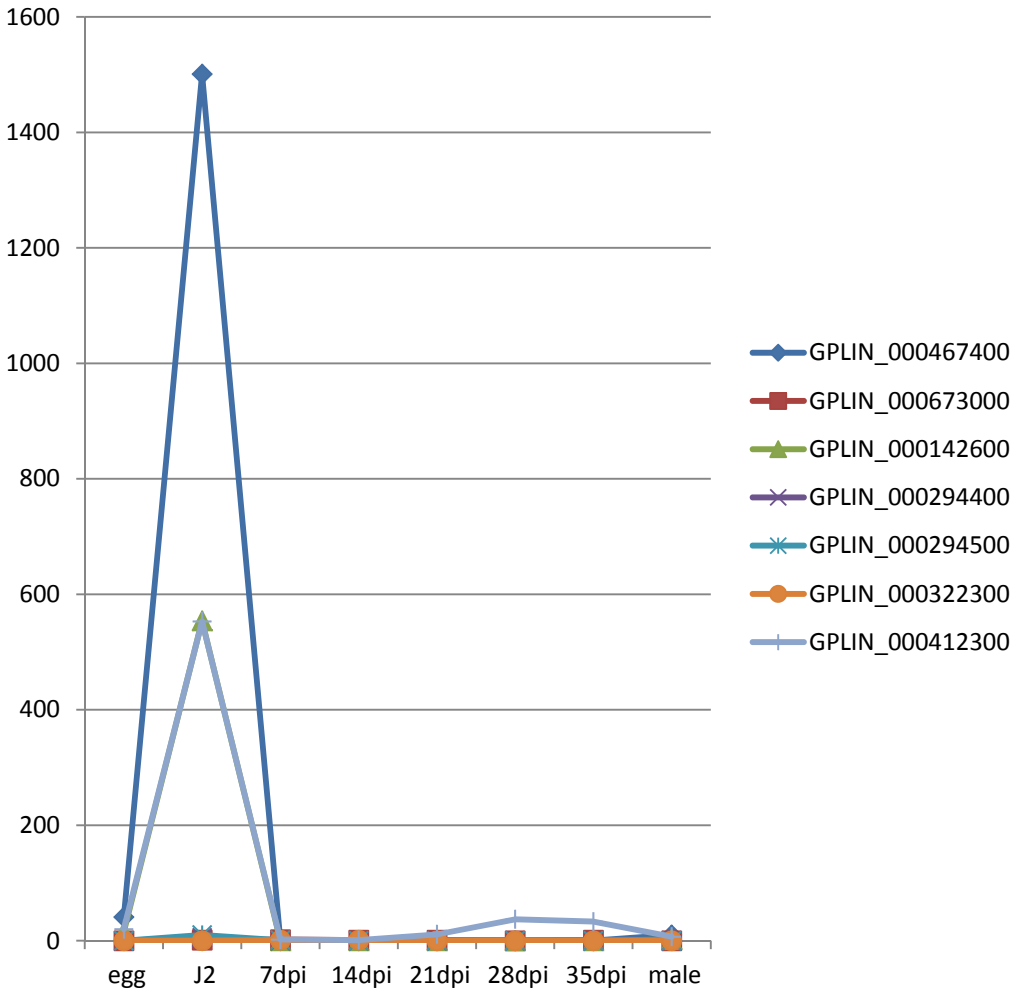


Figure 3.2: Expression profiles of *G. pallida* pectate lyase genes inferred from RNAseq data. Figures on the Y axis are the number of RNAseq reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

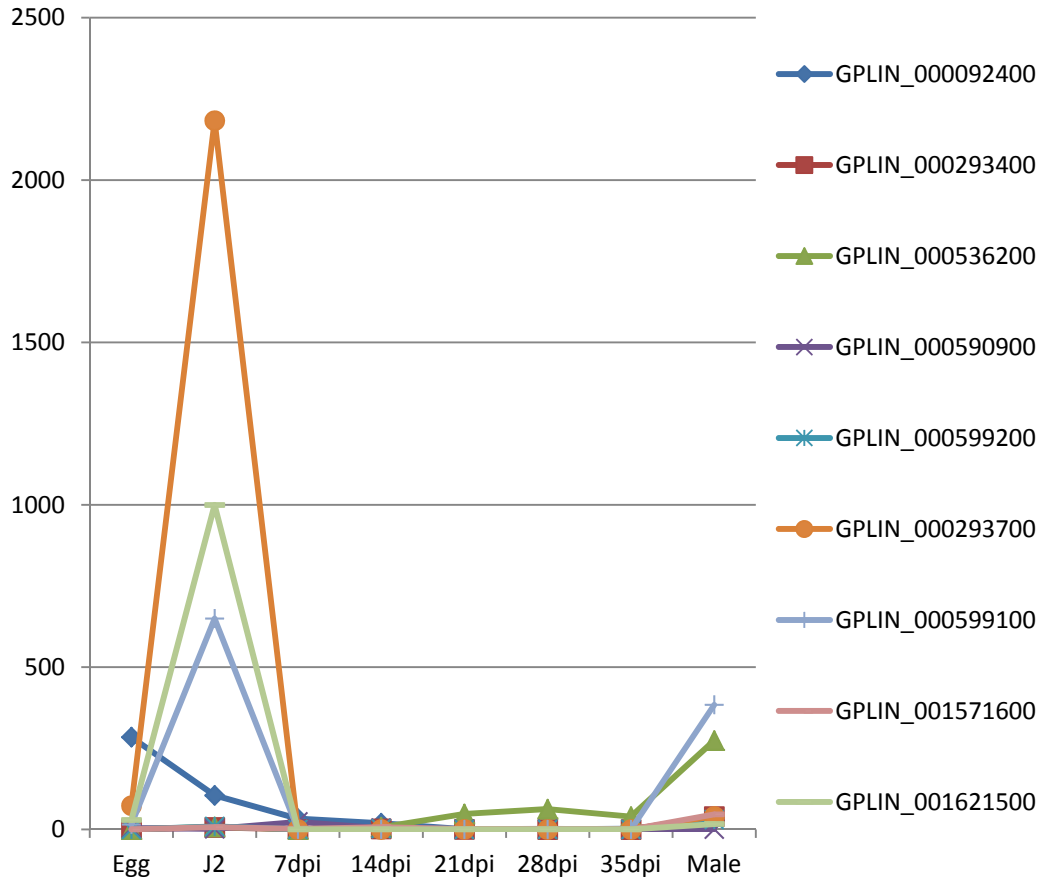


Figure 3.3: Expression of *G. pallida* expansin genes inferred from RNAseq data. Figures on the Y axis are the number of RNAseq reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

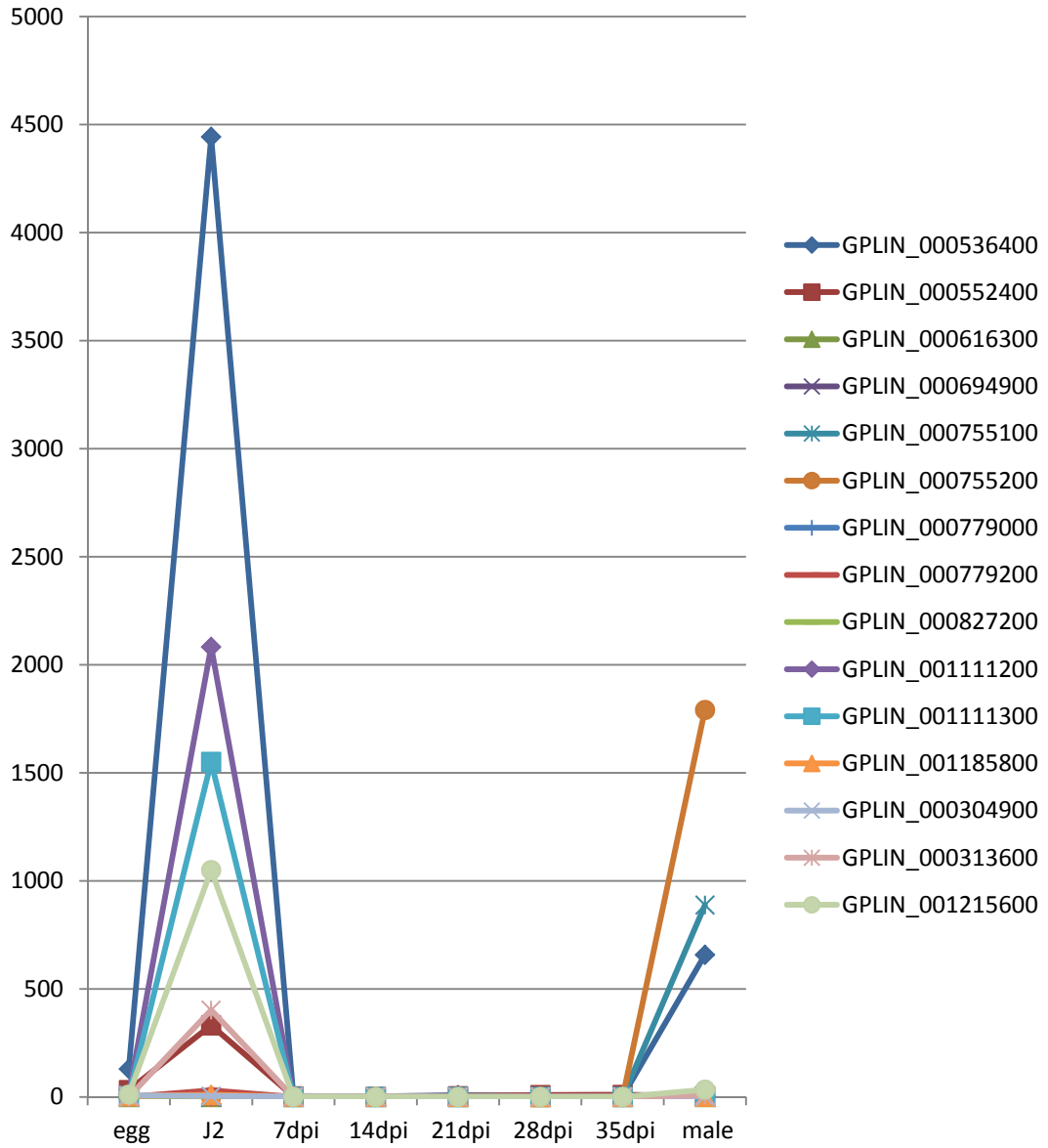


Figure 3.4: Expression profiles of *G. pallida* cellulases (GH5) inferred from RNAseq data. Cellulases are expressed at J2 or J2 and male. Figures on the Y axis are the number of RNAseq reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

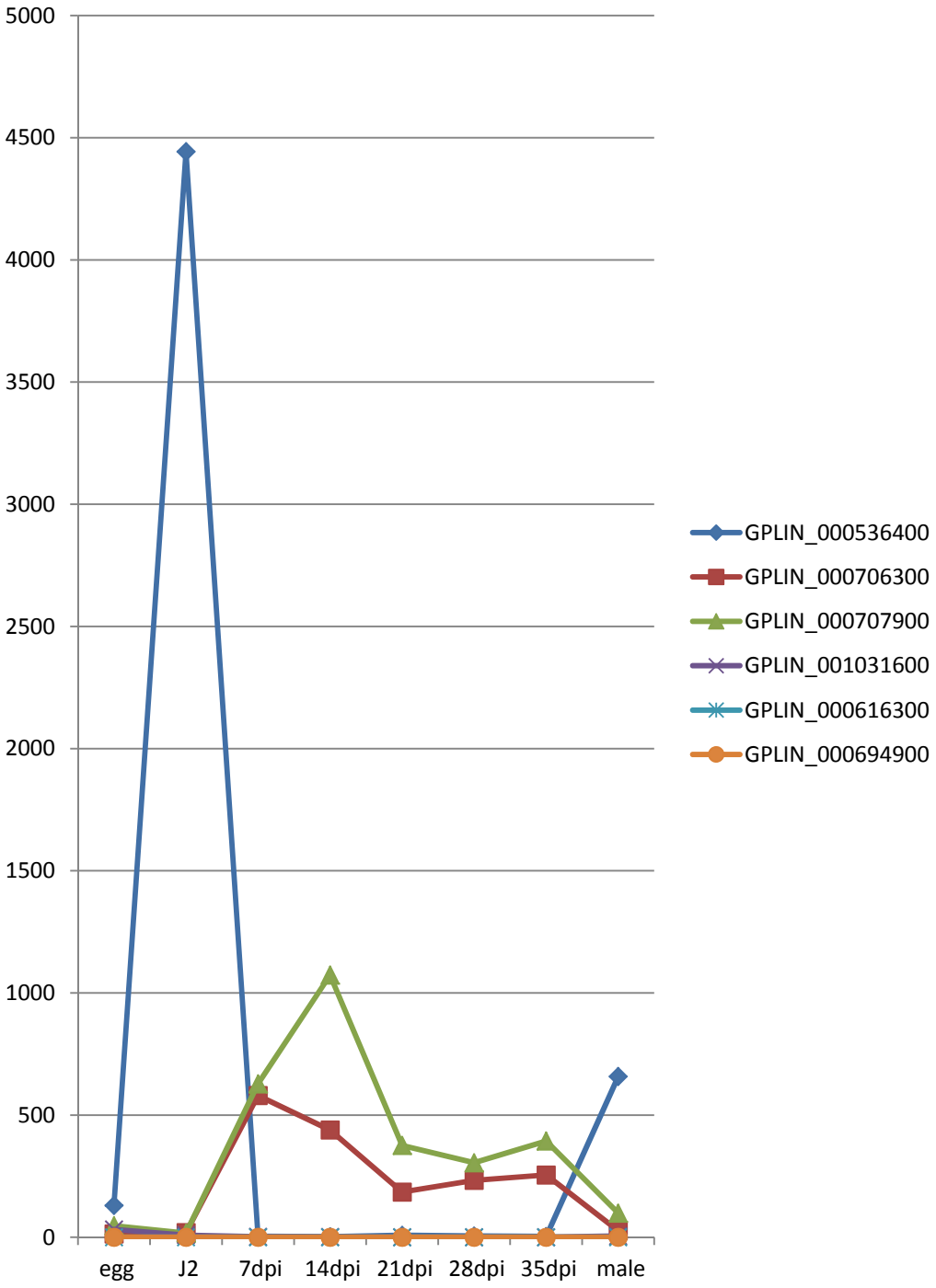


Figure 3.5: Expression profile of *G. pallida* CBM genes inferred from RNAseq data. Figures on the Y axis are the number of RNAseq reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

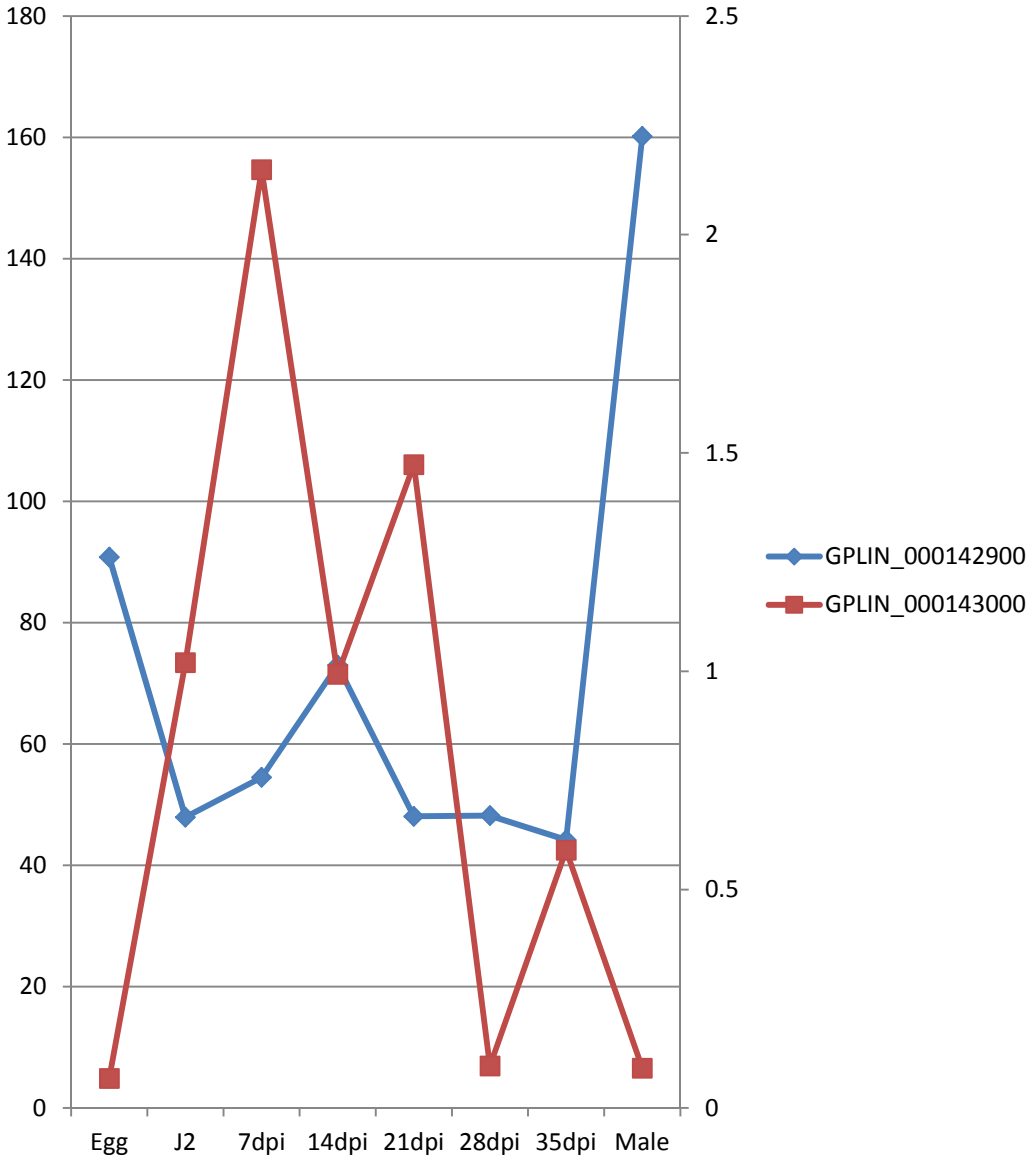


Figure 3.6: Expression of arabinogalactan endo-1,4-β-galactosidase genes (GH53) from *G. pallida* using RNAseq data. GPLIN_000142900 is plotted against the left Y-axis and GPLIN_000143000 is plotted against the right Y-axis, the expression level is relatively very low. Figures on the Y axis are the number of RNAseq reads that map onto the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

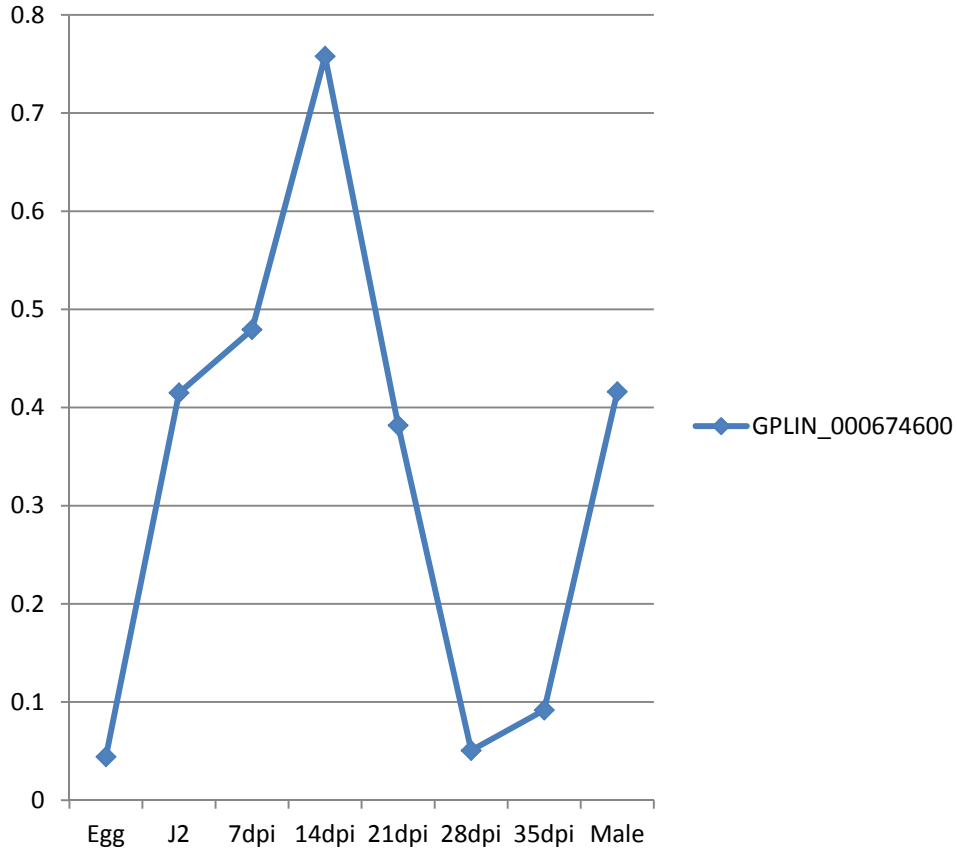


Figure 3.7: Expression of the *G. pallida* glycoside hydrolase (GH43) gene inferred from RNAseq data. Figures on the Y axis are the number of RNAseq reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

3.4.1.2 Other genes potentially acquired by HGT

The *G. pallida* genome contains two predicted chorismate mutase genes. GPLIN_000376700 is highly up-regulated at J2 and is also expressed in unhatched nematodes in eggs, with little expression in parasitic stages or males. GPLIN_000666500 is expressed to some degree at all life stages but is up-regulated at 7 and 14 dpi parasitic stages (Figure 3.8).

Although they are not effectors, genes involved in the synthesis of vitamin B6 have been identified in *H. glycines* (Craig *et al.*, 2009; Craig *et al.*, 2008). Two of these sequences were also identified beside each other on the same contig in the *G. pallida* genome and showed very similar expression profiles (Figure 3.9).

There are two genes that have a PFam cyanate lyase domain. These are GPLIN_000582600 and GPLIN_001292100. These genes have identical sequences at the amino acid level and show identical expression profiles (up-regulated at Egg, J2, 7, 14 dpi and male). The identified cyanate lyase genes have an orthologue in *M. hapla.* This suggests that nematodes in clade 12 may have acquired this gene before divergence, or that there have been separate acquisitions of cyanate lyase via horizontal gene transfer.

3.4.1.3 Functionally characterised effectors from other species

While there are no ascribed functions for most of the effectors identified from cyst nematodes, some cyst nematode effectors have been the subject of detailed functional analysis. Therefore particular attention was paid to whether *G. pallida* also has copies of these genes, as this may imply a conserved and important functional role.

The nematode CLAVATA3/ESR-related peptides are the only reported occurrence of these peptides outside plants (Lu *et al.*, 2009). These peptides may redirect and maintain the differentiation of the root vascular cells into feeding cells (Wang *et al.*, 2005). Although CLAVATA related effectors have been identified from several *Heterodera* and *Globodera* species, the full extent of the gene family will not be known until full genome sequence data is available for these species. In *G. rostochiensis* 16 family members have been identified to date. However, only 5 genes that could encode CLAVATA/ESR peptides were identified in the *G. pallida* genome (GPLIN_000697600, GPLIN_001090600, GPLIN_001090500,

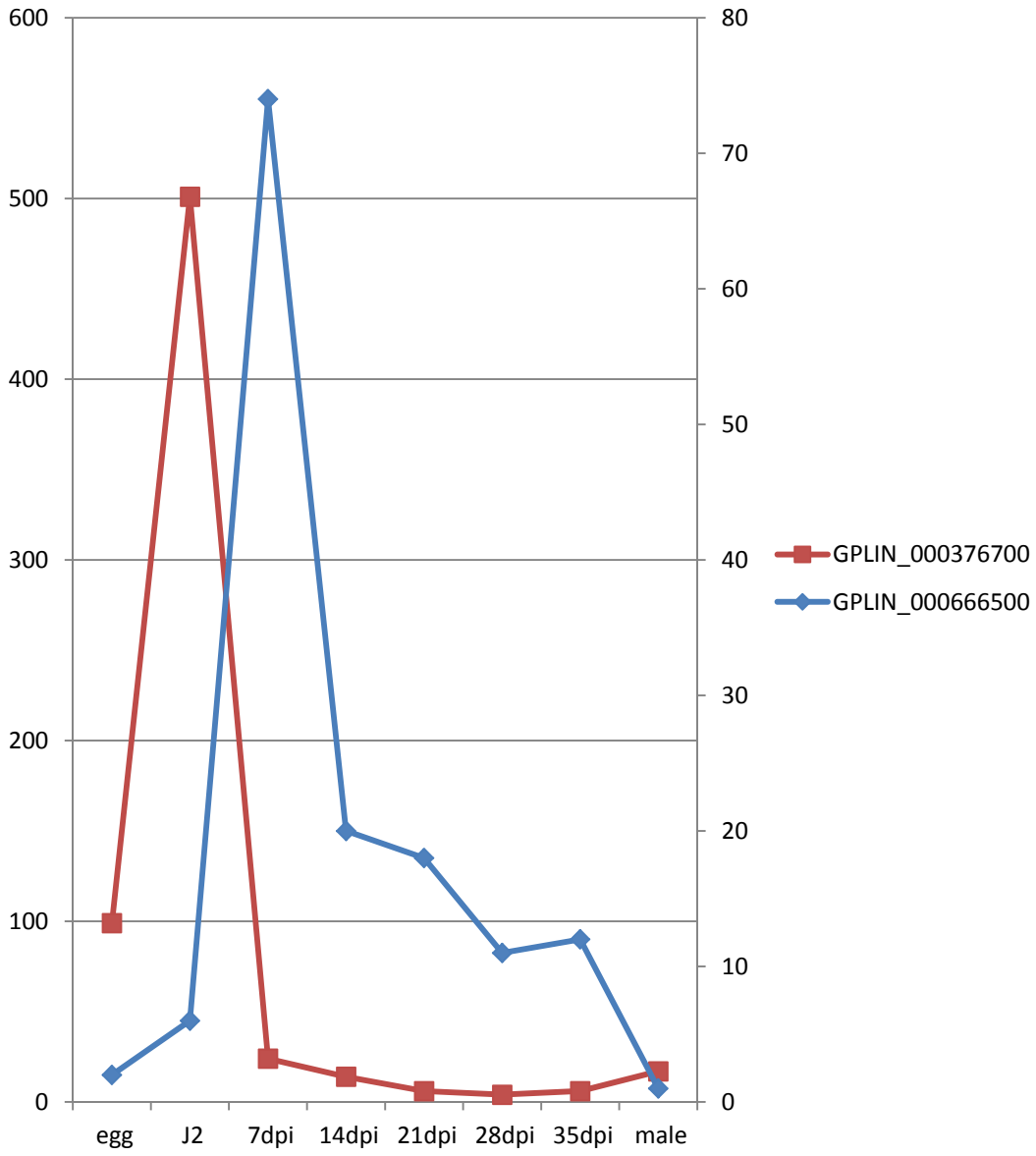


Figure 3.8: Expression profile based on RNAseq data for the 2 identified chorismate mutase genes in the *G. pallida* genome. GPLIN_000376700 is plotted against the left Y-axis and GPLIN_000666500 is plotted against the right Y-axis. Figures on the Y axis are the number of reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

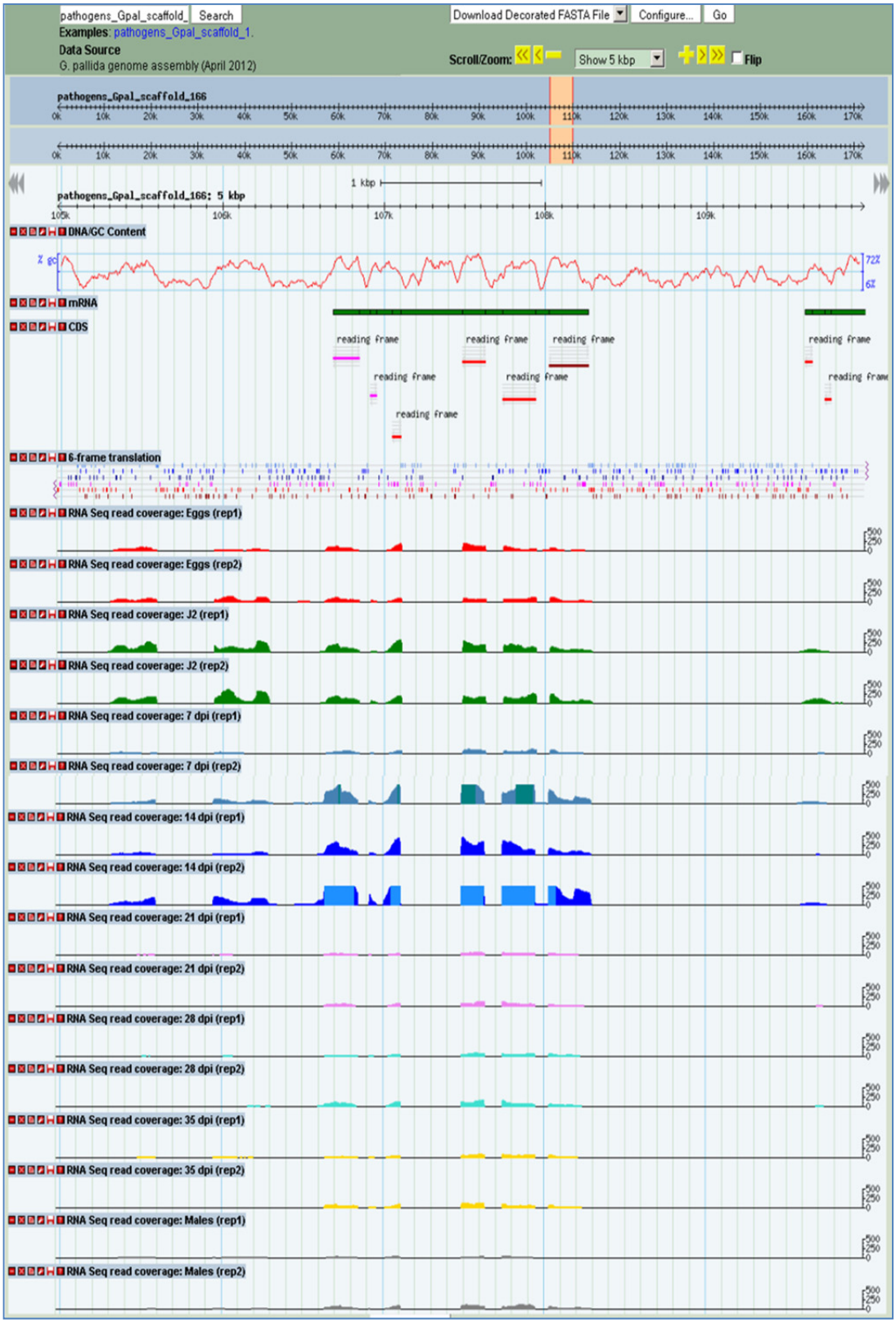


Figure 3.9: Screen shot from GBrowse indicating the part of the assembly encoding the vitamin B6 operon showing normalised RNAseq expression data for various life stages. Expression graphs are capped at 500 reads indicated by the colour change. Duplicated samples for all life stages (egg: red, 7dpi: light blue, 14dpi: dark blue, 21dpi: pink, 28dpi: turquoise, 35dpi: yellow, Male: black) shows the operon is expressed in all life stages with a peak in expression around 7 and 14 dpi.

GPLIN_000950900 and GPLIN_000950800). Four of the CLAVATA effector genes from *G. pallida* possess one CLE domain. However 2 of these genes appear to have not been predicted correctly by the gene finding software as the CLE domain is located in an exon with RNAseq support that has not been called by the gene finding software. One *G. pallida* CLAVATA gene has 4 repeats encoding CLE domains (GPLIN_000697600).

The 19C07 effector from *Heterodera* (see section 1.3) is present in the *G. pallida* genome and is predicted to be gene GPLIN_00078600. This effector has been shown to interact with an auxin efflux protein and may play an important role in feeding site biology. The 10A06 effector from *H. glycines* interacts with spermidine synthase and promotes susceptibility to nematodes when over-expressed in plants. The only *G. pallida* sequence (GPLIN_000730300) similar to the effector had a match with an E-value of 10^{-10} and the length of the match was only 85 out of 250 amino acids. No higher scoring match was found when searching the assembled nucleotide sequence or *de-novo* assemblies of the J2 or 7dpi transcriptome. Therefore the presence of this effector in the *G. pallida* is uncertain. Similarly, there was no significant hit in *G. pallida* for the effector Hg30C02 from *H. glycines.*

Effectors similar to ubiquitin extension proteins (UBI-EP) have been identified in *Heterodera* species, *G. rostochiensis* (Tytgat *et al.*, 2004) and *G. pallida* (Jones *et al.*, 2000). Two sequences similar to UBI-EPs, that have a conserved ubiquitin domain with different C-terminal extensions, were identified in the *G. pallida* genome. A detailed analysis of GPLIN_000167700, which has a 12 amino acid C-terminal extension (ICGHGPNECENQ), is presented in Chapter 5. Another UBI-EP was also present (GPLIN_000642100) which has a different 12 amino acid C-terminal extension (GSMMDYFEDDAM). GPLIN_000167700 is highly up-regulated at J2 whereas GPLIN_000642100 (the un-characterised UBI-EP) is up-regulated at 7dpi and 14 dpi (Figure 3.10), suggesting a different functional role. It should be noted that the levels of expression of GPLIN_000167700 are considerably higher than those for GPLIN_000642100.

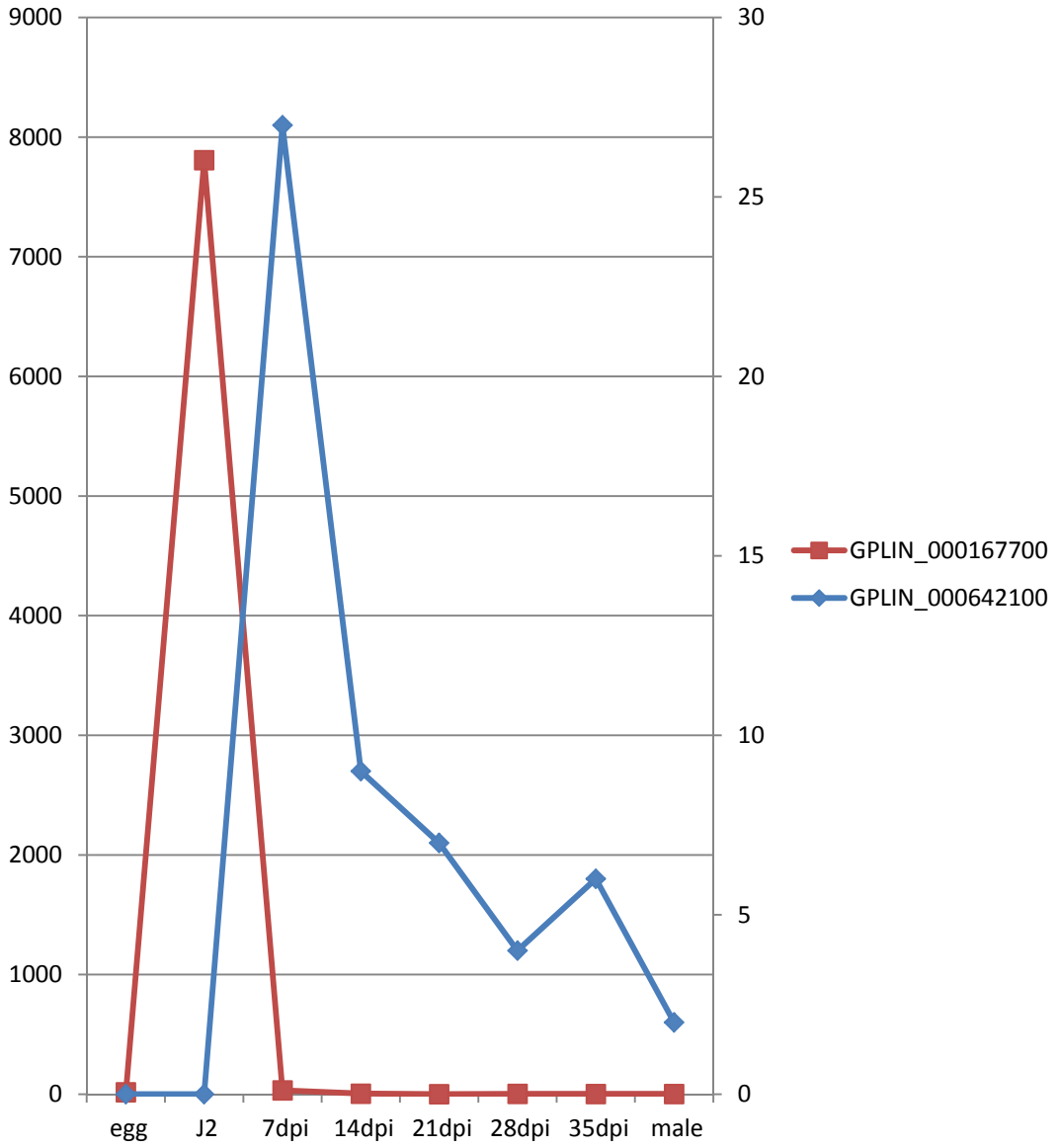


Figure 3.10: Expression profile of the two UBI-EPs identified in the *G. pallida* genome. GPLIN_000167700 (UBI_Del see Chapter 5) is plotted against the left Y-axis and GPLIN_000642100 is plotted against the right Y-axis. Figures on the Y axis are the number of reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

3.4.1.4 Effector gene families

Previous work has revealed the presence of a family of effectors in both *H. glycines* and *G. rostochiensis.* In *H. glycines* seven sequences are present (G30E03, G24A12, 29D09, G11A06, G18H08, G16B09 and G32E03) while in *G. rostochiensis* eighteen similar genes (termed "448") have been identified. Similar effectors also exist as a family (named here the 448 family) in *G. pallida* (Figure 3.11). Analysis of the expression profiles of the 448 family members in *G. pallida* showed that almost all the sequences are up-regulated in parasitic stages (Figure 3.12). A MEME analysis identified a common motif present in all the *G. pallida* 448 proteins (Figure 3.13 and Table 3.5).

Other effectors were also present as substantial gene families in *G. pallida.* These included DGL1 (6 genes), 747 (3 genes), SPRYSECs (see below) and 1106 (11 genes).

3.4.1.5 Cluster analysis of expression profiles of *G. pallida* effectors

Although the expression profiles of some sequences and gene families were examined when characterising the individual genes or families, the entire list of orthologues was analysed to determine whether clusters of similar expression profiles were detectable which may reflect conserved functional roles. This analysis revealed that five different clusters were present: J2 (30 sequences), J2 and male (5 sequences) parasitic stages (61 sequences), constitutive (20 sequences) and parasitic and male (4 sequences).

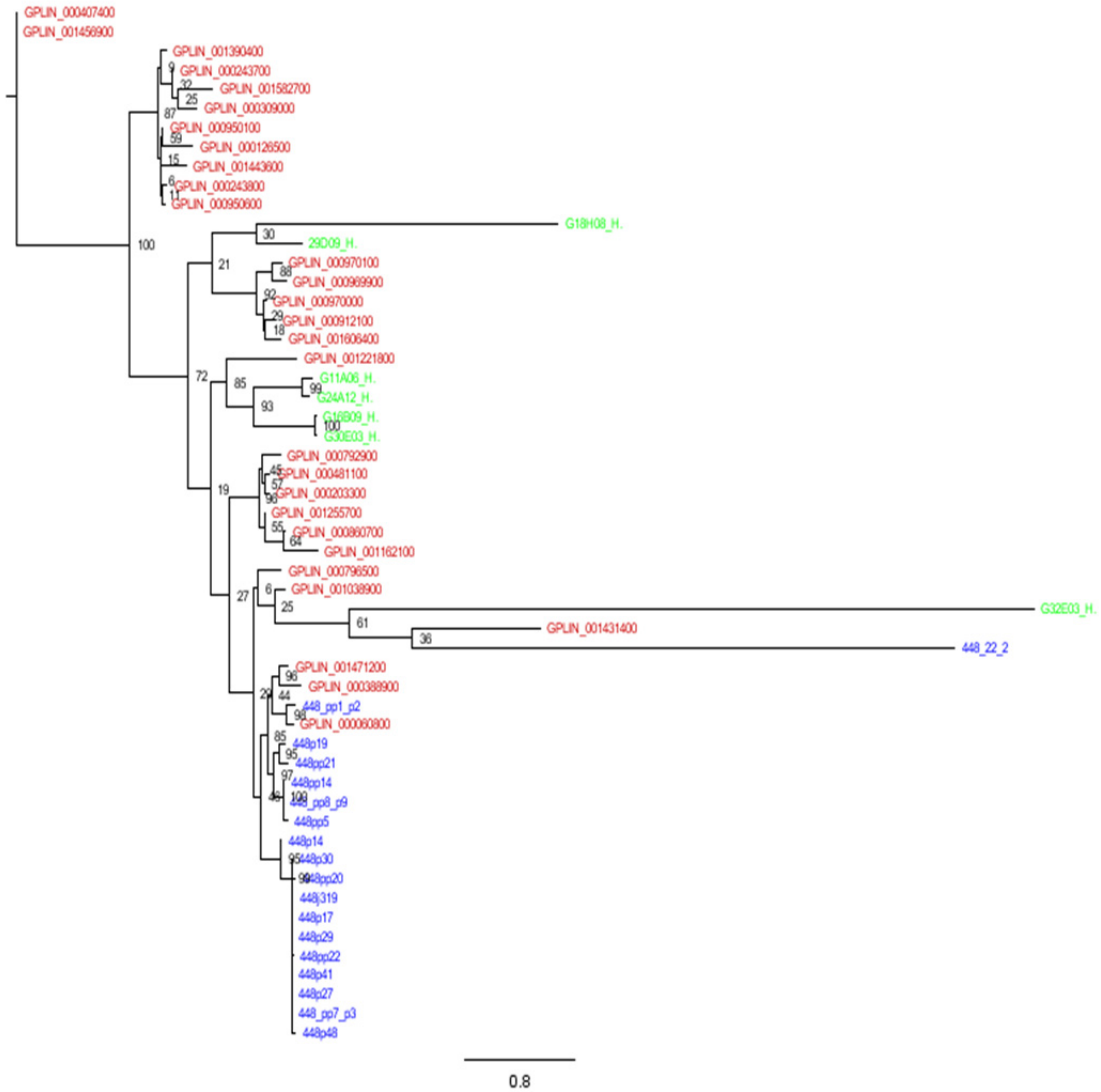


Figure 3.11: A phylogenetic analysis of genes that are similar to 448 effectors. A maximum likelihood phylogenetic analysis of all 448 family members from *G. pallida* - red, *H. glycines* – green and *G. rostochiensis* – blue, with 100 bootstraps.

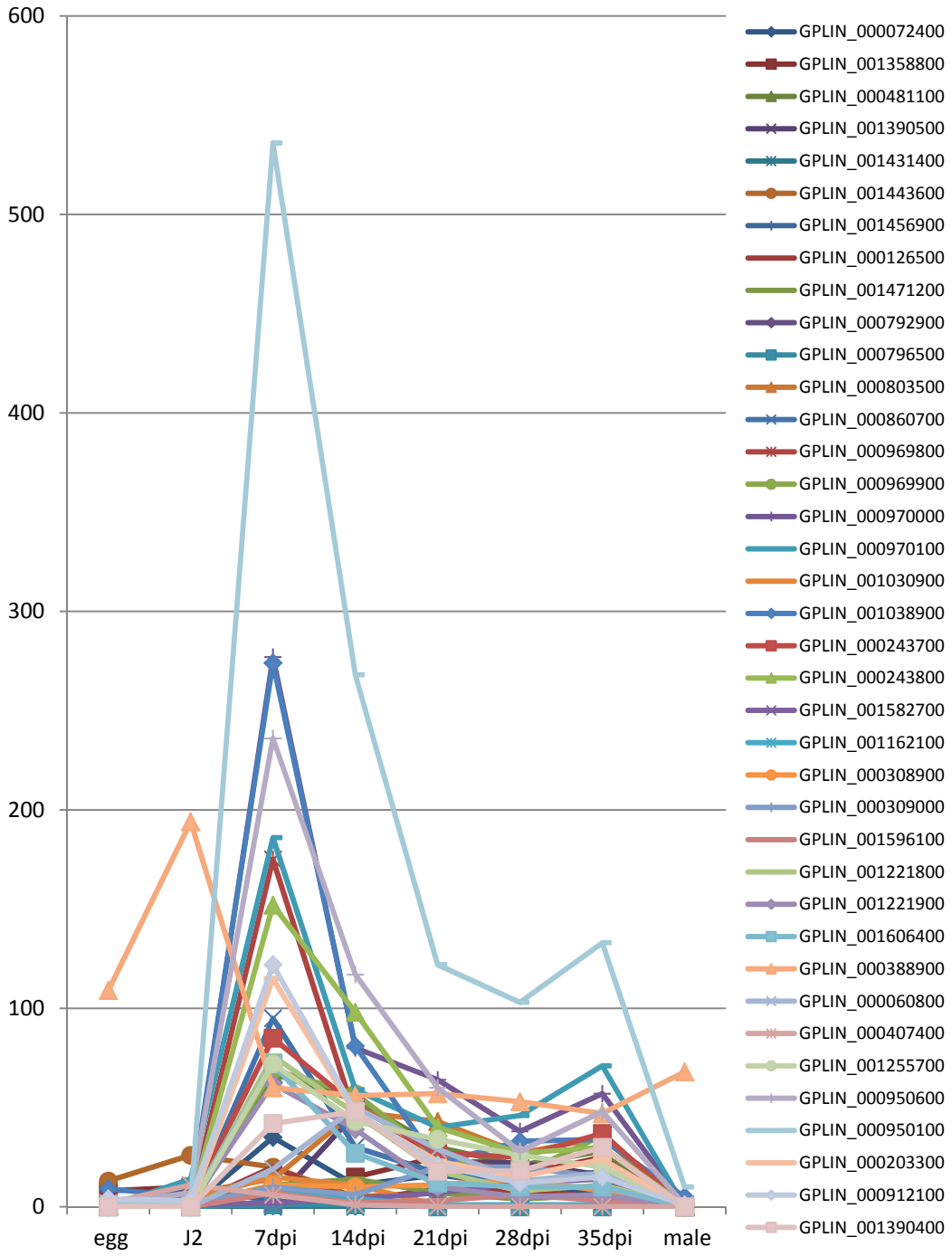


Figure 3.12: Expression profile of 448 family members from *G. pallida* inferred by RNAseq data. Figures on the Y axis are the number of reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

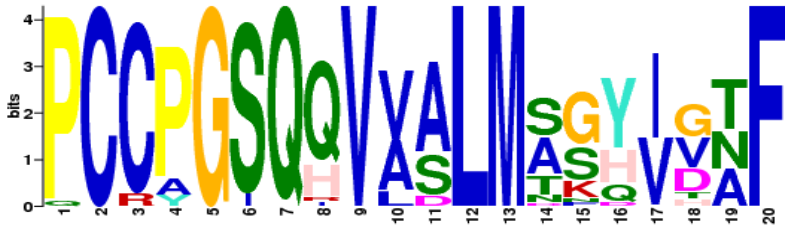


Figure 3.13: MEMEs output showing the relative frequency of amino acids at specific positions within the common motif identified in the 448 effector genes family

Name	Sta-rt	*p*-value	Motif		
GPLIN_000950600	30	5.29E-35	SHLCLAASRF	PCCPGSQQVAALMSGYIVNFTNSVDTDDKQTL	CGSVIEDVQC
GPLIN_000309000	30	5.29E-35	SHLCLAAPKF	PCCPGSQQVAALMSGYIVNFTNSVDTDDKQTL	CGSVIEDHDK
GPLIN_001582700	28	5.29E-35	SHVCQAAPSF	PCCPGSQQVAALMSGYIVNFTNSVDTDDKQTL	CGRVIEDVQ
GPLIN_001390400	30	5.29E-35	SHLCLASPRF	PCCPGSQQVAALMSGYIVNFTNSVDTDDKQTL	CSSVIEDVQC
GPLIN_000950100	30	5.29E-35	SHVCLAAPRF	PCCPGSQQVAALMSGYIVNFTNSVDTDDKQTL	CSSVIEDVQC
GPLIN_000243800	30	5.29E-35	SHLCLAAPRF	PCCPGSQQVAALMSGYIVNFTNSVDTDDKQTL	CGSVIEDVQC
GPLIN_000243700	30	5.29E-35	SHLCLASPRF	PCCPGSQQVAALMSGYIVNFTNSVDTDDKQTL	CGSVIEDVQC
GPLIN_001443600	30	2.09E-30	SHLCLAAPRF	PCCPGSQHVAALMSGYIVNFTNSVDTDDKQNA	RELEAMNNCE
GPLIN_000970100	27	7.15E-30	LFNSCFAAPH	PCCPGSQHVVALMTKYIGTFSAGEDESTVCAS	AENVVNAIKS
GPLIN_000481100	28	1.45E-29	CKCCISAPQF	PCCPGSQQVVSLMAFHVDAFASTMTESTACKN	ANDVETAVKS
GPLIN_000203300	28	1.63E-29	CKCCISAPQF	PCCPGSQQVVSLMASHVDAFASTMTESAACKN	ANDVKNAVKS
GPLIN_001255700	28	1.63E-29	CKCCISAPQF	PCCPGSQQVVSLMASHVDAFASTMTESAACKN	ANDVENAVKS
GPLIN_001162100	28	1.63E-29	CKCCISAPQF	PCCPGSQQVVSLMASHVDAFASTMTESAACKN	ANDVENAVKS
GPLIN_000860700	28	1.63E-29	CKCCISAPQF	PCCPGSQQVVSLMASHVDAFASTMTESAACKN	ANDVENAVKS
GPLIN_001471200	28	5.07E-29	CKCCISAPQF	PCCPGSQQVLSLMAGDVGTFSSEMSESTACET	AEIVANSVKS

Name	Sta-rt	*p*-value	Motif		
GPLIN_000969900	27	1.21E-28	LFNNCFAAPH	PCCPGSQHVVALMTKYIGTFSAGEAESTVCAS	AENVVNAIKS
GPLIN_000060800	28	2.07E-28	CKCCTSAPQF	PCCPGSQRVLALMNGQIGTFSSEMSESEACQT	AENVANDVKS
GPLIN_000970000	27	2.83E-28	LFNSCFAAPH	QCCPGSQHVVALMTKYIGTFSAGEDESTVCAS	AENVVNAIKS
GPLIN_000912100	27	5.26E-28	LFNSCFAAPH	PCCPGSQHVVALMTNYIGTFSADEAESTVCAR	AENVVNAIKS
GPLIN_001038900	30	7.90E-28	CKYCTSAPQF	PCCAGSQQVVALMAGQVDAFTSKMSESKTCET	ADNVANAVKK
GPLIN_000388900	28	3.48E-27	CKCCISAPQF	PCCPGSQQVLSLMAGHVGTFSSEMSESKASPM	AAVPEFVAEI
GPLIN_000796500	18	1.31E-26	VAPWLLLAEF	PCCAGSQQVVALMDSQVHAFSSEMSKSEACTK	AENVANAVRS
GPLIN_001606400	27	8.11E-26	LFNSCFAAPH	PCCPGSQHVVALMTKYIGTFSAEAGEDESTVC	ANAENVVNAI
GPLIN_001221800	20	5.47E-25	WQSSISAPQY	PCCPGSQIVVSLMNSHIGTFSSSMSQTELCSS	AEELERNLRS
GPLIN_000407400	27	1.63E-24	QRSIASPIRY	PCRYGIQQVADLMSSYVTTFKNSVEHNERLAL	CKNVIEDMKD
GPLIN_001456900	27	1.63E-24	QRSIASPIRY	PCRYGIQQVADLMSSYVTTFKNSVEHNERLAL	CKNVIEDMKD
GPLIN_001431400	28	4.11E-22	CKYCTSAPQF	PCCAGSQQVVALMAGQVHAFTPKCPSRRLAQP	PTLSQRPILK

Table 3.5: MEMEs data summarising the motif identified in the 448 effector family. 2000bp upstream of all 448 family members were analysed for motifs.

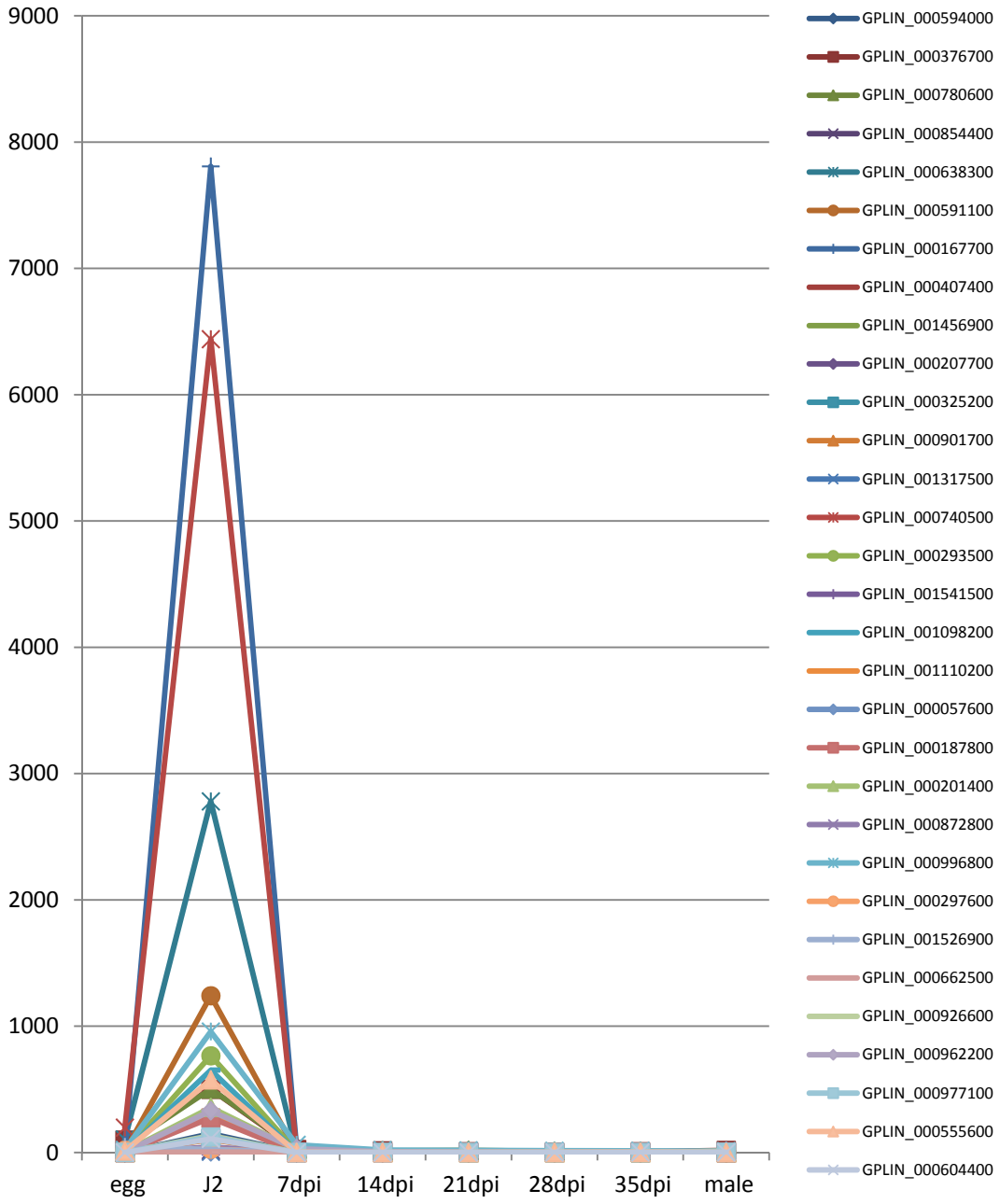


Figure 3.14: Expression profile of orthologous effectors which cluster in the 'J2 only' category, inferred by RNAseq data. Figures on the Y axis are: The number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

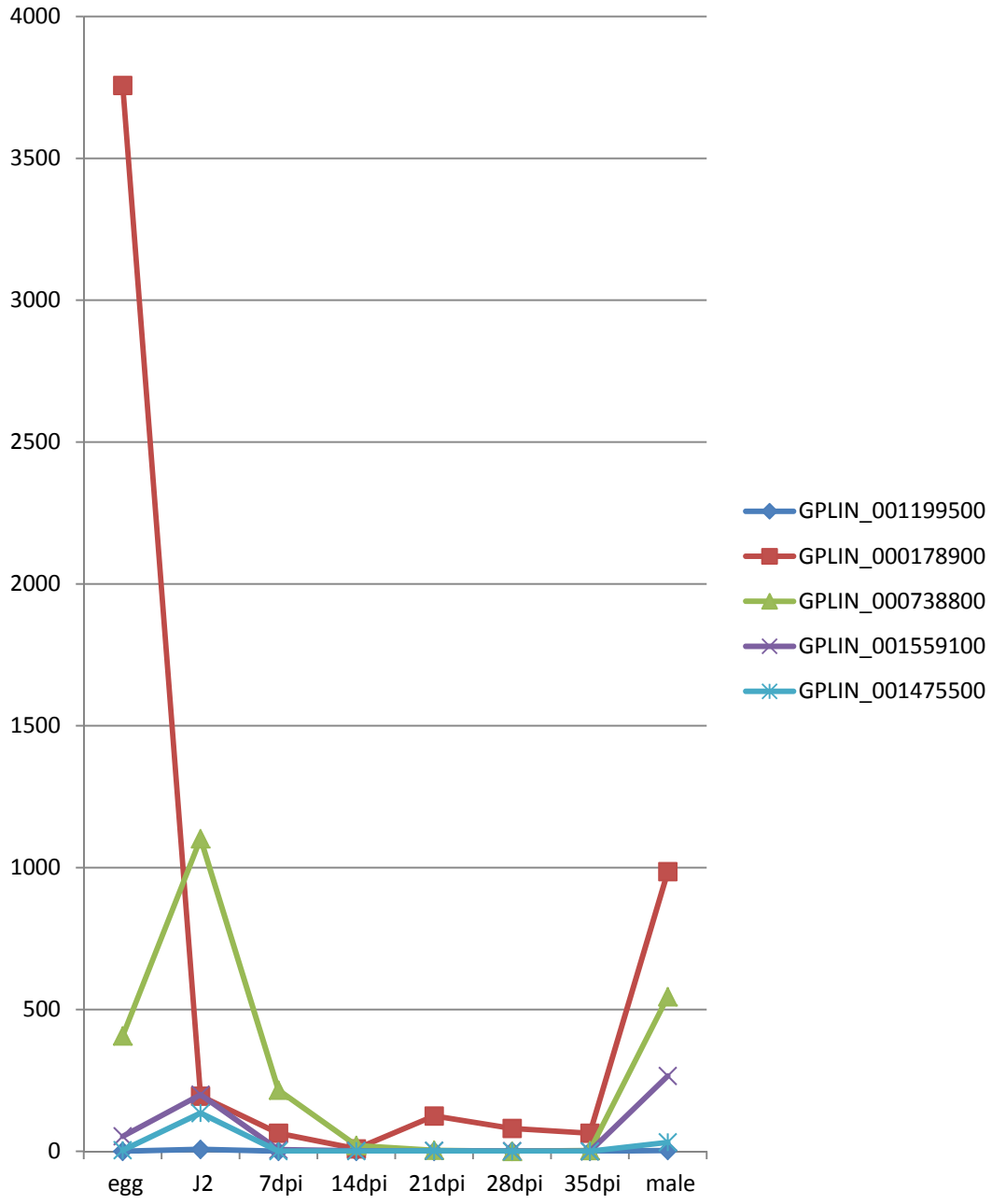


Figure 3.15: Expression profile of orthologous effectors that cluster in the 'J2 and male' category, inferred by RNAseq data. Figures on the Y axis are the number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

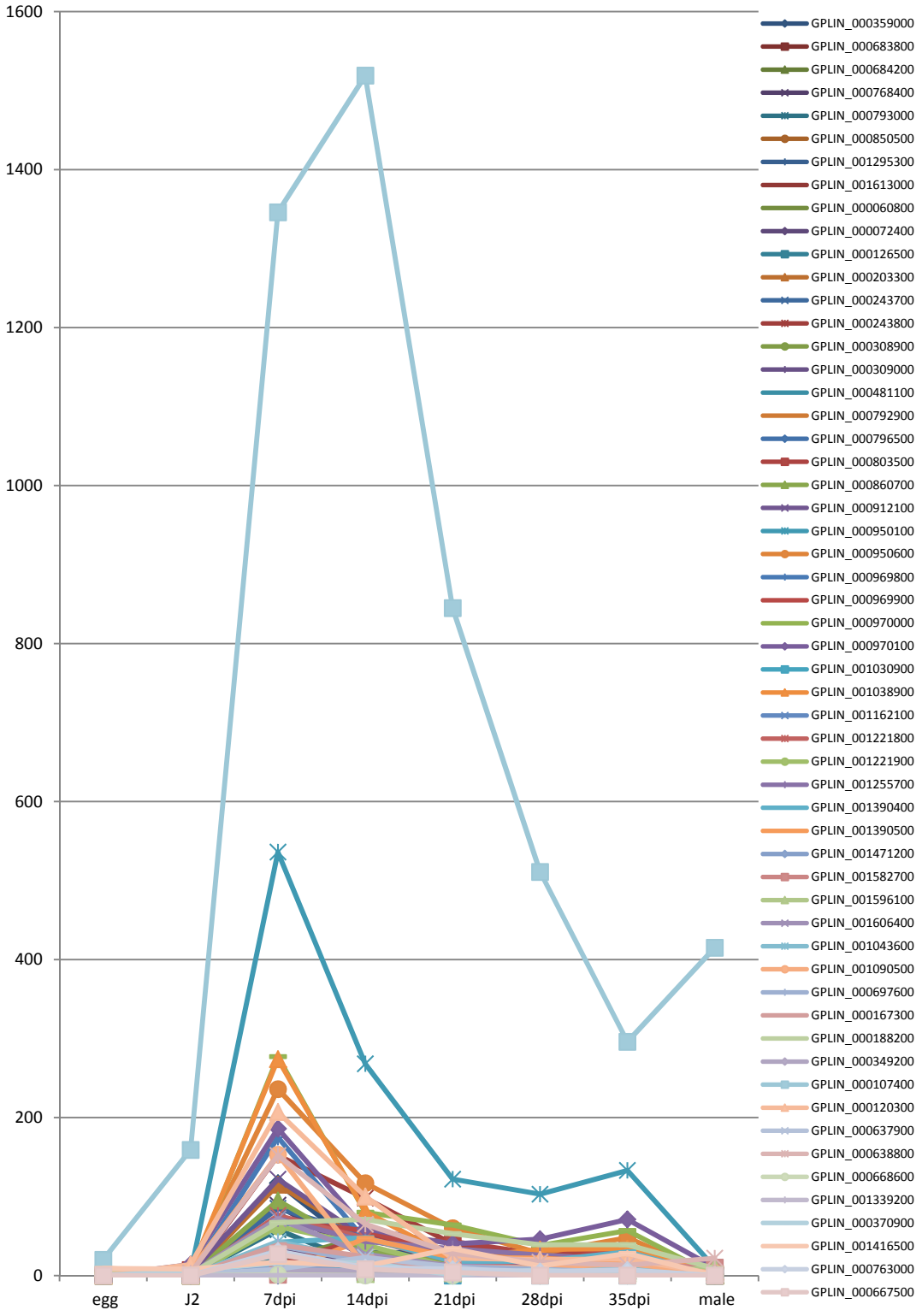


Figure 3.16: Expression profile of orthologous effectors that cluster in the 'parasitic' category, inferred by RNAseq data. Figures on the Y axis are the number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

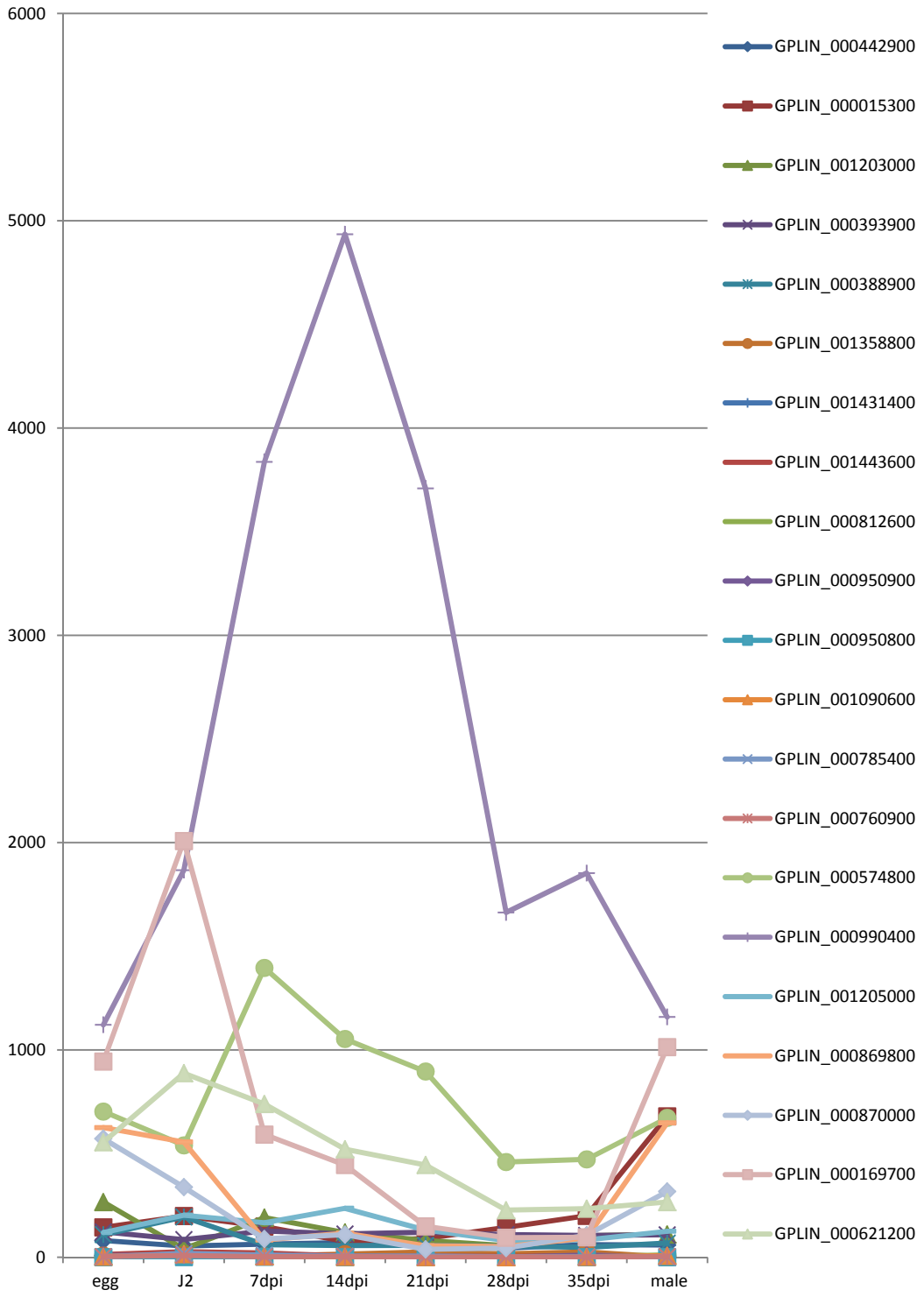


Figure 3.17: Expression profile of orthologous effectors that cluster in the 'constitutive' category, inferred by RNAseq data. Figures on the Y axis are the number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

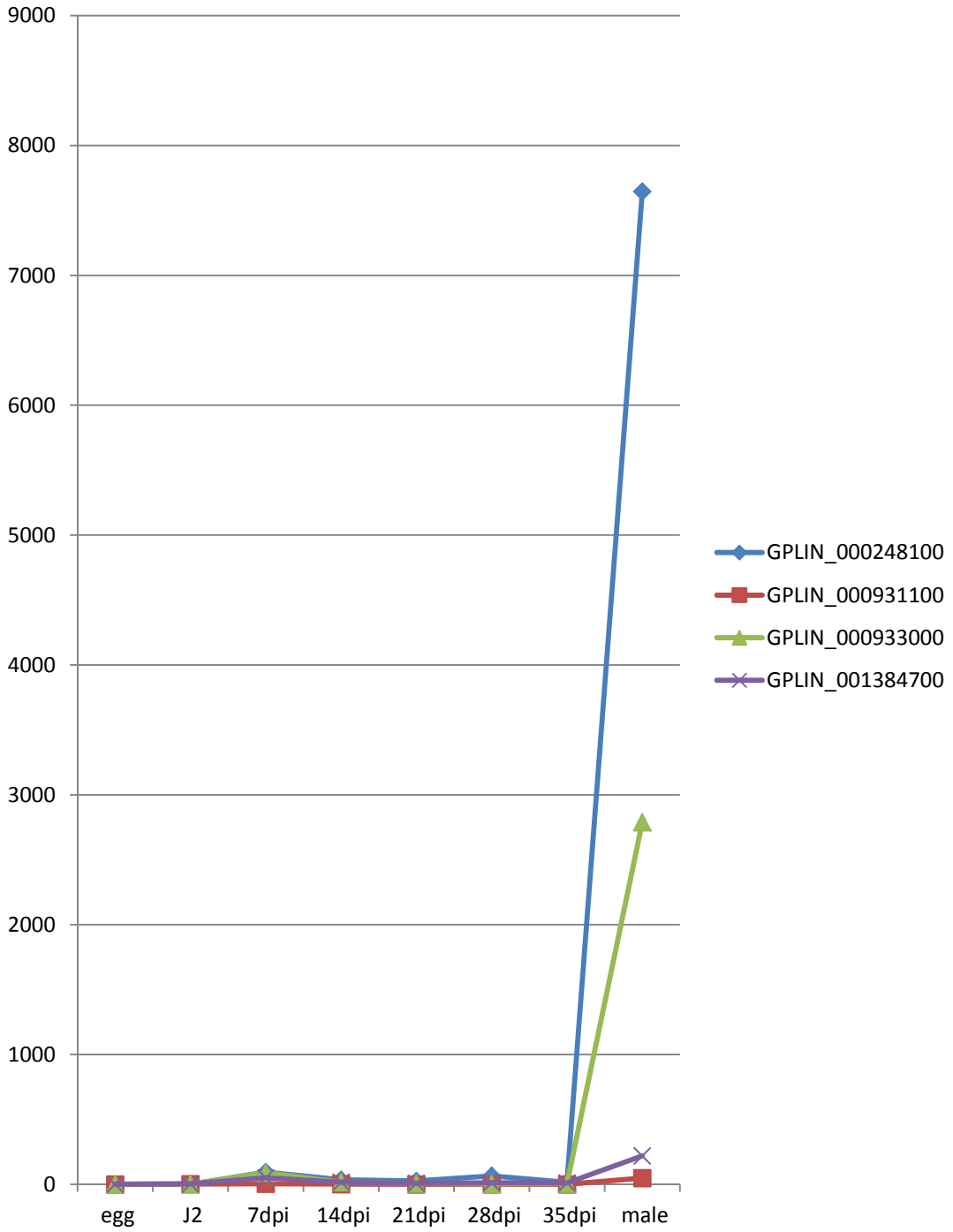


Figure 3.18: Expression profile of orthologous effectors that cluster in the 'parasitic and male' category, inferred by RNAseq data. The expression in the "male" life stage is relatively high, which makes the "parasitic" expression look very small. However, all these genes have significant expression in parasitic stages. Figures on the Y axis are the number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

3.4.2 SPRY domain containing proteins

The B30.2 domain was first identified in an exon, called B30-2 from a human major histocompatibility complex region. The SPRY domain was first identified in a kinase splA of *Dictyostelium* and ryanodine receptor. The two domains share sequence similarity such that the SPRY domain can be identified within all B30.2 domain containing proteins. The function of the B30.2/ SPRY domain is unknown, although some assays suggest that the domain(s) mediates interactions between proteins. The SPRY and B30.2 domains together form an immunoglobulin-like fold (Woo *et al.*, 2006). Previous work has been conducted on SPRYSEC effectors from *Globodera* which has shown some of their localisation patterns within plant cells, their involvement in plant-nematode R-AVR gene interactions and their involvement in the suppression of host defences (Jones *et al.*, 2009b; Postma *et al.*, 2012; Rehman *et al.*, 2009b; Sacco *et al.*, 2009a; Sacco *et al.*, 2007).

299 SPRY domain containing proteins were identified from the *G. pallida* genome sequence. By contrast, *C. elegans* has 8 proteins that contain a SPRY domain(s) while *B. xylophilus* and *M. incognita* have 12 and 27 respectively. Analysis of 30 randomly shuffled protein sets from *G. pallida* generated no matches when searching for SPRY domains, strongly indicating that the predicted SPRY domains do not occur by chance. Therefore the SPRY domain protein family is hugely expanded in *G. pallida* compared to other nematodes.

3.4.2.1 Phylogenetic analysis of the SPRY domain proteins

Phylogenetic analysis was used to determine the relationships between SPRY domain containing proteins from various plant parasitic nematodes. SPRY domain containing proteins were included from humans, yeast and *C. elegans* for comparison. The tree can be summarised as having a clade containing SPRY domain proteins from all species at the top of the tree with a large *Globodera* specific expansion at the bottom of the tree. Reciprocal BLAST hit (RBH) analysis shows that *G. pallida* orthologues of SPRY proteins from other species are all present in the clade at the top of the tree. This suggests these SPRY domain containing proteins have similar function within nematodes and possibly other organisms. All SPRYSEC effectors identified to date are present within the *Globodera* specific expansion (Appendix 3). This shows that the SPRY domain containing proteins may have been

under strong selection pressure as would be anticipated for SPRY domain containing proteins that function as effectors (Figure 3.19).

A phylogenetic analysis of all the SPRY domain containing protein in Nembase4, including *G. pallida, G. rostochiensis* and *H. glycines* was performed (Appendix 3). SPRY domain containing proteins from *G. pallida* and *G. rostochiensis* cluster together in their own clade that expands out of the tree. Known SPRYSECs can be seen in this expanded *Globodera* specific clade. However, another clade can be seen that only contains proteins from *G. pallida, G. rostochiensis* and *Heterodera* (cyst nematodes). This clade is in-between the *Globodera* specific and "all other nematode" clades, seen higher up in the tree. There may be more SPRY domain containing protein in *Heterodera* and *G. rostochiensis* than is presently known due to their partial characterisation. Once sequence data is available the full extent of the family in these nematodes will be realised.

3.4.2.2 Expression analysis of SPRY domain proteins

115 of the 299 *G. pallida* SPRY domain containing proteins were not expressed based on an expression cut off threshold of "2" from the normalised RNAseq data. 7 were expressed in "egg only", 23 were expressed in "Egg and J2" life stages, 36 genes were up-regulated at "J2", 41 genes show constitutive expression, 65 were up-regulated in parasitic stages and 12 were expressed in "male only". Out of the 299 *G. pallida* SPRY domain containing proteins 22 had more than one SPRY domain.

3.4.2.3 SPRY gene expression – qPCR

In order to confirm the expression profiles inferred from RNAseq analysis, primers were designed to SPRY genes that had various specific expression profiles from gpal.201201.Aug_hints.NT.fa. The genes in the old assembly and their current gene matches are summarised along with their RNAseq expression profiles in Table 3.6. For all SPRY domain containing genes analysed by qPCR, the expression was very low (Figure 3.20). However, 8 out of 11 genes analysed had similar or identical expression patterns to that observed via RNAseq analysis (Table 3.6).

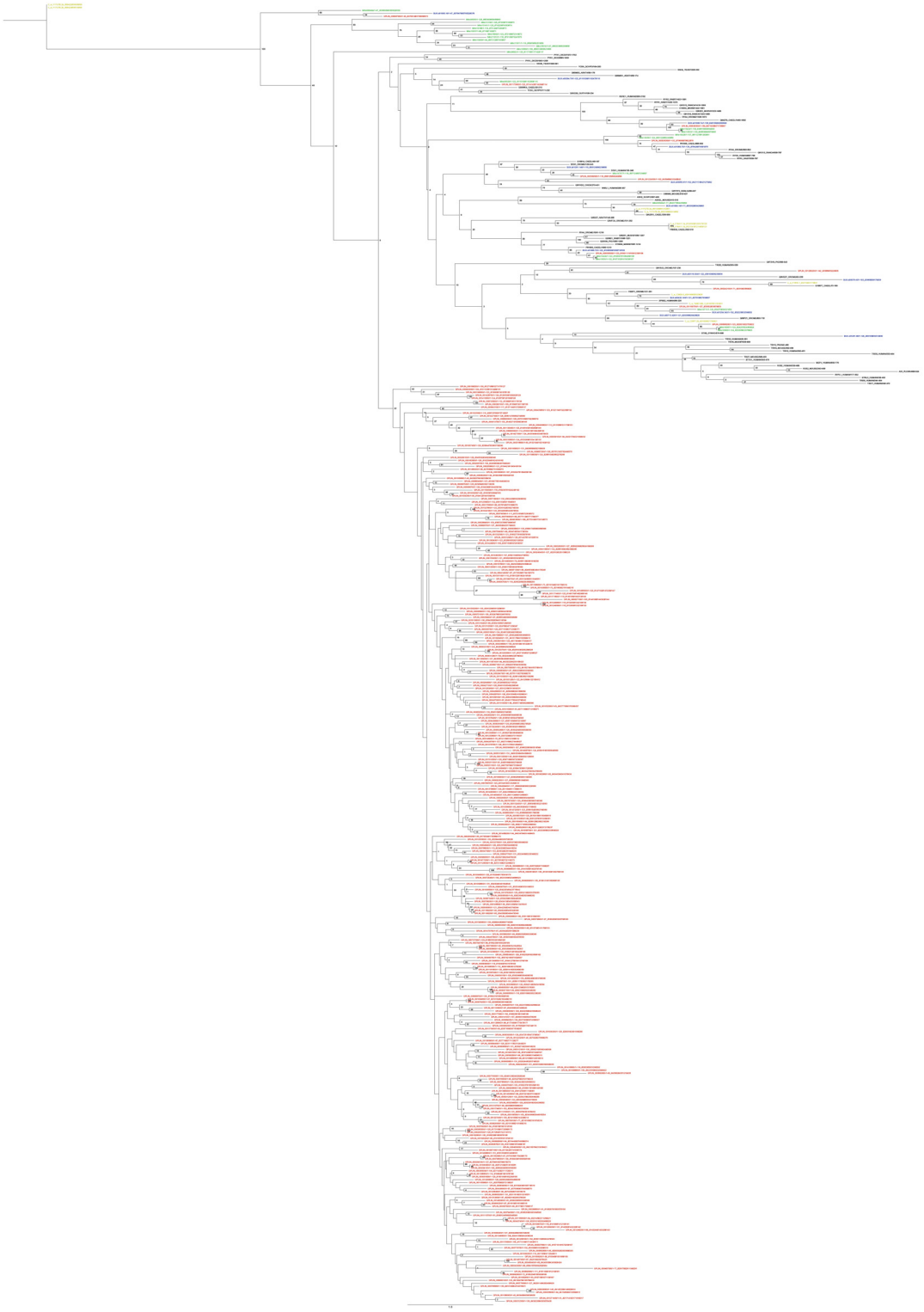


Figure 3.19: A phylogenetic tree of SPRY domain containing proteins. A maximum likelihood phylogenetic analysis of all SPRY domain family members from *G. pallida* - red, *M. incognita* – green, *B. xylophilus* – blue, *C. elegans* – yellow and seeds from PFam - black, following alignment back onto the PFam SPRY definition, with 100 bootstraps. This tree shows a common set of SPRY domain containing proteins found in parasitic nematode and other eukaryotes which have no signal peptides and a huge expansion of the gene family in *G. pallida* of which some have signal peptides and are employed as effectors.

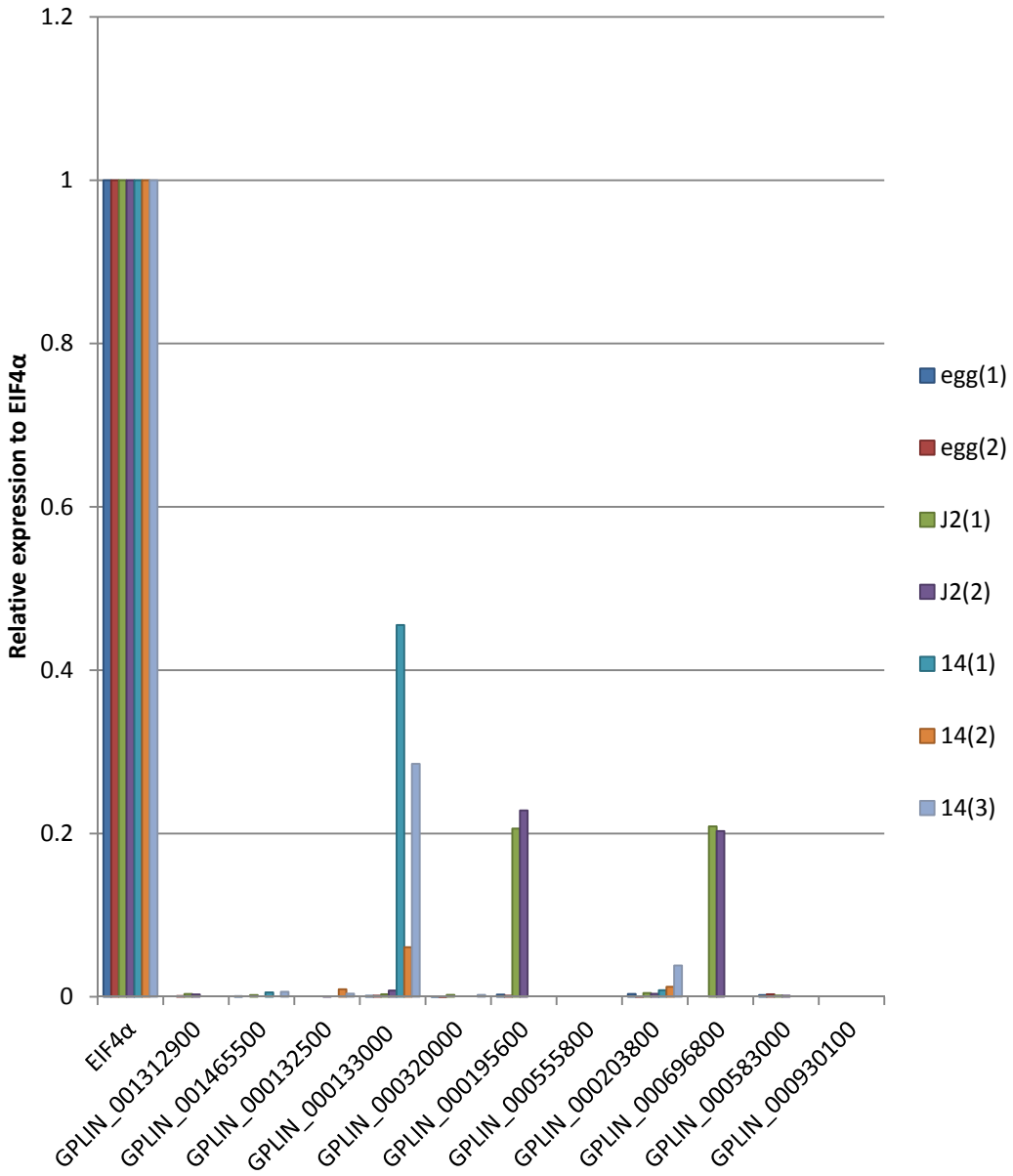


Figure 3.20: qPCR expression of SPRY domain containing proteins relative to EIF4α for life stages egg, J2 and 14 dpi with biological replicas. Specific primers were designed to amplify SPRY domain containing proteins to validate RNAseq data.

Gene name	Expression - prediction based on RNAseq	qPCR expression
GPLIN_000320000	constitutive	very low: parasitic and J2
GPLIN_000555800	constitutive	No expression
GPLIN_000195600	J2	J2
GPLIN_000696800	J2	J2
GPLIN_000133000	J2 and parasitic stage	J2 and parasitic stage
GPLIN_000203800	J2 and parasitic stage	J2 and parasitic stage
GPLIN_000132500	Later parasitic stage	Low expression in parasitic and J2
GPLIN_001312900	Later parasitic stage	very low J2 only
GPLIN_001465500	Later parasitic stage	Later parasitic with very low J2
GPLIN_000583000	no expression	no expression/ very low in J2
GPLIN_000930100	No expression	No expression

Table 3.6: The expression profile of SPRY domain containing genes analysed by qPCR. A summary of the predicted expression profiles based on RNAseq data, the expression profile pattern generated by qPCR data and the gene names are displayed in the table.

3.4.2.4 Secretion of SPRY domain proteins

G. rostochiensis and *G. pallida* are known to secrete SPRY domain containing proteins as effectors and these have been termed SPRYSECs (Rehman *et al.*, 2009b; Sacco *et al.*, 2009b). However, only 30 of the SPRY domain containing proteins (10% of the gene family) from *G. pallida* are predicted by SignalP to have signal peptides in the current predicted gene models. 17 of these are up-regulated at J2, 3 are constitutively expressed, 2 are up-regulated at Male life stage, 5 are up-regulated at parasitic life stages and 3 have little or no expression. The majority of these secreted sequences are expressed in J2 and/or parasitic stage nematodes, at a time when they could play a role in the interaction with the host. All of the secreted SPRY domain proteins are present in the *Globodera* specific expansion of the phylogenetic tree.

In order to determine whether more of the *G. pallida* SPRY domain containing proteins have signal peptides than predicted in genome, the upstream regions of those genes predicted not to have a signal peptide were analysed for the presence of an open reading frame in the same orientation as the gene that could encode a signal peptide. This analysis revealed that a further 19 of the SPRY proteins could have a region encoding a signal peptide upstream. However, there was little RNAseq support for the newly identified coding regions. As a control, analysis of 1000bp downstream of the genes of interest was performed and this analysis also identified 19 signal peptides. The positive and false positive rates were equal, therefore the data generated from this approach were rejected.

3.4.3 Generation and analysis of novel effector list

A bioinformatic pipeline which produced a list of genes encoding secreted proteins up-regulated at J2 or 7dpi parasitic nematodes was developed in order to identify novel candidate effectors. 1705 secreted proteins were predicted from the 16417 predicted proteins from the genome sequence. Based on RNAseq expression clustering, 526 genes were identified as being up-regulated at J2 vs. Egg and 612 genes were up-regulated at 7 dpi vs. J2. Of the secreted proteins, 161 were up-regulated at J2 Vs. egg, 129 were up-regulated at 7dpi Vs. J2 and three were up-regulated both at J2 and again at 7dpi. This gave a total of 293 genes encoding

secreted proteins up-regulated at J2 or 7dpi that could encode novel effectors (Figure 3.21).

BLASTP Vs. NR was then used to remove genes that were clearly not effectors from this list. These included genes encoding digestive proteins and/ or proteins found in *C. elegans*. This allowed 176 proteins to be removed, leaving 117 novel putative effectors. The majority of these novel putative effectors had no similarity to any known proteins except for two sequences with similarity to E3 ligases (GPLIN_000589200 and GPLIN_000271900), two sequences similar to E2 conjugating enzymes (GPLIN_001268500 and GPLIN_001318000) and one with similarity to a zinc-finger like domain (GPLIN_000713500).

The identified putative novel effectors were subjected to Pfam domain analysis (Table 3.7). This showed that 11 of the novel effectors contained 27 domains but the majority (106) had no known domains and coupled with no BLAST matches are therefore completely unknown proteins.

3.4.3.1 Expression analysis of novel effectors

The bioinformatic tool used to identify the novel effectors, required up-regulation at J2 or 7dpi but did not include expression profiles at other life stages. The expression profile of the 117 putative novel effectors were therefore analysed and clustered. The sequences clustered in the following categories, with the number of genes for each category is in brackets: Parasitic and male (31) (Figure 3.22), parasitic (4) (Figure 3.23), J2 only (28) (Figure 3.24), J2 and parasitic (46) (Figure 3.25) and J2 and male (8) (Figure 3.26).

Gene	Pfam domain	accession	Gene length	Domain Length	E-value
GPLIN_000948600	EF_hand_3	PF13202.1	137	25	9.30E-15
GPLIN_000948600	EF_hand_3	PF13202.1	137	25	9.30E-15
GPLIN_000948600	EF_hand_4	PF13405.1	137	31	8.30E-10
GPLIN_000948600	EF_hand_4	PF13405.1	137	31	8.30E-10
GPLIN_000948600	EF_hand_6	PF13833.1	137	54	1.60E-10
GPLIN_000948600	EF_hand_6	PF13833.1	137	54	1.60E-10
GPLIN_000776900	Gal-bind_lectin	PF00337.17	926	133	2.50E-13
GPLIN_000208700	Homeobox	PF00046.24	164	57	2.30E-08
GPLIN_000510600	Pkinase	PF00069.20	320	260	4.20E-57
GPLIN_001391000	Pkinase	PF00069.20	374	260	1.70E-08
GPLIN_000510600	Pkinase_Tyr	PF07714.12	320	259	7.20E-32
GPLIN_001318000	UQ_con	PF00179.21	182	140	2.70E-42
GPLIN_001268500	UQ_con	PF00179.21	305	140	1.50E-30
GPLIN_000075700	VWA	PF00092.23	195	179	4.70E-09
GPLIN_000075700	VWA_2	PF13519.1	195	172	1.50E-10
GPLIN_000713500	zf-C2H2	PF00096.21	161	23	4.30E-13
GPLIN_000713500	zf-C2H2	PF00096.21	161	23	4.30E-13
GPLIN_000713500	zf-C2H2	PF00096.21	161	23	4.30E-13
GPLIN_000713500	zf-C2H2_4	PF13894.1	161	24	2.90E-10
GPLIN_000713500	zf-C2H2_4	PF13894.1	161	24	2.90E-10
GPLIN_000713500	zf-C2H2_4	PF13894.1	161	24	2.90E-10
GPLIN_000589200	zf-C3HC4_3	PF13920.1	544	50	2.30E-13
GPLIN_000713500	zf-H2C2_2	PF13465.1	161	26	2.50E-16
GPLIN_000713500	zf-H2C2_2	PF13465.1	161	26	2.50E-16
GPLIN_000713500	zf-H2C2_2	PF13465.1	161	26	2.50E-16
GPLIN_000271900	zf-rbx1	PF12678.2	297	75	8.00E-10
GPLIN_000271900	zf-RING_2	PF13639.1	297	46	2.50E-11

Table 3.7: PFam analysis of novel effectors. Novel putative effectors identified from the genome sequenced were analysed for known domain. The identified domains are summarised in the following: EF_hand, are helix-loop-helix domains that is thought to be involved in calcium binding. Gal-bind_lectin domains are thought to be involved in binding β-galactoside. Homeobox domains are thought to bind to DNA or RNA to act as transcription factors. Pkinase are protein kinase domains that move a phosphorus group on to proteins, a process called phosphorylation. UQ_con are ubiquitin E2 conjugating domains, see 5.2.2 for an overview of the ubiquitin-proteasome pathway. VWA are known as von Willebrand factor domains that are thought to have functions in transcription, DNA repair, ribosomal and membrane transport and the proteasome. Zf are zinc finger domains that are thought to bind to DNA or RNA to act as transcription factor. Also they can bind to proteins or other small molecules to alter their binding specificity. Zf-rbx1 domains are thought to have implication in cell cycle control. Zf-RING_2 domains are thought to have roles in the ubiquitin-proteasome pathway (http://pfam.sanger.ac.uk/).

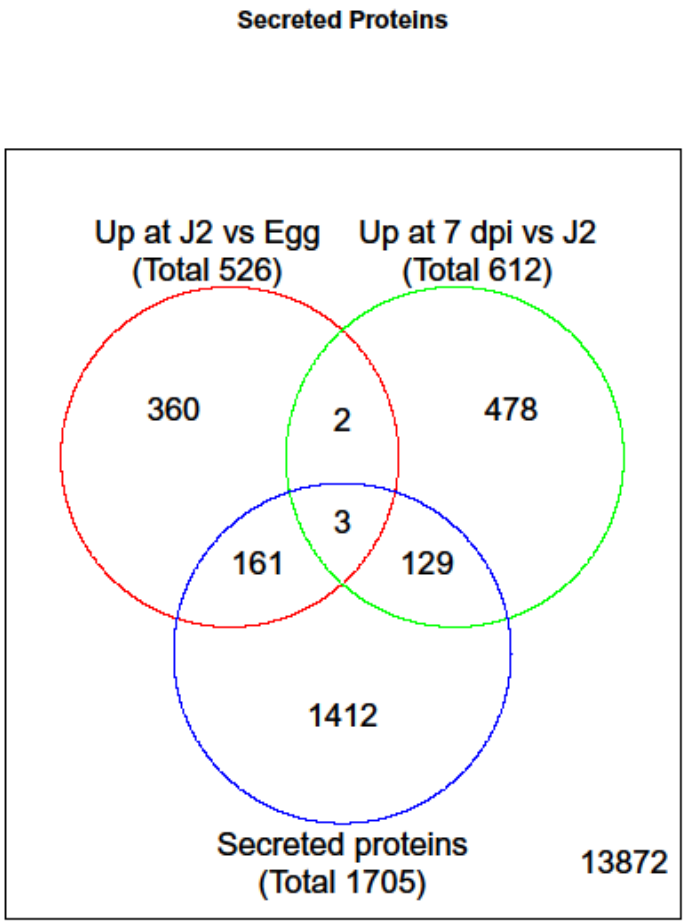


Figure 3.21: Venn-diagram showing expression categories of the putative secreted proteins. All predicted secreted proteins that had a signal peptide and did not have a transmembrane domain from the *Globodera pallida* genome were grouped according to their expression profile. The expression profile was inferred by RNAseq data. A total of 13872 predicted secretory proteins were identified. The number of proteins for each category are diagrammatically represented.

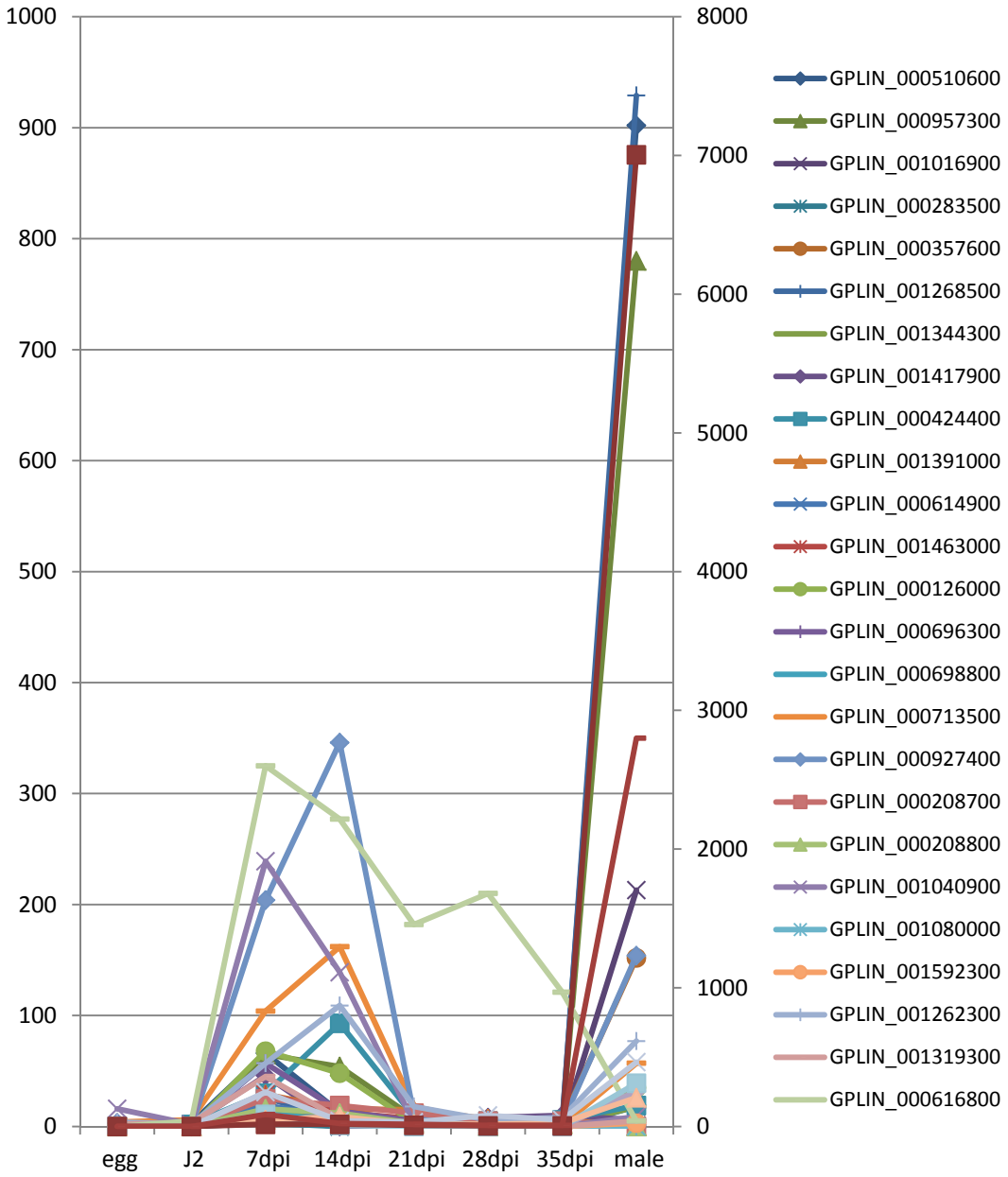


Figure 3.22: Expression profile of novel effectors which cluster in the 'parasitic and male' category, inferred by RNAseq data. GPLIN_000948600 and GPLIN_001318000 are plotted against the left Y-axis due to their high expression. Figures on the Y axis are the number of reads that map on to the gene on per base following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

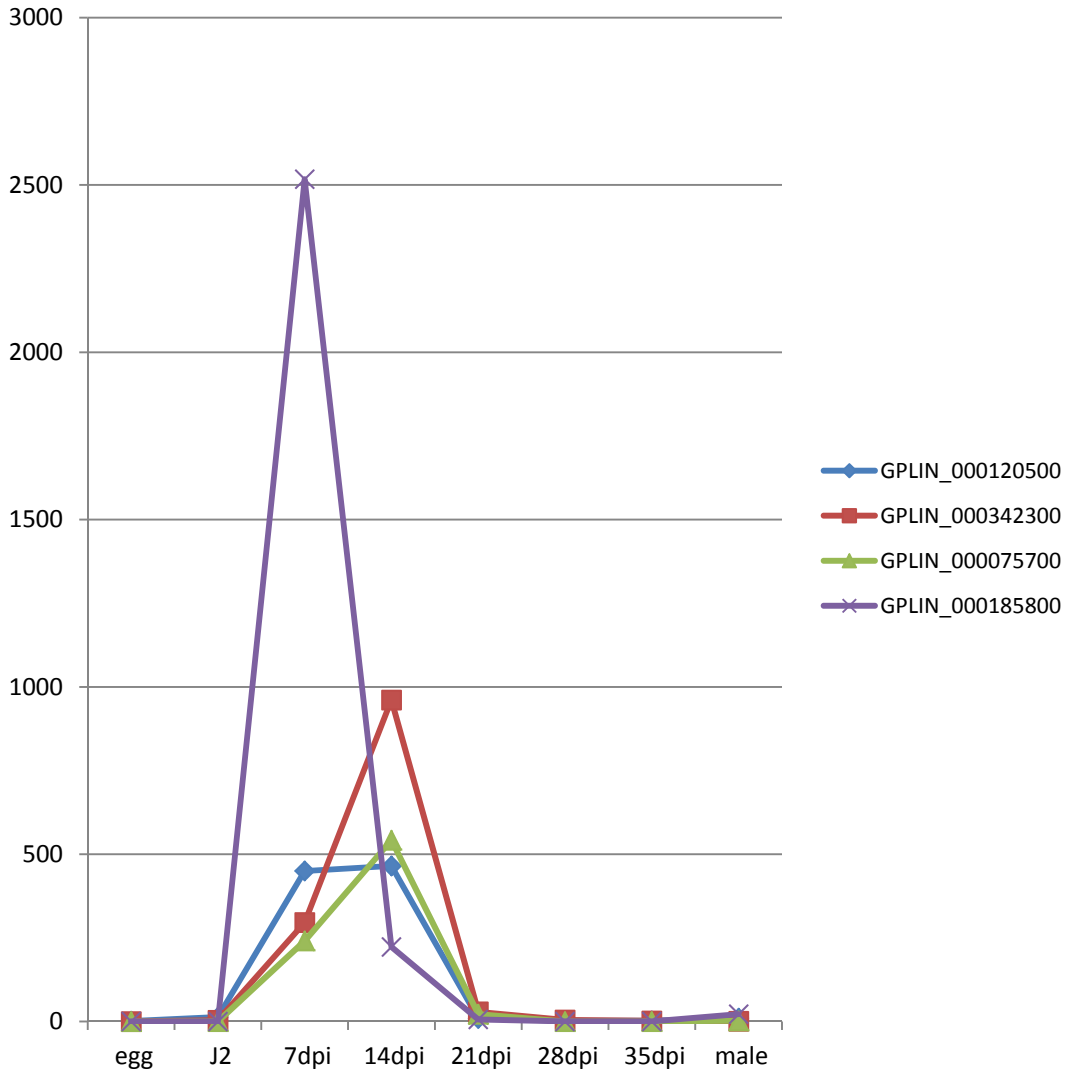


Figure 3.23: Expression profile of novel effectors which cluster in the 'parasitic' category, inferred by RNAseq data. Figures on the Y axis are the number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

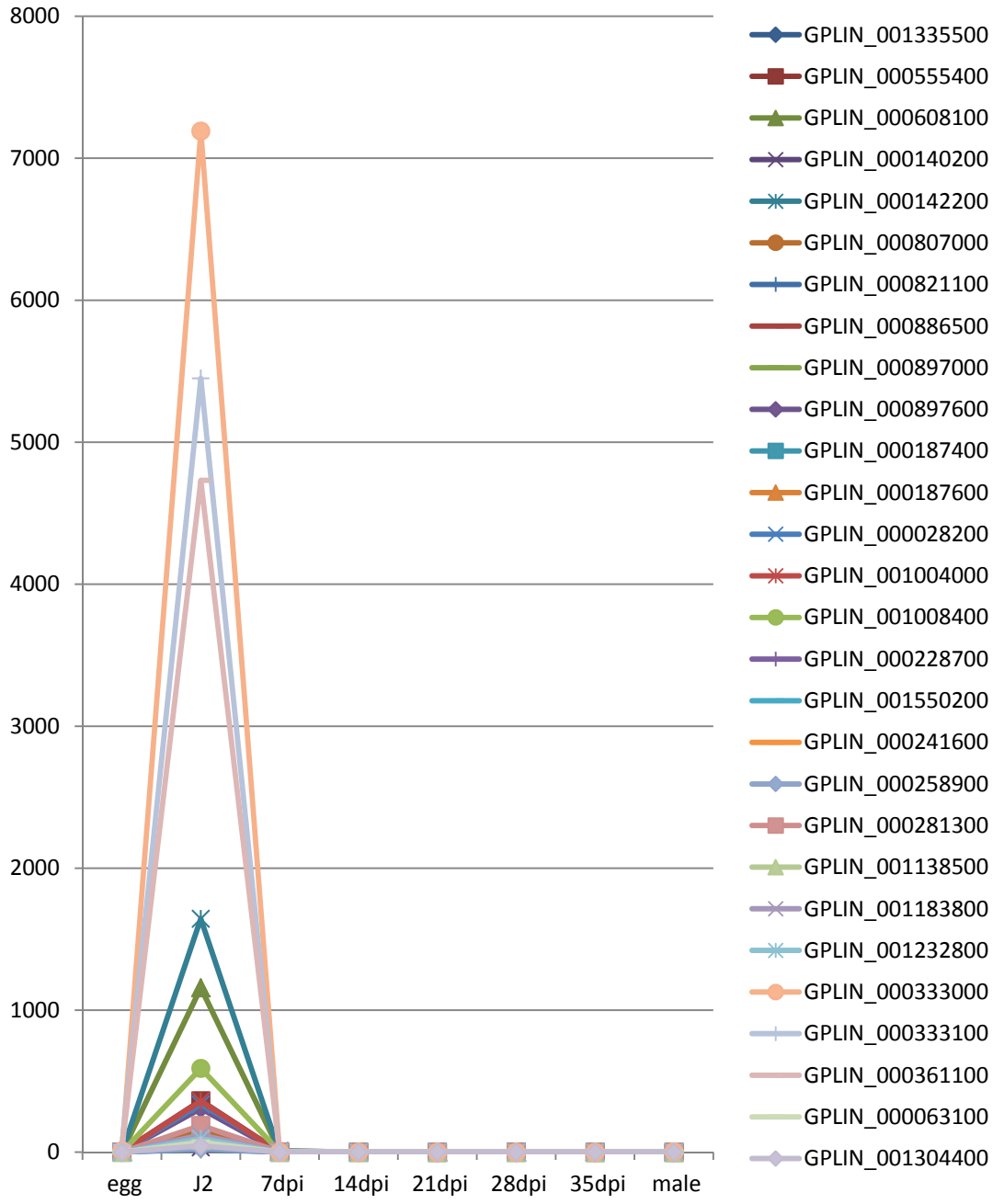


Figure 3.24: Expression profile of novel effectors which cluster in the 'J2 only' category, inferred by RNAseq data. Figures on the Y axis are the number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

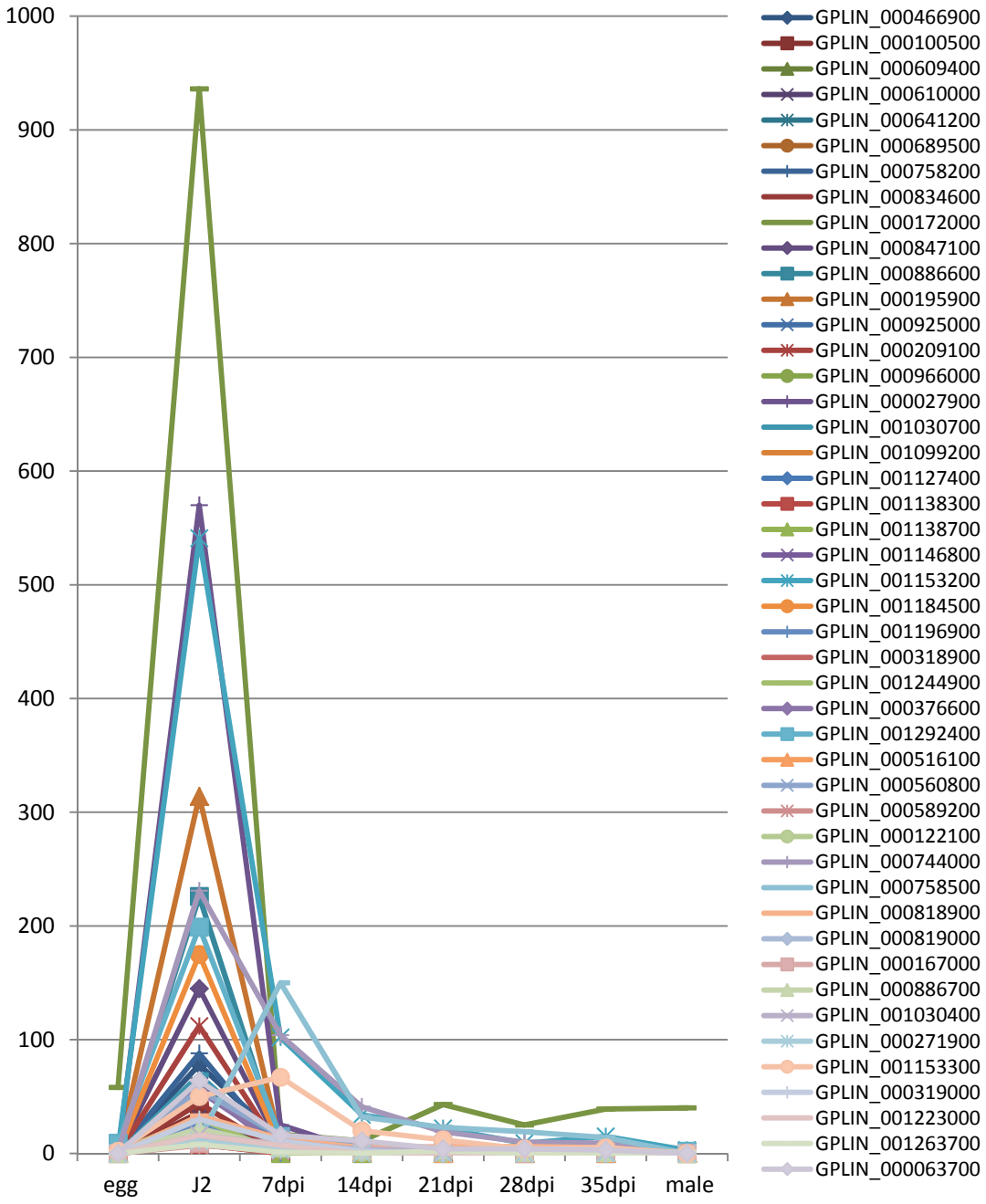


Figure 3.25: Expression profile of novel effectors which cluster in the 'J2 and parasitic' category, inferred by RNAseq data. Figures on the Y axis are the number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

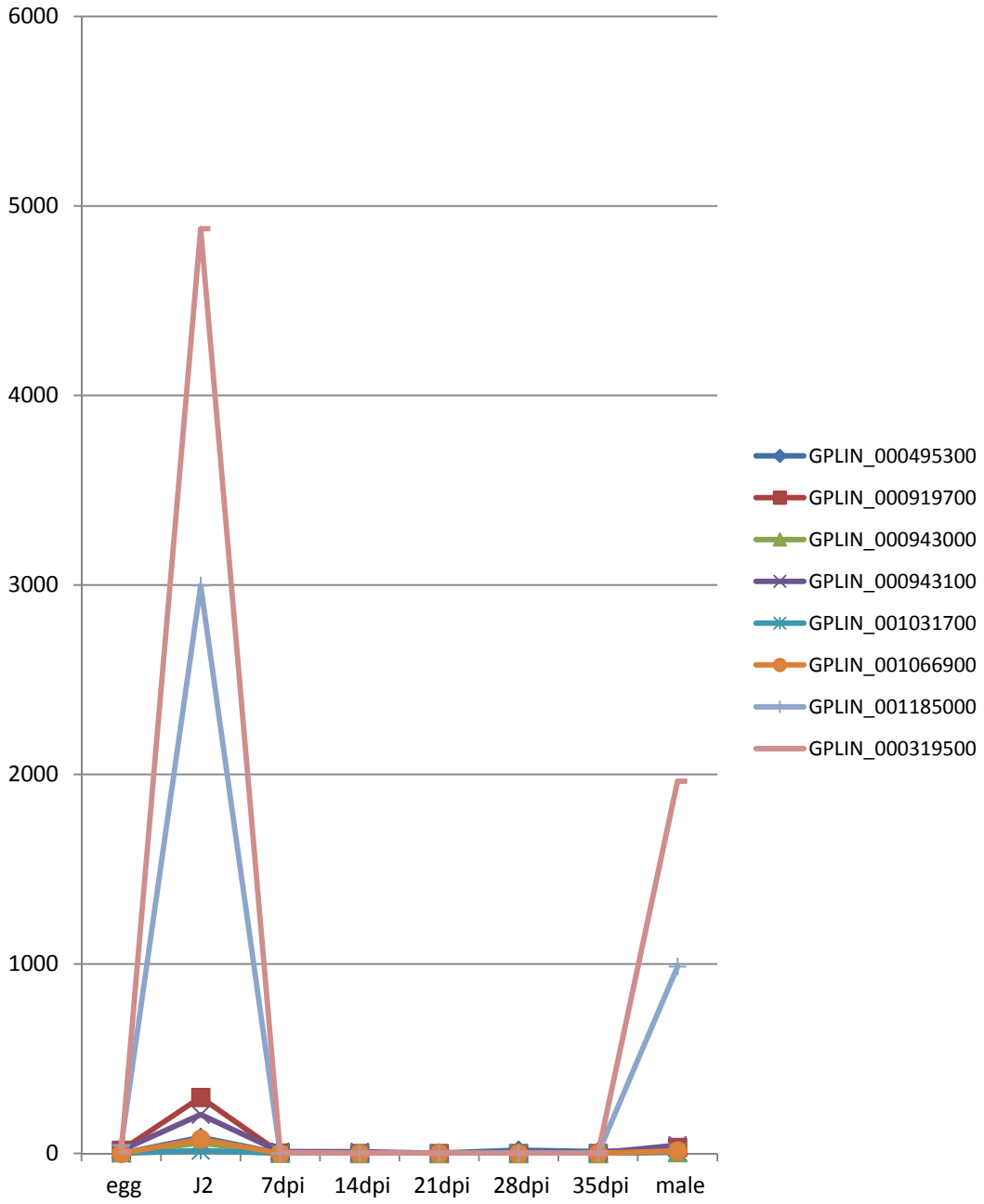


Figure 3.26: Expression profile of novel effectors which cluster in the 'J2 and male' category, inferred by RNAseq data. Figures on the Y axis are the number of reads that map on to the gene on per base, following normalisation. Figures on the x-axis are the time points at which *G. pallida* mRNA was sequenced. Each time point was duplicated (two biological replicas). The data are presented as an average of both biological replicas for each life stage/ time point.

3.4.4 Putative promoter motifs

Little is known about promoters in *G. pallida.* A 745 nucleotide region upstream of the GAPDH gene of *G. rostochiensis* was sufficient to allow constitutive expression of GFP in *C. elegans,* suggesting that the regulatory control elements for this gene are in this region (Qin *et al.*, 1998). In *C. elegans* the majority of genes are located in gene dense regions and the minimal upstream region required for appropriate expression of most transcripts is 2Kb (Okkema, 2006). We therefore analysed 2Kb of sequence upstream of each gene for motifs using MEMEs.

Analysis of upstream sequences of all genes known to be expressed in the dorsal gland cell did not reveal motifs that had conserved position for all sequences. Three motifs were identified which were present upstream of 47 of the 57 genes, though their placement was not in a conserved place in the sequences. Conserved position may not be a requirement for transcriptional regulation. Many other motifs were identified with statistical significance by the software and are shown as blocks on their sequences in Appendix 3.

Analysis of the upstream regions of all known subventral gland cell genes identified a motif that was in 32 of the 33 sequences. The placement of this motif was not conserved and appeared to be randomly placed in the 2000bp region. This was again the pattern for several other motifs which were found in the majority of the sequences. Many significant motifs were identified and are shown as blocks on their sequences in Appendix 3.

In general, it is interesting to note that there are patterns of motifs that commonly follow each other in related gene families. For example, in the dorsal gene MEMEs analysis motif 3 is often followed by motif 2. Motifs 1 or 2 are usually located just before the start of the gene. The 448 family (see 3.4.4.1) has six genes that contain motifs 8, 3, 2 then 1 as a single block. Two expansin genes have motifs 5, 1, 3, 8, 2, and then 6 as a single block. This could be due to recent gene duplication, however, there are differences in the upstream regions which can be seen by the presence/absence of other motifs. The combination of several motifs as a single block is very interesting and should be investigated further. It could be that these combinations of motifs may be a single larger motif.

3.4.4.1 448 genes family putative promoter motifs

MEMEs analysis of the upstream regions (2000bp) for all the genes in the 448 gene family (25 genes) reveals a motif (1) that is 41 nucleotides in length, common in 21 sequences and has an e-value of 10^{-149}. Motif (1) has conserved position and is located approximately 100bp before the start of each of the genes (Table 3.8). Interestingly, analysis of all predicted dorsal gland genes (this includes all 448 family members) also identified this motif in 25 sequences. A further motif (2) was identified in 11 upstream sequences and has an evalue of 10^{-83}, the motif is mainly found around 400bp upstream of the start of the gene (Appendix 3). Motif (3) was also identified in 11 members of the gene family and is mainly located around 400bp upstream of the start of the gene with an evalue of 10^{-66}. Motif (5) is found in 18 of the 25 upstream regions and is located between 300-100bp upstream of the start of the gene and has an evalue of 10^{-35}. Other motifs were identified, these are shown with their position in the sequence in Appendix 3.

3.4.4.2 CBM and expansin putative promoter motifs

MEMEs analysis of the upstream regions (2000bp) for all the genes in the CBM gene family (6 genes) reveals a motif (1) that was present upstream of all the genes (Appendix 3). The placement of this motif was not conserved in the sequences. MEMEs identified several motifs in expansin upstream regions. As mentioned above Expansin genes have motifs 5, 1, 3, 8, 2, and then 6 as a single block (Appendix 3).

3.4.4.3 IVG9 putative promoter motifs

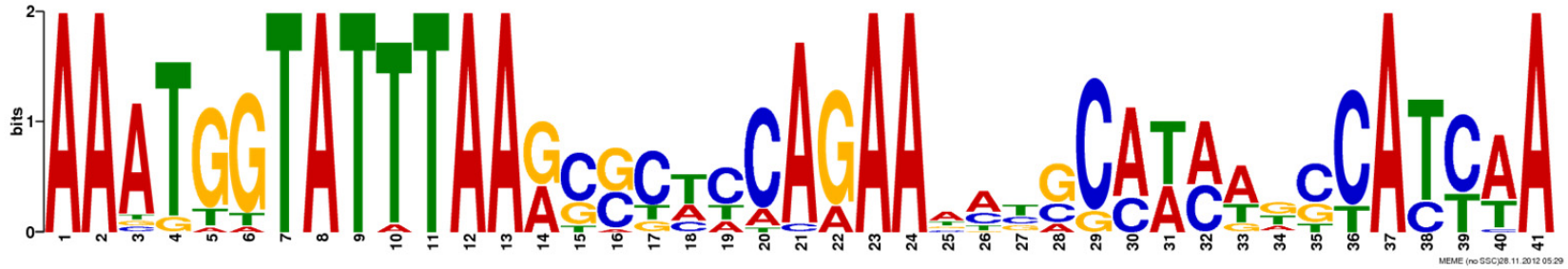
There were 2 motifs identified in all 5 family members. These were motifs 1 and motif 4 (Appendix 3). Motif 3 is found in 3 of the 4 sequences, however one sequence was not long enough to determine if this was present or not. Most upstream regions have a pattern of motif 2 closely followed by 4, which is located just before the start of the gene.

3.4.4.4 1106 putative promoter motifs

There were 2 motifs predicted to be in 6 of the 9 genes. Four of the genes have motifs 3, 4 and 2 all in series just before the start of the gene. Three of the other genes have motifs 1, 7, 2 and 5 all in series before the start of the gene.

3.4.4.5 SPRYSEC putative promoter motifs

Analysis of 16 SPRYSEC upstream regions revealed a motif (TAAGCCAGCGATTAAAGCCGTATAAAAGCGGTGGCAAATGC[AC][AG]C A[GA]AAAG) that is similar in 6 sequences and has a n evalue of 10^{-70}. The motifs in these sequences are similar and are all located within 100bp of the start of the gene. Motifs arranged as blocks on the sequences show that 4 genes have a conserved pattern: Blocks 7, 6, 5, 2, 3, 4 and two of these genes end with blocks 4 and 8.



Name	Start	*p*-value	Sites		
GPLIN_000950100	1922	7.08e-21	TGCCGTCGTC	AAATGGTATTTAAGGCCCCCAGAAACTGCATATGCCATCAA	CGAATTCCAT
GPLIN_001443600	1921	2.77e-20	TGCCGTTGTC	AAATGGTATTTAAGGCCCCCAGAATCGGCATATGCCATCAA	CCAATTCCAA
GPLIN_000243700	1921	3.37e-19	TATGGGCACT	AAATGGTATTTAAGGCCCTCAGAATCGGCATATGCCATCAA	CCAATTCCAT
GPLIN_000203300	1924	9.34e-19	TCTGACCATC	AAATGGTATTTAAACGCTCCAGAACATCCATCAGGCATTTA	AATCACTCAG
GPLIN_000792900	1924	9.34e-19	TGATCTTATC	AAATGGTATTTAAACGCTCCAGAACATCCATCAGGCATTTA	AATCACTCAT
GPLIN_000481100	1923	9.34e-19	TCTGATCATC	AAATGGTATTTAAACGCTCCAGAACATCCATCAGGCATTTA	AATCACTCAG
GPLIN_000388900	1920	2.75e-18	TCGCTCGTCA	AAATGGTATTTAAGCGTAACAGAAAATGCCAAATCCATCAA	CCAACAAAAA
GPLIN_001255700	1923	5.87e-18	TCTCATTATC	AAATGGTATTTAAACGCTCCAGAAAACCCAACAGTCACTTA	ATTTCATTTA
GPLIN_001390400	1921	8.44e-18	TATGGGCACT	AAATGGTATTTAAGGCTTTCAGAATCGGCATATTCCATCAA	CCAATTCCAT
GPLIN_001471200	1902	1.52e-17	TCGCTCGTCA	AAATGGTATTTAAGCGTAACAGAAAATGCCAAATCCATTAA	CCAACAAAAA
GPLIN_000309000	1921	3.32e-17	TGCCGTCGTC	AAATGGTATTTAAGGCCCCTAGAATTGGCATATGCCATCAA	CCGATTCCAT

Name	Start	*p*-value	Sites		
GPLIN_001431400	1900	7.70e-17	CGCTTGTCAA	AAGTGGTATTTAAGCGCATCAAAAAATGCCAAATCCACCAA	CCAAATAAAT
GPLIN_000060800	1910	1.40e-16	TCGCTTGTCA	AAATGTTATTTAAGCGCAACAGAAAGTGCCAAAACCATCAA	CCAAATAAAT
GPLIN_001162100	1922	7.97e-16	TCTCATCATC	AAATGGTATTTAAATGCTCCAGAAAACCCAACAGTTACTTA	AATTTCACTT
GPLIN_000860700	1925	1.68e-15	TCTCATTATC	AAATGGTATTTAAATGCTCCAGAAAACCCAACATTTACTTA	AATTACTTAG
GPLIN_001038900	1905	1.82e-15	TCACTCGTCA	AAATGATATTTAAGCGCATCCGAAAATGCCAAATCTATCAA	CCAAATAAAT
GPLIN_001221800	1913	9.37e-14	CACTCATAAA	AACGGGTATTTAAGCACAACAGAAAAAGCCAAACCTACCAA	CCAACTAAAT
GPLIN_000970100	1919	1.17e-13	ATAAAAGAAA	AAATTGTATTTAAACCGTCAAAAAGTCAGATCGAGCATCAA	TCCAATTCAT
GPLIN_000970000	1919	1.17e-13	ATAAAAGAAA	AAATTGTATTTAAACCGTCAAAAAGTCAGATCGAGCATCAA	TCCAATTCAT
GPLIN_000912100	1915	1.89e-13	CACTAAAGAA	AAATTGTATTTAAACCGTCAAAAAGCCAGATCGAGCATCCA	TTAATCCAAT
GPLIN_001582700	672	5.62e-13	GCCGTCGTCA	AATGATTATATAAGGCTTTCAGAATCGGCATATTCCATCAA	CCAATTCCAT

Table 3.8: MEMEs analysis of 2000bp upstream of all 448 effector family members from *Globodera pallida*. The motif above the table and highlighted in coloured text in the table was found in 21 of the sequences around 100bp before the start of the gene.

3.5 Discussion

3.5.1 Effectors

The analysis of the *G. pallida* genome has provided a useful basis for the identification of putative parasitism genes. Effectors were identified that have been previously reported as specific to *G. pallida* (IVG9 and IA7), as well as a large number (129) of *G. pallida* orthologues of effector genes from other plant parasitic nematodes, 117 novel putative effectors and 53 genes thought to be involved in cell wall modification.

Although many *G. pallida* orthologues of effectors from other species were identified, there were many sequences that appeared not to be present in *G. pallida* that have been identified in other species. Although it is entirely possible that effector gene families evolve rapidly and these effectors are genuinely absent, caution should be taken when interpreting these results. Some genes may be present but not called by gene finding software. For example, BLAST searches identified a sequence on a scaffold similar to A42 that has not been predicted in the v1.0 gene models. However, we carried out a BLAST search of the DNA sequence of the genome assembly for each effector that was absent from the list of orthologues and were able to identify sequences that could possibly encode missing effectors. Effectors may also be absent due to failures in the assembly process. For example, the "66" gene family is entirely absent from the current assembly but was present in previous versions of the genome assembly. This gene family contains extensive repeated "GGGGYGGGGY" regions, and is also repetitive at the nucleotide level. Such regions are computationally difficult for assembly software to resolve, particularly when using short read lengths, and this could account for its absence in the current genome assembly. A42 and "66" have been successfully cloned from *G. pallida* cDNA proving that the genes are present and expressed. This highlights a common view amongst bioinformaticians that a genome sequence is a prediction which is always in constant refinement.

Several *H. glycines* effectors were not identified in the *G. pallida* genome. The functionally characterised effectors Hg30C02 and 10A06 were not identified in the *G. pallida* genome. If this is true then it could mean that *G. pallida* either does not

need these effectors to function in the manner which they do in *Heterodera,* to successfully parasitize its host, or that *G. pallida* may use an alternative effector that has a similar function. The absent effectors may have been lost through negative selection via host recognition.

3.5.1.1 *G. pallida* orthologues of previously described and characterised effectors

In many cases putative orthologous effectors from other cyst nematodes were identified, that have previously been functionally characterised such as CLAVATA (CLE), SPRYSECs (for example RBP1), 1106, 19C07 and ubiquitin extension protein. Although the sequences are similar experimental validation of predicted function is still required.

3.5.1.2 Cyst nematode effectorome vs. RKN effectorome

Comparison between the identified putative RKN effectors and the effectorome of *G. pallida* revealed that these pathogens share only 4 common effectors (excluding cell wall modifying enzymes). One of these common effectors, GPLIN_001475500 is similar to *M. incognita* gland cell protein 28. GPLIN_001475500 is up-regulated at J2 and male, suggesting a role in migration. In support of this, a similar effector has been identified in migratory endoparasitic nematodes. Chorismate mutase is also common between *G. pallida,* RKN and *P. coffeae.* Therefore chorismate mutase may have a role in suppressing host defences during infection by a range of nematodes. Cell wall modifying enzymes (see below) were present in both *M. incognita* and *G. pallida*. This reflects a crucial role for cell wall modifying enzymes in all plant-parasitic nematodes and implies acquisition from bacteria or fungi before divergence of Clade 12 nematodes (Haegeman *et al.*, 2011a). Since some cell wall modifying enzymes are present in *M. incognita* and not in *G. pallida,* and vice versa, this suggests that other horizontal gene transfer events have occurred since divergence of the ancestral species in Clade 12. There are differences in the numbers of genes in the families of the cell wall modifying enzymes between *M. incognita* and *G. pallida,* which reflects duplication following specialisation for host parasitism.

3.5.1.3 Gene families

It is interesting to note that some cyst nematode effectors exist in gene families, for example, DGL1, 448, 1106, 747 and SPRYSECs, which is not the case for RKN

effectors. Genes that exist as families could have different functions within the host and/or they could target different variations of the same target within the host. However, the RKN effectors are adequate to successfully perform their function in a variety of hosts. *G. pallida* may duplicate their effectors in families as a way of introducing variation to avoid host recognition or/and evolve more efficient versions of the effector. In addition, *G. pallida* replicates sexually whereas RKN reproduces via asexual mitosis. Therefore the existence of gene families in *G. pallida* may be related to the reproductive strategy of this nematode.

An effector family called 448, which are mainly expressed during parasitic life stages was analysed for any motifs common within the effector family. A MEME analysis identified a motif "PCCPGSQQVAALM**"** that is present in 27 of the 29 family members and is located between 18-30 amino acids from the start of the gene. Conserved regions within an effector family could be used as a target for RNAi. The targeting of this motif could in theory down-regulate 27 effector genes at once. If the effector family is essential for the virulence of the nematode, then this target could prove durable over time as the loss of an entire effector family could be difficult to overcome. The effector family is also present in other cyst nematodes (*H. glycines* and *G. rostochiensis*), therefore if this control method proved successful, it could be used as a specific and highly directed method to control cyst nematodes in a number of hosts. The presence of this motif may reflect recent duplication within the gene family or it may be a domain with an as yet uncharacterised but important functional role.

3.5.1.4 SPRY domain containing proteins and SPRYSECs

A large family (299) of SPRY domain containing proteins was identified in the *G. pallida* genome. By contrast, *C. elegans* has 8 SPRY domain proteins while *B. xylophilus* and *M. incognita* have 12 and 27 SPRY domain containing proteins respectively. Phylogenetic analysis showed that this family of proteins is hugely expanded in *Globodera,* where a large *Globodera* specific clade can be seen. *G. pallida* SPRY domain containing proteins that have reciprocal BLAST matches with SPRY domain containing proteins from other PPN species nested together outside of the *Globodera* specific expansion during phylogenetic analysis, suggesting that these proteins could have conserved or similar function within nematodes. The B30.2/

SPRY domain has yet unknown function, however it may be important for protein:protein interactions.

Expression analysis based on RNAseq data shows that 115 *G. pallida* SPRY domain containing proteins were not expressed. 7 were expressed in "egg only", 23 were expressed in "Egg and J2" life stages, 36 genes were up-regulated at "J2", 41 genes show constitutive expression, 65 were up-regulated in parasitic stages and 12 were expressed in "male only". Data obtained using qPCR support the expression profiles observed via RNAseq analysis. The large number of non-expressed SPRY domain containing proteins was unexpected. It would be interesting to repeat the RNAseq analysis on *G. pallida* following parasitic interactions with other hosts, for example tomato or wild Solanaceous species. It may be that these genes are not required for parasitism of *Solanum tuberosum* L. cv. 'Désirée', but are used in other hosts. It is also possible that other populations/pathotypes of *G. pallida* express a different subset of the SPRY domain proteins and the RNAseq data from other *G. pallida* populations will make it possible to investigate this in future. The situation for SPRY domain proteins in *G. pallida* reflects that observed for the RXLR gene family in *P. infestans*. A recent analysis of RXLRs from a range of *P. infestans* isolates showed that just 45 of the 500 RXLRs present in the genome of this species are expressed in all isolates, with other RXLRs expressed in a strain-specific manner (D. Cooke, JHI, pers. comm.).

Out of the 299 SPRY domain containing proteins from *G. pallida* only 30 (10% of the gene family) are predicted by SignalP to have signal peptides. 17 of these are up-regulated at J2, 3 are constitutively expressed, 2 are up-regulated at Male life stage, 5 are up-regulated at parasitic life stages and 3 have little or no expression. The majority of these secreted sequences are expressed in J2 and/or parasitic stage nematodes, at a time when they could play a role in the interaction with the host. This makes sense as some SPRYSECs suppress host defences (Postma *et al.*, 2012; Sacco *et al.*, 2009a) and the nematode needs to suppress host defences in order to establish and maintain a successful feeding site.

Although other plant parasitic nematodes have SPRY domain containing proteins, SPRYSECs are only present in *G. pallida, G. rostochiensis* and *G. mexicana.* A SPRYSEC has been identified in *H. glycines* (John Jones pers. comm.). However,

when the thresholds applied in this work are used analyse this gene, it does not satisfy the thresholds and is rejected as being a SPRYSEC. No SPRY domain containing protein with a signal peptide was identified in any other nematode species, except other *Globodera* species (this includes analysis of all Nembase4 sequences: http://www.nematodes.org/downloads/databases/NEMBASE4/index.shtml). Therefore SPRYSECs are likely to be a *Globodera* specific trait.

The huge expansion of SPRY domain containing proteins in *G. pallida* and that only 10% of these have predicted signal peptides is difficult to explain. Despite repeated attempts to identify more putative signal peptides from transcriptome, genome and *de novo* transcriptome assembly data, no significant increase in the number of SPRYSECs was found. It could be that a signal peptide exon is spliced on to SPRY domain containing genes. Or it may be that the SPRY domain family is still in the process of expanding. The huge expansion of the family of genes could be a way of introducing variability by rapid duplication and diversification as a way of speeding up evolution. SPRYSECs are known to interact with the host, RBP1 is a known AVR gene. Therefore the gene family could be subjected to high selection pressure from its host. Members of the SPRY gene family may even be required for successful parasitic interaction with different hosts. This also opens up an interesting question, how do genes gain signal peptides to become effectors? There may have been other SPRYSECs in the past that could have been selected against due to natural selection through their interaction with the host, which could account for the large number of non-expressed SPRY genes. One other possibility is that *G. pallida* was sequenced from a population of nematodes, as clonal lines were impossible to obtain, therefore, could the large number of genes found be an artefact due to different alleles found and mis-assembled by the assembly software? As already mentioned, limited evidence implicates the B30.2/ SPRY domain domain(s) as a protein-interacting module which can have a wide range of functions and therefore be involved in many different cellular processes. Taking this into account it could also be possible that not all of the identified SPRY genes are ever going to act or function as effectors.

3.5.1.5 Novel effectors

117 putative novel effectors were identified from the *G. pallida* genome sequence. 11 of these genes had 27 known PFam domains, leaving 106 genes encoding completely unknown proteins. The 27 known domains include domains associated

with transcription factors (one up-regulated at J2 and the others up-regulated at parasitic life stages), calcium binding structures (up-regulated at parasitic stages), ubiquitin E2 (up-regulated at early parasitic stages and male) and E3 enzymes (up-regulated at J2), protein kinases (up-regulated at parasitic stages) and domains thought to interfere with the cell cycle (this domain containing gene is up-regulated at J2). The identification of novel effectors with known domains makes them highly interesting targets for future functional analysis. If they are found to be vital for successful parasitism of *G. pallida,* the absence of these genes in the rest of the known sequence database to date could make them good targets for highly directed control strategies.

3.5.2 HGT

3.5.2.1 Cell wall modifying enzymes

Horizontal gene transfer is a rare event. Previous studies have shown a large number of cell wall modifying genes in PPN genomes thought to have been acquired from bacteria or fungi. For example, the *M. incognita* genome includes 90 genes thought to have arisen due to horizontal gene transfer, *M. hapla* 41 and *B. xylophilus* 41. Analysis of the *G. pallida* genome reveals 53 cell wall modifying genes including 16 GH5 β-1,4 endoglucanases, 7 PL3 Pectate lyases, 1 GH43 Arabinase, 9 expansin, 6 CBM and 2 GH53 arabinogalactan endo-1,4-β-galactosidases. All of these enzymes have been previously identified in other PPN, although the only other known occurrence of an arabinogalactan endo-1,4-β-galactosidase is from the closely related cyst nematode *H. glycines* (Vanholme et al., 2009) and the migratory nematode *P. coffeae* (Haegeman *et al.*, 2011b).

The plant cell wall is a complex structure which has to be overcome for an invading parasite to be successful. The range and number of genes encoding enzymes for this purpose demonstrate the complexity of the cell wall. The invading J2 nematode uses its cell wall modifying enzymes as it migrates intracellularly through the zone of elongation to a site near the vascular tissue. Once here, a cell is then selected to form the initial syncytial cell. It is also possible that the nematode will use its cell wall modifying enzymes to manipulate the host structure, to help remodel the cell wall during formation of the syncytium. It was previously thought that plant parasitic

nematodes induce expression of the plant's own cell wall degrading machinery in the syncytium to form and modify the feeding site.

Many cell wall modifying enzymes expression profiles suggest a role in migration (up-regulated in J2 and/or male). However, some genes are up-regulated in parasitic stages, which suggest a role in modification of the syncytial cell wall, for example arabinogalactan and glycoside hydrolase. CBM have two roles in the host. Some CBMs are involved in migration and *G. pallida* has a CBM gene which is up-regulated at J2 and male life stages, consistent with this role. However, one *H. glycines* CBM interacts with a host pectin methylesterase to regulate feeding cell growth and expansion (Hewezi *et al.*, 2008) and *G. pallida* CBMs were identified that are upregulated in parasitic stages that may be the functional analogues of this CBM.

The nematode metabolises arabinogalactan by using enzymes with GH53 and GH43 domains. Interestingly the gene encoding GH43 is up-regulated in migratory and early parasitic life stages, suggesting a role in both migration and syncytium formation. There are 2 genes encoding a GH53 domain, one of which has expression at J2 but a greater expression profile in parasitic stages, suggesting a role in the syncytium. The other has constitutive expression with a peak in male, suggesting both a role in syncytium formation and development and also migration.

3.5.2.2 Other HGT candidates

The *G. pallida* genome contains two predicted chorismate mutase genes. One of these is highly up regulated at J2, with little expression in parasitic stages and males, and another that is expressed to some degree at all life stages but is up-regulated at 7 and 14 dpi parasitic stages. Previous work has shown that chorismate mutase is expressed in J2 in *G. pallida* (Jones *et al.*, 2003), but has also been found to be expressed in parasitic stages in *M. javanica* (Painter and Lambert, 2003). Chorismate mutase has been suggested to have a role in the suppression of host defences. It has also been suggested to have a role in giant cell formation. However, the effector's presence in *G. pallida,* RKN and migratory nematodes suggests a common function between these nematodes. It is also possible that the two chorismate mutases present in *G. pallida* (and by extension other nematodes) have different functional roles. The chorismate mutase expressed during migration and common in root knot nematodes

and migratory nematodes may suppress host defences induced during migration while the other, expressed at parasitic stages, may have another role related to biotrophy.

Genes that are involved in detoxification such as cyanate lyase and genes involved in nutrient synthesis were also identified. Genes involved in the synthesis of vitamin B6 have been identified in *H. glycines* and *G. rostochiensis* (Craig *et al.*, 2008) and RKN (Craig *et al.*, 2009). These sequences were also identified in the *G. pallida* genome. Two of the sequences are located next to each other on the same contig and have identical expression profiles. It is thought that nematodes may need to synthesise their own vitamin B6 as the host may restrict the availability of this essential vitamin as a defence response (Craig *et al.*, 2009). These genes may have been acquired from soil borne bacteria or fungi that have lived in close proximity to nematodes. The mechanism of horizontal gene transfer is unknown and extremely rare.

3.5.3 Promoter prediction

Analysis of upstream regions of genes, previously experimentally shown to be expressed in the dorsal, and separately the subventral, glands did not reveal any obvious motifs that may be gland specific. Therefore it may be necessary to repeat the analysis on a longer upstream sequence, instead of 2000bp to identify putative gland specific motifs. The identification of gland cell specific motifs could be used for the prediction of novel effectors and their expression patterns. *In situ* hybridisation would be needed to confirm that the bioinformatic spatial expression predictions are correct. Such a motif could be included in the bioinformatic analysis pipeline of effectors that could highlight candidates, which have already been shown to have a signal peptide and no transmembrane domain, for further characterisation. Experimental analysis of effectors is expensive in terms of time therefore better predictions for the identification of effectors that are expressed in the gland cell would be highly advantageous. Also, these motifs and their transcription factors could be effective targets for biotechnological control of the pathogen by interfering with the expression of a subset of effector genes.

The identification of gland cell specific motif was an ambitious task. However, analysis of specific gene families revealed some very interesting motifs and in

several cases motifs were found within a few hundred base pairs of the start of the gene for several members of the gene family. Also, motifs and clusters of motifs (that may form one long motif) were found that occur in several genes in a gene family's upstream regions. Motifs contained imperfect palindromic sequences, which are thought to be involved with interactions with transcription factors (Thompson *et al.*, 2003). These could be excellent candidates to functionally characterise although the absence of a transformation system for any plant nematode would mean that a heterologous system would need to be used for these experiments. The MEMEs analysis could be taken further by clustering genes with similar expression profiles and analysing the upstream regions to see if there are any motifs that could be associated with specific temporal expression profiles.

As mentioned above, palindromic sequences were identified in the motifs. Palindromic sequences are of particular interest as they can interact and bind with transcription factors and may therefore be involved in transcriptional regulation (Thompson *et al.*, 2003). A full genome analysis to identify palindromes could be applied to determine their presence, and once identified, sequence similarity and cluster analysis could be performed. An example of this kind of analysis was performed on bacterial genomes in a project undertaken in another laboratory at the start of this PhD programme (Appendix 3). However, due to computational limitations (the *G. pallida* genome size is over 130Mb while the bacterial genomes analysed were each approximately 6Mb) this has not been performed on any nematode genome to date. Identified palindromes that are common in subsets of effectors could be targets of high interest for further characterisation.

Summary

- The *G. pallida* genome assembly contains 129 genes that are orthologues of effectors from other cyst nematodes.
- 117 novel putative effectors were identified. Some of these have recognised PFam domains such as transcription factors, E2 and E3.
- 53 cell wall modifying genes were identified, some of which are up-regulated at parasitic stages.
- Other horizontal gene transfer candidates were identified (chorismate mutase, cyanate lyase, vitamin B6 operon).

- *G. pallida* effectors are present in large gene families, in contrast to *M. incognita* (RKN) effectors.
- There were 4 effectors in common between *G. pallida* and *M. incognita.*
- A large SPRY domain-containing gene family was identified of which 30 genes may encode secreted effectors.

4 Transgenic over-expression of effector genes *in planta*

4.1 Introduction

Developing a better understanding of the mode of action of effectors is a strong focus for the plant pathology community. As well as providing information about how pathogens infect plants, this information can also lead to a better understanding of how plants defend themselves against pathogens. In addition, functionally characterised effectors, if essential for the pathogen, could be used as targets for alternative control strategies. Various experimental procedures are used to help decipher the mode of action of effectors.

Effectors of the potato cyst nematode *Globodera pallida* are synthesised in the dorsal and subventral gland cells, which together are known as the oesophageal glands (Gheysen and Jones, 2006). While bioinformatic approaches can provide many hundreds of candidate effectors, these can only be considered as true effectors if they are shown to be expressed in the oesophageal gland cells using *in situ* hybridisation. Once a protein is verified as an effector, a wide range of techniques can be used to study its function in more detail. The host target of the effector can be sought using yeast-two hybrid analysis followed by split-YFP assays to confirm that interactions detected in yeast can also occur in the plant (Lee *et al.*, 2011). The gene in question can be down-regulated by RNAi to determine how important its presence is for the nematode to establish a successful infection (Bakhetia *et al.*, 2008). If an effector is essential for a successful parasitic interaction, then this effector could prove to be an excellent target for control. The temporal expression profile of an effector can also give clues about its potential role. For example, effectors expressed only at later parasitic stages are unlikely to be involved in motility or establishment of the feeding site. In contrast, effectors expressed at pre-parasitic J2 could be involved in root invasion, feeding site induction and/or suppression of host defences. qPCR and RNAseq can be used to determine the expression profile of the effector (Hewezi *et al.*, 2010b). Confocal microscopy analysis of transiently expressed effectors linked to fluorescent proteins can be used to determine the sub-cellular localisation of the fusion protein once in a plant cell; such information can give clues about the cellular targets of the effector (Jones *et al.*, 2009b). Transient expression of an effector, followed by, or in combination with transient expression of a known inducer of plant defence

pathways can be used to investigate whether or not the effector interferes with host defences (Postma *et al.*, 2012). However, transient expression as an experimental procedure has some drawbacks as well as some advantages.

One of the advantages of transient expression is that it is relatively easy to transiently express an effector. Transient expression assays can therefore be considered as amenable to high throughput analysis and also permit many replications. Gene(s) which may be lethal that could not be characterised via stable transformation can be investigated using a transient approach, although care needs to be applied in analysing the phenotypes induced by such genes. Several genes can be transiently expressed at once allowing the opportunity for the analysis of the combinatorial effect of multiple effectors, or the consequences of expressing an effector in combination with host targets. Insertion effects are likely to be minimised in transient assays as each cell is transformed in a separate transformation event (Rico *et al.*, 2010).

However, transient expression also has some drawbacks. Transient expression only lasts for around 5-7 days and long term effects cannot be assessed. The phenotype of transient expression is local to the area subjected to infiltration, which can be an advantage depending on what the investigator wants to analyse, but in order to determine the effect the effector has on a whole plant stable transformation is required. Infiltration assays cannot be performed on the root system and therefore cannot be used to analyse phenotypes in terms of nematode infection rates. The proteome may be different in the root system compared to the leaf system and this needs to be considered when studying the effector of root parasite effectors in the leaf. Transient expression requires the use, and therefore presence of, *Agrobacterium* which could interfere with plant processes such as PTI, ABA and salicylic acid production leading to misleading results when analysing defence signalling pathways (Rico *et al.*, 2010). Leaves can be easily damaged during the process which may induce plant defence pathways, also potentially interfering with the results.

Experiments using transient expression assays are highly focused in answering specific questions regarding the effector. Another approach, which can be combined with those above, is to stably transform the host and/or a model plant to determine how the pathogen protein interacts with and alters host cellular processes. Microarray analysis can then be used to determine if the presence of the effector protein alters the

expression of host genes (Hauck *et al.*, 2003a). Analysis of differentially expressed genes as a result of the presence of an effector can contribute to the formation of a hypothesis for the functional role of the effector. To complement yeast-two hybrid and split-YFP assays and to provide further evidence for the interaction of two proteins, pull-down assays can be used (Mersmann *et al.*, 2008). Over-expression in a stably transformed line may result in hyper-susceptibility, which can be investigated by nematode infection or inoculation with other plant-pathogens (Hewezi *et al.*, 2010b). Data from phenotypic analysis can implicate interactions with certain host pathways or developmental processes to account for the results observed.

4.1.1 Transgenic expression for functional analysis

Transgenic expression of nematode effectors *in planta* has been used to gain an insight into their potential function by observing the phenotypic effects that occur as a result of over-expression. *Arabidopsis* and *H. schachtii* are frequently used as a model plant–parasitic nematode system in experiments using this approach (Sijmons *et al.*, 2008). For example, over-expression of 10A06 from *H. schachtii* in *Arabidopsis* resulted in plants that were hyper-susceptible to nematode infection, that produced a greater number of leaves, longer roots and flowered later when compared to the control. The transgenic plants had significantly lower *PR-1* gene expression compared to wild type plants, which could account for their increased susceptibility to nematode infection. The over-expression of 10A06 increased the levels of Spermidine Synthase 2 which may have reduced the production of SA accounting for the increased susceptibility. Reduced levels of *PR* gene expression were also recorded in these lines (Hewezi *et al.*, 2010b). Another *H. schachtii* effector, Hs4F01, was also over-expressed in *Arabidopsis* and resulted in hyper-susceptibility to nematode infection (Patel *et al.*, 2010b). In contrast over expression of HS10C07 in *Arabidopsis* was correlated with a decrease in nematode infection. These transgenic lines also produced lateral roots 24 h earlier than control lines (Lee *et al.*, 2011). These studies illustrate the value of using transgenic lines over-expressing effectors in the whole plant.

Arabidopsis has many advantages as a model host for the study of plant-nematode interactions. The speed, ease and economy with which transformants can be produced are far greater compared to crop plants. The *Arabidopsis* genome has been

sequenced (*Arabidopsis*, 2000) and is well annotated (Castelli *et al.*, 2004). A large number of mutants are available which can be used for further experimentation (Alonso *et al.*, 2003). Microarray chips can be used to identify any changes to gene expression due to transgene over-expression (Zimmermann *et al.*, 2004). The resources available for *Arabidopsis* mean that it will continue to be used as a model for plant–nematode interactions and many other fields. However, *Arabidopsis* is not a commercially important crop and questions have been raised about how good a model it is for crop species (Muller and Tester, 2007). For example, an effector that has evolved to interact with potato proteins may not have the same function in *Arabidopsis*. Therefore, where possible, transgenic over-expression of effectors from species that do not infect *Arabidopsis* needs to be performed in the host. Transgenic over-expression in the pathogen's host presents many difficulties. Transformation of crop species can be laborious and limited numbers of lines may be produced due to low transformation efficiency of these plants. Hairy root cultures are frequently used to overcome this limitation. For example, a *Meloidogyne javanica* chorismate mutase gene was expressed in soybean hairy roots, which altered root formation and development of the vascular system. The observed phenotype was recovered by the exogenous application of IAA, suggesting the transgenic plants were auxin deficient (Doyle and Lambert, 2003). A root-knot nematode effector 16D10 was expressed in tobacco h airy roots, which resulted in cell proliferation with normal differentiation. 16D10 is thought to interact with two SCARECROW transcription factor family members (Huang *et al.*, 2006b). This shows that hairy root cultures can be a useful alternative to *Arabidopsis* in the study of effectors, although the inherent variability of nematode reproduction in this system needs to be considered (Plovie *et al.*, 2003). Further barriers may be encountered when using the host for the study of effector over-expression as genomes of a number of crop species are either not published or the annotation is substandard in comparison to *Arabidopsis.* However, characterisation of over-expression in the host is needed to confirm/reject findings found in *Arabidopsis* over-expression studies.

4.1.2 Phenotyping of growth characteristics

The experiments mentioned above are all considered low-throughput and therefore require relatively large amounts of labour to perform. Phenotyping experiments can be relatively variable and therefore large numbers of replicate plants must be

analysed to ensure sufficient data can be obtained. This also requires a large amount of labour. The main problem in determining the phenotypic effects of the expression of a transgene is obtaining reliable and regular measurements for a group that is sufficiently large enough to represent the population in question. When transgenic lines are subjected to phenotypic analysis, local variations in conditions can also affect the growth of plants. Randomisation is therefore required to reduce these effects. These constraints, to some extent can be overcome with a new system that combines accurate digital imagery with automated movement of plants around the glasshouse and automated water dispensing systems to reduce local variations in growth conditions. The system has been successfully used to phenotype plants such as *Arabidopsis* (Arvidsson *et al.*, 2011). The accuracy of the image analysis means that fewer plants need to be used due to a reduced variation in data acquisition and this can reduce costs. Images are acquired for 3D vectorisation of the plant using visible, UV and near infra-red light to obtain information about the physical appearance, water and metabolite distribution. The images are then analysed using pre-made algorithms to determine any significant difference between the groups (www.lemnatec.com).

This phenotyping technology is relatively new and has mainly been used for selecting lines with desirable phenotypes such as salt tolerance, and nutrient and water usage efficiency. Shoot biomass of barley, grown in high salinity conditions was predicted based on 3D vectorisation and analysis of the results were found to be more reliable than 2D vectorisation. This technique could therefore be used to select lines which grow faster in high salinity conditions (Golzarian *et al.*, 2011). The colouration of leaves in response to nutrient deficiency can also be quantified and therefore could be used to select lines that more efficiently use the nutrients available (Berger *et al.*, 2012). Water usage has been measured in soybean based on far-infra red light imaging. This data could be used to select for drought tolerant lines (Pereyra-Irujo *et al.*, 2012). Wheat leaf hairiness/trichome abundance (pubescence) plays an important role in adaption to environmental conditions and resistance to pests. This has been successfully quantified by high throughput analysis and as above, could be used to select for desirable lines (Genaev *et al.*, 2012). A large number of variables can be analysed using light with different wave lengths. Taking these advantages into account, this could identify small phenotypic differences that

may be otherwise missed due to human error in manual phenotyping experiments. Therefore several lines of *Arabidopsis* expressing effectors produced in this project will be analysed using this technology to determine any phenotypic effects due to effector over-expression.

4.2 Aims

The aims of this part of the project were:

- Identify *G. pallida* effector genes from an early draft of the genome sequence using a database of orthologous effectors from other plant parasitic nematodes and clone the identified effectors with a HA tag at the C-terminus.
- Transform *Arabidopsis* and potato with the HA tagged effectors to act as tools for functional characterisation.
- Demonstrate the transgenic status of the transformed plants and select lines for further analysis.
- Analyse the phenotypic effect(s) of effector over-expression within the plant by subjecting the plants to nematode infection, *P. infestans* infection and repeated growth measurements when grown in glasshouse conditions.

4.3 Materials and methods

4.3.1 Identification and cloning of putative *G. pallida* effector genes

A list of known effectors from *G. rostochiensis* and *H. glycines* was produced and subjected to sequence similarity analysis using BLASTn against *de novo* egg, J2, 21 dpi and 28 dpi transcriptome assemblies and an early genome assembly of *G. pallida* (June 2010) (Appendix 4). Identified sequences were analysed for the presence of a signal peptide and absence of a transmembrane domain using Signal P 3.0 (Dyrl Bendtsen *et al.*, 2004) and TMHMM (Krogh *et al.*, 2001). *G. pallida* putative effectors previously identified via EST analysis (Jones *et al.*, 2009b) were also subjected to BLASTn against the sequence data to determine their presence in the transcriptome and genome assemblies.

PCR primers (Chapter 2) were designed to amplify the identified effector sequences from *G. pallida* cDNA (without the signal peptide). RNAseq data was used to determine the life stage at which the effectors of interest were expressed, allowing cDNA from the appropriate life stage to be used for amplification. The effectors were subjected to the Gateway cloning (Invitrogen) procedure detailed in Chapter 2 and cloned into pCR®8 TOPO entry vector using the pCR®8/EW TOPO TA cloning kit (Life Technologies) (Xu and Li, 2008). Once cloned, the effectors were re-amplified with a primer to which a HA tag (TACCCTTATGATGTACCTGATTATGCCTGA) and a stop codon at the C-terminus of the sequence had been added, and were cloned into the pCR®8 TOPO entry vector. The HA-tagged effectors were then cloned into the final expression vector pK7WG2 (Karimi *et al.*, 2002b). The expression vector pK7WG2 confers kanamycin resistance for selection of transgenic plants, spectinomycin resistance for bacterial selection and constitutive expression of the gene of interest using a CaMV 35S promoter. The expression vectors containing the HA-tagged effectors were transformed into *Agrobacterium tumefaciens* GV3101 for *Arabidopsis* transformation and *A. tumefaciens* LBA4404 for potato transformation (Chapter 2). Both strains of *Agrobacterium* contained a helper plasmid pBBRIMCS5-VIGG-N54D, conferring gentamycin resistance (Dr Andrew Love, JHI); this plasmid provides constitutive expression of VirG, an essential gene for Vir gene induction following the perception of phenolic compounds, which is a requirement for T-DNA

transfer (Gelvin, 2009), hence improving transformation efficiency. GFP (green fluorescent protein) (primers: see Chapter 2) (Chalfie, 1995) was also cloned with a HA tag to act as a control construct during subsequent experimental procedures. The cloned effector sequences were subjected to sequencing throughout the cloning process to check sequence orientation, integrity and confirm the presence of the start and stop codon.

4.3.2 Potato transformation

Solanum tuberosum L. cv. 'Désirée' was grown in magenta vessels (Sigma) in aseptic conditions on multiplication media (see below) and maintained via cuttings in tissue culture every 4 weeks. Potato was grown in glasshouse conditions at 18–20°C under 16 h/8 h light/ dark. All tissue culture procedures used were performed in aseptic conditions. Stems were removed from this material and subjected to *Agrobacterium* mediated transformation (Figure 4.1).

A 5ml starter LB (see below) culture containing 50 μg/ml rifampicin and 100 μg/ml spectinomycin was inoculated with *Agrobacterium tumefaciens* LBA4404 containing the construct of interest and incubated at 28°C overnight with agitation. 2ml of this culture was used to inoculate 20 ml of LB, plus antibiotics, which was again incubated overnight at 28°C overnight with agitation. LB: 10 g/l typtone (Oxoid, Basinstoke, UK), 5 g/l yeast extact (Oxoid), 5 g/l NaCl (Fisher).

Stem material from wild type potato plantlets was cut into 1.5 cm sections for internodal transformation. The stem cuttings were incubated in transformation solution (see below) (45ml) and *Agrobacterium* culture (5ml), for 10 mins at room temperature. Transformation solution was composed of Linsmaier and Skoog medium (Duchefa Biochemie, Haarlem, Netherlands), 10 g/l sucrose, pH 5.8, adjusted with 1M KOH. Excess solution was removed using a sterile syringe. The stem cuttings were placed on filter paper on co-culture medium (see below) for 1 day, and were transferred to co-culture medium for 2 more days. The stem cuttings were then transferred to callus induction medium (see below) without antibiotic selection for 3 days, followed by 1 week on callus induction medium with antibiotic selection.

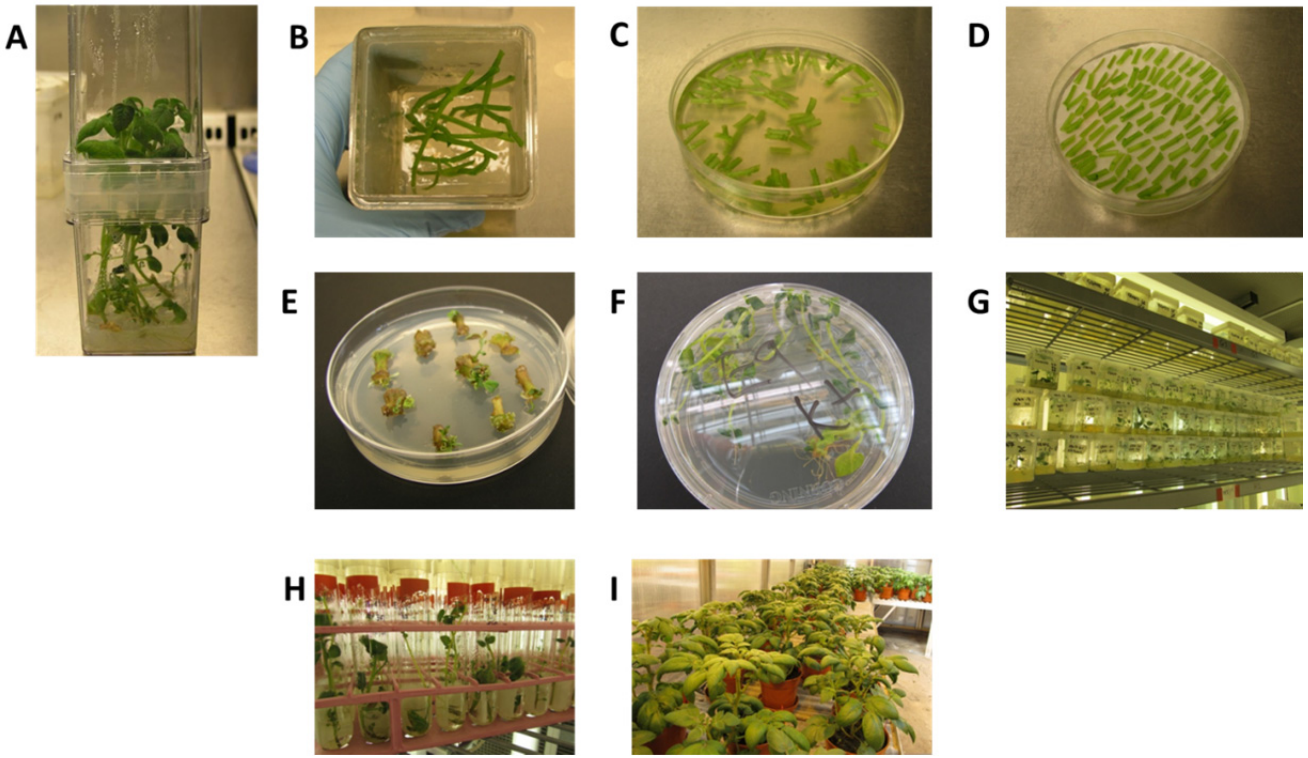


Figure 4.1: Transformation procedure for potato. (A) Wild type potato maintained in tissue culture that is used to provide a stock of stem material for transformation. (B) Stem material obtained from the maintained wild type plantlets which are kept in agar immediately prior to the transformation procedure to avoid desiccation of the material. All leaves and side shoots were removed. (C) Stem cuttings incubating in transformation buffer with *Agrobacterium* for internodal transformation. (D) Explants incubating on co-culture medium after the transformation step. (E) Following partial callus induction the explants are incubated on shoot induction medium. Shoots regenerated at this stage are then transferred to multiplication medium with antibiotic selection. (F) Putative transgenic lines rooting on multiplication medium with antibiotics. (G) Growth chamber full of transgenic lines in magentas. (H) Transgenic lines identified for phenotypic analysis growing in liquid medium in glass tubes (liquid medium was multiplication medium without agar). (I) Transgenic lines of interest in a glasshouse.

The explants that had been subjected to callus induction were then transferred to shoot induction media (see below) with antibiotic selection for 4 - 6 weeks and were transferred to fresh media every 2 weeks to reduce the build-up of phenolic compounds in the medium. During the "shoot induction" period, harvesting of healthy shoots was performed ensuring that subsequent shoots were not taken from the same location, by removing the area already harvested. The harvested shoots were transferred to multiplication medium with antibiotic selection for regeneration.

4.3.2.1 Potato growth and selection media

Multiplication medium

Multiplication medium was used to maintain all wild type and putative transformed lines. Multiplication medium contains LS medium (Linsmaier and Skoog) (Duchefa Biochemie), 30 g/l sucrose (Fisher), 5.5 g/l plant agar (Duchefa Biochemie), 500 µg/ml cefotaxime (Wockhardt) pH 5.8 (adjusted with 1M KOH). A high concentration of cefotaxime was used to reduce the risk of *Agrobacterium* from the original transformation persisting and giving false positives in PCR analysis. To reduce this risk further, all plants were maintained for 2 months on multiplication medium prior to molecular analysis. For antibiotic selection 100 µg/ml kanamycin was added

Co-culture medium

0.43 g/l Murashige and Skoog basal salts (Duchefa Biochemie), 0.108 g/l Nitsch vitamins (Duchefa Biochemie), 5.5 g/l plant agar, 30 g/l sucrose, pH 5.8 (adjusted with 1M KOH).

Callus induction medium

0.43 g/l Murashige and Skoog basal salts (Duchefa Biochemie), 0.108 g/l Nitsch vitamins (Duchefa Biochemie), 5.5 g/l plant agar, 30 g/l sucrose, 0.1 mg/l NAA, 5 mg/l zeatin riboside, 500 µg/ml cefotaxime, 10mg/ml $AgNO_3$ pH 5.8 (adjusted with 1M KOH) (silver nitrate is used to suppress host defences and therefore improve the transformation efficiency).For antibiotic selection 100 µg/ml kanamycin was added.

Shoot induction medium

0.43 g/l Murashige and Skoog basal salts (Duchefa Biochemie), 0.108 g/l Nitsch vitamins (Duchefa Biochemie), 5.5 g/l plant agar, 30 g/l sucrose, 0.3 mg/l gibberellic

acid, 5 mg/l zeatin riboside, 500 µg/ml cefotaxime, 10 mg/ml $AgNO_3$, 100 µg/ml kanamycin pH 5.8 (adjusted with 1M KOH).

4.3.3 *Arabidopsis* transformation

A. thaliana Col-0 seeds were sown in compost and grown in a glasshouse at 20°C with a 16 h light and 8 h dark photo-period. The first primary inflorescence was removed to promote growth of multiple secondary inflorescences and at approximately growth stage 5.1 where unopened flower buds can be seen (Boyes *et al.*, 2001) a minimum of 9 *Arabidopsis* plants per construct were subjected to transformation.

A 20 ml LB culture of *Agrobacterium* GV3101 harbouring the construct of interest, was grown overnight with spectinomycin, gentamycin and rifampicin at 28^{o}C with agitation. The following day the 20 ml culture was used to inoculate 200 ml of LB (containing appropriate antibiotics; see above). This was incubated at 28 °C with agitation until an optical density of A_{600nm} 0.5-0.8 was reached (approximately 4-5 hours). The culture was subjected to centrifugation (2455 × *g* for 10 mins) to pellet the bacteria. The supernatant was discarded and the pellet was re-suspended in 200 ml of 5 % sucrose solution. 100 µl of Silwet (Lehle Seeds, Texas, USA) was then added and gently mixed with the solution (Silwet aids binding of the bacterial cells to the plant material, hence improving transformation efficiency). The *Agrobacterium* solution was transferred to a 2 litre beaker. The inflorescences of the *Arabidopsis* plants were dipped into the solution and gently swirled for 30 seconds, ensuring that all flower buds were adequately covered in the transformation solution. The plants were then incubated under a propagator lid for 24 hours to maintain high humidity to increase the transformation frequency. The plants subjected to transformation were then grown in the glasshouse, seeds were collected and subjected to *in vitro* antibiotic selection to identify transformants.

Seeds obtained from plants subjected to the transformation protocol were sterilised and grown in tissue culture as described in section 2.2. The medium used contained 40 µg/ml kanamycin and 50 µg/ml cefotaxime to select for transformed seeds that contained the antibiotic resistance gene. Plantlets (T1) that survived antibiotic selection were transferred to soil and grown to maturity to produce self-fertilised seeds (T2). Approximately 300 T2 seeds were sterilised and subjected to antibiotic

selection in tissue culture. The frequency of the plantlets that survived antibiotic selection (segregation rate) was used to determine the number of unlinked transgenic copies. Lines showing a 3:1 survival ratio were considered to have a single transgene insertion. Siblings for each transgenic line were again transferred to soil and grown to produce seed (T3). T3 seeds were once again subjected to antibiotic selection and the segregation ratio was used to confirm transgene insertion number. During this final selection, those batches of seeds that had a 100% survival rate under antibiotic selection were considered to be homozygous with a single insertion.

4.3.4 Molecular analysis of potato and *Arabidopsis* transgenic lines

Putative transgenic lines, identified by antibiotic resistance phenotype, were subjected to molecular analysis to confirm their transgenic status.

4.3.4.1 Amplification and sequencing of DNA insert

For the analysis of transgenic plants a direct PCR kit was used Plant PCR kit (Sigma). DNA was extracted from putative transgenic plants using the Plant PCR kit, and the resulting DNA was then subjected to PCR, as described in chapter 2, using reagents from the Plant PCR kit and 35S promoter/ terminator primers (Chapter 2), with a Tm of 61 °C to amplify the DNA insertion. The PCR products were then analysed via gel electrophoresis to determine the presence or absence and size of the amplified DNA. PCR products from positive transgenic lines were cleaned, as described in section 2.5.3 and sequenced to ensure the inserted sequence was correct.

4.3.4.2 Western blot analysis of effector protein production

Proteins were extracted from leaf tissue from PCR positive transgenic lines and subjected to colorimetric and chemiluminescence western blot analysis (see section 2.7) in an attempt to demonstrate transgenic protein production in both potato and *Arabidopsis*. 20 µg of protein was loaded per lane per sample. 24 western blots using different antibodies, membranes and variations on procedures were performed. Millipore membranes with a pore size of 0.45 µm (nitrocellulose) and 0.2 µm (Immobilon) (PVDF) were used to determine if the small effector proteins could be retained and more efficiently identified using a membrane with a smaller pore size. This was unsuccessful. Blocking times were adjusted from 1 hour to overnight in an attempt to remove non-specific bands (~100 kDa). The anti-HA antibody, made by GenTex (Irvine, USA) (HA.C5) raised in mouse using a secondary IGG anti-mouse

antibody produced a lot of non-specific bands and was not sensitive enough to detect any HA-tagged proteins. Therefore a high affinity monoclonal antibody coupled with a peroxidase (Roche, Welwyn Garden City, UK) was used in subsequent experiments. This was more sensitive and did not produce as many non-specific bands. Benchmark pre-stained ladder (Invitrogen) was used as a molecular weight marker for all colorimetric western blots. Chemiluminescence detection was performed using Super Signal West Femto kit (Thermo scientific). All Chemiluminescence western blots used Novex Sharp Pre-stained Protein Standard (Invitrogen) as a reference molecular weight marker. Ponceau S acid red staining was performed to confirm the presence of proteins on membranes during the western blot procedure (section 2.7.4).

4.3.4.3 RNA analysis for quantification of transgene expression

Due to a lack of data from western blot analysis, selection of transgenic lines to use in subsequent phenotyping assays was performed via RNA expression analysis. Total RNA was extracted from leaf tissue from twelve potato lines per construct and separately from six *Arabidopsis* lines per construct using a Plant RNA Mini kit (Bioline) following the manufacturer's instructions, except the final wash step was repeated to improve the quality of the final RNA. The extracted RNA was quantified using a Nanodrop spectrometer. An equal quantity (500 ng) of RNA for each transgenic line was then treated with DNase (Promega RQ1) to remove DNA contamination. 8 µl of DNase-treated RNA was converted to 1st strand cDNA using a cDNA synthesis kit (Bioline) following the manufacturer's instructions except that the synthesis reaction was allowed to proceed for 60 mins instead of 40 mins. Finally 1 µl of cDNA was subjected to PCR analysis using gene specific primers (section 2.9.3) and 25 amplification cycles to allow differential expression to be clearly seen when analysed using gel electrophoresis. Amplification of a reference gene sequence was also performed (18S for *Arabidopsis* and EF1α for potato) to allow visual normalisation of transgene expression and confirm that equal quantities of cDNA were present in all samples.

4.3.5 Phenotypic analysis of transgenic potato lines

To determine any phenotypic effects of transgene over-expression, two high and one lower expressing line of transgenic potato (for each effector construct) were grown in the glasshouse and subjected to a range of analyses. The control in this experiment

was 3 independent lines of non-transformed plants that had undergone the same regeneration procedure as the transgenic lines but without antibiotic selection and without *Agrobacterium*-transformation. These are termed transformation controls. The three independent control lines were collectively treated as one control group during analysis.

Potato explants were synchronised in tissue culture and each grown in 5 ml of liquid multiplication medium (without agar) in glass tubes (Sigma) for 3 weeks, until they were well rooted. 12-14 individual plantlets were set up for each transgenic or control line. The plantlets were then transferred to a compost/Perlite mixture in pots (10 cm) and grown in glasshouse conditions (section 2.4) under a propagator lid for 2 weeks to allow the plantlet to adjust to the change in humidity. Four weeks after planting in soil the plants were transferred to bigger pots (17.8 cm) containing compost; this was considered the start point for all phenotyping measurements. For each line 8 plants of comparable size were selected for transfer. At regular time points length and width of one comparable terminal leaflet per plant and plant height were measured (every 2-3 days). The youngest terminal leaf >1 cm was selected for measurements. At the final time point, when natural senescence of the control plant had begun, the number of leaf nodes on the main stem, any other growth abnormalities, wet and dry weight (above ground biomass), number and weight of tubers were recorded for each plant. The number of days to complete petal opening of the first flower was recorded for each plant. The transgenic plants and controls were ordered in a Latin square in the glasshouse, to minimise local environmental factors affecting growth. Statistical analysis (parametric: ANOVA with Tukey post hoc correction. Non-parametric: Kruskal-Wallis with Bonferroni post hoc correction) was performed using R (Team, 2008).

4.3.6 Nematode infection assay

Transgenic potato lines were challenged with *G. pallida* to determine any alteration in their susceptibility to the pathogen as a result of over-expression of each effector in potato. One high and one lower expressing line for each effector was used for this assay. 12 plantlets for each line of interest were multiplied and synchronised in tissue culture for 3 weeks on multiplication medium (with plant agar). The resulting plantlets were then transferred to soil in multi-pot trays (5 cm per pot) and grown in

glasshouse conditions. Three weeks after planting in soil 600 J2 nematodes were applied to each plant. Four 1 ml aliquots of *G. pallida* J2s resuspended in water (150 nematodes ml^{-1}) were applied to the soil surface at locations 1.5 cm around the base of the plant stem. Two weeks post infection three plants from each transgenic line and nine plants from the control group were subjected to acid fuchsin staining (section 2.3) to determine infection rates. Six weeks post infection, five plants from each line and 15 plants from the control group were subjected to acid fuchsin staining (section 2.3). The control plants in this experiment included a transformation control and a GFP expressing line. All identified nematodes were counted and their developmental stage was recorded. The sizes of the root systems for the transgenic lines were visibly different, therefore the root systems were weighed prior to staining, once all soil had been removed, to gain a quantitative measure for the number of nematodes per gram of root.

4.3.7 *Phytophthora infestans* CS-12 detached leaf assay

Phytophthora infestans CS12 which has a transgene insertion leading to stable down-regulation of the effector AVR3a by RNAi (Bos *et al.*, 2010b) was used to determine if any of the transgenic potato lines were more susceptible to *P. infestans* than the control plants. The control plants in this experiment included a transformation control and GFP expressing lines. The transgenic line of *P. infestans* was used as preliminary experiments showed that it was extremely difficult to detect a plant showing increased susceptibility to the wild type oomycete under the experimental conditions being used. *P. infestans* CS12 was maintained on rye sucrose agar plates with 20 µg/ml of gentamycin and 1% v/v Pimaricin (an antifungal compound). At 12 days post inoculation of the agar plates, sporangia were harvested from the plates and the detached leaf assay was performed as described by (Whisson *et al.*, 2007b). Briefly, 3 leaves of equal ages located within 10 cm from the growing tip, from each transgenic line, grown in glasshouse conditions, were inoculated with four 10 µl spots, each containing 125 sporangia of *P. infestans* CS12. The leaves were incubated at room temperature for 6 days in a sealed box in a sealed autoclave bag. Each box contained a moist paper towel to maintain humidity. To minimise the variability between boxes, each box had its own set of control leaves to allow direct comparison between effector gene expressing plants and control plants. Each day the leaves were imaged twice at the same time points to record the infection symptoms.

4.3.7.1 Scoring *Phytophthora* CS-12 detached leaf assay

A computer program "*Phytophthora* CS-12 detached leaf assay Autoit program" was written to quantify the area of infection as a percentage of the whole leaf (Appendix 4). The first step in this process was to transform the images using Photoshop (Adobe) by reducing "red" by 70%, "blue" by "100%" and increasing "green" by 200%. The transformed images then had their background removed using PowerPoint (Microsoft). The leaves are in grey scale where the darkest regions correspond to the infection zones. The program then quantifies the brightness of each pixel. Any pixel that is black or up to 60 shades lighter is considered infected leaf tissue, any pixel lighter than this is considered non-infected background leaf (Figure 4.2). Shades of variation is an internal built in function in the computer language. The results are returned as a percentage of the leaf infected with *Phytophthora*. The results from this output were analysed using a non-parametric Krustal-Wallis test with Bonferroni correction in the package R.

The time taken for first signs of infection was also recorded, including the number of infection points that were positive for infection for every spot on every leaf.

4.3.8 Phenotypic analysis of transgenic *Arabidopsis* expressing *Globodera pallida* effectors

Homozygous, single insertion lines expressing each effector were generated for *Arabidopsis*. Four expressing lines for each construct, identified via RNA expression analysis were, subjected to growth analysis of T3 plants. Plants expressing effectors: GpG7E05, GpE9, Gp66p1, GpA42, GpG20E03b, GpSCN1120 and Gp1106 were sent for analysis using a Lemnatec system at Keygene (Wageningen, Netherlands). Controls for this experiment were two lines of GFP and a transformation control group.

4.3.8.1 Lemnatec data

High throughput phenotypic analysis of transgenic *Arabidopsis* was performed by PhenoFab (http://www.phenofab.com) using Lemnatec.

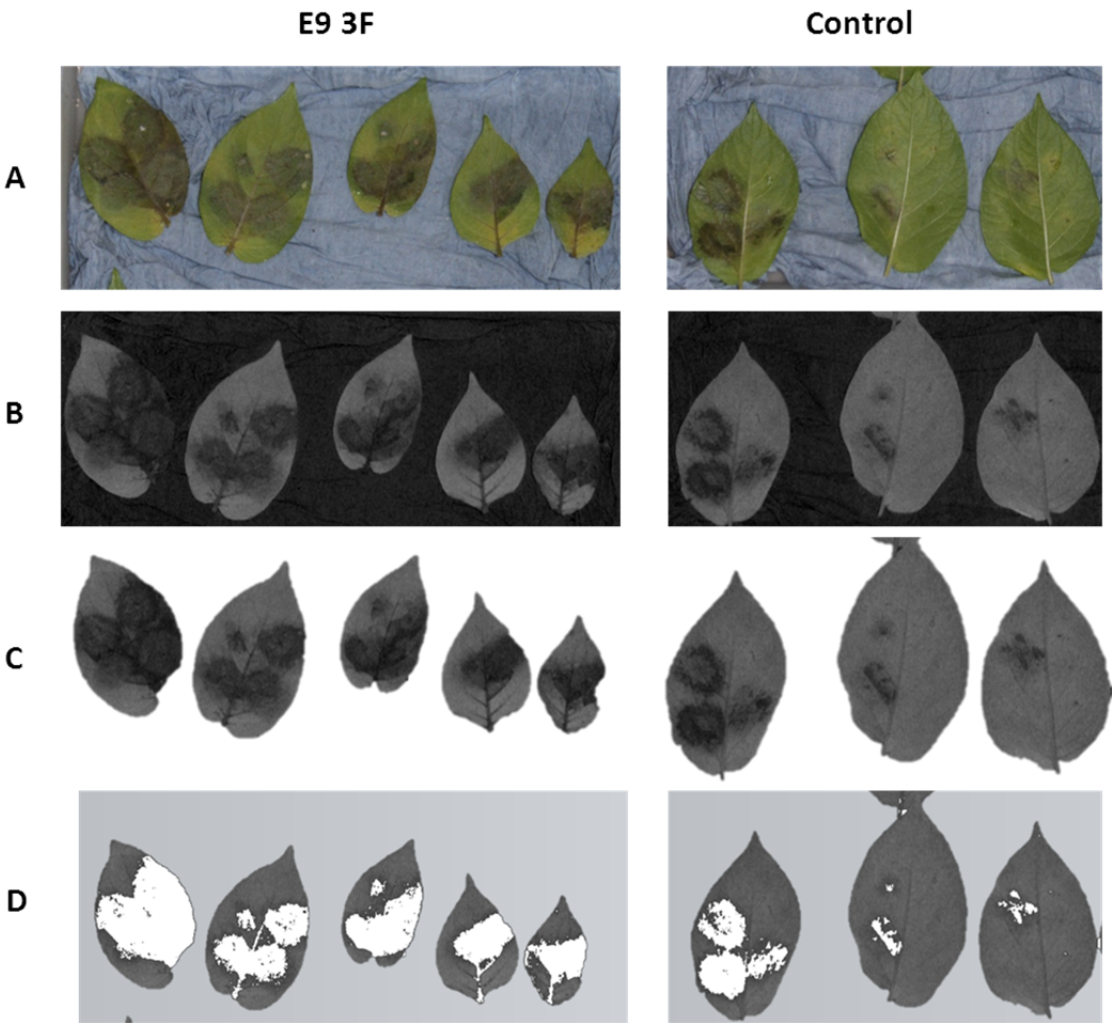


Figure 4.2: Procedure for quantifying *Phytophthora infestans* infection level on detached potato leaves. Detached leaf assay for potato transgenic line E9 3F and a control line 5 days post inoculation with *P. infestans* CS-12. (A) Image of leaf which has not been altered. 4 spots of inoculum were put on each leaf and the leaves were incubated in the dark in a moist environment at room temperature. (B) Photoshop transformed image: Each pixel was altered by reducing "red" by 70%, "blue" by "100%" and increasing "green" by 200%. (C) The background has been removed using PowerPoint so that only the leaf is analysed in the next step. (D) A computer program quantifies the amount of disease symptom, replacing the pixel that is defined as infection with a white pixel. The result is returned as a percentage of the leaf showing symptoms.

4.3.9 Statistical analysis

Normality was tested with a Kolmogorov-Smirnov test either on raw data or following log transformation. If the data were normally distributed an Analysis of variance (ANOVA) (Shapiro and Wilk, 1965) was used to test for significance, if the variation was not significantly different between the groups being analysed, and if there were 3 or more means being sampled. The ANOVA test was developed to avoid type 1 errors which may be encountered by the use of multiple t-tests. Least significance values defined by the user (for example, $P<0.05$, $P<0.01$ or $P<0.001$) can be used to determine if population A is significantly different from population B. A post hoc analysis can then be used to determine the "honest significant difference" by adjusting the P-values according to the number of means being sampled, avoiding type 1 error, which produces the P-values for multiple comparisons within the experiment being tested. Tukey is an example of a post hoc analysis (Keselman, 1976). Tukey post hoc analysis is more stringent than an ANOVA (using least significant values). Therefore Tukey post hoc correction was used for stringent analysis of the data when multiple populations were examined to identify significant differences with greater confidence. If the data was not normally distributed a non-parametric version of an ANOVA called Kruskal-Wallis with Bonferroni post hoc correction was used.

4.4 Results

4.4.1 Cloning of *Globodera pallida* effector genes

45 out of 54 effectors identified using a bioinformatics approach on data generated from the *G. pallida* genome and transcriptome assemblies were cloned from *G. pallida* cDNA into pCR®8/EW TOPO entry vector. Twenty one of these effectors (Table 4.1) were subsequently cloned with a HA tag at the C-terminus into the destination vector pK7WG2. The sequences of these effectors are listed in Appendix 4. The availability of RNAseq data was extremely helpful in identifying when/if the effectors are expressed, which allowed cDNA from appropriate life stages to be used for PCR reactions. All effector constructs were sequenced and only those with the expected sequence were used in plant transformation.

4.4.2 Potato transformation

Seventeen effector constructs and one containing GFP were used for transformation of potato cv. 'Désirée', resulting in over 310 transgenic lines. More lines were produced than the total number in the table, for example G20E03 produced several more un-harvested shoots. However, enough lines had already been generated for this effector and therefore these were discarded. The *G. pallida* effectors transformed into potato were GpIVG9, GpIA7, GpDGL1, Gp448-3, Gp448-4, GpA42, GpG20E03b, GpG20E03, GpHg10C02, GpG16H02, Gp66p1, GpE9, GpG8A07, GpG7E05, GpHgSEC4, GpSCN1120, Gp1106 and two controls: GFP and transformation control. A range of transformation efficiencies were observed, for example when transforming with 66p1, 33 independent transgenic lines were produced, whereas A42 produced no transgenic lines, even after repeating the transformation process. This could show that some effectors could have a deleterious effect on the plant preventing regeneration of viable transgenic plants (Table 4.2). Kanamycin (100 μg/ml) was a robust selection method as assessed by the proportion of plants that rooted on the antibiotic containing media (see below) (Figure 4.3). Only shoots that successfully rooted on kanamycin media considered putative transformants, and therefore subjected to molecular analysis to confirm their transgenic status.

Effector name	Information and accession number	species	location	Reference
Gp448-1	secretory protein 4D06 (AAN32892.1)	*H. glycines*	DG	Gao *et al*., 2003
Gp448-2	secretory protein 4D06 (AAN32892.1)	*H. glycines*	DG	Gao *et al*., 2003
Gp448-3	Putative gland protein 29D09 (AAP30755.1)	*H. glycines*	DG	Gao et al., 2003
Gp448-4	Putative gland protein 29D09 (AAP30755.1)	*H. glycines*	DG	Gao et al., 2003
GpG20E03b	similar to Putative gland protein G20E03 (AAO85459.1)	*H. glycines*	SV	Gao et al., 2003
GpG20E03	Putative gland protein G20E03 (AAO85459.1)	*H. glycines*	SV	Gao et al., 2003
GpA42	A42 protein (CAD60975.1)	*G. rostochiensis*	DG	Qin *et al.*, 2000
GpDGL1	-	*G. rostochiensis*	-	-
GpHg10C02	gland-specific protein (AAO33473.1)	*H. glycines*	SV	Gao *et al.*, 2003
GpSCN1120	Oesophageal gland protein scn1120 (AAK94491.1)	*H. glycines*	-	Boer *et al*., 2002
Gp66P1		*G. rostochiensis*	-	
GpG7E05	Putative gland protein G7E05 (AAP30762.1)	*H. glycines*	DG	Gao et al., 2003
GpE9	E9 protein (CAD60977.1)	*G. rostochiensis*	-	Qin *et al.*, 2000
Gp747	-	*G. rostochiensis*	-	-
Gp4D06	secretory protein 4D06 (AAN32892.1)	*H. glycines*	DG	Gao *et al.*, 2003
GpG8A07	Putative gland protein G8A07 (AAP30833.1)	*H. glycines*	DG	Gao *et al.,* 2003
GpHgsec4	Oesophageal gland cell secretory protein 4 (AAG21334.2)	*H. glycines*	Lateral/ ganglia	Gao *et al*., 2003
Gp1106-2	(AFH68236)	*G. rostochiensis*	-	
GpG16H02	Putative gland protein G16H02 (AAP30769.1)	*H. glycines*	-	Gao *et al*., 2003
GpIVG9	IVG9 (ABF51007.1)	*G. mexicana*	DG	Blanchard *et al.*, 2007
GpIA7	IA7 (ABF51008.1)	*G. pallida*	SV	Blanchard *et al.*, 2007

Table 4.1: Orthologues of effectors identified from an early draft of the *Globodera pallida* genome sequence and cloned from cDNA. The sequences are presented in Appendix 4. Some of the effectors show sequence similarity with each other and therefore belong to a family, for example Gp448 members. The species column indicates which species the effectors were first identified in, and the location column summarises the cellular localisation of these effectors in the species they were first identified along with a reference for the information. DG (dorsal gland). SV (subventral gland).

Transgene	Total number of lines	Number of positives for insertion	Number of negatives for insertion	Number of lines not tested
GpIA7	11	9	0	2
GpSCN1120	15	12	2	1
Gp448-3	35	21	2	12
GpG16H02	11	8	0	3
Control	5	0	5	0
Gp66p1	33	13	1	19
GpG20E03b	15	13	1	1
GpHgSec4	33	22	1	12
GpE9	14	12	0	2
GpG7EO5	14	14	0	0
GpG8A07	18	16	2	0
GpHg10C02	3	3	0	0
GpDGL1	28	15	0	13
GpIVG9	9	6	0	3
GFP	12	4	1	7
GpA42	10	0	2	8
Gp448-4	22	14	1	7
GpG20E03	19	13	1	5

Table 4.2: The number of transgenic potato lines generated for each effector construct. The table displays information about the total number of lines generated for each effector construct transformed into potato. The putative transgenic lines were screened by PCR using promoter and terminator primers. The PCR products were then analysed by gel electrophoresis. The positive lines identified were taken forward to subsequent transgene screening procedures.

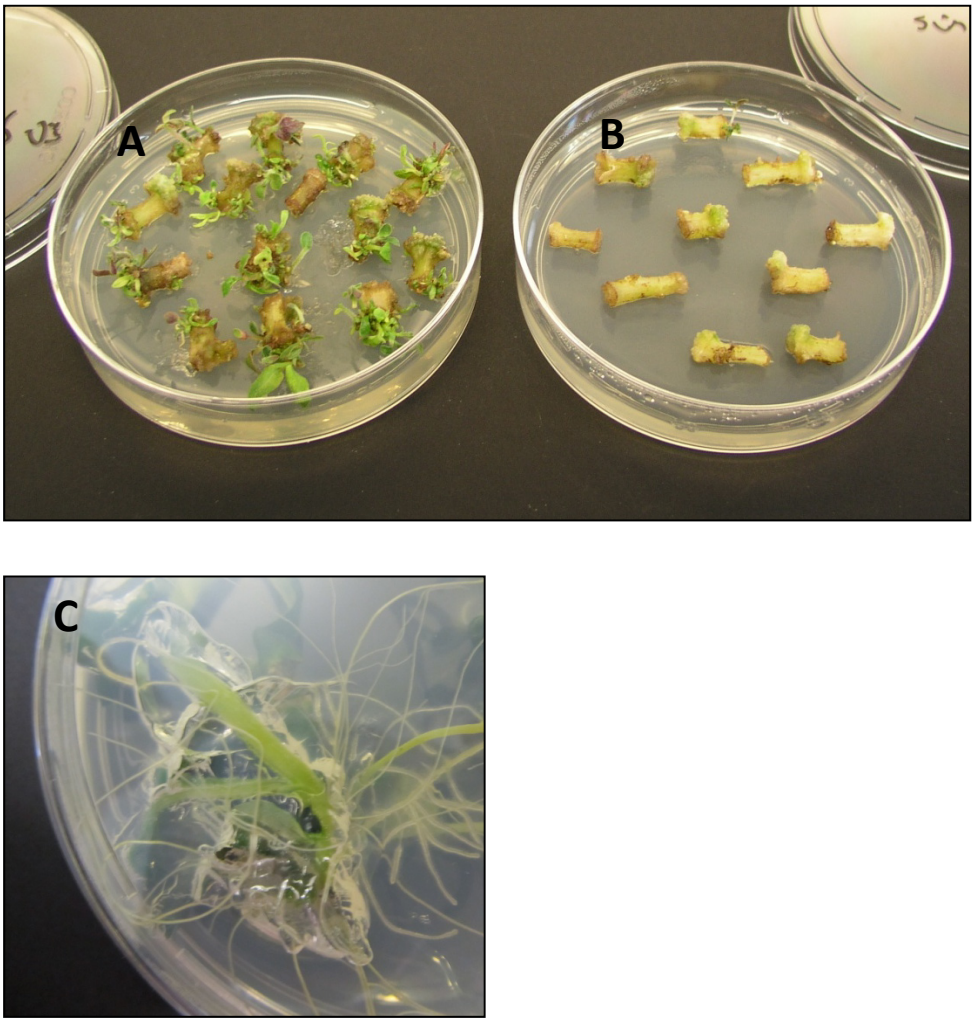


Figure 4.3: A range in transformation efficiency was observed during potato transformation with *Globodera pallida* effector gene constructs. (A): Potato internodal sections transformed with the 66p1 effector produced 33 lines of transgenic potato. (B): Potato internodal sections transformed with the A42 effector produced zero transgenic plants; even when repeated no transgenic lines expressing A42 survived. (C): GFP-expressing transgenic potato rooting on medium containing kanamycin.

4.4.1 *Arabidopsis* transformation

Twenty one effectors and GFP were transformed into *Arabidopsis*. The effectors transformed into *Arabidopsis* were GpIVG9, GpIA7, GpDGL1, Gp448-1, Gp448-2, Gp448-3, Gp448-4, GpA42, GpG20E03b, GpG20E03, GpHg10C02, GpG16H02, Gp66p1, GpE9, GpG8A07, GpG7E05, GpHgSEC4, GpSCN1120, Gp1106, Gp4D06, Gp747 and controls (GFP, transformation control).

4.4.2 Screening of transgenic lines

Transgenic lines selected on antibiotic containing media for both potato and *Arabidopsis* (generation T2) were subjected to further analysis to confirm their transgenic status. Transgenic lines were characterised for their production of the transgenic protein via western blot analysis and the expression of the transgene via RNA analysis to determine which lines would be maintained in tissue culture (potato), subjected to phenotypic analysis (potato and *Arabidopsis*) and which lines will be taken through further generations to obtain homozygous lines (*Arabidopsis*).

4.4.2.1 Amplification of transgene sequence by PCR

Primers (35S promoter and 35S terminator) were used to amplify the effector coding region from the putative transgenic lines (Figure 4.4). *Arabidopsis* putative transformants were not screened by PCR due to the observed reliability of antibiotic selection, instead they were screened by semi quantitative RT-PCR (section 4.4.2.3). More than 90% of potato lines analysed by PCR were positive for an insertion of the expected size, based on gel electrophoresis analysis. If an analysed line/sample did not produce a gel electrophoresis product of the expected size (for example lane 7 in Figure 4.4), this line was discarded along with any lines that did not produce a strong electrophoresis product, (for example lane 23 in Figure 4.4). Any electrophoresis products that did not match the expected size could have arisen due to a mislabelled sample or because some of the insertion region may have been lost during the transformation procedure. Lines that generated products of an incorrect size were also discarded. To confirm that the DNA insertion was correct, four transgenic lines were chosen for sequence analysis, two transformed with IA7 and two transformed with DGL1. The results showed, for all four lines, that the inserted effector gene sequence was correct with the start codon, coding region, HA tag and stop codon all correctly maintained (Figure 4.5).

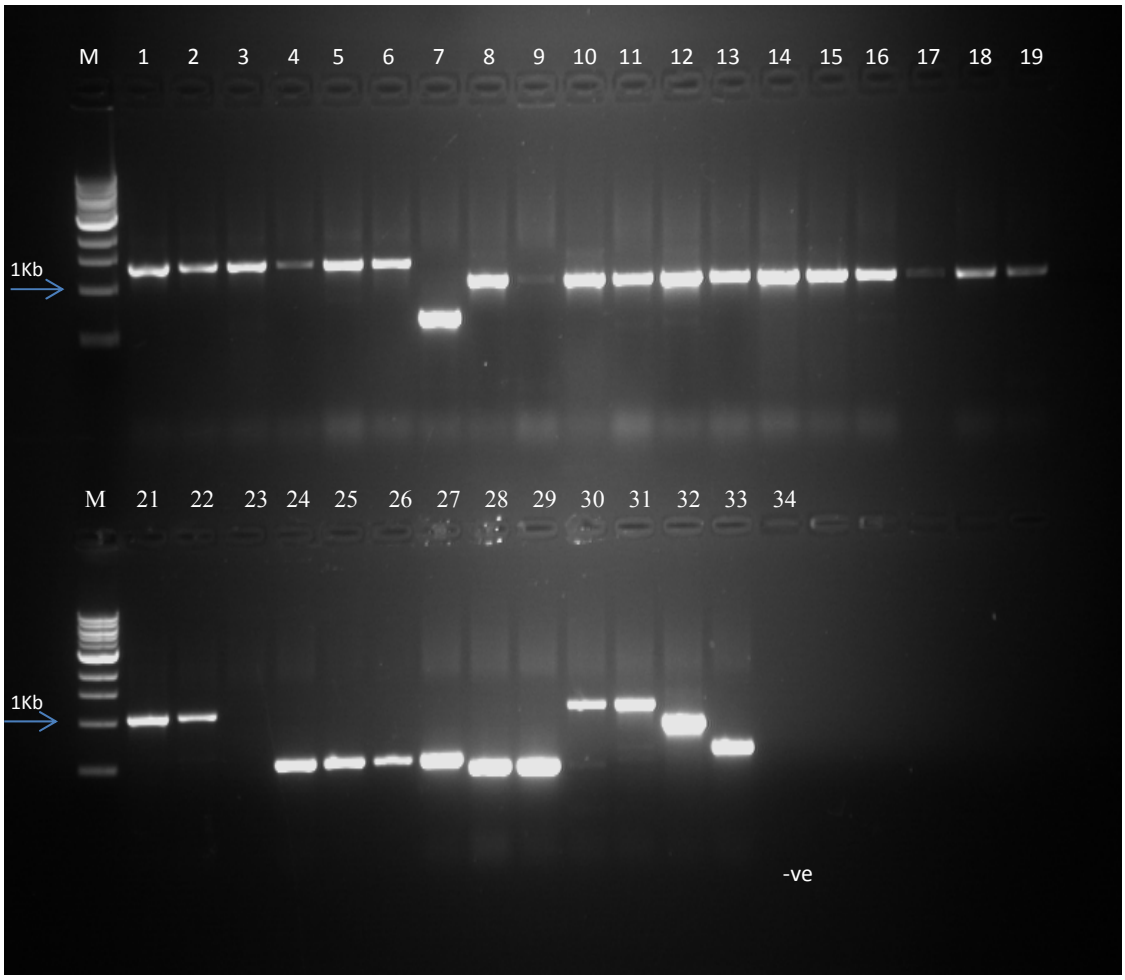


Figure 4.4: PCR analysis of transgenic potato lines. Gel electrophoresis image of transgenic potato lines analysed by PCR using 35S promoter and terminator primers. All product sizes were correct, except lane 7. Any line which did not have the correct product size or produce a strong band was discarded and not taken forward for RNA screening. Lanes: M: Bioline 1 kb Ladder, 1-6: E9 (1352bp), 7-14: G8A07 (1070bp), 15-22: Hgsec4 (1073bp), 23-27: IA7 (527bp), 28,29: DGL1 (464bp), 30,31: GFP (1070bp), 32: Hg10CO2 (602bp), 33: G20E03b (881bp), 34 (-ve).

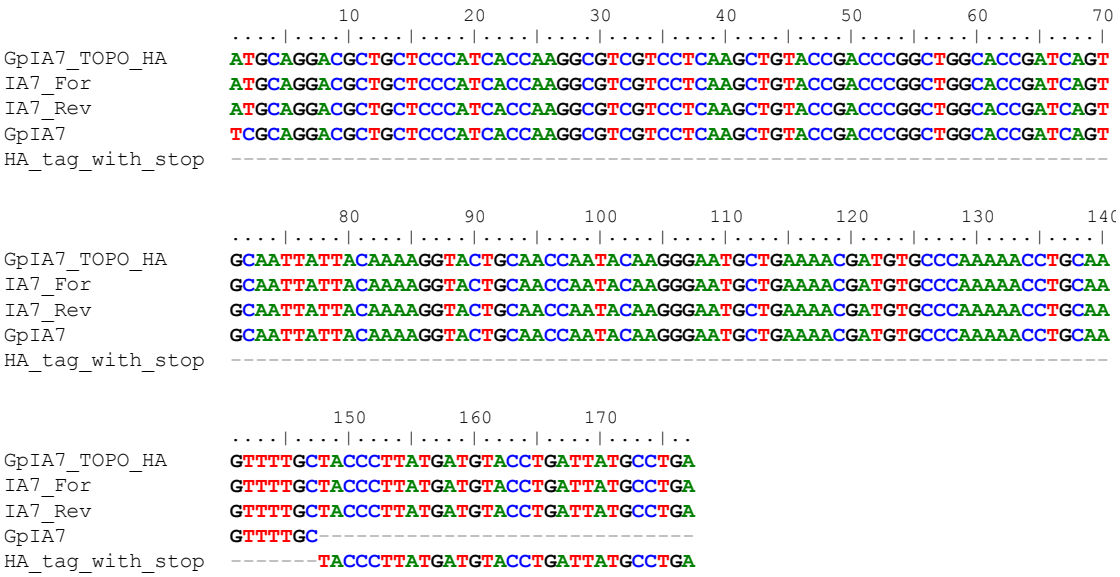


Figure 4.5: Sequence confirmation of the transgene in a potato line transformed with the IA7 effector construct. An alignment of the DNA insertion sequence amplified from potato for an IA7 transgenic line with the IA7 effector gene sequence. IA7_For and IA7_Rev sequences were obtained by Sanger sequencing the transgenic DNA insert using forward and reverse gene specific primers. The original effector sequence (GpIA7) without the signal peptide (note the start codon for this gene is at the start of the signal peptide which is missing in the GM lines, hence this sequence does not start with ATG), the TOPO-entry vector clone sequence (GpIA7_TOPO_HA) and the HA-tag sequence with stop codon are shown aligned with the transgenic insert.

4.4.2.2 Western blot analysis of transgenic protein expression

Detection of effector protein production via western blot analysis proved difficult. Colorimetric detection (see section 2.7) of GFP proteins from extracts of transgenic potato with an anti-HA antibody was successful (Figure 4.6). However, no effector proteins were identified using colorimetric detection for protein extracts from either *Arabidopsis* or potato lines, using an anti-HA antibody. These experiments used a primary anti-HA antibody, made by GenTex (HA.C5) raised in mouse in combination with a secondary anti-mouse IGG antibody. This combination resulted in many non-specific bands and was not sensitive enough to detect any HA-tagged proteins with this anti-HA antibody (other than GFP-HA). Therefore a high affinity monoclonal primary antibody coupled with a peroxidase (Roche) was used in subsequent experiments. This was more sensitive and did not produce as many non-specific bands. The more sensitive monoclonal antibody and chemiluminescence detection using a maximum sensitivity detection kit (see section 2.7) was used to detect the HA tag in subsequent analysis of potato lines. This method also had limited success. Proteins were detected for high and low expressing potato lines transformed with Gp448-3 (identified via RNA expression analysis), and the quantity of protein present was proportional to the level of RNA expression levels detected by RT-PCR (Figure 4.7). No other effector protein was successfully identified from transgenic potato or *Arabidopsis* lines.

Due to the lack of detection of transgenic protein, a western blot using chemiluminescence detection was performed on proteins extracted from leaf regions transiently expressing effectors in *N. benthamiana.* This was performed to determine if transiently expressed effectors could be detected. However, the only HA-tagged proteins successfully detected from both stable transformed potato and transient expression in *N. benthamiana* were Gp448-3, Gp448-4, GFP and GpE9 (Figure 4.8).

4.4.2.3 Transgene expression via RNA analysis

RNA was extracted from PCR positive transgenic lines of potato and *Arabidopsis*. 500ng of RNA for all lines of interest was converted to cDNA, after the removal of genomic DNA. 25 cycles of PCR were then used to amplify the effector sequence gene of interest, using gene specific primers. The relatively low PCR amplification cycle number allowed the visual identification of differential expression between the lines examined. Although an equal quantity of RNA was used for analysis of all

lines, a further PCR reaction to amplify a reference gene EF1α, was used to give an indication of the relative amount of cDNA in each PCR reaction for each gene. After analysis of PCR products relative intensities using gel electrophoresis, 6 high and 1 lower expressing potato lines were chosen for further analysis and maintenance in tissue culture. This process was performed for all transgenic potato lines and 6 *Arabidopsis* lines per construct. A typical result for the analysis of the transgenic potato lines is shown in Figure 4.9. Twelve lines of GpG20E03 were analysed via RT-PCR and differential levels of gene expression can be seen in the gel electrophoresis image. A typical result for the analysis of the transgenic *Arabidopsis* lines is shown in Figure 4.10.

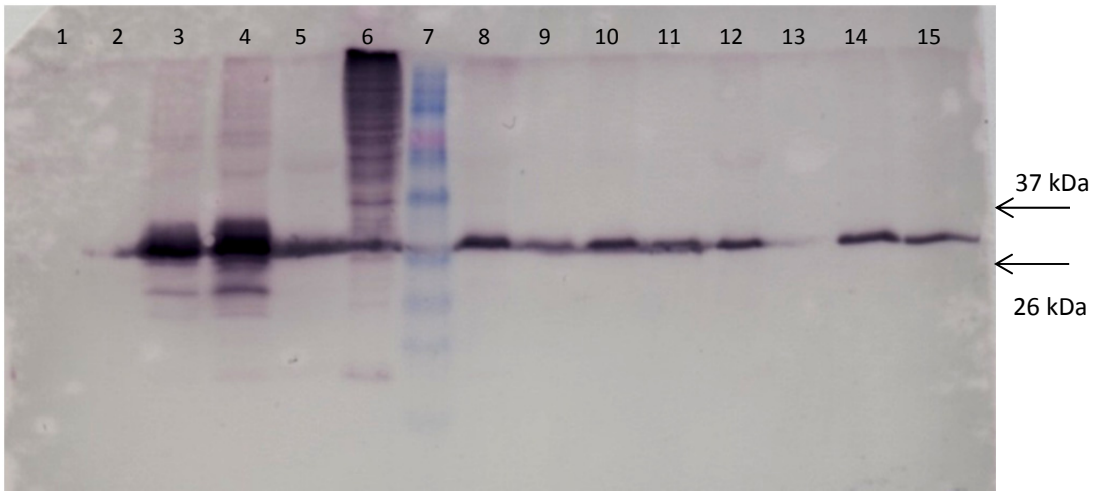


Figure 4.6: Colorimetric western blot analysis of potato lines transformed with a construct expressing GFP. Colorimetric western blot analysis of eight transgenic GFP-Ha (28 kDa) potato lines using an anti-HA tag antibody. 20 µg of soluble protein was loaded per lane. Lane 1: transformation control (–ve), 2: transformation control (–ve), 3: transformation control with GFP protein (+ve), 4: *Agrobacterium* ¬containing GFP in plasmid pK7WG2, 5: *Arabidopsis* transformation control with GFP protein (+ve), 6: *Arabidopsis* transformed with GFP (+ve) (existing lab material), 7: Ladder, 8: GFP-1, 9: GFP-3, 10: GFP-4, 11: GFP-5, 12: GFP-7, 13: GFP-9, 14: GFP-10, 15: GFP-12.

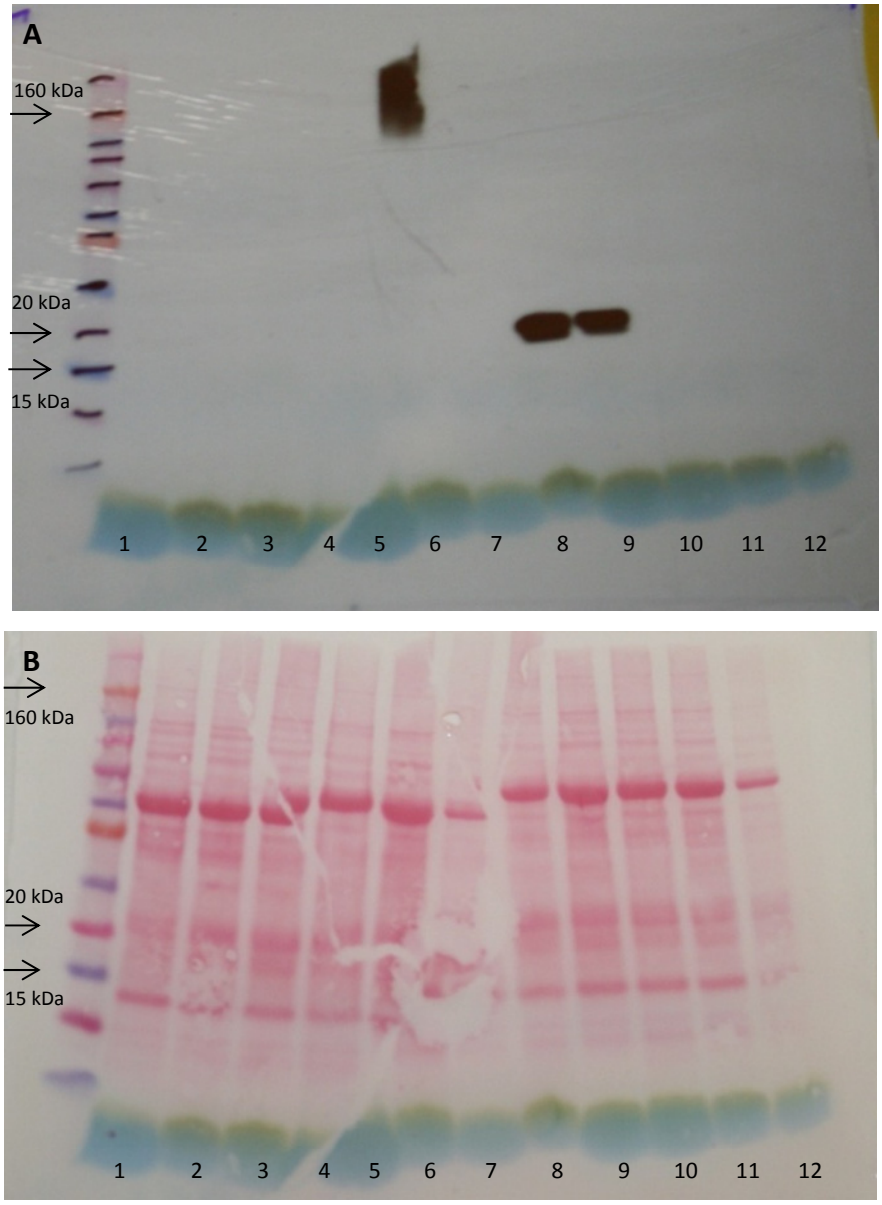


Figure 4.7: Western blot detection of HA-tagged effector proteins from transgenic potato lines. (A) Western blot detection of HA-tagged effector proteins from transgenic potato using chemiluminescence detection and a monoclonal HA antibody coupled with horse-radish peroxidase. 20 µg of protein was loaded per lane. Bands can be seen in extracts from effector 448-3 high (5C) and low expressing (15C) construct. (B) Ponceau S acid red staining of total protein on the nitrocellulose membrane was conducted to prove protein transfer to the membrane. Lanes: 1: Ladder, 2: Gp66p1-4 (11 kDa), 3: Gp66p1-7 (11 kDa), 4: GpIVG9-4C (10 kDa), 5: GpIVG9-8G (10 kDa), 6: GpG8A07-2 (30 kDa), 7: GpG8A07-3 (30 kDa), 8: Gp448-3-5C (18 kDa), 9: Gp448-3-15C (18 kDa), 10: GpG16H02-2EX (16 kDa), 11: GpG16H02-8C (16 kDa), 12: transformation control.

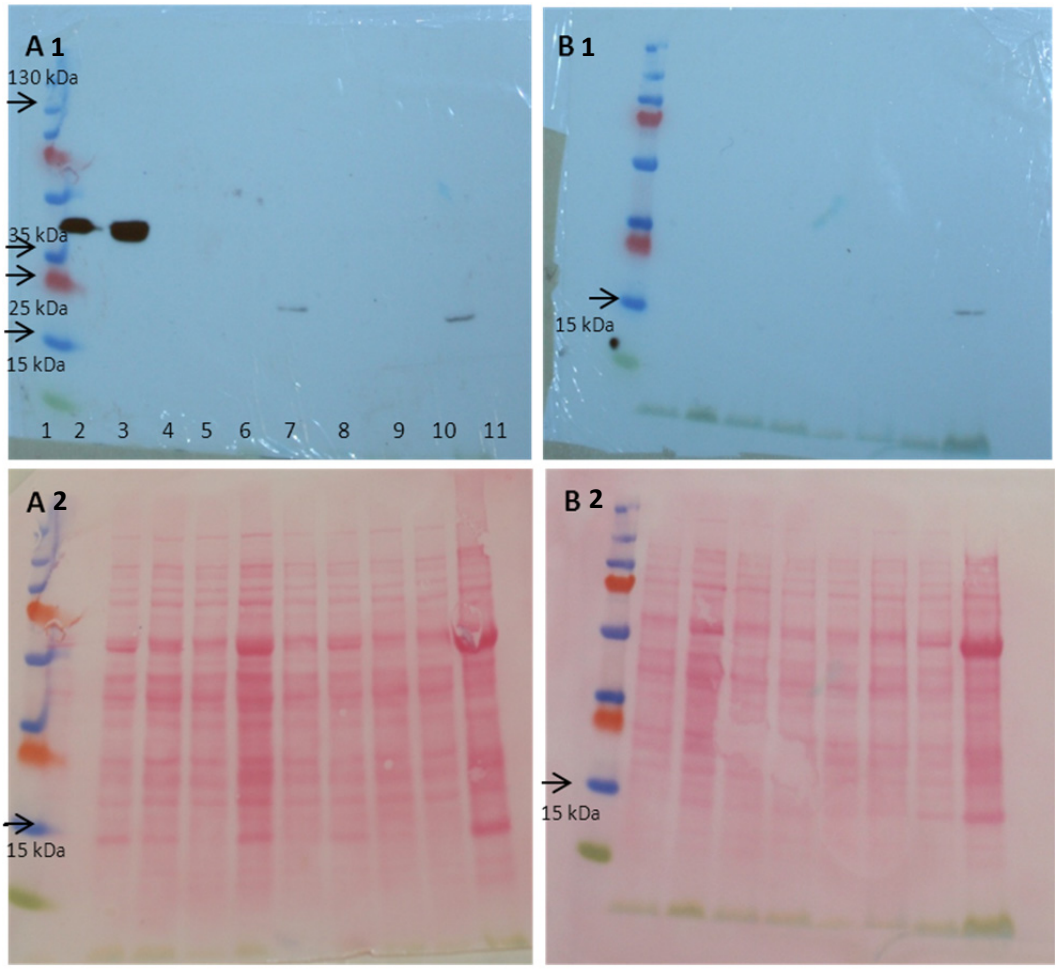


Figure 4.8 Western blot detection of HA-tagged effector proteins transiently expressed in *N. benthamiana.* Western blot detection of HA-tagged effector proteins transiently expressed in *N. benthamiana* using chemiluminescence detection and a monoclonal anti-HA antibody coupled with horse-radish peroxidase. 20 µg of soluble protein was loaded per lane. Ponceau S acid red staining of total protein on the nitrocellulose membrane to prove protein transfer to the membrane. (A) Lanes: 1: Ladder, 2: GpE9 (38.0 kDa), 3: GpE9 (38.0 kDa), 4: GpSCN1120 (6.3 kDa), 5: GpG20E03 (17.8 kDa), 6: GpG8A07 (26.3 kDa), 7: Gp448-4 (16.5 kDa), 8: GpG20E03b (17.5 kDa), 9: GpG16H02 (12.7 kDa), 10: +(Ve) (18.0 kDa), 11: transformation control. (B) Lanes: 1: Ladder, 2: GpHgsec4 (26.2 kDa), 3: GpDgl1 (3.1 kDa), 4: GpIVG9 (6.6 kDa), 5: Gp66p1 (9.1 kDa), 6: GpHg10C02 (8.0 kDa), 7: GpIA7 (5.4 kDa), 8: GpA42 (4.6 kDa), 9: +(ve) (18.0 kDa).

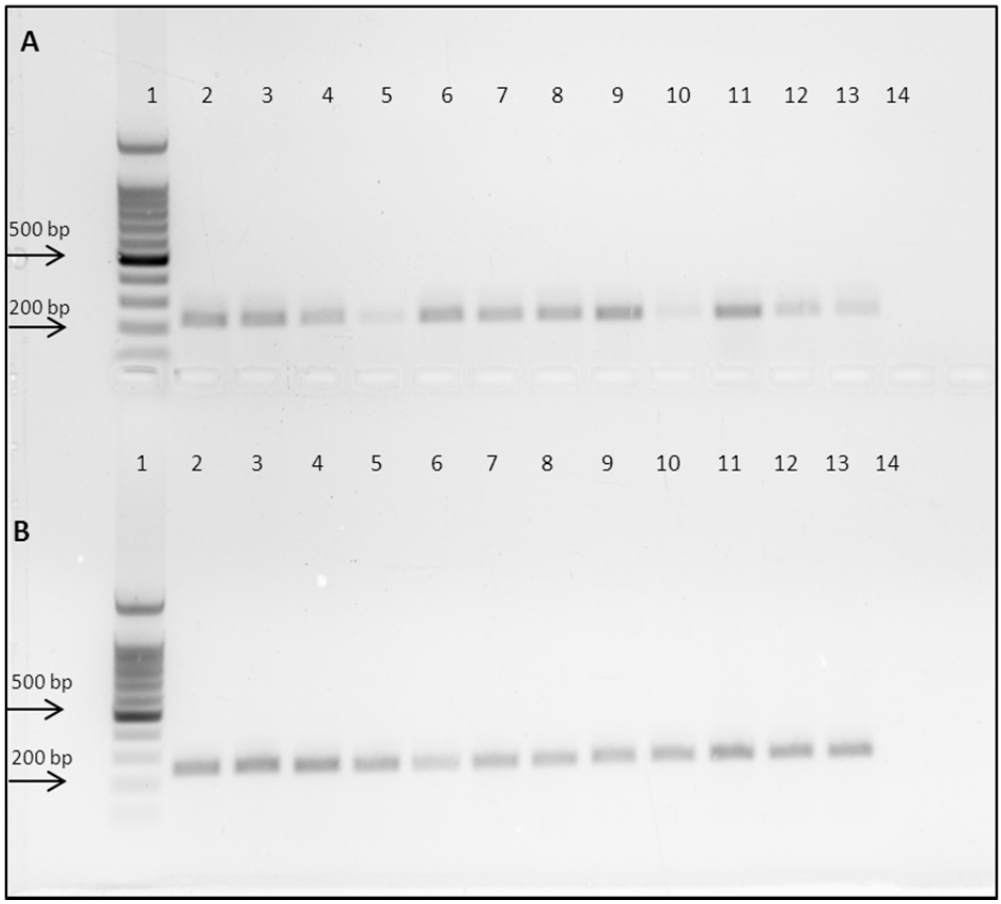


Figure 4.9: Semi Quantitative RT-PCR analysis of effector expression in transgenic potato lines. Gel electrophoresis image for the characterisation of RNA expression for transgenic lines of potato. For all effector constructs transformed into potato, 12 lines were analysed using gene specific primers and imaged together with the amplification of a reference gene EF1α from the same cDNA sample, for relative expression determination. Image above (A) cDNA from twelve GpG20E03 lines was amplified using gene specific primers, expected product size was 225bp. (B) Amplification of EF1α from the corresponding lines to show the relative amount of starting cDNA present in the PCR reaction. Expected product size was 230bp. Using this information 6 high and 1 low expressing line was selected for further phenotypic analysis and maintenance in tissue culture. Lanes: 1: Ladder, 2: 3F, 3: 4F, 4: 5F, 5: 6F, 6: 7F, 7: 8F, 8: 5G, 9: 6G, 10: 7G, 11: 8G, 12: 9G, 13: 10G and 14: (-ve). This analysis was conducted with representatives of all construct expressing potato lines.

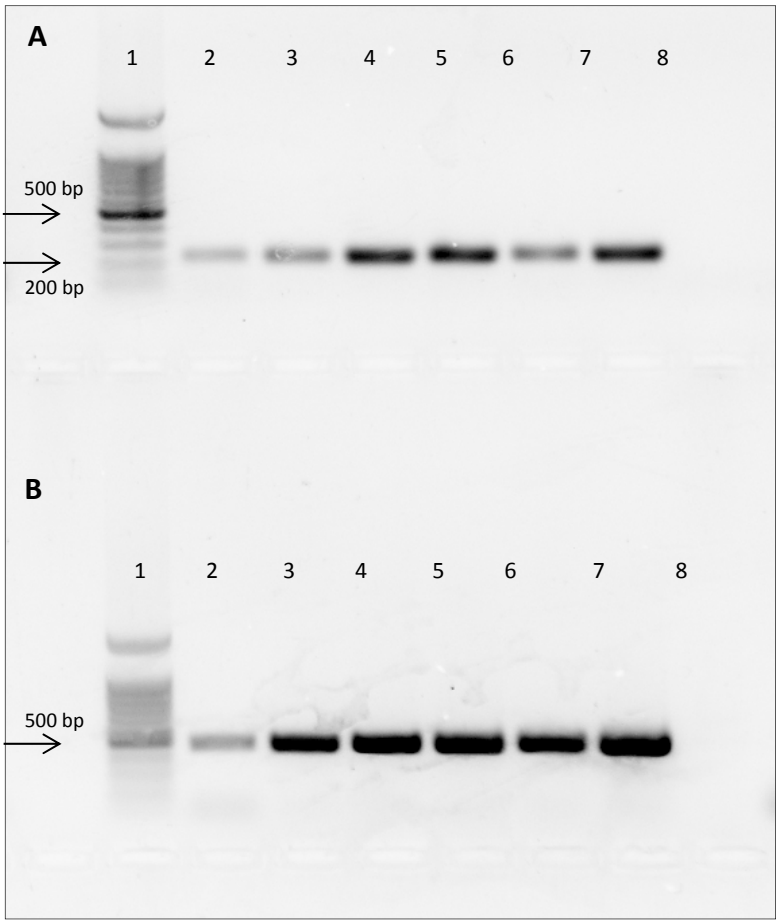


Figure 4.10: Semi quantitation RT-PCR gel electrophoresis image for the characterisation of RNA expression for transgenic lines of *Arabidopsis*. Six lines for each construct at T2 generation were analysed using gene specific primers and imaged together with the amplification of a reference gene 18S, for relative expression determination. For example in the image above (A) cDNA from six GpG7E05 lines was amplified using gene specific primers, expected product size 270bp and directly below (B) the amplification products from the corresponding lines for the 18S gene, expected product size 452bp to show the relative amount of starting cDNA present in the PCR reaction. Using this information, four lines were selected for further phenotypic analysis. Lanes: 1: Ladder, 2: GpG7E05_line 5, 3: GpG7E05_line 6, 4: GpG7E05_line 1, 5: GpG7E05_line 12, 6: GpG7E05_line 2, 7: GpG7E05_line 9, 8: (-ve) transformation control. This analysis was conducted on representatives of the following transformed *Arabidopsis* expressing constructs: GpIA7, GpG8A07, GpG7E05, GpE9, Gp66p1, GpHg10C02, GpA42, GpG20E03b, GpSCN1120 and Gp1106.

4.4.3 Phenotypic analysis of transgenic potato lines

4.4.3.1 Growth phenotypes

Two high and one lower expressing line for each effector construct were grown in glasshouse conditions. Data were collected to determine if there was any phenotypic effect(s) as a result of transgene over-expression for the following parameters at the final time point, when natural senescence of the control plant had begun: The number of leaf nodes on the main stem, any other growth abnormalities, wet and dry weight (above ground biomass), number and weight of tubers. During the experiment regular measurements of the same terminal leaf length, width and the overall plant height was recorded. Also, the number of days to complete petal opening of the first flower was recorded for each plant.

Due to the amount of glasshouse space required to carry out phenotypic analysis on sufficient potato plants to allow rigorous statistical analysis, lines for only 3-4 constructs could be analysed in one glasshouse at the same time. Therefore comparisons were only performed with controls in the same glasshouse as the transgenic line in question. The complete phenotypic analysis was carried out in two glasshouses over two separate occasions. In the first experiment lines expressing GpIA7, Gp448-3, GpG8A07, GpDGL1, GpG16H02 and GpG20E03b were subjected to phenotypic analysis. An initial ANOVA showed that GpG16H02 ($P<0.01$) and GpIA7 ($P<0.05$) had longer leaves when compared to the control. GpIA7 plants were also significantly smaller than control plants ($P<0.05$). GpG16H02 leaves were wider than the control ($P<0.05$). GpG8A07 ($P<0.05$), GpIA7 ($P<0.05$), GpG16H02 ($P<0.01$) and GpG20E03b ($P<0.01$) produced heavier tubers than the control. GpG16H02 produced more tubers than the control ($P<0.05$) and GpG8A07 had more leaf nodes than the control ($P<0.05$). This analysis did not take expression levels into consideration and took all plants expressing a gene as a single group.

If all lines are statistically analysed separately, therefore separating out high (High) and lower (Low) expressing lines in the analysis, with a post hoc analysis (ANOVA, with Tukey post hoc). GpIA7 -5C (High) ($P<0.01$), GpIA7 -6C (High) ($P<0.01$) and GpG8A07 -3 (Low) ($P<0.05$), GpG8A07 -1 (High) ($P<0.01$) had heavier tubers than the control (Figure 4.11). GpHg10C02_1E (High) had more leaf nodes ($P<0.05$) and a greater dry weight mass compared to the control ($P<0.001$) (Figure 4.14).

GpG8A07 -2 wet weight was significantly less than the control (P<0.05) (Figure 4.11). GpG8A07 -3 (Low) (P<0.05), GpG8A07 -1 (High) (P<0.01) (Figure 4.11), GpG16H02 -8E (Low) (P<0.05) (Figure 4.13) and GpG7E05_4 (High) (P<0.01) had a significantly greater number of tubers (Figure 4.14).

Also GpG16H02 -8E (Low) (P<0.05), GpG16H02 -7E (High) (P<0.01) and GpG20E03B 11 (High) (P<0.05), GpG20E03B -13 (High) (P<0.001) (Figure 4.13) had heavier tubers than the control. Whereas tubers from GpHgSEC4_2E (High) (P<0.05) and GpHgSEC4_16E (High) (P<0.01) weighed significantly less than the control (Log transformed due to unacceptable residual variation) (Figure 4.14). GpG20E03b -5b (Low) (P<0.05) (Figure 4.13) and GpSCN1120 -1F (High) (P<0.001) dry weight was significantly less than the control (Figure 4.16). GpIA7 -4C (Low) (P=0.001) and 5C (High) (P<0.0001) flowered significantly later than the control (Figure 4.11) (Tukey post hoc analysis).

At the final time point all three lines of GpG16H02 (5E and 7E P<0.05, 8E P<0.01) (Figure 4.13) had longer leaves than the control, GpG16H02 -5E (High) had wider leaves than the control (P<0.001), all three lines of GpIA7 were significantly smaller than the control, 4C (P<0.01), 5C (P<0.01), 6C (P<0.01) (Figure 4.12) and GpG8A07 -2 (High) was also smaller than the control (P<0.05). GpIVG9 -2F (High) and GpHgSEC4 -2E (High) were significantly taller than the control (P<0.01) (Figure 4.15). GpSCN1120 -1G (Low) had a significantly lower length/width ratio than the control (P<0.05) (Figure 4.16) (Tukey post hoc analysis).

Over the course over the experiment certain variables were measured regularly and analysed for significant difference. Recording of the measurements over time showed GpG16H02 -5E and GpG16H02 -8E had longer leaves than the control (P<0.05 and P<0.01 respectively). Recording of the leaf width measurements over time showed that GpG16H02 -5E, GpG16H02 -8E and GpHg10C02 -1E had wider leaves than the control (P<0.01, P<0.05 and P<0.05, respectively). GpIA7 -6C (P<0.05), GpIA7 -4C and GpIA7 -5C (P<0.01) grew slower and were therefore smaller than the control. Whereas GpIVG9 -2F (High) and GpHgSEC4 -2E (High) grew significantly faster than the control (P<0.01 and P<0.05, respectively). GpIVG9 -2F grew leaves with a lower width/ length ratio compared to the control (P<0.05) (Repeated measurements ANOVA).

4.4.3.2 Growth phenotypic observations

During phenotypic analysis of potato plants grown in glasshouse conditions, a few observations were seen regarding their physical phenotypes. Leaves from plants transformed with a construct containing GpIA7 seemed to contain a greater number of white spots on the leaves (Figure 4.17). The white spots could be a stress response.

Leaves from plant transformed with a construct containing GpIVG9 seemed to have a curly leaf phenotype (Figure 4.18). The morphology of the leaves was changed and the stems seemed weaker when handling.

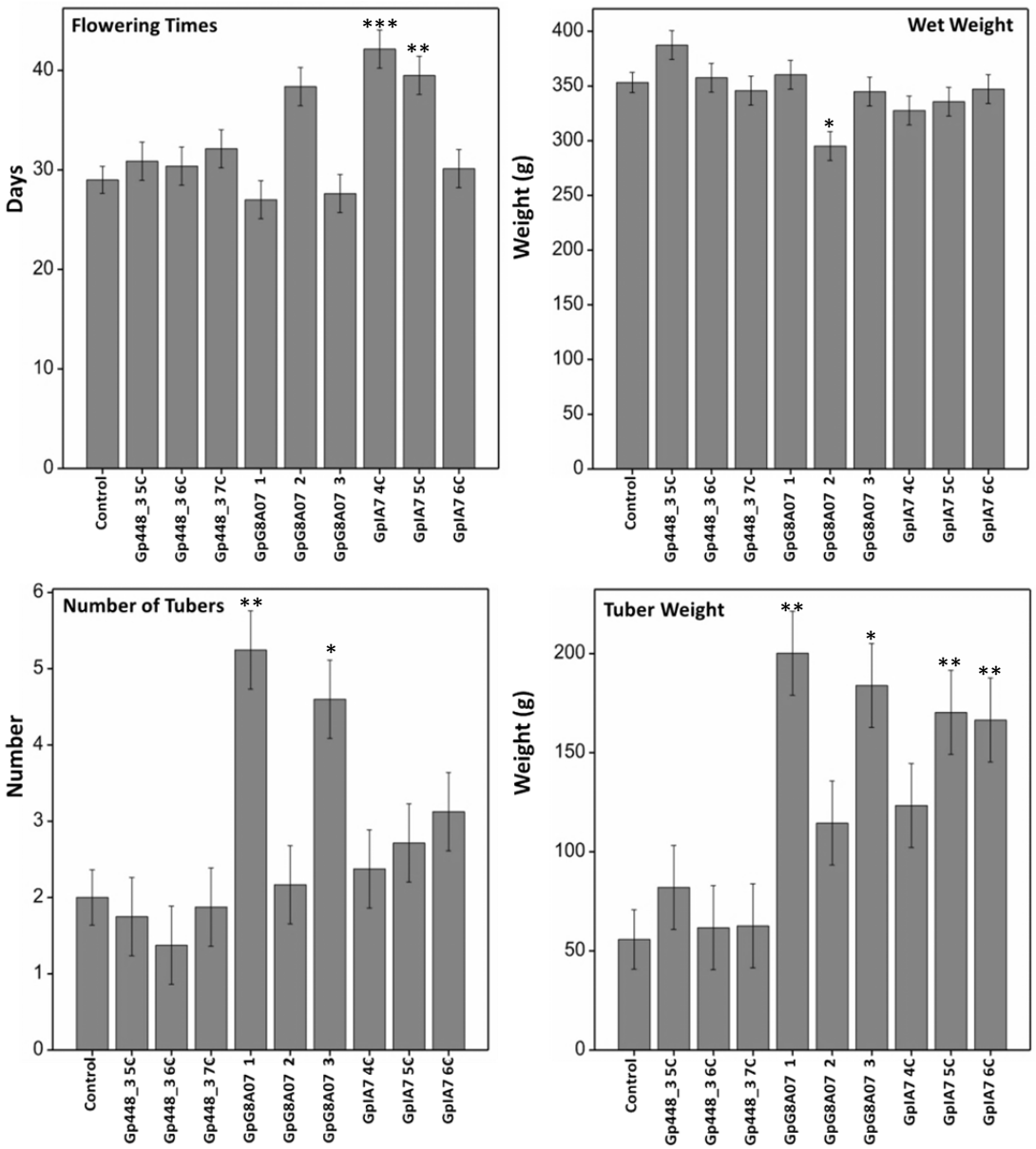


Figure 4.11: Growth phenotypes observed for transgenic potato lines.. Graphical representation of the mean and standard error of the mean for variables measured during growth phenotypic analysis of potato lines expressing Gp448-3, GpG8A07 and GpIA7. Two high and one lower expressing line were chosen to represent the effector construct expressing potatoes, expression was determined via semi-quantitative RT-PCR. N=8 for each effector expressing line and N=24 for the control group. Significance was determined using a highly stringent Tukey post hoc analysis. **Gp448-3:** -5C (Low), -6C (High) -7C (High). **GpG8A07:** -3 (Low), -1 (High), -2 (High). **GpIA7:** -5C (High), -6C (High),-4C (Low).)

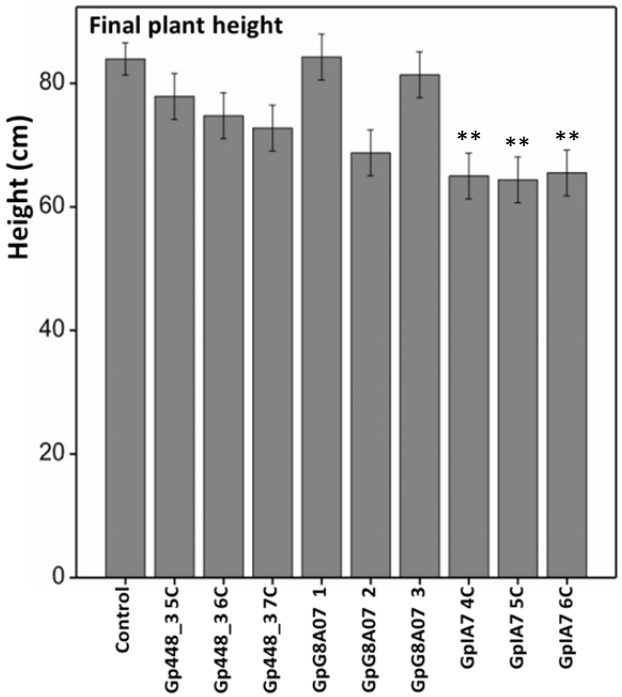


Figure 4.12: Final plant height for transgenic potato lines expressing Gp448-3, GpG8A07 and GpIA7. Graphical representation of the mean and standard error of the mean for the final plant height measured during growth phenotypic analysis of potato lines expressing Gp448-3, GpG8A07 and GpIA7. Two high and one lower expressing line were chosen to represent the effector construct expressing potatoes, expression was determined via semi-quantitative RT-PCR. N=8 for each effector expressing line and N=24 for the control group. Significance was determined using a highly stringent Tukey post hoc analysis. **Gp448-3:** -5C (Low), -6C (High) -7C (High). **GpG8A07:** -3 (Low), -1 (High), -2 (High). **GpIA7:** -5C (High), -6C (High),-4C (Low).

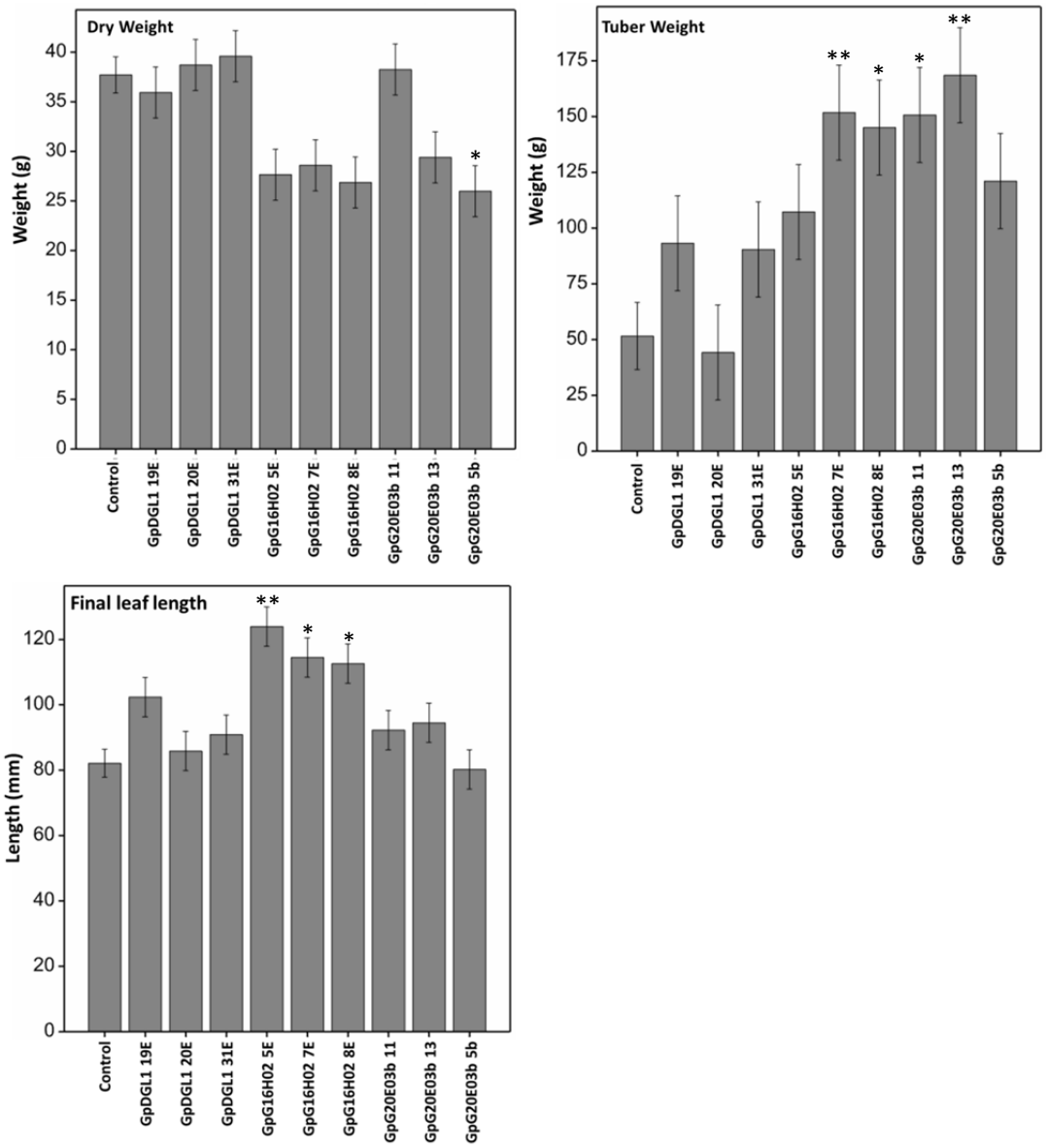


Figure 4.13: Growth phenotypes observed for transgenic potato lines. Graphical representation of the mean and standard error of the mean for variables measured during growth phenotypic analysis of potato lines expressing GpDGL1, GpG16H02 and GpG20E03b. Two high and one lower expressing line were chosen to represent the effector construct expressing potatoes. N=8 for each effector expressing line and N=24 for the control group. Significance was determined using a highly stringent Tukey post hoc analysis. **GpDGL1**: -19E (Low), -20E (High), 31E (High). **GpG16H02**: -8E (Low), -7E (High) -5E (High). **GpG20E03B**: -11 (High), -13 (High), -5b (Low).

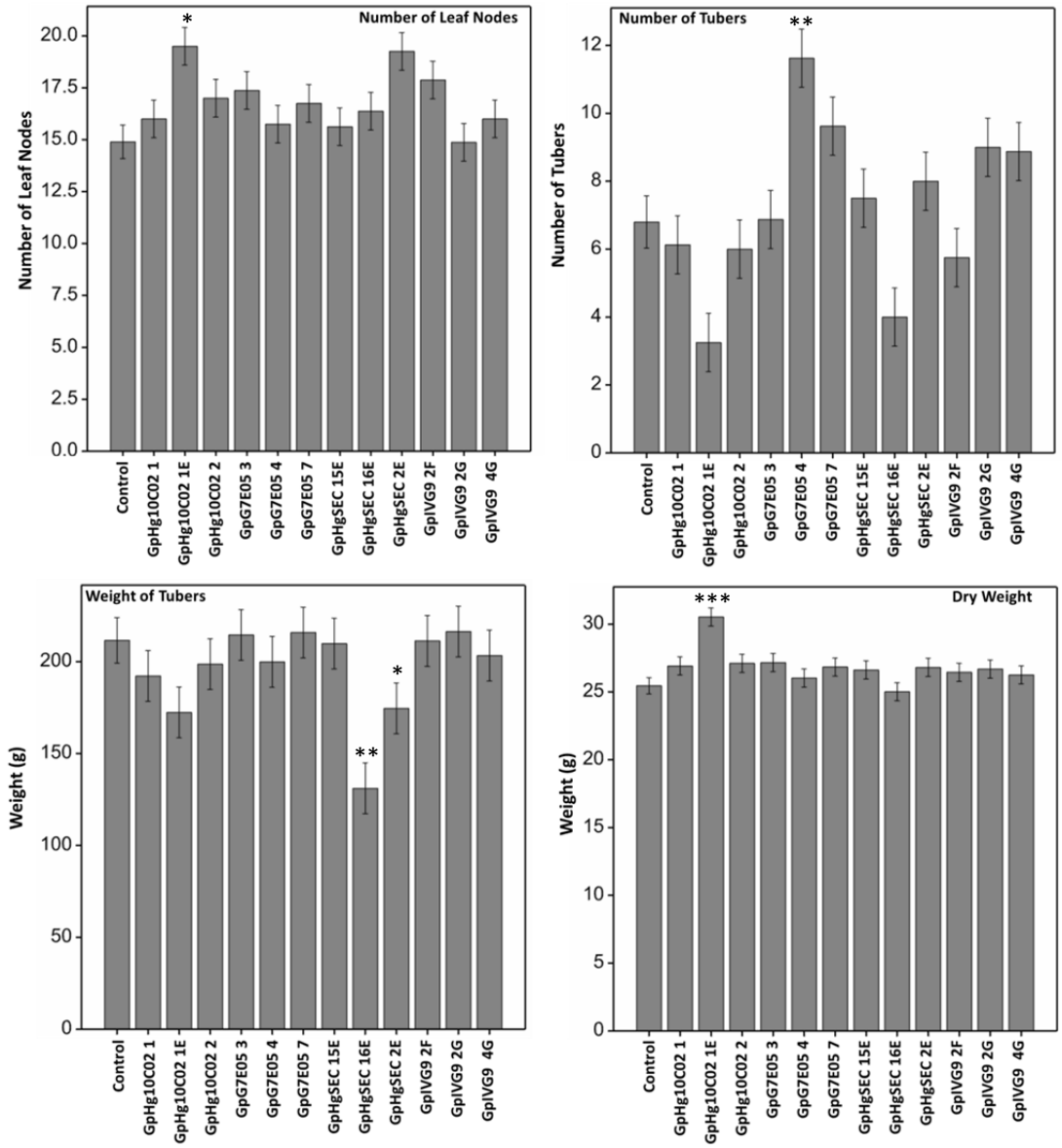


Figure 4.14: Growth phenotypes observed for transgenic potato lines. Graphical representation of the mean and standard error of the mean for variables measured during growth phenotypic analysis of potato lines expressing GpHg10C02, GpG7E05, GpHgSEC4 and GpIVG9. Two high and one lower expressing line was chosen to represent the effector construct expressing potatoes. N=8 for each effector expressing line and N=24 for the control group. Significance was determined using a highly stringent Tukey post hoc analysis. **GpHg10C02**: -1 (High), -1E (High), -2 (Low). **GpG7E05**: -3 (High), -4 (High), -7 (Low). **GpHgSEC**: -15E (Low), -2E (High), -16E (High). **GpIVG9**: -2F (High) -2G (High), 4G (Low).

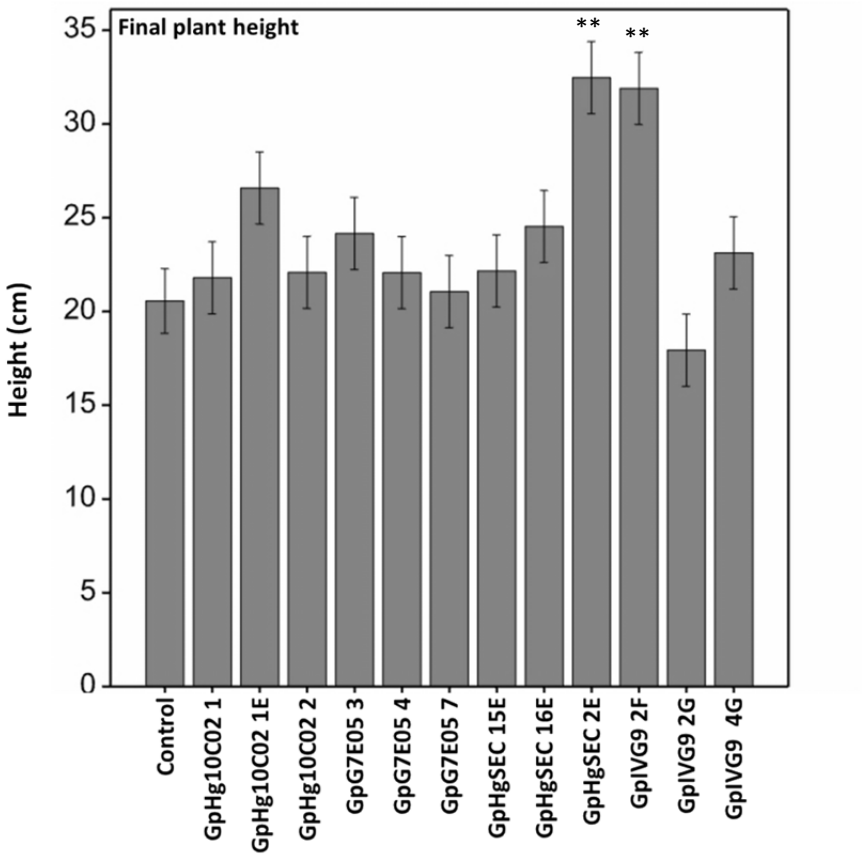


Figure 4.15: Final plant height for transgenic potato lines expressing GpHg10C02, GpG7E05, GpHgSEC4 and GpIVG9. Graphical representation of the mean and standard error of the mean for the final plant height measured during growth phenotypic analysis of potato lines expressing GpHg10C02, GpG7E05, GpHgSEC4 and GpIVG9. Two high and one lower expressing line were chosen to represent the effector construct expressing potatoes. N=8 for each effector expressing line and N=24 for the control group. Significance was determined using a highly stringent Tukey post hoc analysis. **GpHg10C02**: -1 (High), -1E (High), -2 (Low). **GpG7E05**: -3 (High), -4 (High), -7 (Low). **GpHgSEC**: -15E (Low), -2E (High), -16E (High). **GpIVG9**: -2F (High) -2G (High), 4G (Low).

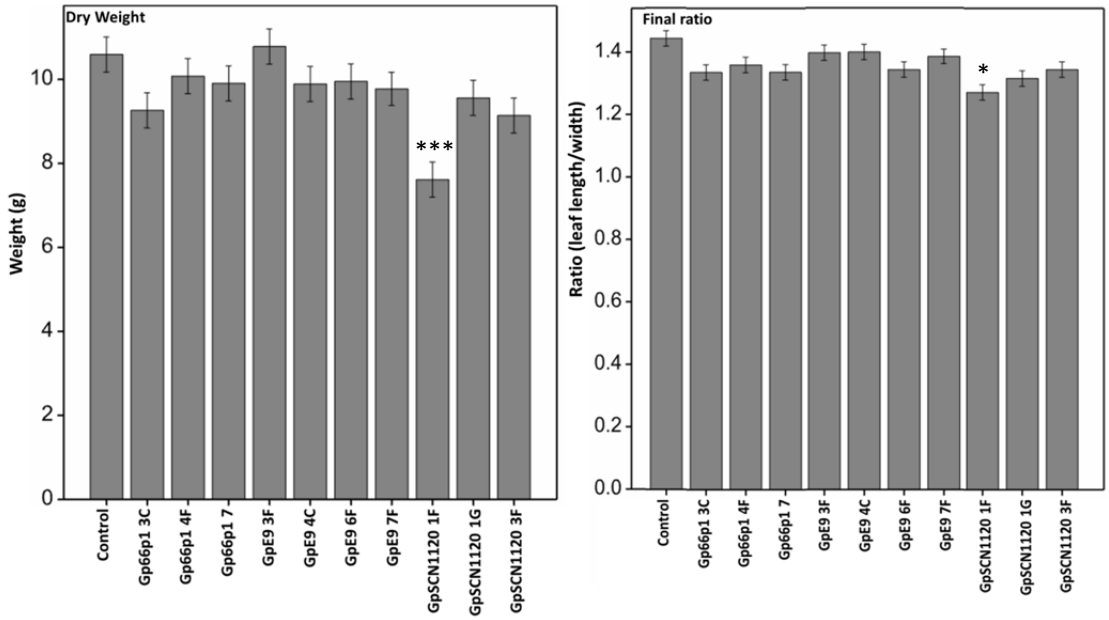


Figure 4.16: Growth phenotypes observed for transgenic potato lines. Graphical representation of the mean and standard error of the mean for variables measured during growth phenotypic analysis of potato lines expressing Gp66p1, GpE9 and GpSCN1120. Two high and one lower expressing line were chosen to represent the effector construct expressing potatoes. N=8 for each effector expressing line and N=24 for the control group. Significance was determined using a highly stringent Tukey post hoc analysis. **Gp66p1**: -3C (High), -4F (Low), -7 (High). **GpE9**: -4C (Low), 6F (High), 7F (High). **GpSCN1120**: -1F (High), -5C (High), -1G (Low).

Figure 4.17: White leaf spot phenotype observed for GpIA7 transgenic lines. (A) Transgenic lines expressing a construct containing GpIA7 had leaves that had more white spots, as an observation that the control plant leaves. (B) Transformation control plant showing leaves of normal morphology.

Figure 4.18: Curly leaf phenotype observed for GpIVG9 transgenic lines. (A and C) Transgenic lines expressing a construct containing GpIVG9 had leaves that showed a curly leaf–like phenotype. (B and D) Transformation control plants showing leaves of normal morphology.

4.4.3.3 Nematode infection assay

A high and a lower expressing line for all effector gene expressing potatoes were subjected to a nematode infection assay. This was to determine if the expression of the effector in potato altered its susceptibility to nematode infection, and if the change in susceptibility was proportional to the expression of the transgene. The data (number of nematodes per gram of root) was log10 transformed due to an unacceptable level of residual variation.

There was no significant difference in the number of nematodes present in the root at 14 dpi (ANONVA, Tukey post hoc) or in their stage of development per gram of root (Kruskal-Wallis with Bonferroni post hoc correction). At six weeks post infection, line GpHgSEC4 -2E (High) had significantly fewer nematodes ($P<0.05$) and line GpDGL1 20E (High) had significantly more nematodes per gram of root ($P<0.01$) when compared to the control. No other significant difference was observed (Tukey post hoc analysis).

Using a less stringent statistical analysis (ANOVA) of the number of nematodes per weight of root 2 weeks post infection showed that transgenic potato expressing GpG7E05 had lower numbers of nematodes ($P<0.05$). No other significant data was found in this experiment (Figure 4.19). The life stage of the observed nematodes was recorded. No nematode had yet developed beyond J3 stage. There was no significant difference in the proportions of J2/J3 (Figure 4.20).

Statistical analysis (ANOVA) of the number of nematodes per weight of root 6 weeks post infection showed that high expressing lines for GpDGL1 and GpIVG9 were more susceptible to nematode infection ($P<0.01$). Interestingly, the GpG7E05 high expressing line ($P<0.05$), the GpHgSEC4 high expressing line ($P<0.01$), the Gp448-4 high expressing line ($P<0.05$), the Gp448-3 high ($P<0.01$) and the low expressing line ($P<0.05$) were all less susceptible to nematode infection. Nematode infection rates on transgenic GFP plants were not statistically different to the transformation control group (Figure 4.21). The life stage of the observed nematodes was recorded. There was no significant difference in the proportions of the life stages observed (Figure 4.22).

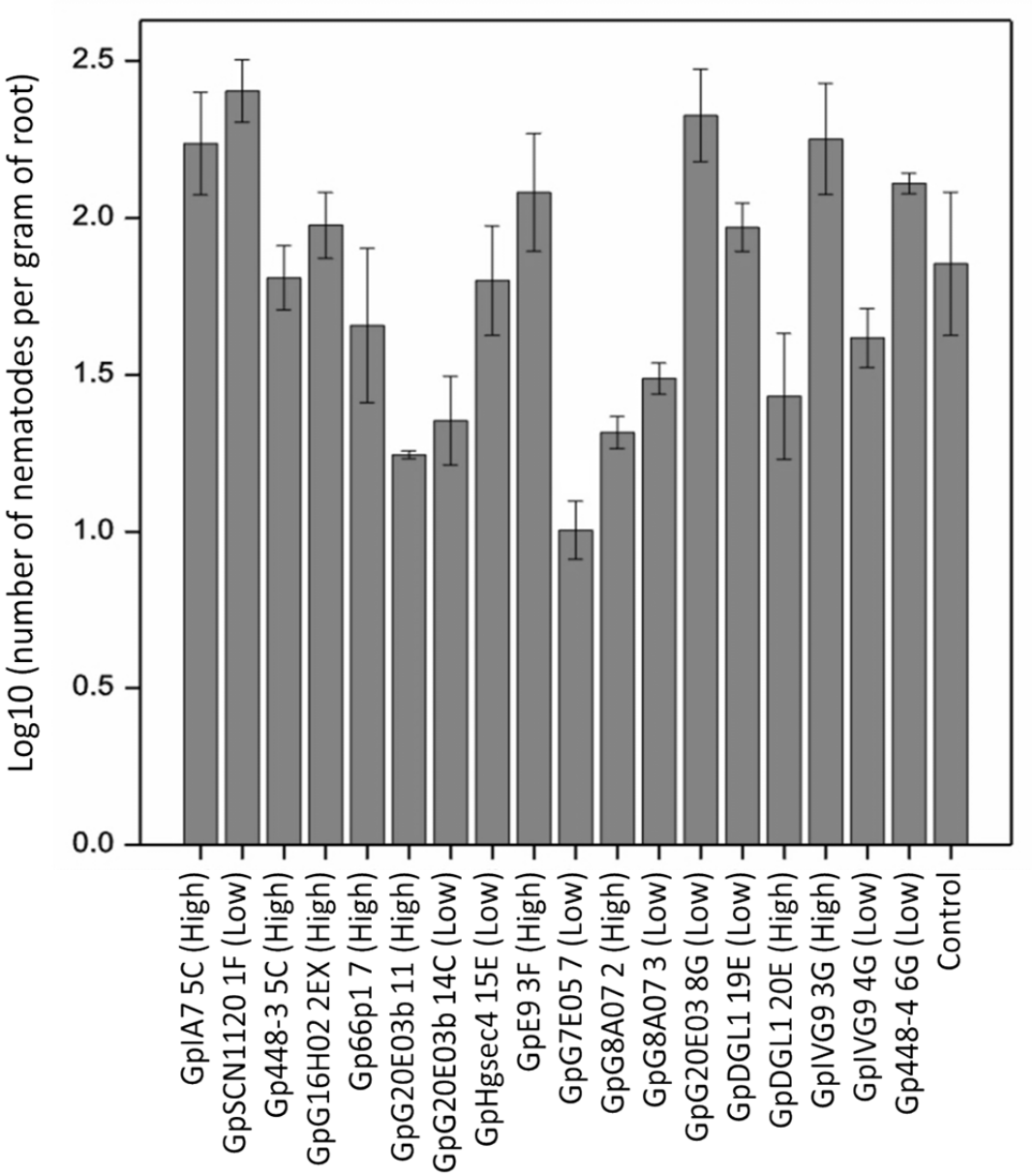


Figure 4.19: Total number of nematodes in transgenic potato roots 14 days post infection. Graph representing Log10 number of nematodes per gram of root for transgenic potato lines 14 days after infection with 600 nematodes. Error bars represent the standard error of the mean. One high and one lower expressing line was chosen to represent the effector construct expressing potatoes. N=3 for each effector expressing line and N=15 for the control group. Significance was determined using a highly stringent Tukey post hoc analysis.

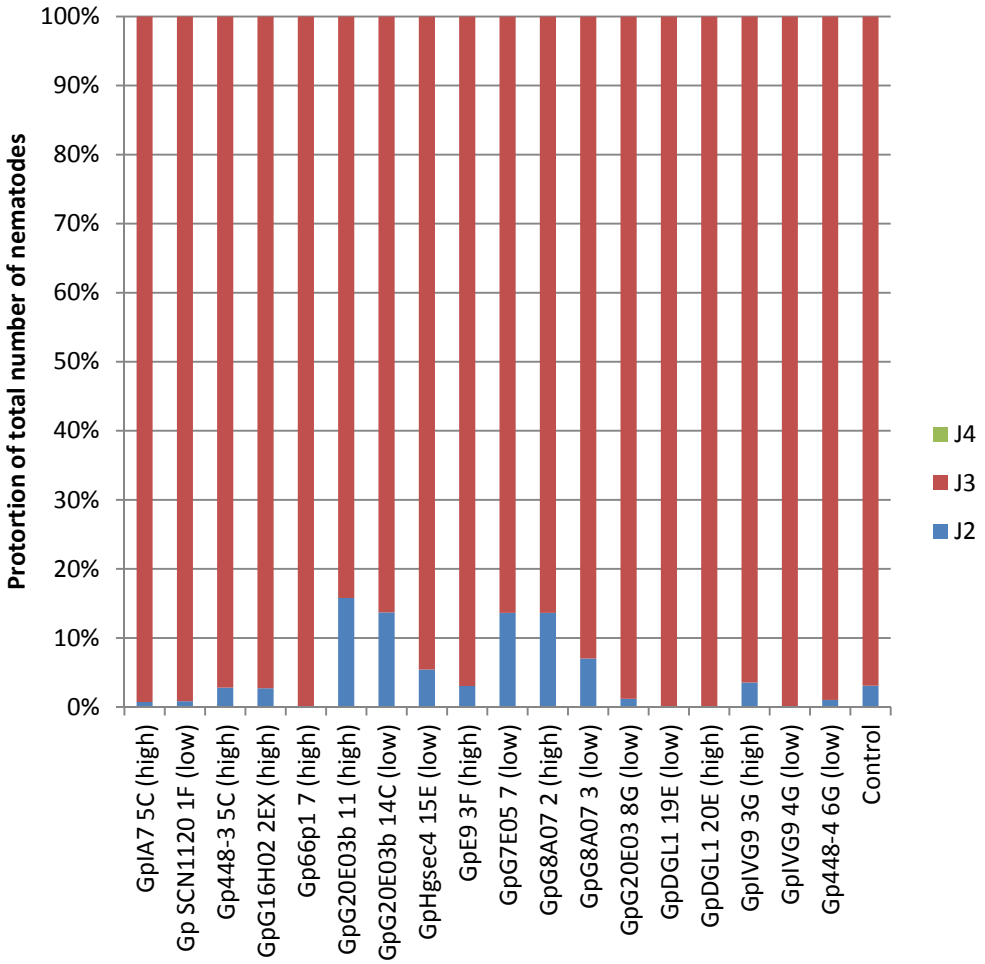


Figure 4.20: Graph representing the relative proportions of the nematodes identified in each life stage, plotted as a percentage of the total number found 14 days post infection. Life stages for nematodes identified in the infection assay were recorded. The data are represented as an average per line, as a percentage of the average total number per line. No identified nematodes had reached J4 life stage. No significant difference was observed. One high and one lower expressing line was chosen to represent the effector construct expressing potatoes. N=3 for each effector expressing line and N=15 for the control group. Significance was determined using a highly stringent Kruskal-Wallis analysis with Bonferroni post hoc correction.

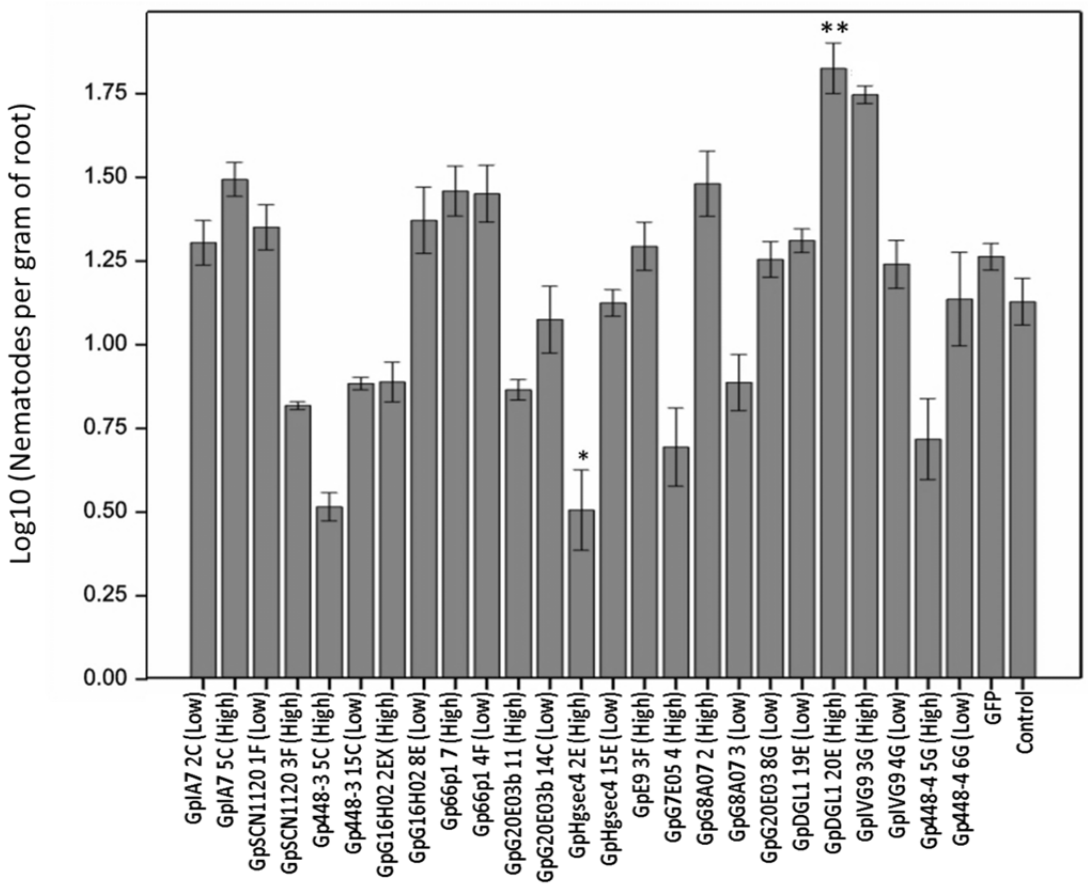


Figure 4.21: Total number of nematodes in transgenic potato roots six weeks post infection. Graph representing Log10 number of nematodes per gram of root for transgenic potato lines six weeks after infection with 600 nematodes. Error bars represent the standard error of the mean. One high and one lower expressing line was chosen to represent the effector construct expressing potatoes. N=5 for each effector expressing line and N=15 for the control group. Significance was determined using a highly stringent Tukey post hoc analysis.

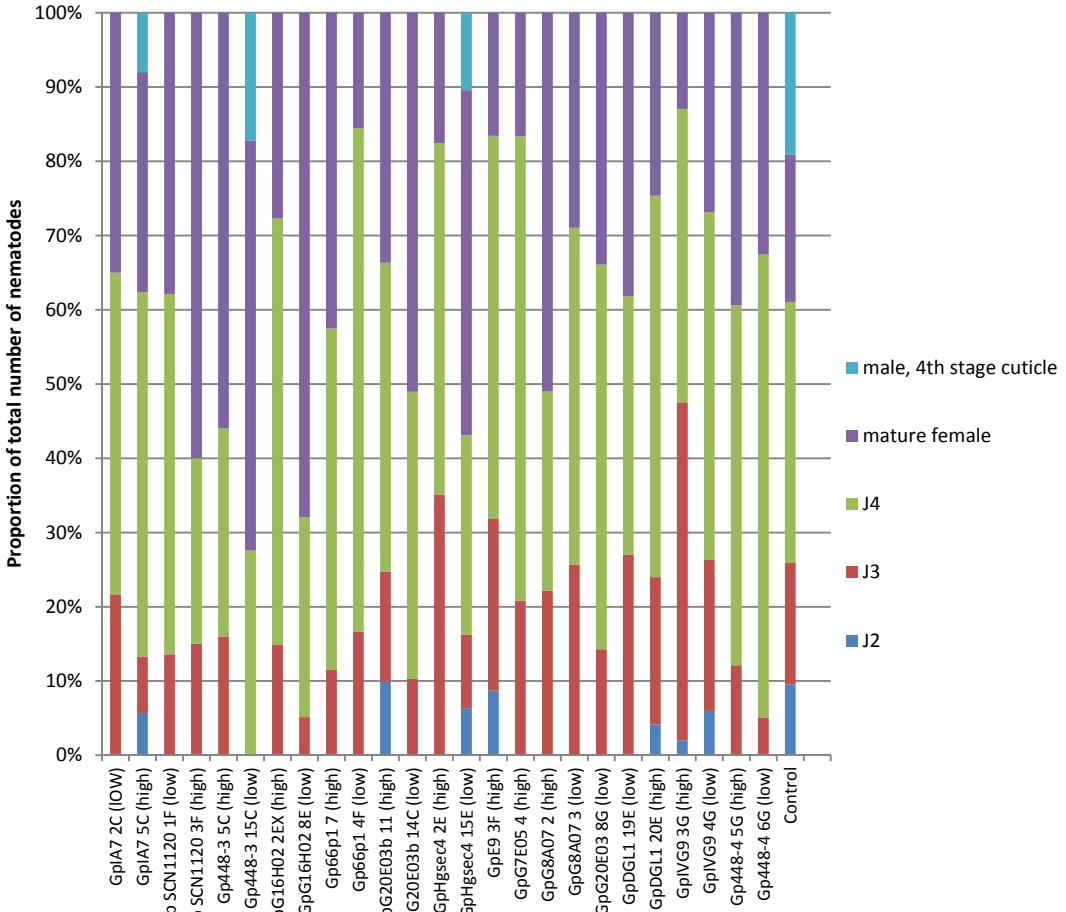


Figure 4.22: Graph representing the relative proportions of the nematodes identified in each life stage, plotted as a percentage of the total number found six weeks post infection. Life stages for nematodes identified in the infection assay were recorded. The data are represented as an average per line, as a percentage of the average total number per line. No significant difference was observed. One high and one lower expressing line was chosen to represent the effector construct expressing potatoes. N=5 for each effector expressing line and N=15 for the control group. Significance was determined using a highly stringent Kruskal-Wallis analysis with Bonferroni post hoc correction.

.

4.4.3.4 *Phytophthora* CS-12 detached leaf assay

An attenuated strain of *Phytophthora infestans,* in which the AVR3a effector gene is down-regulated by the transgenic expression of a double hairpin construct (Bos *et al.*, 2010b), was used to determine if any of the potato lines expressing effector genes were more susceptible to the attenuated *P. infestans.* A computer program was developed and used to quantify the infection zone in the *Phytophthora* CS-12 detached leaf assay as a percentage of the leaf showing symptoms. A high and lower expressing transgenic potato line representing each effector construct was used in the assay.

The data output from the computer program was not normally distributed, and the data was not normally distributed following Log10 normalisation. Therefore a non-parametric test was used (Kruskal-Wallis with post hoc Wilcoxon-Bonferroni correction). Due to the significant variation between controls from different boxes, each box had its own control and analysis was only performed on a per box basis.

No significant differences were found between most transgenic lines and their corresponding control. However, the high expressing line IA7 -5C had significantly greater disease symptoms (as quantified by the computer program) when compared to the control leaves in the same box ($P<0.001$) (Figure 4.23). Line E9 3F appeared as though it was more susceptible to the pathogen. However, statistical analysis of the disease symptoms as quantified by the program, showed it was not ($P=0.053$) (Figure 4.24).

The time taken for first sign of disease symptoms was recorded. Using a non-parametric test (Kruskal-Wallis with post hoc Wilcoxon-Bonferroni correction) E9 3F showed symptoms significantly earlier than the control leaves in the same box ($P<0.05$). IA7 -5C which had significantly more disease symptoms (as shown above) also showed disease symptoms earlier than the control ($P=0.05$) (Figure 4.25). Images of the leaves from the detached leaf assay can be visually seen with their corresponding control leaves are presented for IA7 -5C (Figure 4.26) and E9 3F (Figure 4.27).

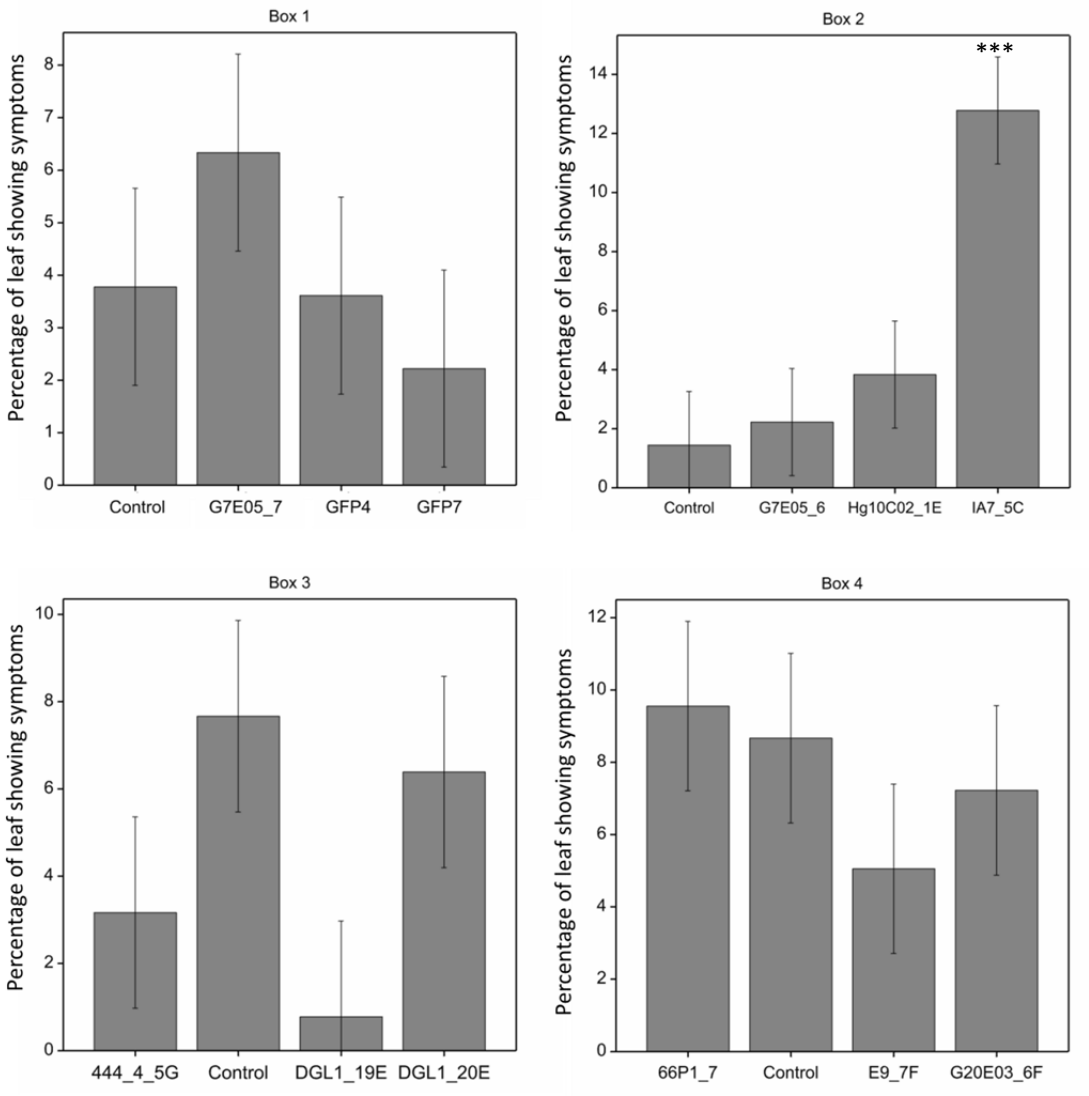


Figure 4.23: Graphs representing the area of the leaves quantified as showing disease symptoms when inoculated with *Phytophthora infestans* CS-12 five days post inoculation. Graphs representing the mean and standard error of the mean for boxes 1-4, for the output of a computer program that quantified the number of pixels in an image relating to disease symptoms. The results were returned as a percentage of the leaf that was showing disease symptoms. The leaf images were scored over 6 time points. One high and one lower expressing line was chosen to represent the effector construct expressing potatoes. N=3 (12 spots on 3 leaves) for each effector expressing line and for the control group. Significance was determined using a highly stringent Kruskal-Wallis analysis with Bonferroni post hoc correction.

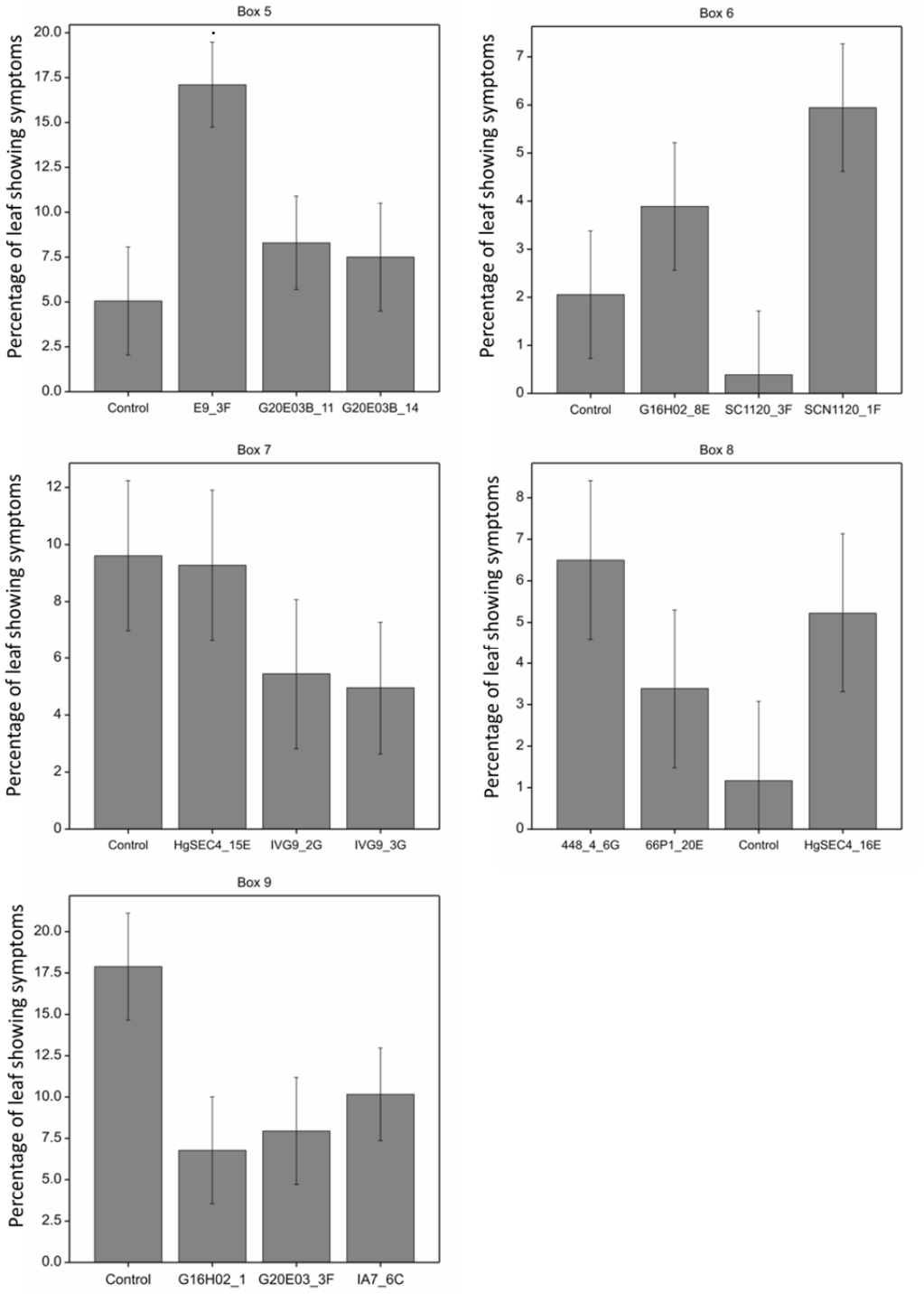


Figure 4.24: Graphs representing the area of the leaves quantified as showing disease symptoms when inoculated with *Phytophthora infestans* CS-12 five days post inoculation. Graphs representing the mean and standard error of the mean for boxes 5-9, for the output of a computer program that quantified the number of pixels in an image relating to disease symptoms. The results were returned as a percentage of the leaf that was showing disease symptoms. The leaf images were scored over 6 time points. One high and one lower expressing line was chosen to represent the effector construct expressing potatoes. N=3 (12 spots on 3 leaves) for each effector expressing line and for the control group. Significance was determined using a highly stringent Kruskal-Wallis analysis with Bonferroni post hoc correction.

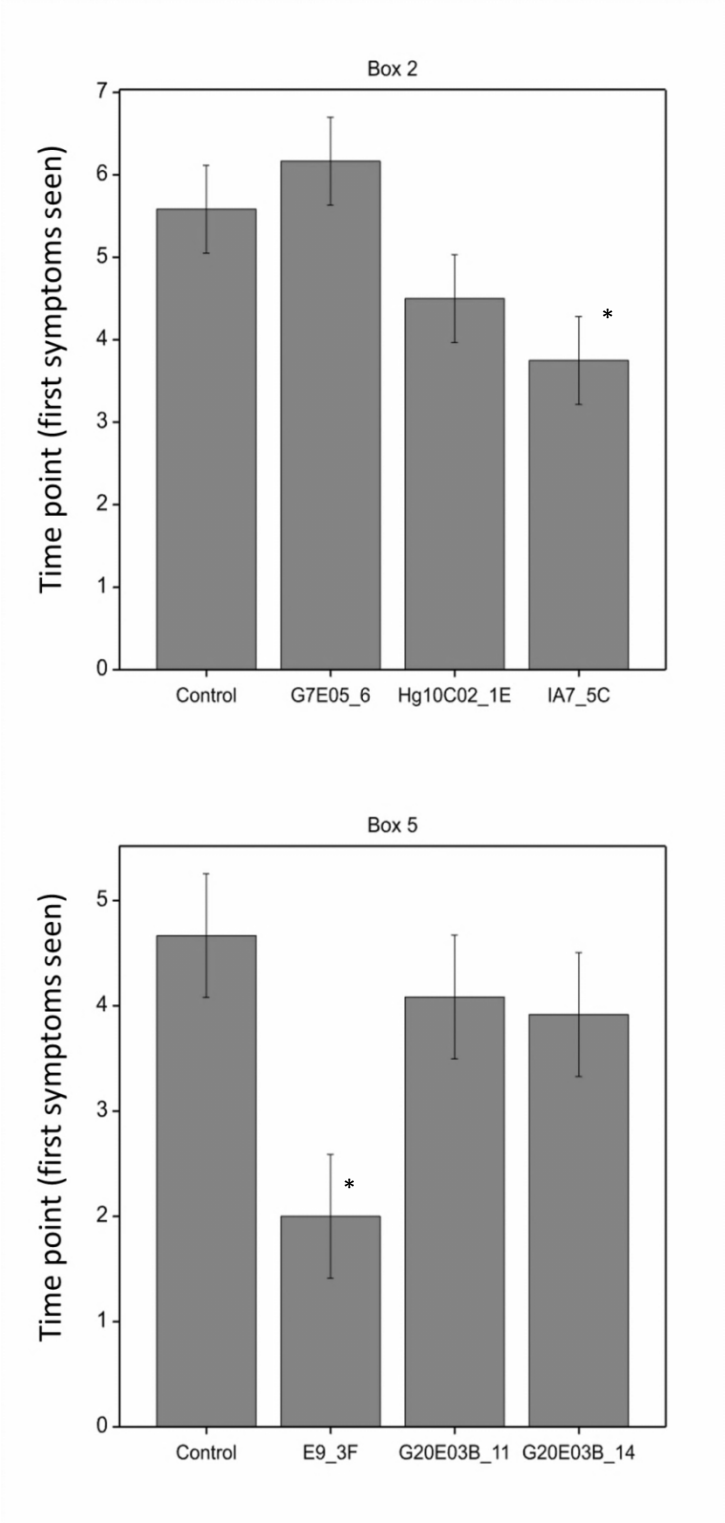


Figure 4.25: Graphs representing the time taken for first symptoms to be seen for transgenic and control lines inoculated with *Phytophthora infestans* CS-12. Data is presented as the mean and standard error of the time taken for first symptoms of disease to be seen for boxes 2 and 5 where a significant difference was observed. One high and one lower expressing line was chosen to represent the effector construct expressing potatoes. N=3 (12 spots on 3 leaves) for each effector expressing line and for the control group. Significance was determined using a highly stringent Kruskal-Wallis analysis with Bonferroni post hoc correction.

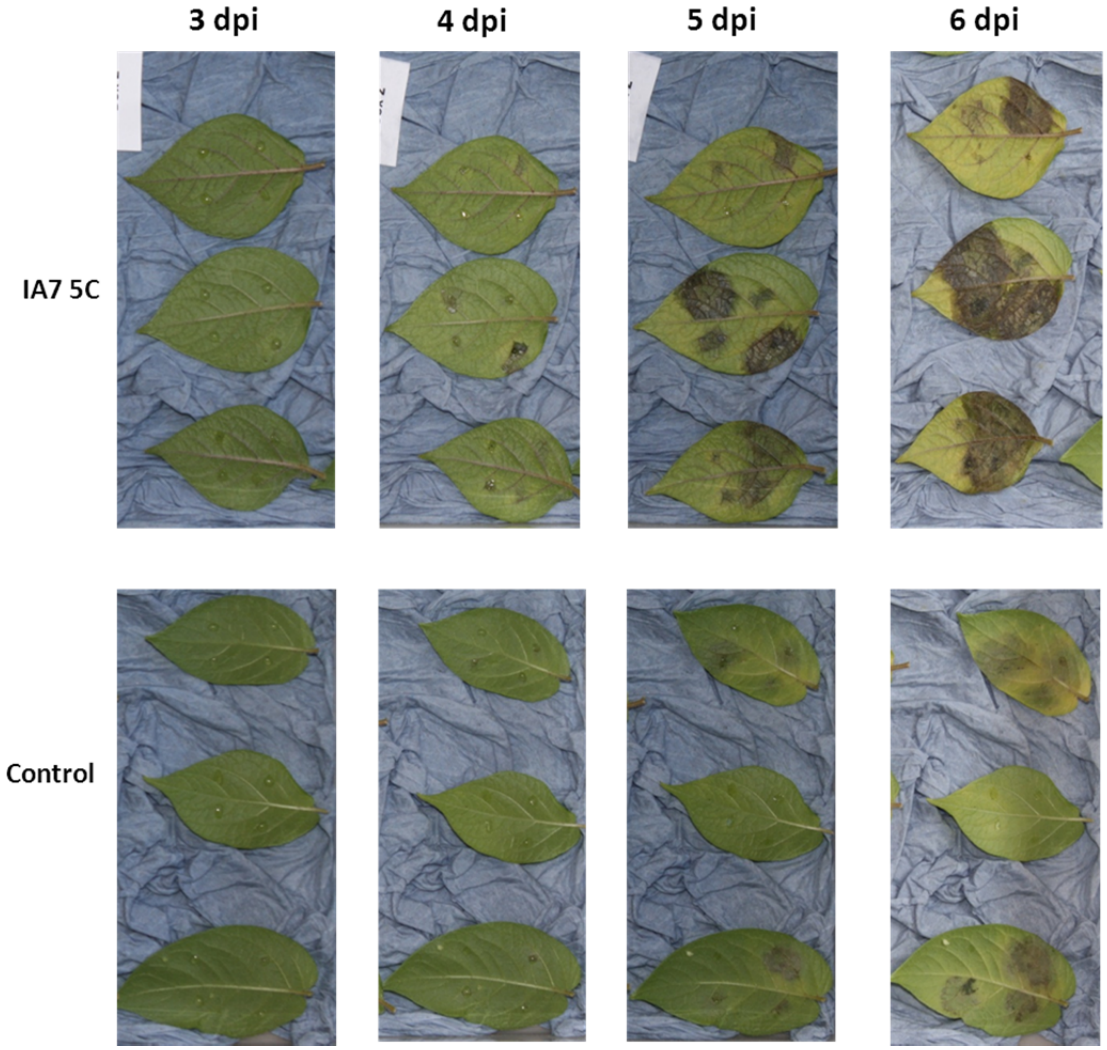


Figure 4.26: Images from *Phytophthora infestans* CS-12 detached leaf assay for transgenic line IA7 5C and the control leaves five days post infection. Four 10 µl spots, containing 150 sporangia were inoculated per leaf. The inoculated leaves were incubated at room temperature.

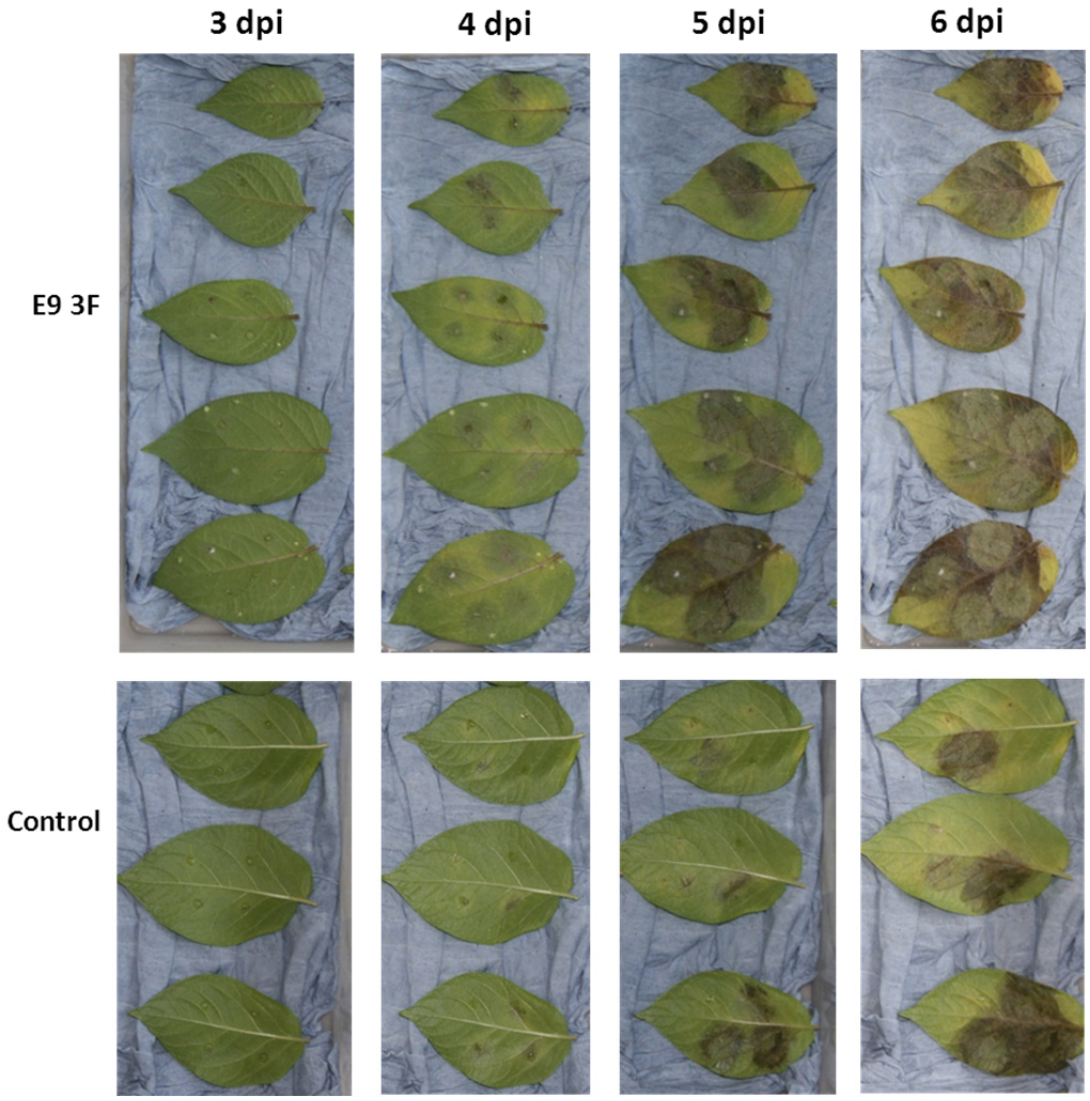


Figure 4.27: Images from *Phytophthora infestans* CS-12 detached leaf assay for transgenic line E9 3F and the control leaves five days post infection. Four 10 µl spots, containing 150 sporangia were inoculated per leaf. The inoculated leaves were incubated at room temperature.

4.4.4 *Arabidopsis* phenotypic analysis

4.4.4.1 Lemnatec data

Four lines from each of the following effector constructs were subjected to high throughput phenotypic analysis. These lines were GpIA7, GpG7E05, GpE9, Gp66p1, GpA42, GpG20E03b, GpSCN1120 and Gp1106. Controls for this experiment were a GFP expressing line and a transformation control group. Although the seeds for this experiment were sent to the Lemnatec service provider many months before the scheduled write up time for this project, at the time of writing the company had not completed the work and provided the data.

4.5 Discussion

Transgenic expression of plant-parasitic nematode effector genes in plants has been shown to be a useful tool in characterising their mode of action (Doyle and Lambert, 2003; Hamamouch *et al.*, 2012b; Hewezi *et al.*, 2010b; Huang *et al.*, 2006b; Patel *et al.*, 2010a). During the cloning process a number of incorrect amplification products of varying sizes and sequences were observed. *G. pallida* effectors exist in gene families and therefore these amplification products, which were of incorrect size and sequence, could well have been different members of the gene family. Out of 45 effectors cloned from *G. pallida*, 17 were transformed into potato and 21 were transformed into *Arabidopsis* with the aim of gaining information to help identify their modes of action by observing the phenotypic effects in the plant when such genes are over-expressed.

PCR was used to confirm the transgenic status of putative potato transformants by amplification of the insertion region using promoter and terminator primers, after antibiotic selection. The PCR products were analysed by gel electrophoresis for the absence/presence of an amplification product and the product size. Over 90% of the analysed potato samples were positive for an insertion of the correct size, therefore the antibiotic selection during the transformation procedure was sufficiently rigorous for successful selection of transgenic lines. This is consistent with other published internodal potato transformation experiments (89%) (Banerjee *et al.*, 2006). Although a kanamycin concentration of 75 µg/ml has been shown to delay shoot regeneration by around 14 days (Banerjee *et al.*, 2006), the reduced escape frequency of non-transgenic lines correlated with a high kanamycin concentration was deemed advantageous. PCR was not used to screen *Arabidopsis* putative transformants. These were subjected to subsequent selection procedures.

Over 310 lines of transgenic potato were generated. This is an unsustainable number to maintain. Therefore selection of lines based on transgenic protein production was the next logical step, in order to reduce the number of lines to maintain and choose those lines which were to be subjected to further analysis and characterisation. All effector proteins (if produced) would have a HA tag on their C-terminus. The HA tag is in frame with the C-terminus of the coding region of the effector with the stop codon located at the end of the HA tag. Western blot analysis using a HA-antibody

could, in theory, be used to detect and quantify the amount of transgenic protein being produced by each line. However, demonstrating production of the effector proteins was extremely challenging. Transgenic potato expressing a HA-tagged protein has previously been detected by western blot analysis (Bendahmane *et al.*, 2002). Identification of transgenic protein was only successful for the large (in comparison to the other effector proteins that were not identified) effector protein (Gp448-3 - 18 kDa) and GFP (28 kDa). An explanation for this could be that small proteins (<10 kDa for example, GpDGL1) may have been difficult to detect as these can pass through the membrane (Kurien and Scofield, 2006), although a membrane which was specifically designed for small proteins was tested and this did not aid in colorimetric detection. It is also possible that effector proteins could be subjected to rapid degradation by the ubiquitin-proteasome pathway. It is known that transgenic proteins can be subjected to rapid degradation and some attempts have been made to increase the stability of transgenic protein production (Jang *et al.*, 2012). The extraction method used may not have been suitable, a protocol that was suitable for insoluble proteins should have been tested (Hurkman and Tanaka, 1986). Effectors may have subcellular localisation that may not be suitable for the extraction method. For example, some *G. pallida* SPRYSECs are localised to the nucleus and nucleolus (Jones *et al.*, 2009b) and therefore effectors should be extracted with an extraction method suitable for their sub-cellular localisation (Komatsu, 2007). Other possibilities are that the HA-tag may be cleaved off therefore making detection impossible, this has been previously observed (Liefhebber *et al.*, 2010), or simply there may be no effector proteins produced. Western blot analysis of transiently expressed effector-HA tag constructs in *N. benthamiana* only detected two out of 14 effectors, which were the largest proteins tested; these two were Gp448-4 and GpE9. There may be a common reason for the lack of effector proteins detected in the transiently and stably expressed transgenic effectors in the western blot analysis. Using a different protein tag may improve detection. For example, FLAG-tag (Baumberger *et al.*, 2007; CHIU *et al.*, 2010), HIS-tag (AHN and Zimmerman, 2006) and myc-tag (Artsaenko *et al.*, 1998) have been successfully detected by western blot analysis of transgenic plants.

The insertion region was sequenced for two GpDGL1 and two GpIA7 transformed potato plants. All sequences contained the correct start and stop codon, effector sequence and HA tag, all in frame. Therefore the lack of identified protein was not

due to problems with the coding sequence, for these lines at least. Given the problems encountered with Western blotting, selection of lines of interest was performed using semi-quantitative RT-PCR to identify low and high expressing lines for each construct. Semi-quantitative RT-PCR to identify transgenic lines has been used for plants expressing nematode effectors (Doyle and Lambert, 2003; Hewezi *et al.*, 2010b; Lee *et al.*, 2011).

Although there was a lack of evidence for transgenic protein production, significant phenotypic differences were observed during phenotyping experiments, when compared to the controls. These data cannot alone elucidate the mode of action of the effector, as there could be many explanations for the phenotypes observed. However, when yeast-two hybrid data becomes available along with subsequent characterisation data, the identification of pathogen-host/protein-protein interaction(s) and further experimental data could help explain the observed phenotypes.

Transgenic potato plants were grown under glasshouse conditions and subjected to phenotypic analysis by recording measurements of a number of growth variables. All comparisons stated below are significant compared to the control group. No significant differences were observed for Gp66p1, Gp448-3, GpE9 and GpDGL1 expressing plants in the potato phenotyping glasshouse experiment. Lines representing GpG16H02 transformed plants had significantly longer and wider leaves, a greater number (total yield) and heavier tuber mass. This shows that expression of GpG16H02 significantly alters growth of a number of variables in potato, maybe by altering plant hormones levels that stimulate growth. Plant hormones are known to be altered during a plant–nematode interaction (Hermsmeier *et al.*, 1998; Hewezi *et al.*, 2010b; Puthoff *et al.*, 2003; Szakasits *et al.*, 2009a). Lines representing GpG8A07 transformed plants had less wet weight, were smaller and produced a greater number of, and heavier, tubers than the control. Expression of GpG8A07 reduced above ground biomass but increased tuber-mass, therefore this effector could redirect plant resources to root systems. The benefit to the nematode of an effector that can achieve this is clear. Lines representing GpIA7 transformed plants flowered later, produced heavier tubers, grew slower and showed differences in pathogen susceptibility. GpIA7 transgenic leaves were observed to have more white spots than the control. Although it is difficult to speculate as to the meaning of the observed phenotypes for GpIA7

expressing plants, these data suggest that this effector should be prioritised for further investigations. Lines expressing GpG20E03b transformed plants produced heavier tubers and had a lower dry weight. A line expressing GpHg10C02 had a greater number of leaf nodes, had a greater dry mass and over the time course of the experiment grew wider leaves than the control. Cytokinins are thought to be involved in internodal length regulation (Koch and Durako, 1991). PPN are known to secrete a biologically active cytokinin into their host (De Meutter *et al.*, 2003). It would be interesting to determine if GpHg10C02 alters hormone levels within the host. Tubers from lines expressing GpHgSEC4 weighed more, the plants were taller and grew faster over the experiment in comparison to the control. This could implicate GpHgSEC4 in altering plant hormones that control plant growth. The dry weight of a line expressing GpSCN1120 was less and has a lower leaf length/width ratio than the control. A transgenic line expressing GpG7E05 produced a greater number of tubers but had no other significant phenotype. A line expressing GpIVG9 was taller, grew faster and had a lower leaf length/width ratio compared to the control. Interestingly, leaves from GpIVG9 transformed plants had a curly leaf-like phenotype. The CURLY-LEAF gene in *Arabidopsis* controls both division and elongation of cells. *Arabidopsis* CURLY-LEAF mutants (*clf-25*) show a curly leaf phenotype and have altered leaf length to width ratio. This is hypothesised to be due to the reduction in cell expansion, due to the lack of expression of the gene responsible for regulating it (Kim *et al.*, 1998). *G. pallida* alters the cell cycle within the feeding site (Gheysen and Jones, 2006). Further investigation is required to determine if GpIVG9 has an effect on the cell cycle in transgenic plants. These data alone are not enough to determine the function of effectors. However, when data from future experiments are combined with the observations documented here, a hypothesis may be made which can then be experimentally tested.

It was interesting to observe that during the potato-nematode infection assay only high expressing line GpDGL1 supported a greater number of nematodes six weeks post-infection while GpHgSEC4 -2E supported fewer nematodes. No significant difference in the number of nematodes was observed two weeks post infection, suggesting that the observed differences were not due to differences in the ability of the nematodes to invade the plants. Such a low number of lines supporting a greater number of nematodes was unexpected as the expression of effectors should, theoretically, aid the

nematode infection process. For example, over-expression of *H. schachtii* effectors 10A06 and Hs4F01 in *Arabidopsis* resulted in an increase in susceptibility to nematodes (Hewezi *et al.*, 2010b; Patel *et al.*, 2010a). In contrast to this, Hs19C07 over-expression in *Arabidopsis* decreased nematode infection, and the reduction in infection rates was proportional to transgene expression (Lee *et al.*, 2011). This experiment was a pilot screen to identify any candidate effectors for further work. The results from the nematode infection assays therefore suggest GpDGL1 and GpHgSEC4 should be subjected to further characterisation.

If a less stringent statistical analysis (ANOVA without post hoc analysis) is used to analyse the potato-nematode infection assay data, members of the Gp448 (Gp448-3 high and low, and Gp448-4 high expressing potato lines) effector family supported fewer nematodes than the control and the significance of the data was proportional to the expression levels of the gene. These observations make the Gp448 gene family a target for further study. Potato lines expressing GpG7E05 supported fewer nematodes in both 2- and 6- weeks post infection assays, suggesting that the nematodes were less able to invade these lines. High expressing lines for GpDGL1 and GpIVG9 supported a greater number of nematodes. Caution should be applied to the interpretation of these results due to the statistical analysis used, therefore a repeat assay with a greater replication number is needed. It would be interesting to see if the presence of these effectors alters plant defences, which could explain the observed result.

A detached leaf assay (Whisson *et al.*, 2007b) using an attenuated strain of *P. infestans* CS-12*,* which has its AVR3a effector down-regulated by RNAi due to the transgenic expression of a double hairpin construct (Bos *et al.*, 2010a), was used to determine if the over-expression of effectors in potato rendered the transgenic leaves more susceptible to this pathogen. Wild type *P. infestans* was shown in a pilot assay to be too virulent and therefore could not be used for these experiments as we aimed to score an increase in susceptibility. In the assay system used wild type *P. infestans* caused extensive and rapid disease and identifying increased susceptibility would be challenging. This experimental set up was used to allow screening of a large number of potato lines in a high throughput manner to identify candidates for further work. A computer program was made to quantify the area of infection. High expressing (as determine by semi quantitative RT-PCR) transgenic potato lines GpIA7 -5C and

GpE9 -3F were significantly more susceptible to this pathogen than the control ($P<0.001$ and $P=0.05$, respectively). Screening transgenic lines for increased susceptibility to other pathogens is a relatively new concept in the nematology field. However, a few successful examples have been published. Transgenic potato over-expressing SPRYSEC-19 were more susceptible to the fungal pathogen *Verticillium dahliae* (Postma *et al.*, 2012) and *Pseudomonas syringae* was used to screen transgenic lines of *Arabidopsis* expressing 10A06 for an altered susceptibility to pathogens (Hewezi *et al.*, 2010b). Therefore the use of other pathogens is a powerful way of identifying an altered ability to resist pathogen parasitism.

The detached leaf *P. infestans* assay and the nematode infection assays did not identify the same effector-expressing potato lines for further characterisation. An explanation for this could be that the *P. infestans* CS-12 pathogen has an important effector down-regulated (AVR3) (Bos *et al.*, 2010a). Therefore, the assay system used here could be identifying a nematode effector that could either recover the *Phytophthora* phenotype by targeting the same pathway or the same target as AVR3 (CMPG1) (Bos *et al.*, 2010b), or identifying a nematode effector that targets and interferes with other defence pathways rendering the potato incapable or having a reduced ability to defend against pathogen invasion. The nematode infection assay could be identifying effectors that could have a wide range of roles, such as rendering the infection more efficient, aiding in feeding site development/maintenance or suppressing host defences. Potato expressing nematode effector GpE9 was more susceptible to *P. infestans* CS-12. No physical phenotype was observed for lines expressing this effector. Effectors that interfere with host defences may not produce a physical phenotype but maybe more susceptible to pathogens. Therefore GpE9 would be an interesting effector to further characterise to determine its potential role in suppressing host defences. GpIA7 transformed plants had a number of physical phenotypes (see above) and GpIA7 transformed leaves were more susceptible to *P. infestans* CS-12. Although physical phenotypes were observed this does not mean the effector could not be interfering with host defences, therefore this gene should also be further characterised.

The use of stable transformation allows the identification of phenotypes due to the expression of an effector that would not otherwise be observed by transient

expression. However, there are some negatives associated with stable transformation that need to be considered. Firstly, proving that the effector protein is being produced is difficult, although this may also be true for transient expression. For any phenotypic effect observed without the proof of transgenic protein production it could be argued that the phenotype may be due to transgene insertion, especially as in many of the above experiments both high expressing lines did not result in the same phenotype. The production of stably transformed plants is time consuming, and analysing these plants is laborious. It may be wise to subject effectors to higher throughput assays to identify interesting effectors that could then be subjected to detailed transgenic characterisation. This would allow the focus to be concentrated on a reduced number of effector expressing lines, therefore allowing an adequate number of repeats to be carried out in assays. An example of a high throughput experimental procedure that could be used to determine if the effectors suppress host defences is the Effector Detector Vector system which determines the effect an effector has in a model bacterial pathogen, for example, *Pseudomonas syringae,* when inoculated on a host. The effect of the effector can be assessed and can then be further examined using bacterial mutants that have key effectors mutated to determine any recovery phenotypes (Sohn *et al.*, 2007).

These data presented above are preliminary results from large scale screen used to identify lines that may be of interest and therefore be the focus of further experimentation. The experiments described here help prioritise effectors for future studies; the GpIA7 effector in particular. The assays used are variable and therefore large differences were observed between the controls. Such variation in the control group could interfere with identifying truly interesting effector genes. However, enough data has been produced to implicate several effectors as being targets for further analysis. In addition, the transgenic lines produced here will provide a valuable resource for further, larger scale phenotyping studies in the future, should any of the effectors used be identified as potentially important in the host-parasite interaction.

Summary

- 45 putative *G. pallida* effectors were cloned for functional characterisation.

- 21 effectors were constitutively expressed in potato and/or *Arabidopsis* to aid functional analysis.
- Several significant phenotypes were observed. Most interestingly, *GpIA7* expressing plants showed delayed flowering, stunted growth and an increased susceptibility to *P. infestans* CS-12. *GpIVG9* expressing plants showed distorted leaves, accelerated growth and an increased susceptibility to nematode invasion.

5 Functional analysis of a *Globodera pallida* effector similar to ubiquitin extension proteins

5.1 Introduction

Globodera pallida second stage juveniles (J2s) invade a host plant through the root tip in the zone of elongation and migrate intracellularly through the inner cortex layers to the site of initial feeding site formation. During the migration through the root, nematodes have been shown to induce responses in the plant due to host tissue damage (Grundler *et al.*, 1997). Damaged host tissues may release damage-associated molecular patterns (DAMPs) that are able to induce defence responses in neighbouring cells (Lotze *et al.*, 2007). Migration through the root by cyst nematodes is destructive and could induce production of DAMPs (Smant and Jones, 2011). Presumably the induction of DAMPs is either not a significant factor in terms of preventing successful nematode invasion or the nematodes can migrate faster than a DAMP induced defence response. When the nematode reaches a suitable potential feeding site its behaviour changes (Wyss *et al.*, 1992; Sobczak and Golinowski, 2011). The nematode gently probes cells to determine if they are responsive and can be manipulated into a feeding site. If the cell detects the invading nematode, host defence responses are activated. These responses may include callose deposition (Sobczak *et al.*, 1999) and production of reactive oxygen species (oxidative burst; Felix *et al.*, 1999; Jones and Dangl, 2006). The nematode will sample alternative cells until a receptive cell is identified. Therefore it is essential for the survival of the invading nematode to suppress host defence signalling pathways during selection of the feeding site and subsequently for as long as the feeding site is required (Haegeman *et al.*, 2012; Smant and Jones, 2011). Both the suppression of host defences and initiation of the feeding site are thought to be mediated by molecules produced in the nematode oesophageal gland cells and secreted through the stylet into the plant (Gheysen and Jones, 2006).

5.1.1 Ubiquitin (UBI) and the ubiquitination-proteasome pathway

It is known that plant defence signalling pathways are often controlled by ubiquitination and that pathogens possess effectors that manipulate the ubiquitination system to suppress host defences (Zhou and Chai, 2008). Ubiquitin belongs to a family of polypeptides that all possess a characteristic ubiquitin fold which acts as a

recognition promoter for protein–protein interactions and is used by all eukaryotes to influence various cellular processes. Proteins can become mono-ubiquitinated, leading to changes in trafficking or protein function, or can become polyubiquitinated leading to degradation through the 26S proteasome. Ubiquitin is a highly conserved 76 amino acid protein that terminates in a double glycine and has lysine residues at positions 6, 11, 27, 29, 33, 48 and 63. Chain topology is determined by the attachment of a UBI molecule's terminal GG to another UBI's lysine site and this ultimately determines the fate of the tagged protein. For example, chain topology through lysine 48 is the main signal for protein degradation and that through lysine 63 changes protein function (Kaiser and Huang, 2005). The ubiquitin proteasome pathway may rival transcription as the dominant cellular regulatory mechanism (Vierstra, 2009).

5.1.2 Enzymes required for Ubiquitination: E1, E2, E3 and DUBs

The ubiquitination process involves the sequential action of three classes of enzymes: E1 (ubiquitin activating enzymes), E2 (ubiquitin-conjugating enzymes) and E3 (ubiquitin ligases) (Ye and Rape, 2009). There are 2 E1 encoding genes, 37 E2 encoding genes, more than 1400 E3 encoding genes and 64 DUBs (deubiquitinating enzymes) predicted in the *A. thaliana* genome (Mazzucotelli *et al.*, 2006; Vierstra, 2009). This abundance of genes and the fact that ubiquitin is conserved across the eukaryotic kingdom highlights the complexity and importance of the ubiquitin–proteasome pathway. E1 is required to form a high energy bond between the E1 and the C-terminal glycine residue of the ubiquitin. The activated ubiquitin is then transferred to an E2 by *trans*-esterification from the UBI-E1 complex (Vierstra, 2009). It has been suggested that E2 enzymes are the main mediator in determining the chain assembly in the ubiquitination process. Therefore chain initiation, elongation and topology, which determine the fate of the protein, are determined by E2 enzymes (Ye and Rape, 2009). E3 enzymes recognise target substrates and confer specificity in the transfer of the ubiquitin from the UBI-E2 complex to the target protein (Deshaies and Joazeiro, 2009), where the ubiquitination sites on the protein are predicted to be exposed on the surface in a stem loop structure (Catic *et al.*, 2004). Poly-ubiquitin chains can be formed by E2-E3 complexes to control the fate of the tagged protein (Deshaies and Joazeiro, 2009). Ubiquitination can be reversed by DUBs. These are protease enzymes that release ubiquitin molecules from their targets thus reducing degradation of a target protein (Vierstra, 2009; Hartmann-Petersen *et*

al., 2003). In addition, DUBs are responsible for proteolysis of poly-ubiquitin precursors and cleavage of ubiquitin extension proteins. Cross-talk between ubiquitin and DUBs, which is regulated by phosphorylation and other post-translational modifications, is used to control the abundance of a protein within the cell. If more protein is required then DUBs will remove UBI from the tagged protein preventing its degradation; if too much protein is present DUBs are not activated and the protein will be recycled via the 26S proteasome pathway (Hunter, 2007). Further to this, DUBs have been implicated in defence against pathogens. For example, in a study investigating *Arabidopsis thaliana* DUB genes UBP12 and UBP13 various lines of evidence implicated DUBs in plant defence. Using individual *Arabidopsis* mutants for *ubp12* and *ubp13*, the individual mutants alone were not more susceptible to *Pseudomonas syringae* than wild type controls. *Arabidopsis* UBP12 and UBP13 were shown to share functional redundancy with each other. Therefore RNAi was used to reduce the expression of both UBP12 and UBP13 and this resulted in a significant reduction in *Pseudomonas syringae* growth. This suggests that UBP12/13 collectively have a repressive role in defence responses and since they are de-ubiquitinating enzymes they may be involved in regulating levels of important defence signalling compounds. The *Arabidopsis* DUB UBP12 has a Solanaceous homologue from tobacco, NtUBP12, which also functions as negative regulator of the Cf9 dependant hypersensitive response (Ewan *et al.*, 2011).

5.1.3 Nematode effectors that target the ubiquitin-proteasome pathway

Effectors which may target the ubiquitination pathway have been identified in the secretome of a number of cyst nematode species. Ubiquitin (UBI) extension proteins have been identified from *Heterodera schachtii* and *H. glycines* (Tytgat *et al.*, 2004; Gao *et al.*, 2003). The genes encoding these proteins are expressed in the dorsal gland cell and have a C-terminal extension coupled to the conserved ubiquitin-like sequence. In *H. glycines* the protein is cleaved at the junction between the UBI region and the C-terminal extension and the C-terminal extension is targeted to the nucleolus. The C-terminal extension is highly variable between nematode species and has no sequence similarity to known proteins (Tytgat *et al.*, 2004; Gao *et al.*, 2003). Analysis of ESTs has shown that *G. pallida* also produces a similar effector and *in situ* hybridisation indicates that it is expressed in the dorsal gland (Jones *et al.*, 2009). It has been proposed that the ubiquitin domain acts as a chaperone for the C-terminal extension and

that after the protein is cleaved by DUBs within the host cell the C-terminal extension could function alone. No similar effector has been identified in root knot nematodes (RKN). Since RKN induce giant cells, which are functionally and developmentally distinct from syncytia, it has been proposed that the C-terminal extension may be involved in the formation of the syncytium (Elling *et al.*, 2009; Tytgat *et al.*, 2004). However, no functional evidence has been produced in support of this claim.

Bioinformatic analysis of genome and EST data from cyst nematodes has revealed several other putative effectors that are similar to proteins involved in the ubiquitination pathway (Elling *et al.*, 2009b; Gao *et al.*, 2003b). The *H. glycines* effector candidate 8H06 is similar to SKP1 (S-phase kinase-associated), which is a component of the multi protein E3 ubiquitin ligase SCF complex. 8H06 also contains a predicted nuclear localisation sequence (Gao *et al.*, 2003). Although detailed functional analysis is lacking, it has been suggested that the 8H06 protein could interact with the cell cycle mechanism to maintain the syncytium in repeated S-phase as SKP1 is a key component of the SCF complex that provides ubiquitin-protein ligase activity required for cell cycle progression (Bellafiore *et al.*, 2008).

Nematode effectors have also been identified that contain a RING domain. The *H. glycines* 10A06 effector contains a predicted RING-H2 zinc finger (Elling *et al.*, 2007). RING H2 proteins are single component E3 ligases. Detailed functional analysis has shown that 10A06 interacts with Spermidine Synthase 2, a key enzyme in polyamide biosynthesis. As a result of this interaction, polyamine oxidase activity is increased, stimulating the induction of antioxidant genes in the syncytium. Transgenic plants expressing 10A06 are more susceptible to nematode infection (Hewezi *et al.*, 2010a). The mechanism by which the 10A06 RING-H2 protein could stimulate activity of spermidine synthase is not clear.

Although the genome sequences of two RKN species have been published (Abad *et al.*, 2008; Opperman *et al.*, 2008), there are no examples of characterised effectors from RKN that target or exploit the ubiquitination pathway. However, twenty SUMOs (small ubiquitin like modifiers) were identified in the *M. incognita* genome (Abad *et al.*, 2008). Direct analysis of RKN secretions has also identified several potential ubiquitination related proteins including ubiquitin carboxyl-terminal hydrolase,

ubiquitin-activating enzyme and a ubiquitin-like protein but their functions still need to be investigated (Bellafiore *et al.,* 2008).

5.1.4 Effectors that interfere with the ubiquitin proteasome pathway from other pathosystems

Effectors that interact with the ubiquitin proteasome pathway are present in a wide range of pathogens (*e.g* Birch *et al.*, 2009; Goehre and Robatzek, 2008; Jones and Dangl, 2006; Schrammeijer *et al.*, 2001). It is fascinating to note that bacteria, which do not use the ubiquitination pathway in their own internal cellular processes, have evolved effectors that mimic or exploit components of the ubiquitin–proteasome pathway of their hosts. For example, the *Pseudomonas syringae* AVRPtoB effector is an E3 ubiquitin ligase that suppresses defence responses of tomato by promoting the ubiquitination and degradation of the PAMP receptor FLS2 (Goehre *et al.*, 2008a; Goehre *et al.*, 2008b). During the transfer of the T-DNA from *Agrobacterium tumefaciens* an F-box protein (VirF) is secreted and is thought to interact with E3 ligases (Schrammeijer *et al.*, 2001a). The *P. syringae* effector HopPtoM interacts with ubiquitination pathway components to target MIN7 in order to reduce vascular flow and suppress callose deposition. Furthermore AvrBsT, AvrRxv, XopD and AvrXv4 from *Xanthomonas campestris* pv. *vesicatoria* are all YOP-J like SUMO proteases with unknown functions (Goehre and Robatzek, 2008). *Ralstonia solanacearum* secretes an effector called GALA which is an F-box protein which, together with other components, forms an E3 ligase essential for the pathogenic fitness of the pathogen (Angot *et al.*, 2006a).

Oomycete and fungal effectors have also been identified that target the ubiquitination pathway. The *Phytophthora infestans* effector AVR3a targets and stabilises a host E3 ligase called CMPG1 which is essential for downstream signal transduction during PAMP triggered immunity induced by INF1 (a *P. infestans* elicitor or PAMP) (Birch *et al.*, 2009b; Bos *et al.*, 2010a). CMPG1 is also involved in downstream signalling following activation of PTI induced by Cf4/ AVR4. This demonstrates the importance of these types of effectors to pathogens and also the importance of the ubiquitin proteasome pathway to the host organism.

A deeper understanding of effector targets is essential in order to understand the biology of the nematode feeding site as well as being important for the potential

biotechnological targeting of effector proteins in the development of alternative pathogen control methods. The aim of this work was to functionally characterise an ubiquitin-extension protein identified from *G. pallida* using assays to determine if this effector interferes with host defence responses.

5.2 Materials and Methods

5.2.1 Amplification and cloning of ubiquitin extension protein (UBI-EP)

The *G. pallida* UBI-EP was amplified from DNA and cDNA using sequence information derived from an EST study of this nematode (Jones *et al.,* 2009). For analysis of variation in the sequence of the GpUBI-EP, total DNA was extracted from 20 cysts of each of two populations of *G. pallida* using a DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's instructions. The two populations were Ca 1998 (a population grown on a susceptible potato cultivar) and Ca6 1998 (a population repeatedly grown on the partially resistant potato line 62-33-3). The original source nematode material for Ca6 1998 had undergone 4 generations of selection on 62-33-3 (Turner *et al.*, 1983) and a subsequent 4 generations of selection on 62-33-3 at SCRI/JHI. PCR was used to amplify GpUBI-EP with *Taq* DNA polymerase (Promega) using primers UBI_F and UBI_R. PCR products were purified as in section 2.5.3 and cloned using the pGemT Easy Vector System (Promega). Colonies were prepared using GeneJet plasmid preparation kit (Fermentas) and sequenced by the JHI sequencing service.

Three forms of the GpUBI-EP (UBI_WT, UBI_Del and UBI_only) were amplified by PCR from plasmid clones generated using a proof reading DNA polymerase (Novagen KOD Hot Start DNA polymerase). Primers UBIpDONR221F and UBIpDONR221R were used to amplify both the full version and the version of the gene with the deletion in the C-terminal extension. Primers UBIpDONR221F and UBItruncpDONR221R were used to amplify just the UBI domain of the gene without the C-terminal extension.

The PCR products were cloned into the Gateway destination vector pMDC32 (Invitrogen). Plasmids that contained the desired insert were transformed into *Agrobacterium tumefaciens* GV3101 to enable transient expression of the desired gene *in planta* for characterisation of the expressed gene.

5.2.2 Scoring method for infiltration assay used to characterise suppression of host defences

Infiltration experiments were conducted as previously described (Sacco *et al.*, 2007). Once a HR had been induced (see below) infiltration zones were scored based on a 4 point scoring system: 0 = no suppression of HR, 1 = little suppression of HR, 2 =

evidence of suppression, 3 = very high suppression of HR (a crescent of HR may be seen outside the gene of interest infiltration zone). The resulting data were statistically analysed using Genstat 12th edition.

5.2.3 Assay for suppression of PTI (PAMP Triggered Immunity) induced by INF1

The *Phytophthora infestans* INF1 protein was used in order to induce PTI in *N. benthamiana* (Bos *et al.*, 2006). Exposure of *N. benthamiana* to INF1 induces an extremely strong PTI response culminating in death of cells exposed to the protein. *Agrobacterium* cultures containing the gene of interest or control samples containing empty vector only or GFP (O.D. 0.2) were co-infiltrated into leaves with *Agrobacterium* containing INF1 (O.D. 0.5).. Alternatively, cultures containing the gene of interest (O.D. 0.2) were infiltrated into leaves and *Agrobacterium* containing INF1 (O.D. 0.5) was infiltrated one or two days after infiltration of the effector containing construct, overlapping the previous infiltration. The resulting infiltrations were subjected to the scoring procedure detailed above over a 7 day period.

5.2.3.1 Callose deposition assay for suppression of PTI

Four week old *N. benthamiana* leaves were infiltrated as previously described (Sacco *et al.*, 2007) with *A. tumefaciens* containing the effector UBI_WT in plasmid pMDC32. 24 hours post infiltration the leaves were subjected to a second infiltration with liquid culture medium in which *P. infestans* had previously been grown. This filtered culture supernatant contains large quantities of INF1, which is the most abundant protein produced by *P. infestans* in culture (S. Whisson, pers. comm.). Culture supernatant that had not contained *P. infestans* was used as a control. In addition, infiltration zones of *A. tumefaciens* containing empty pMDC32 plasmid were infiltrated with INF1 supernatant. 24 hours later the infiltrated regions were infiltrated with 0.01% aniline blue in 0.1 M Sørensen's phosphate buffer at pH 7.5 (Hauck *et al.*, 2003b). The leaves were then incubated at room temperature for 10 mins. The infiltrated zone was cut out using a cork borer and mounted on a microscope slide with a drop of 10% glycerol. Each leaf disk was imaged in four places, avoiding any areas of mechanical damage, under UV light using a Nikon UV2A filter block with a 5× lens. All images were captured using a 200.5 ms exposure time with 2× gain adjustment. In a separate experiment callose was also induced by PTI induction following FLG22 perception (S. Whisson, pers. comm.) in

the same experimental setup as described above where the infiltration of FLG22 replaced INF1.

Callose deposition images were automatically scored by a custom made Autoit computer program (http://www.autoitscript.com/site/autoit/) Autoit script 001 (Appendix 5). The program was written to quantify stained callose within an image and count the number of pixels within an image allowing 100 shades of variation (shades of variation is an inbuilt function) that matched a user defined colour. For this experiment callose colour code was defined as 0x5EA2FB.

5.2.4 Infiltration assays to characterise suppression of Effector Triggered Immunity (ETI)

The interaction between the *S. tuberosum* R3a resistance gene against *P. infestans* and its cognate avirulence gene AVR3aKI was used to induce ETI in *N. benthamiana* (Bos *et al.*, 2006). Cultures containing the gene of interest or control plasmid (O.D. 0.2) were co-infiltrated with *Agrobacterium* containing R3a and *Agrobacterium* containing AVR3aKI (O.D. 0.5) into leaves. Alternatively, cultures containing the gene of interest (O.D. 0.2) were infiltrated into leaves and 2 days later *Agrobacterium* containing R3a and *Agrobacterium* containing AVR3aKI (O.D. 0.5) were infiltrated overlapping the previous infiltration. The resulting infiltrations were subjected to the scoring procedure detailed above over a 7 day period. Further suppression assays were performed as detailed above using the *S. tuberosum* R gene Gpa2 and AVRGpa2 (also called RBP1), a SPRYSEC effector from *G. pallida* (Sacco *et al.*, 2009a) and using a fungal effector AVR4 from *Cladosporium fulvum* with its corresponding R gene Cf4 from tomato (Van der Hoorn *et al.*, 2000).

5.2.5 Assay to determine if GpUBI-EP stabilises a CMPG1-YFP fusion protein

CMPG1 is an E3 ubiquitin ligase that is involved in defence responses invoked by pathogen perception. AVR3aKI has been shown to suppress INF1 and Cf4/AVR4 induced defence responses by stabilising a CMPG1-YFP fusion (Bos *et al.*, 2010a). *A. tumefaciens* containing a vector encoding potato CMPG1 fused at the C-terminal with a yellow fluorescent protein (Bos *et al.*, 2010a) was co-infiltrated with *A. tumefaciens* containing pMDC32 GpUBI-EP (UBI_Del) at a final concentration of O.D. 0.5 (600nm) into 4 week old *N. benthamiana* leaves. 48h post infiltration the leaves were visualised using a confocal microscope to determine if GpUBI-EP stabilises CMPG1.

5.2.6 Sub-cellular localisation

TRV (tobacco rattle virus) expression constructs (Liu *et al.*, 2002) were made, infiltrated and analysed as described by Jones *et al.* (2009). Briefly TRV has a bipartite dsRNA genome; for experimental exploitation the two RNAs have been cloned separately into binary vectors (RNA1 and RNA2) which are co-infiltrated separately into plants. The RNA2 sequence has been modified to contain GFP with a Gateway recombination cassette cloned at the N or C terminus (Valentine *et al.*, 2004). For sub-cellular localisation the UBI-EP sequences were cloned into pDONR221 (Invitrogen) and then transferred into the appropriate TRV RNA2 using LR recombination. GFP-UBI fusion proteins were imaged using a Leica SP1 Confocal laser scanning microscope and analysed using LCS software (Leica, Milton Keynes, UK).

5.2.7 Western blotting

Leaf material displaying GFP fluorescence was harvested and frozen in liquid nitrogen from *N. benthamiana* leaves infiltrated with TRV UBI-GFP 3' or TRV UBI-GFP 5'. These constructs allowed expression of the UBI wild type or deleted form of the protein as N- or C-terminal fusions with GFP. Proteins were extracted for all harvested material and from control material containing TRV expressing free GFP and from stably transformed *N. benthamiana* plants expressing GFP constitutively. Frozen leaf material was ground to a fine powder using a mortar and pestle.

Western blotting was carried out according to section 2.7 and the blots were analysed by colorimetric detection using appropriate primary antibody (Anti-GFP rabbit serum at 1:2000 dilution – Invitrogen) and secondary antibody (Anti Rabbit IgG Alkaline Phosphatase Conjugate at 1:10 000 dilution – Sigma).

5.2.8 qPCR and bioinformatic analysis of *GpUBI-EP* expression profile

Primers were designed using http://www.premierbiosoft.com/qPCR and http://frodo.wi.mit.edu/primer3 to produce qPCR products approximately 100bp in length with a Tm of 59°C. Primers (section 2.9.4) were designed to amplify two isoforms of the *GpUBI-EP*; the reverse primer was designed to span the C-terminal extension to confer specificity. UBI_For_qPCR (forward primer) paired with UBI_Rev_all_qPCR (reverse primer) amplified both the full version of the GpUBI-EP and the GpUBI-EP which contained the 3 amino acid deletion in the C terminal

extension. UBI_For_qPCR (forward primer) paired with UBI_R_WT_SP_qPCR (reverse primer) amplified only the full version of the GpUBI-EP. It was not possible to design primer sets that only amplified the deleted form but this strategy allowed expression of the two isoforms of the GpUBI-EP to be distinguished. Primers were used to amplify EF-1α and EIF4α for the purpose of normalising the data. qPCR was performed as described in 2.5.1.

Transcriptome data from the *G. pallida* genome sequencing project (http://www.sanger.ac.uk/resources/downloads/helminths/globodera-pallida.html) were used to determine the expression profile of *GpUBI-EP* based on a bioinformatic approach. Sequence reads were mapped on to the genome sequence: super-contigs generated in assembly 201011. The UBI_WT sequence was identified via local BLAST (Altschul *et al.*, 1990) within supercontig sno959cc1wk1 and UBI_Del was contained within sno13601cc1wk1. Using Python script 001 (Peter Cock, James Hutton Institute 2011) (Appendix 5) expression values were mapped to each base within the given contig. Python script 002 (Appendix 5) was made to return a list of values that represent the average number of reads that map on to the gene per base for each life stage following a normalisation process against the expression of housekeeping gene EIF4α (the same gene used for the normalisation of the qPCR data). (EF-1α was not present in the genome assembly used for this analysis and could not therefore be included.) All scripts were made and run on Python 2.6 with Biopython 1.57 (Cock *et al.*, 2009). Transcriptome data were viewed using Gbrowse http://ppcollab.hutton.ac.uk/gb2/gbrowse/Gp_ass_2010_11/ .

5.2.8.1 RNA extraction

RNA was extracted from nematodes using a Dynabead kit (Invitrogen) following the manufacturer's instructions except that the final re-suspension volume was 10 µl. Eight µl of the RNA extraction was treated with RQ1 DNase (Promega) to remove any contaminating gDNA. The DNase-treated RNA was converted to cDNA using a SuperScriptIII First Strand kit (Invitrogen) following the manufacturer's instructions. cDNA synthesis was primed using an oligo(dT) primer. PCR was performed to demonstrate the specificity of the qPCR primers for nematode genes.

5.2.9 Nematode infection assay on *Arabidopsis thaliana* DUB mutant lines

Arabidopsis seeds, wild type (Col 0) and DUB mutant lines *ubp12.1* (from GabiKat Gk244_E11) and *ubp13.1* (Nottingham *Arabidopsis* Stock centre N628312) (Ewan *et al.*, 2011) were grown in tissue culture as described in section 2.2. The resulting plants were infected at 21 days post sowing. Thirty five sterile J2s of *H. schachtii* (see below for sterilisation procedure) were applied directly on to the plant roots at 3 infection points per plant. The infection zones were then covered with GF/A paper for 24h to aid infection. The infected plants were incubated in a growth chamber at 20°C with 16h per day photo-period for 17 days. Fifteen wild type Col 0, ten *ubp12.1* and thirteen *ubp13.1* plants were used in these experiments.

H. schachtii cysts were collected and incubated in 0.1% malachite green for 1 hour with rotation. The cysts were then washed with tap water overnight followed by a 24-hour incubation in an antibiotic solution at 4°C. The antibiotic solution contained: 8.0 mg ml^{-1} streptomycin sulphate, 6.0 mg ml^{-1} penicillin G, 6.13 mg ml^{-1} polymixin B, 5.0 mg ml^{-1} tetracycline and 1.0 mg ml^{-1} Amphotericin B (Urwin *et al.*, 1997). The sterilised cysts were then washed in sterile distilled water and incubated at 20°C in the dark in 3 mM $ZnCl_2$ to stimulate hatching. Hatched J2s were pelleted in non-stick centrifuge tubes (Axygen) and incubated in 0.1% v/v chlorhexidine digluconate and 0.5 mg ml^{-1} hexadecyltrimethylammonium bromide (CTAB) for 30 mins with rotation (Goverse *et al.*, 1999). The sterilised J2s were rinsed 3 times in sterile distilled water and resuspended at a concentration of 1 J2 μl^{-1}. Infected plants at 17 days post infection were subjected to acid fuchsin staining (section 2.3).

5.3 Results

5.3.1 *Globodera pallida* UBI extension protein

The *G. pallida* UBI extension protein (GpUBI-EP) is a 107 amino acid protein consisting of a 19 amino acid signal peptide for secretion, a 76 amino acid ubiquitin domain and a 12 amino acid C-terminal extension peptide (Figure 1A). The lysine amino acids within the GpUBI-EP ubiquitin domain are located at correct functional positions. The ubiquitin domain ends with two glycine amino acids (GG); both of these features are consistent with normal functional ubiquitin proteins (Kaiser and Huang, 2005). The ubiquitin domain has a few amino acid changes when compared to other cyst nematode UBI domains and *Pinus sylvestris* polyubiquitin (Tytgat *et al.*, 2004). The GpUBI-EP (if 1 is considered the start of the UBI domain) contains valine at amino acid 23 instead of isoleucine, threonine for alanine at position 28, methionine instead of leucine at position 56 and tyrosine instead of aspartic acid at position 58. These changes may impact on the function of the UBI domain. The C terminal extension is known to be variable between species; *Globodera* and *Heterodera* C-terminal extensions show no similarity. The GpUBI-EP is very similar to a ubiquitin extension protein from *G. rostochiensis* with the C terminal extensions of these proteins sharing 9 of the 12 amino acids (Figure 5.1).

5.3.2 Variation in C terminal extension

Two forms of the GpUBI-EP are present in *G. pallida.* The UBI_Del isoform (see below) occurs within nematode populations at low frequency in unselected lines (less than 5%) but if nematodes are cultivated on partially resistant lines of potato which contain the *H3* or *Gpa5* R-gene then the UBI_Del allele frequency significantly increases. This analysis was extended by comparing sequences of *GpUBI-EP*s from *G. pallida* population Ca 1998 (a nematode population cultivated on susceptible potato lines) and Ca6 1998 (a nematode population repeatedly grown on the partially resistant potato line 62-33-3). Clones generated from each population were sequenced and compared. All 20 clones from the Ca 1998 population contained the same UBI_WT sequence, whereas 25% of the 20 clones sequenced from the Ca6 1998 selected population contained a 3 amino acid deletion at the start of the C-terminal extension and terminated with a histidine residue instead of a glutamine. This isoform will be referred to as UBI_Del (Figure 5.1).

5.3.3 *GpUBI-EP* expression profile

Two methods were used to determine the expression profile for *GpUBI-EP*: qPCR and bioinformatics on the transcriptome data from the *G. pallida* sequencing project. As detailed in the material and methods section Python scripts were written to return the average number of reads that map on to a given gene per nucleotide for each life stage (life stage meaning days post infection), following normalisation against EIF4α (UBI_WT: [egg= 1, J2 = 400, 14dpi = 1, 21dpi = 1, 28dpi = 3, 35dpi = 4, male=0]. UBI_Del: [egg= 1, J2 = 287, 14dpi = 1, 21dpi = 0, 28dpi = 1, 35dpi =, 1, male= 0]). The results show that *GpUBI-EP* is highly up-regulated at J2 with little or no expression at other life stages (egg, 7dpi, 14dpi, 21dpi, 28dpi, 35dpi and male) (Figure 5.2 – UBI_WT and Figure 5.3 – UBI_Del).

qPCR was used to confirm the expression profile indicated by transcriptome analysis. The experimental data was normalised using EF-1α and EIF4α *GpUBI-EP* is highly up-regulated at the J2 life stage with little expression at 7dpi, 14dpi, 21dpi and 28dpi (these experimental time points were each performed with 2 biological replicates). In addition, UBI-WT including the UBI-Del expression profile follows the same pattern (Figure 5.4). It was not possible to design primers that would amplify the UBI_Del alone that would satisfy the criteria for qPCR primer guidelines (http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/general documents/cms_041440.pdf).

The qPCR and bioinformatics expression profile provide complementary evidence that *GpUBI-EP* (UBI-WT and UBI-Del) is highly up-regulated at the J2 stage.

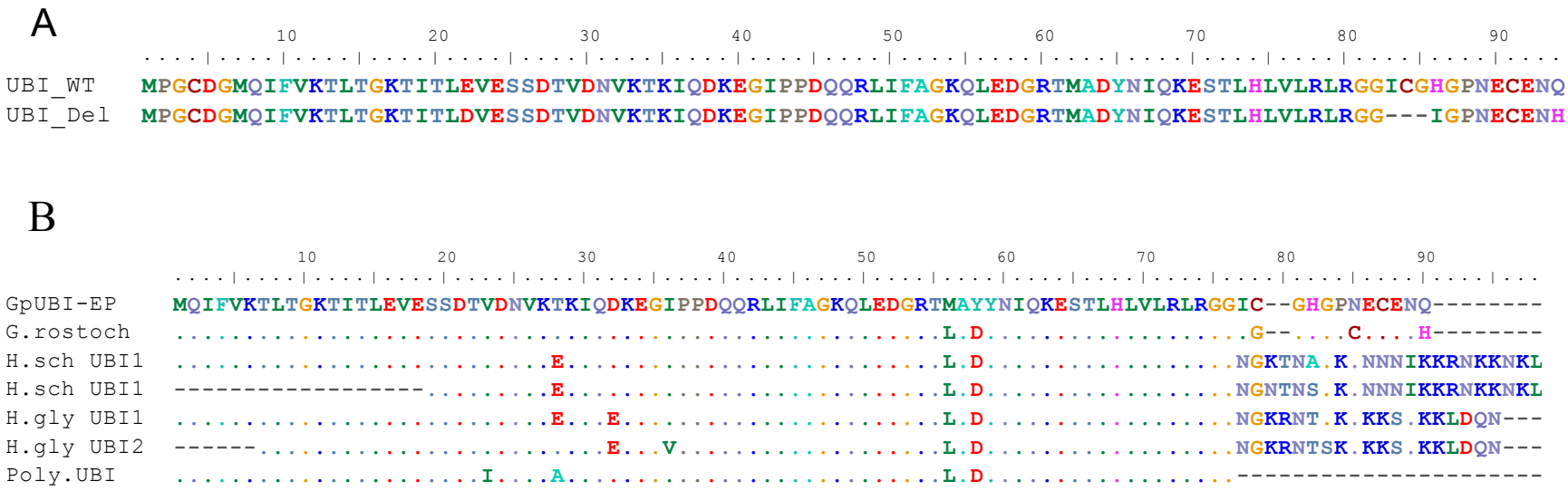


Figure 5.1: The sequence of GpUBI-EP. (A) Amino acid sequence of GpUBI-EP with a diagrammatic representation of the gene's components. UBI_Del refers to the isoform of the gene with the deletion in the C-terminal extension, UBI_WT refers to the full isoform of the gene. The 3 amino acid deletion can be observed at amino acid 84. All UBI_Del sequences terminate in an H residue, whereas all UBI_WT sequences terminate in a Q residue. (B) An alignment of ubiquitin extension proteins from cyst nematodes, including a polyubiquitin sequence. GpUBI-EP (accession number GR367886) and *G. rostochiensis* (accession number BM355031) effector sequences show a high degree of similarity within the C-terminal extension. *H. schachtii* UBI1 (accession number AY286305), *H. schachtii* UBI2 (accession number AY288520, *H. glycines* UBI1 (accession number AF469060_1), *H. glycines* UBI2 (accession number AF473831_1) show high similarity between each other but have a different C-terminal extension from *Globodera* sequences. All UBI domains are highly conserved when aligned with Polyubiquitin (accession number Q39940).

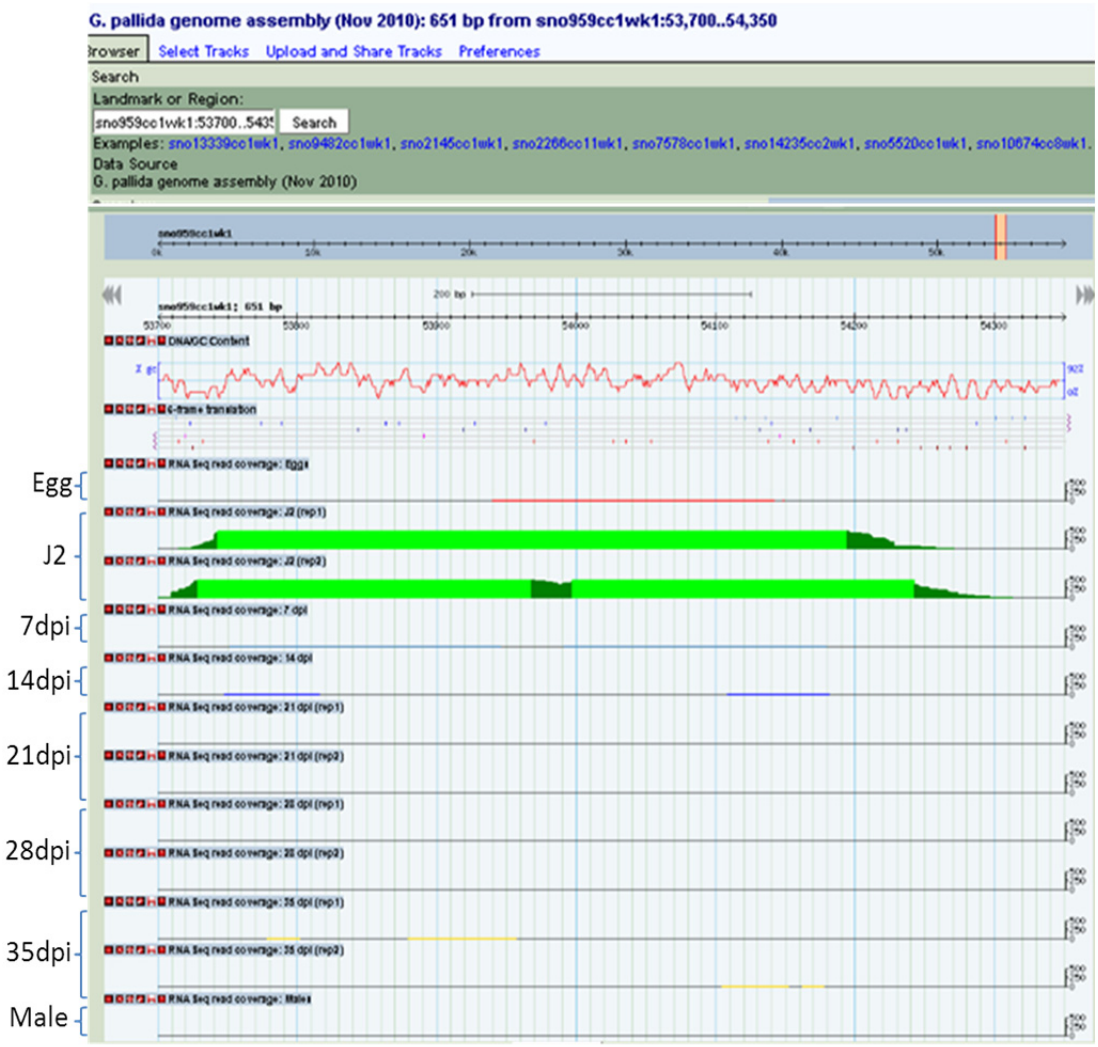


Figure 5.2 GpUBI-EP (UBI_WT) expression profile. Screen shot from GBrowse showing the part of the assembly encoding GpUBI-EP showing unnormalised RNAseq expression data for various life stages. Expression graphs are capped at 500 reads indicated by the colour change. Duplicated samples for the J2 life stage (green) show over a read depth of over the 500 threshold, other life stages (egg: red, 7dpi: light blue, 14dpi: dark blue, 21dpi: pink, 28dpi: turquois, 35dpi: yellow) other life stages show a low read map number. Although not normalised, this gene is highly expressed and specific at J2. The Gbrowse graphical representation above shows the relative amount of reads that map on to the contig for each life stage. The dense green region clearly shows up-regulation of UBI_WT in J2 with very little expression at other life stages.

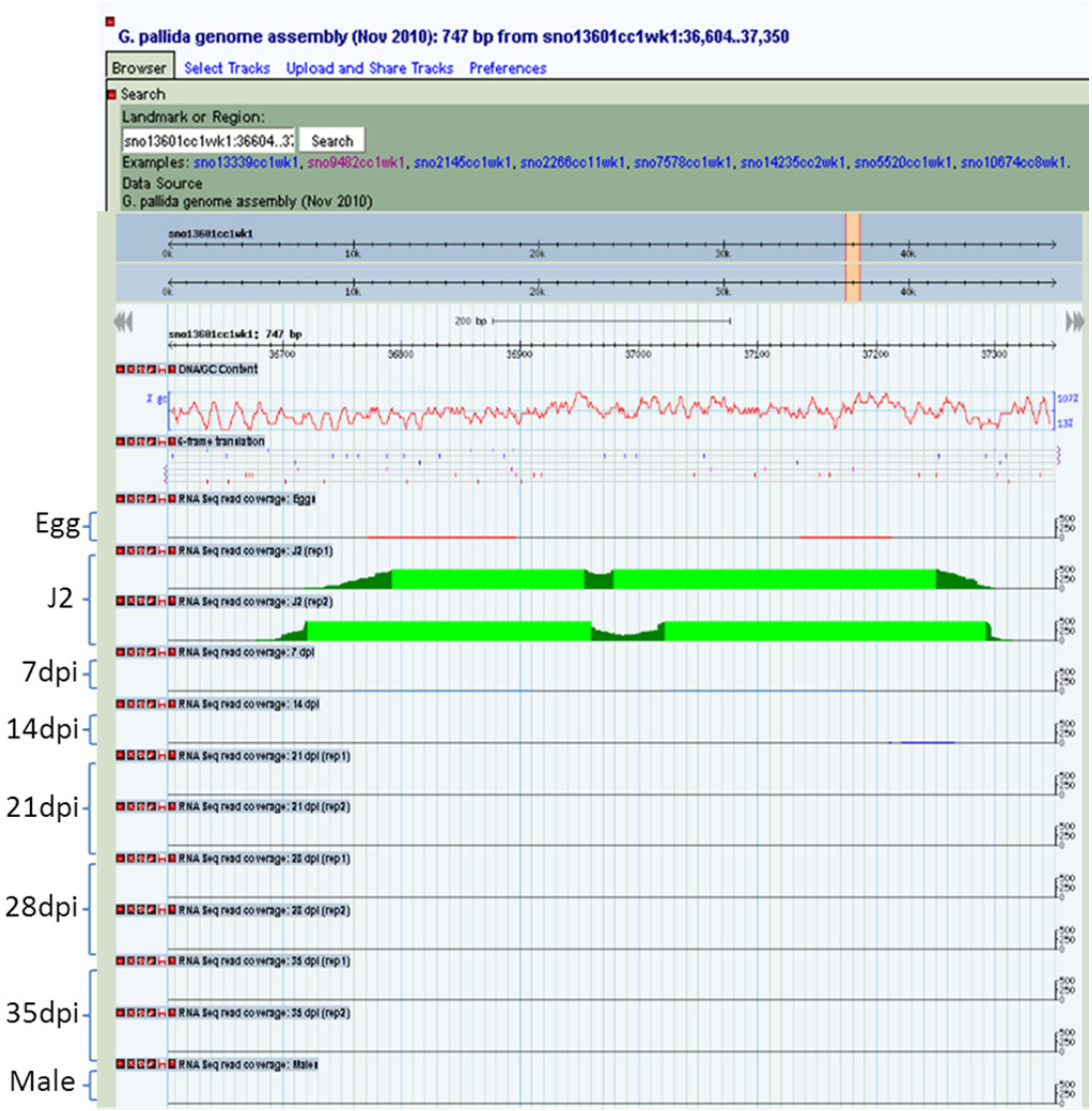


Figure 5.3 GpUBI-EP (UBI_Del) expression profile. Screen shot from GBrowse showing the part of the assembly encoding GpUBI-EP showing unnormalised RNAseq expression data for various life stages. Expression graphs are capped at 500 reads indicated by the colour change. Duplicated samples for the J2 life stage (green) show over a read depth of over the 500 threshold, other life stages (egg: red, 7dpi: light blue, 14dpi: dark blue, 21dpi: pink, 28dpi: turquois, 35dpi: yellow) other life stages show a low read map number. Although not normalised, this gene is highly expressed and specific at J2. The Gbrowse graphical representation above shows the relative amount of reads that map on to the contig for each life stage. The dense green region clearly shows up-regulation of UBI_WT in J2 with very little expression at other life stages.

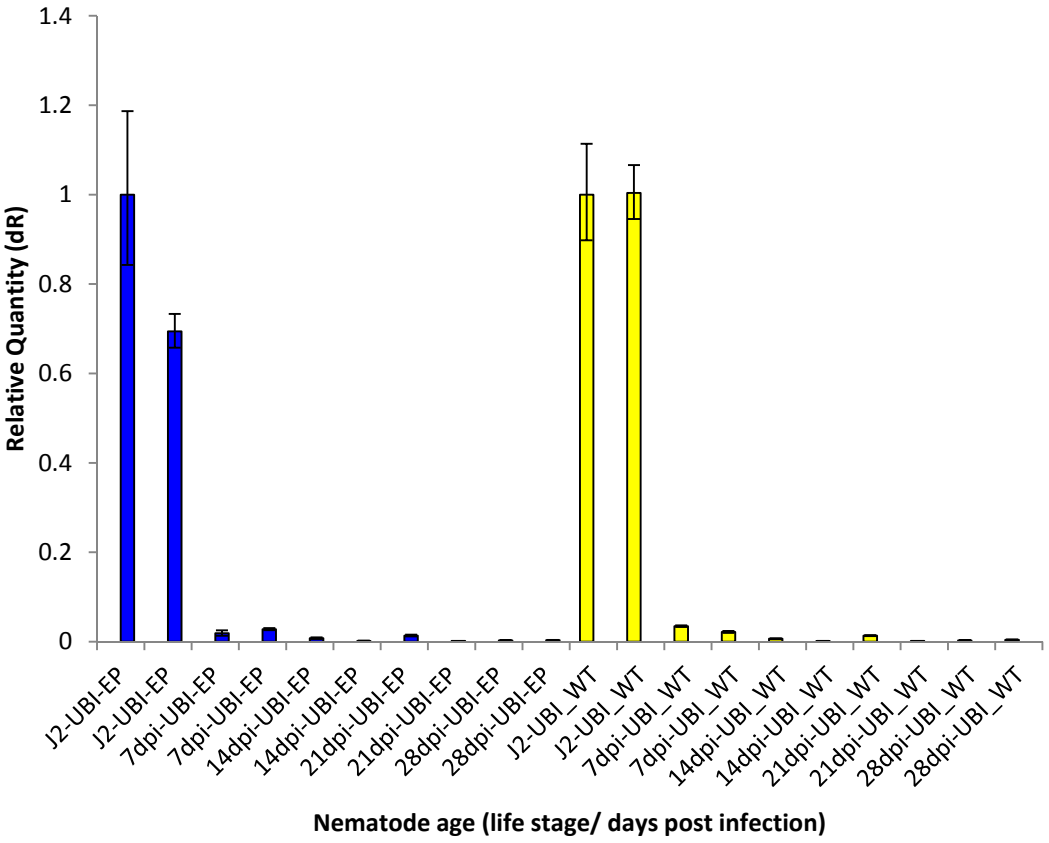


Figure 5.4 Graphical representation of GpUBI-EP qPCR data.
The graphs show the relative quantity of expression for both UBI_WT and UBI_Del (blue bars). Both these gene products were amplified by the same primer pair. It proved unsuccessful to amplify UBI_Del alone. However, amplification of UBI_WT was possible due to the successful design of primer specific to this version of the effector. The expression profile for UBI_WT is shown on the left side of the graph (yellow bars). The life stages sample were: J2, 7 dpi, 14 dpi, 21 dpi and 28 dpi. All life stages are shown in duplicate. The results were normalised using the expression profile for EF1α and EIF4α.

5.3.4 GpUBI-EP cleavage in plants and sub-cellular localisation

N. benthamiana leaves infiltrated with TRV UBI-GFP 3' or TRV UBI-GFP 5' that showed GFP fluorescence under a hand-held UV lamp were analysed by confocal microscopy. These TRV vector constructs allowed expression of the GpUBI-EP wild type or deleted form of the protein as N- or C-terminal fusions with GFP. For comparison, the *H. glycines* UBI-EP in similar GFP fusion constructs was also observed. The TRV UBI-GFP 3' and TRV UBI-GFP 5' from both UBI_Del and UBI_WT were all localised in the cytoplasm. By contrast, the *H. glycines* UBI domain (*H. glycines* 3'-GFP) localised to the cytoplasm whereas the *H. glycines* C-terminal extension (*H. glycines* 5'-GFP) localised to the nucleolus (Figure 5.5).

Cleavage of the expressed GpUBI-EP *in planta* was confirmed by western blotting of protein extracts from fluorescent leaf material. If the effector is cleaved after the GG of the UBI domain then the C-terminal extension-GFP fusion would be 28.3 kDa and the GFP-UBI fusion would be 35.6 kDa. (GFP: 27 kDa, Ubiquitin: 8.6 kDa and the C terminal extension: 1.3 kDa; Molecular weight calculated using http://expasy.org/tools/pi_tool.html). Western blotting showed that there is the expected molecular weight difference between TRV UBI-GFP 3' and TRV UBI-GFP 5' for both UBI_WT and UBI_Del suggesting that the GpUBI-EP may be cleaved at the GG (Wing, 2003) (Figure 5.6). The cleavage product from the TRV UBI-GFP 5' construct migrated slightly more slowly than free GFP, confirming that cleavage did not occur between GFP and the effector. Interestingly UBI_Del with GFP fused to the C-terminal extension was detected as a doublet suggesting that the cleavage may be inefficient in this variant of the protein.

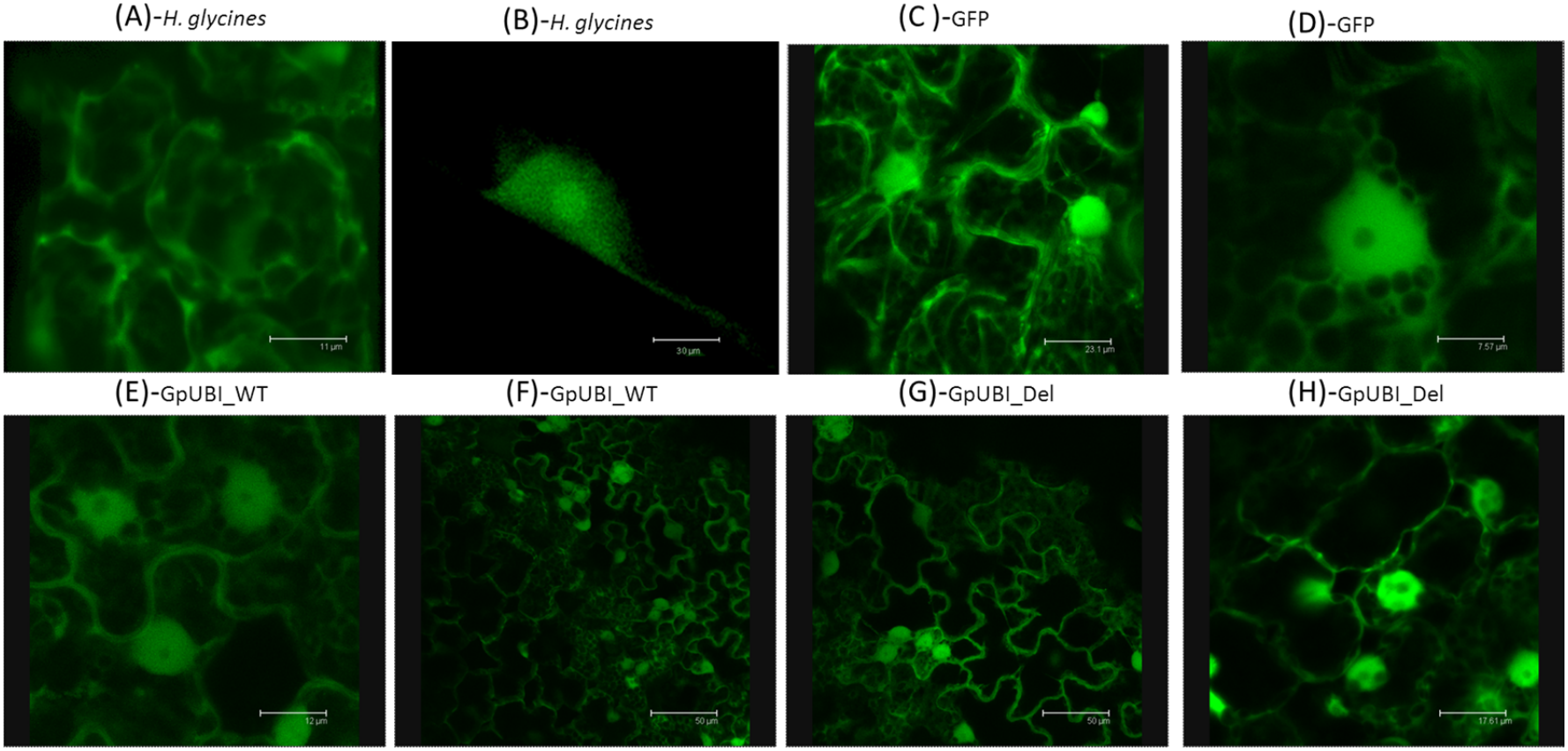


Figure 5.5: Confocal images showing cellular localisation of GpUBI-EP components by using GFP fusion proteins. Confocal images used to determine cellular localisation of *Heterodera glycines*, UBI_WT and UBI_Del effector components using GFP-effector fusions. Using a GFP fusion preceding the UBI domain (TRV UBI-GFP 5') it is shown that all UBI domains in question remain cytoplasmic once cleaved (A, E, G respectively). A GFP fusion at the C-terminal end of the effector (TRV UBI-GFP 3') shows that once cleaved the C-terminal extension of *H. glycines* (B) (image from John Jones) is targeted to the nucleolus and the C-terminal extension from UBI_WT (F) and UBI_Del (H) remains cytoplasmic once cleaved. Note the green nucleolus in the *H. glycines* C-terminal-GFP fusion picture (B), free-GFP is excluded from the nucleolus (see control images C and D). These images show a different localisation of the C-terminal extension from *G. pallida* and *H. glycines*. Scale bars: 11 µm, 30 µm, 23.1 µm, 7.57 µm, 12 µm, 50 µm, 50 µm, 17.51 µm for pictures A, B, C, D, E, F, G and H respectively.

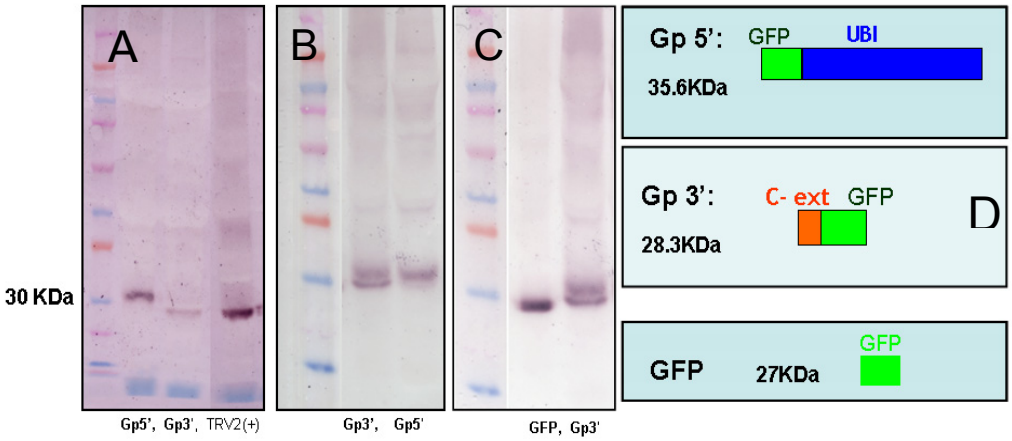


Figure 5.6: Western blot images demonstrating cleavage of the effector between the UBI domain and the C-terminal extension. (A): Western blot image of *G. pallida* Ubiquitin extension effector protein tagged with GFP using effector-GFP fusion (TRV UBI-GFP 3') and GFP-effector fusion (TRV UBI-GFP 5') constructs and a positive GFP control (TRV2). (B): UBI_Del tagged with GFP using effector-GFP fusion (TRV UBI-GFP 3') and GFP-effector fusion (TRV UBI-GFP 5') constructs. (C) image where free GFP has been blotted against the UBI_Del C-terminal extension-GFP blot. All gel images shown were manipulated for clarity where extra lanes were removed that were not relevant to the picture. Ladders and proteins of interest were analysed on the same gel. (D) Diagrammatic representations of the fusion products with their molecular weights displayed. Interestingly UBI_Del C-terminal extension-GFP blots as a doublet suggesting inefficient cleavage.

5.3.5 Suppression of host defences by GpUBI-EP

A variety of systems were used to induce host defence responses. The *P. infestans* protein INF1 was used to induce PTI in *N. benthamiana* which responds to this protein with a strong cell death response. ETI was induced by the co-expression of *P. infestans* protein AVR3aKI and its corresponding *R*-gene called *R3a* (Bos *et al*., 2006). ETI was also induced by the co-expression of *G. pallida* RBP1 (AVRGpa2) and potato Gpa2 (Sacco *et al.*, 2009). Host defences were also provoked by the transient expression of AVR4 from *Cladosporium fulvum* and its corresponding resistance gene *Cf4* from Tomato (Van der Hoorn *et al.*, 2000). The suppression of host defences was analysed by expressing an empty vector control, UBI_only domain, the full version of the effector UBI_WT and the isoform that contains the deletion UBI_Del. The constructs were either expressed before infiltration of the inducers, or co-infiltrated with the inducers. The infiltrated zones were analysed over a 7 day period.

Using the scoring method detailed in section 5.2.2, the results were analysed to determine if there was any significant suppression difference between the data-populations. Due to uneven distributions between the data in all experiments, a Mann-Whitney two sample non-parametric test was used to determine statistical significance; see Appendix 5 for detailed analysis.

5.3.5.1 GpUBI-EP suppression of PTI induced by INF1

When INF1 and the GpUBI-EPs or controls were co-infiltrated no suppression of host defence, as visualised by cell death, was observed (n=22, $P>0.05$). However, when *A. tumefaciens* containing a plasmid allowing expression of INF1 was infiltrated one day after infiltration of the effector constructs, data analysis suggested that UBI_only did not suppress PTI (n=22: $P=0.738$) while significant data (n=22: $P=0.738$ UBI_only, $P=0.006$ UBI_WT, $P<0.001$ UBI_Del) suggest that the both isoforms of the protein with the C-terminal extension did suppress PTI at time points 3, 5 (Figure 5.7) and 7 dpi.

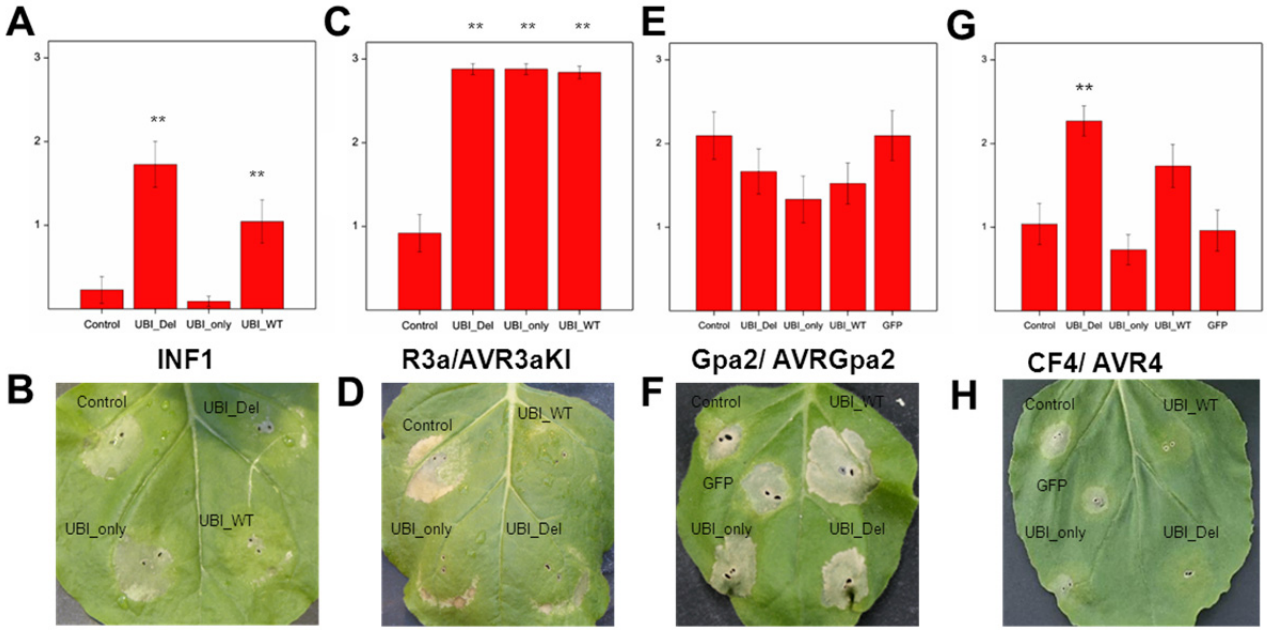


Figure 5.7: Statistical and visual representation of host defence-challenger assay experiments where GpUBI-EP was shown to interfere with host defences. Graphical bars represent standard error of the mean. (A) Represents 5 days post infiltration with INF1 (applied one day after effector infiltration). UBI_Del and UBI_WT were statistically significantly different to the control, whereas UBI_only was not significantly different to the control. This suggests the presence of the C-terminal extension is important for the suppression of PTI induced by INF1. (B) Image of *N. benthamiana* leaf 5 days post infiltration with INF1 applied one day after effector infiltration, showing a strong HR in the control and UBI_only infiltrated zones. Strong suppression of the HR can be seen in UBI_Del and UBI_WT infiltrated zones. (C) Graph representing the mean and standard error of the mean for each population of the GpUBI-EP isoforms three days post infiltration with AVR3aKI and R3a applied two days after effector/ control infiltrations. All populations were statistically significantly different to the control suggesting the ubiquitin domain is important for suppression of ETI induced by R3a/AVR3aKI. (D) Image of *N. benthamiana* leaf 5 days post infiltration with R3a and AVR3aKI infiltrated two days after effector/ control infiltrations showing a strong HR in the control and high levels of suppression in all UBI infiltration zones. (E) Graph representing the mean and standard error of the mean for each population of the GpUBI-EP isoforms 7 days post infiltration with Gpa2 and AVRGpa2 (RBP1) two days after effector/ control infiltrations (HR induced by this challenger combination takes longer to develop), showing a HR in all infiltration zones. No populations were statistically significantly different to the control. (F) Image of *N. benthamiana* leaf 7 days post infiltration with Gpa2 and AVRGpa2 (RBP1) two days after effector/ control infiltrations, showing a HR in all infiltration zones. (G) Represents 7 days post infiltration with CF4/ AVR4 (applied one day after effector infiltration). UBI_Del and UBI_WT were statistically significantly different to the control, whereas UBI_only was not significantly different to the control. This suggests the presence of the C-terminal extension is important for the suppression of defences induced by CF4/ AVR4. (H) Image of *N. benthamiana* leaf 5 days post infiltration with CF4/ AVR4 applied one day after effector infiltration, showing a HR in the control and UBI_only infiltrated zones. Suppression of the HR can be seen in UBI_Del and UBI_WT infiltrated zones.

5.3.5.1.1 Callose deposition assay

When components of the PAMP-triggered immunity pathway are activated by the recognition of PAMPs such as INF1 various 'first lines' of defence are activated including, for example, hydrogen peroxide production and callose deposition on the cell surface. An initial experiment was undertaken to determine whether or not callose deposition was increased when leaves were infiltrated with INF1 supernatant as compared to filtrate without INF1. In both cases leaves were previously infiltrated with *Agrobacterium* containing an empty pMDC32 vector. The presence of INF1 protein significantly increased the number of callose deposits revealed by aniline blue staining ($P<0.001$). This also suggested that *Agrobacterium* containing the plasmid pMDC32 did not interfere with callose deposition. Therefore, in the presence of INF1 callose is deposited on the cell surface.

The assay was subsequently repeated to determine if the presence of GpUBI-EP (UBI_WT) suppresses callose deposition induced by INF1. The presence of GpUBI-EP significantly reduced callose deposition induced by INF1 supernatant ($P<0.05$ ANOVA). In a direct comparison for the data sets which had been infiltrated with INF1 (PMDC32 with INF1 vs. GpUBI-EP with INF1), using a two-sample non-parametric Mann-Whitney test, the data suggest that GpUBI-EP reduces the amount of callose deposition when compared to an empty vector construct ($P<0.001$) (Figure 5.8). Raw data are provided in Appendix 5 – callose deposition results.

The assay was used to determine if GpUBI-EP (UBI_WT) interfered with callose deposition induced by the infiltration and subsequent perception of FLG22 in the same manner already described above. There was a significant difference between the pMDC32 with FLG22 vs. pMDC32 with filtrate ($P<0.05$, ANOVA). These data suggest that in the presence of FLG22 callose is deposited and the empty vector control did not interfere with this process. However, in the presence of GpUBI-EP there was no significant reduction in callose deposition ($P>0.05$, ANOVA).

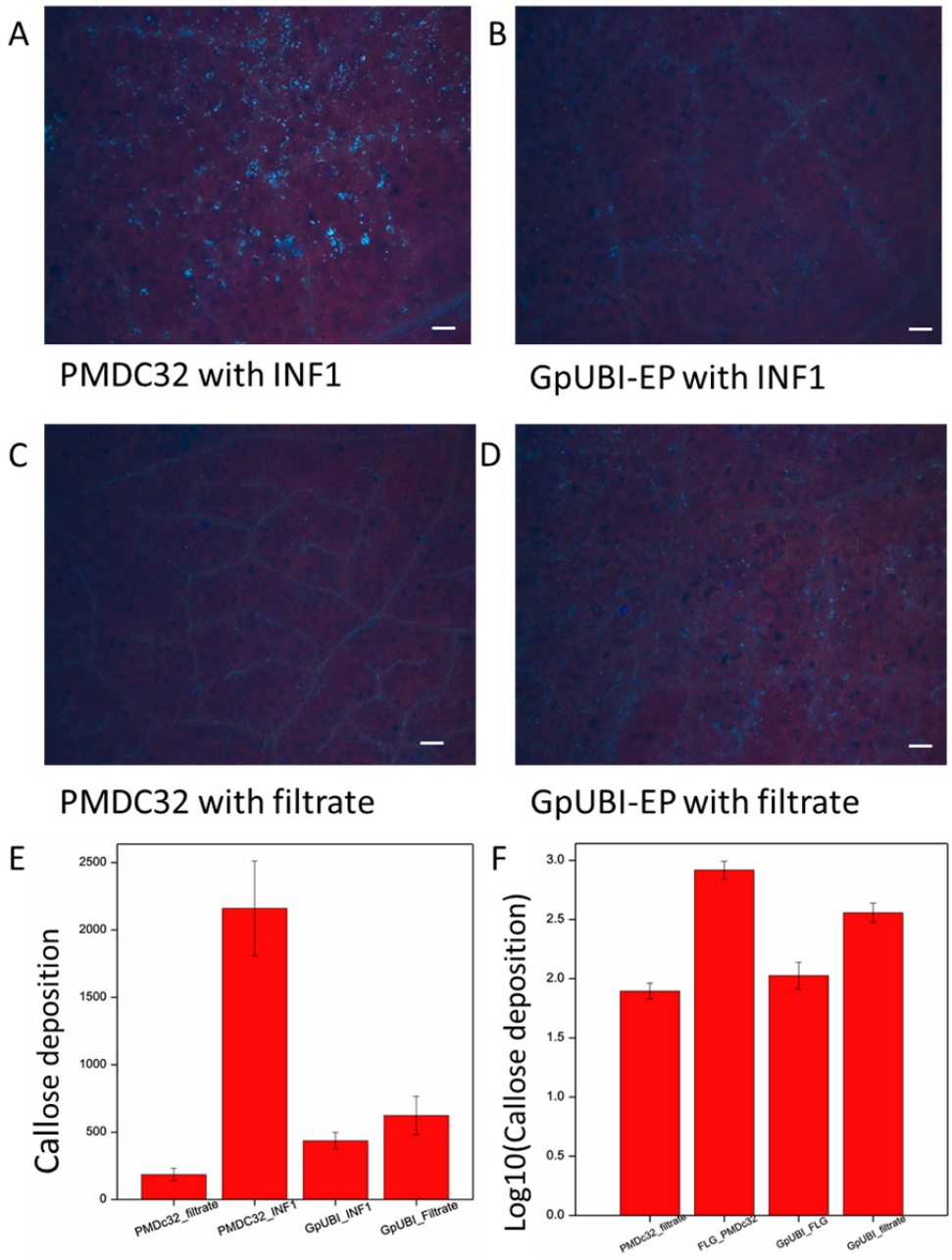


Figure 5.8: Example images from the callose deposition assay used to characterise GpUBI-EP. (A) Empty vector control infiltrated with INF1 filtrate: callose deposition can be seen as white/ blue spots as a result of the induction of PTI by the recognition of INF1. Callose deposition was quantified by a custom computer script that counted the number of pixels in the image that matched the colour of callose. (B) GpUBI-EP with INF1: when this data set was compared to pMDC32 with INF1 there was a significant reduction in callose deposition implying that GpUBI-EP may interfere with PTI induced by INF1. (C) Empty vector control with filtrate control: there was no inducer of PTI in this experiment and therefore callose deposition cannot be seen. There was a significant increase in callose deposition between pMDC32 with INF1 vs. pMDC32 with filtrate. This shows that INF1 did significantly increase callose deposition. (D) GpUBI-EP with filtrate control. (E) Graph representing the results from the callose deposition assay from INF1 induced PTI. The presence of GpUBI-EP significantly reduced INF1-induced callose deposition ($P<0.05$, ANOVA). (F) Graph representing the results from the callose deposition assay from FLG22 induced PTI. GpUBI-EP did not reduce callose deposition induced by FLG22 ($P>0.05$, ANOVA). Images for callose deposition induced by FLG22 are not shown. Scale bar is 100μm.

5.3.5.2 GpUBI-EP suppression of defences induced by CF4 and AVR4

One day after the infiltration of the effector constructs the combination of CF4 and AVR4 was infiltrated to induce cell death (Van der Hoorn *et al.*, 2000). The response induced by Cf4/ AVR4 is not clearly defined as PTI or ETI, see discussion. The results for 3, 5 and 7 dpi show that effector constructs that contain the C-terminal extension suppress defence responses induced by the CF4 and AVR4 combination (n=26: P=0.712 GFP, P=0.553 UBI_only, P=0.060 UBI_WT, P<0.001 UBI_Del) (Figure 5.7). The constructs containing the UBI domain alone did not suppress Cf4/AVR4 induced cell death.

5.3.5.3 GpUBI-EP does not stabilise CMPG1-YFP fusion protein in *N. benthamiana*

The data shown here implicate GpUBI-EP in suppressing PTI induced by INF1 and Cf4/AVR4. The signalling pathways induced by both of these elicitors are thought to function through CMPG1. The *P. infestans* effector AVR3aKI suppresses host defences induced by INF1 and Cf4/AVR4 by stabilising CMPG1 (Gilroy *et al.*, 2011; Bos *et al.*, 2010a; Gilroy *et al.*, 2011). When GpUBI-EP and CMPG1-YFP proteins were co-expressed no accumulation of fluorescent protein could be visualised within the nucleus, whereas co-expression of AVR3aKI and CMPG1-YFP led to an accumulation of fluorescence signal in the nucleus. Therefore GpUBI-EP does not stabilise CMPG1 in the same way as AVR3aKI (Figure 5.9).

5.3.5.4 GpUBI-EP suppression of ETI induced by R3a and AVR3aKI

Two days after the infiltration of the effector constructs the combination of R3a and AVR3aKI was infiltrated to induce a HR. The results for 3, 5 (data shown) and 7 dpi show that all effector constructs highly suppressed ETI induced by the R3a and AVR3aKI combination. The control was significantly different from UBI_only, UBI_Del and UBI_WT (n=26: P< 0.001 in each case). UBI_only was not significantly different to UBI_Del and UBI_WT (P=0.745 and P=0.572). UBI_Del was not significantly different from UBI_WT (P=0.230) (Figure 5.7). This suggests that the UBI domain is important for suppressing defence responses provoked by R3a and AVR3aKI and that the presence of the C-terminal extension is not important for this suppression.

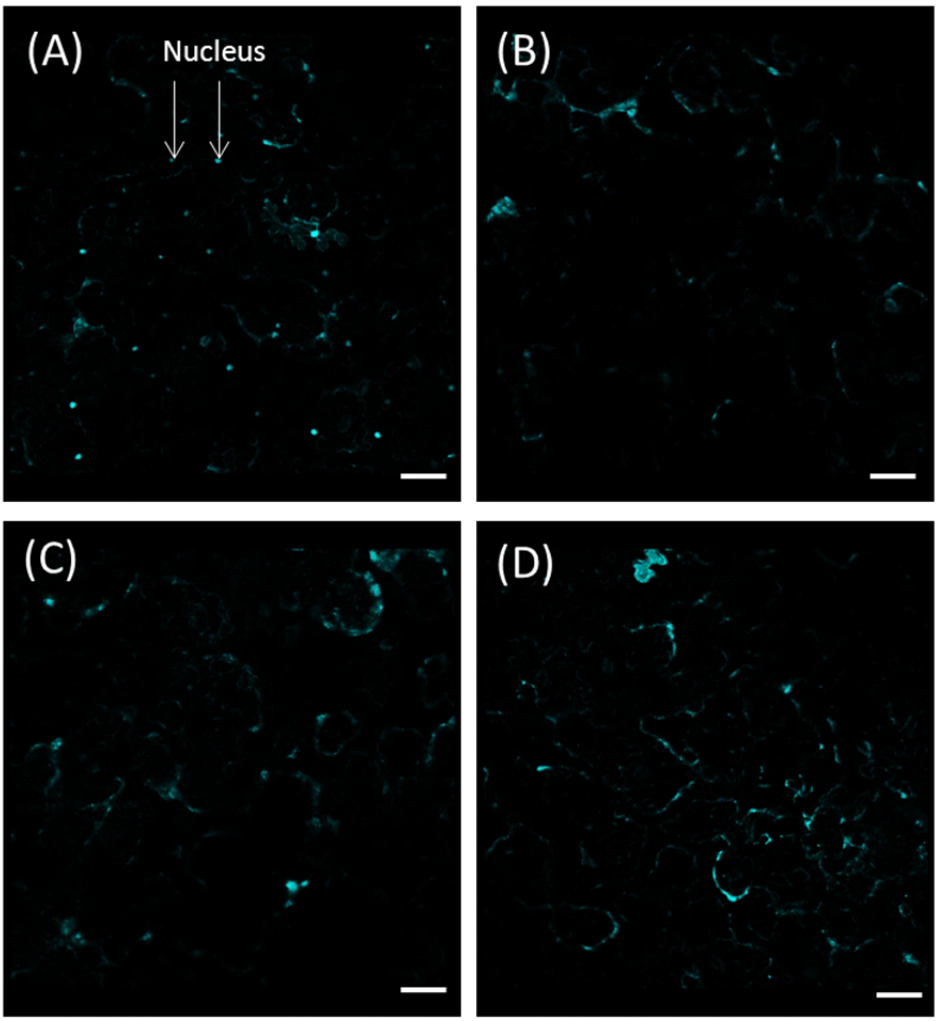


Figure 5.9: Confocal images of *N. benthamiana* leave cells co-transformed with CMPG1-YFP and effector's to determine the effector's ability to stabilisation the CMPG1-YFP fusion protein. (A) *P. infestans* effector AVR3aKI co-expressed with CMPG1-YFP, stabilisation can be visualised by the accumulation of the fusion protein in the nucleus (+ve control). (B) UBI_Del co-expressed with CMPG1-YFP, no stabilisation can be seen here hence no accumulation within the nucleus. (C) UBI domain of GpUBI-EP co-expressed with CMPG1-YFP, no stabilisation can be seen here hence no accumulation within the nucleus. (D) CMPG1-YFP expressed alone as a negative control. These data suggests that GpUBI-EP does not stabilise CMPG1. Scale bar is 60μm

5.3.5.5 GpUBI-EP does not suppress ETI induced by Gpa2 and AVRGpa2

Two days after the infiltration of the effector constructs the potato R gene *Gpa2* and its corresponding AVR gene *RBP1*, a SPRYSEC effector from *G. pallida,* were used to induce HR. After 7 days (*Gpa2/AVRGpa2* was infiltrated one day after effector infiltration) the control was not significantly different from UBI_only, UBI_Del and UBI_WT (n=21: P>0.05 in each case). Therefore the GpUBI-EP constructs did not suppress ETI induced by *Gpa2* and *AVRGpa2* (Figure 5.7).

Further infiltration experiments were conducted to assess the suppression of ETI by GpUBI-EP. The *P. infestans* AVR2 and its corresponding R genes *BLB3* from *Solanum bulbocastanum* and *R2* from *S. tuberosum*, the *Cladosporium fulvum* AVR9 proteins and the corresponding R gene *Cf9* from tomato were used in order to induce ETI in *N. benthamiana*. Unfortunately no consistent HR was induced in the controls for any of these systems even after several attempts. Therefore these combinations could not be used in the characterisation of GpUBI-EP.

5.3.5.6 GpUBI-EP infiltration does not inhibit further *Agrobacterium* transformation

Evidence was needed to show that GpUBI-EP did not inhibit subsequent or co-*Agrobacterium* transformation. Existing laboratory clones of a *G. pallida* SPRYSEC fused to GFP in two plasmids, pGWB5 and pGWB6, in *A. tumefaciens* GV3101 cells were used to determine if GpUBI-EP inhibited co-*Agrobacterium* transformation. UBI_WT was either co-infiltrated with these SPRYSEC-GFP constructs or infiltrated one day before infiltration with the SPRYSEC-GFP constructs. GFP was shown to be present by confocal microscopy in all experiments (as well as the empty vector pMDC32 control) thus confirming that the presence of the GpUBI-EPs does not prevent subsequent transformation from *A. tumefaciens* (Figure 5.10).

5.3.6 SPRYSEC effectors do not suppress ETI or PTI

To determine the validity of the infiltration assay, alternative effectors were characterised in the manner already described. This was important to determine whether or not a consistent induction of a defence response was possible in the presence of an effector protein. If the induction of a HR was not supressed then this would provide evidence that the observed effects are specific to the GpUBI-EP.

SPRYSECs are a large family of effectors found within *G. pallida*. Four of these SPYSECs, 12N3-3, 44D10, GPS-3 and RBP1, were used for the suppression infiltration assays detailed above that were used to characterise GpUBI-EP. All assays were performed using both co-infiltration of the effectors and inducers as well as the SPRYSEC effector being infiltrated 1 day before the infiltration of the defence response inducer.

For each of the inducers used (INF1, R3a/AVR3aKI and Cf4/AVR4) a strong response was found in all infiltrated regions. Therefore 12N3-3, 44D10, GPS-3 and RBP1 did not interfere with the defence responses invoked by INF1, R3a/AVR3aKI or Cf4/AVR4 (Figure 5.11). These results show that it was possible to provoke defence responses using INF1, R3a/AVR3aKI and Cf4/AVR4 in the presence of a *G. pallida* effector protein. These data provide supporting evidence for the results observed in the GpUBI-EP assay.

5.3.7 Analysis of infected *Arabidopsis* DUB mutant lines

It is known that ubiquitin extension proteins are cleaved by DUBs, and the data presented above have shown that the GpUBI-EP is cleaved in plants. An experiment was therefore undertaken to determine if DUB mutants are altered in their susceptibility to nematode invasion.

Arabidopsis thaliana wild type (Col 0) and DUB mutant lines *ubp13.1* and *ubp12.1* were infected with *H. schachtii,* a related cyst nematode which also has a ubiquitin extension protein amongst its complement of effectors. Both *ubp12.1* and *ubp13.1* mutant lines were more susceptible to nematode infection than the wild type control (Figure 5.12) ($P<0.001$). Raw data are provided in Appendix 5. DUBs may therefore have an important role in defence against nematodes as well as the bacterial pathogen *Pseudomonas syringae* (Ewan *et al.*, 2011).

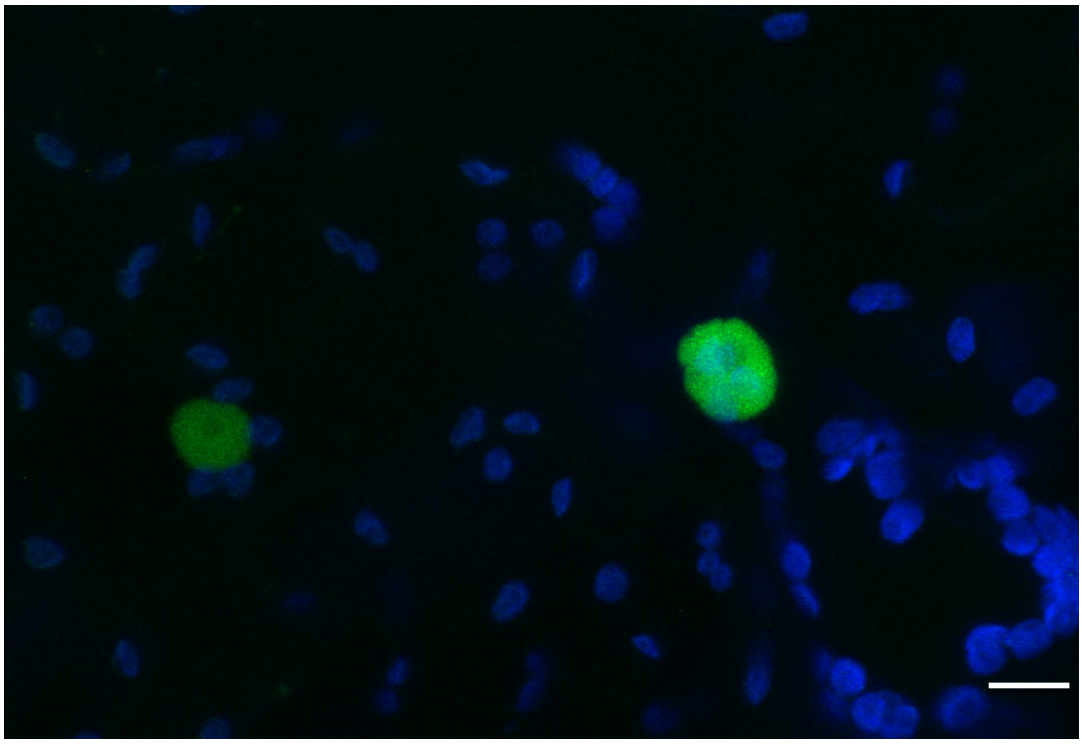


Figure 5.10: Confocal image of GFP transiently expressed post effector expression. The presence of GFP provides evidence that further transformation was possible. Confocal image analysis of *N. benthamiana* leaves infiltrated with *Globodera pallida* SPRYSEC fused to GFP (in plasmids pGWB5 and pGWB6 in *Agrobacterium tumefaciens* GV3101 cells) co-infiltrated with and infiltrated one day after infiltration with UBI_WT. GFP can be seen in the nucleus of the cell. Blue seen in the image is auto fluorescence from chloroplasts in the plant cell. Scale bar is 6μm

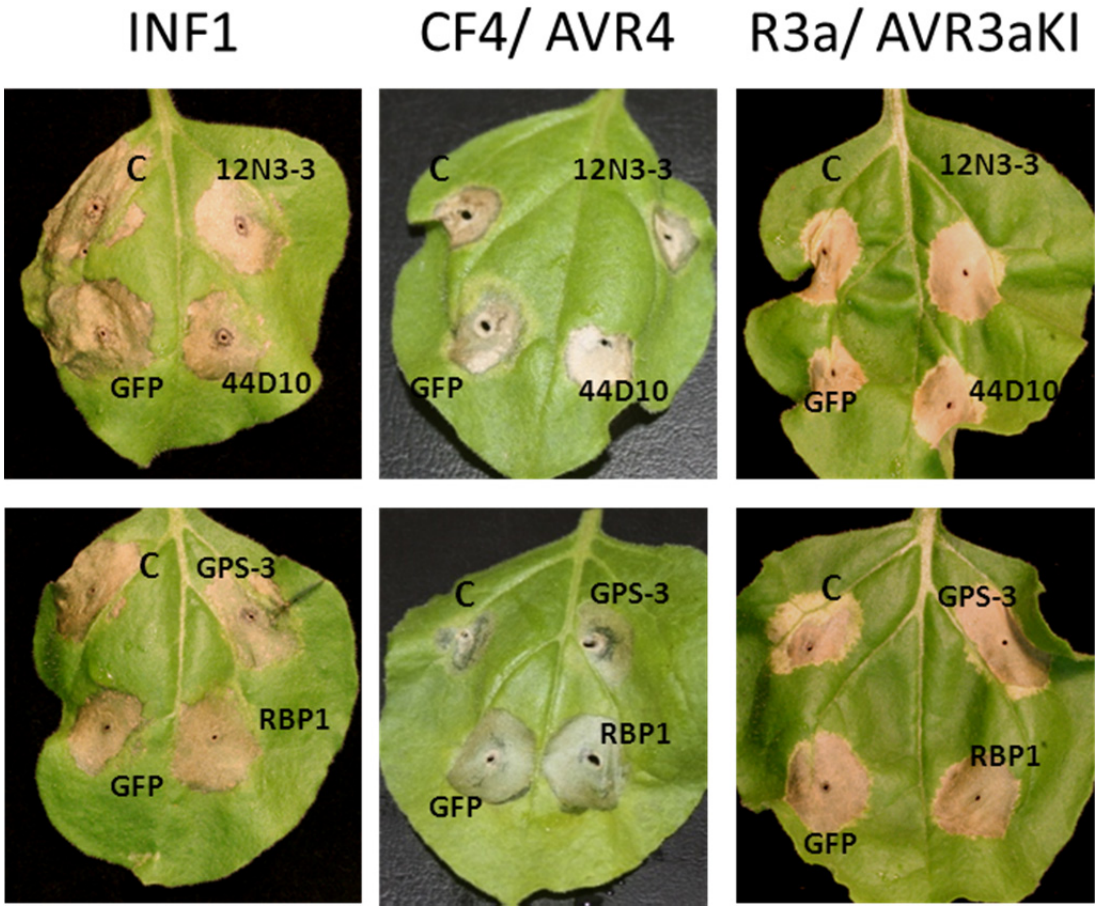


Figure 5.11: Suppression assay using SPRYSEC effectors 12N3-3, 44D10, GPS-3 and RBP1. The infiltration assay was performed on four SPRYSEC effectors using the defence response inducers used to characterise GpUBI-EP. None of the SPRYSECs assayed suppressed any of the induced defence response. This shows that the induction of a hyper sensitive response was possible in the presence of an effector protein. These findings support the observed data for the characterisation of GpUBI-EP.

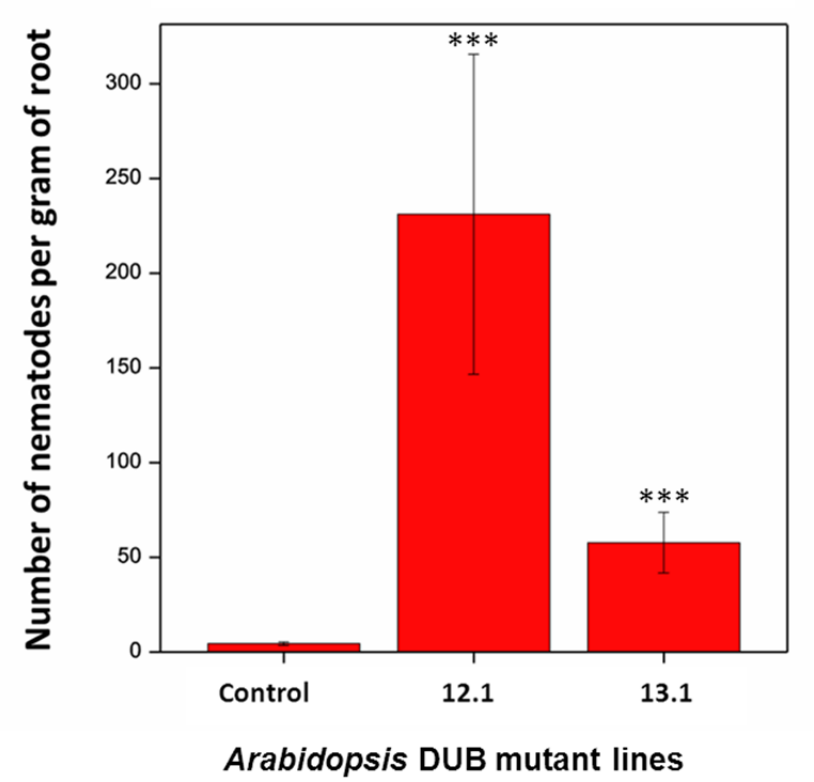


Figure 5.12: Nematode infection assay using *Arabidopsis* wild type and DUB mutants *12.1* and *13.1* with *Heterodera schachtii*. *Arabidopsis* WT and mutant lines *ubp13.1* and *ubp12.1* were infected with *H. schachtii to* determine if there was an increase in susceptibility when compared to the controls. Data suggest that *ubp13.1* and *ubp12.1* were significantly more susceptible ($P<0.001$) (ANOVA). Graph represents mean and standard error of the mean**.**

5.4 Discussion

The *G. pallida* UBI extension protein is comprised of three domains: a signal peptide, a highly conserved ubiquitin (UBI) domain and a C-terminal extension which shows no similarity to any known protein. Similar proteins have been identified in several species of cyst nematodes. It has been suggested that the C-terminal extension of *H. schachtii* may play a role in syncytium formation (Tytgat *et al.*, 2004). Circumstantial evidence in support of this is provided by the absence of a RKN homologue of this effector. However, the C-terminal extensions of the UBI_EPs from various cyst nematodes show that these sequences vary considerably between species, making it unlikely that they have a conserved role in syncytium formation throughout the Genus.

A difference in the 12 amino acid C-terminal extension of the *G. pallida* effector protein was found between nematode populations cultivated on susceptible and partially resistant host material. The frequency of UBI_Del isoform, which has a 3 amino acid deletion in the C-terminal extension, was increased in nematode populations cultivated on partially resistant material. Such an increase in the allele frequency suggests that the C-terminal extension has an important role in the function or recognition of the effector. Early statistical analysis suggests that there is a fitness penalty for carrying the deletion in the C-terminal extension (Dr Vivian Blok – unpublished data). In the work presented here the nematodes were grown on plants containing the *H3* resistance gene. Intriguingly, the frequency of the UBI_Del form also increases in nematodes grown on another resistance source, *Gpa5* (V. Blok, pers. comm.), suggesting that this selection is not specifically associated with recognition by a specific resistance protein. The ubiquitin domain of some of the UBI_Del proteins terminated in serine–glycine at the C-terminus instead of the usual glycine–glycine. Such ubiquitin domains are rare, although another example is a protein called HUB, in which the ubiquitin domain terminates in tyrosine–tyrosine but is still a functional ubiquitin (Jentsch and Pyrowolakis, 2000).

The UBI domain of the GpUBI-EP contains lysine residues at positions 6, 11, 27, 29, 33, 48 and 63 (all positions in relation to the start of the ubiquitin domain). These residues are essential for normal ubiquitin function. However, the ubiquitin domain does show some amino acid changes when compared to other ubiquitin sequences.

Ubiquitin is very highly conserved across species and it is therefore possible that these changes are functionally significant. The GpUBI-EP contains a valine at amino acid 23 instead of an isoleucine, a serine in place of threonine at position 28 and a methionine instead of a leucine at position 56. Although all of these amino acid substitutions are functionally conserved, they may be significant in altering the function of the UBI domain. Therefore a functional ubiquitination assay using the ubiquitin domain of the Gp_UBI_EP is required to confirm that the UBI domain can act as a functional ubiquitin.

Since the signal peptide will be cleaved when the GpUBI-EP is secreted from the nematode, the UBI and C-terminal extension will be the mature protein released into the plant. Ubiquitin extension proteins are known to be inactive, in terms of ubiquitination, due to the presence of the C-terminal extension (Kaiser and Huang, 2005). DUBs cleave the C-terminal extension thus releasing the UBI domain. The role of the C-terminal extension in the GpUBI_EP is not clear. It may have a function itself within the host, as suggested for the *H. schachtii* UBI_EP. The C-terminal extension of this protein is cleaved and localises to the nucleolus of the host cell while the UBI domain remains within the cytoplasm (Tytgat *et al.,* 2004). Unfortunately it proved technically challenging to clone the C-terminal extension for the purpose of functional analysis in the absence of the ubiquitin domain. If the ubiquitin domain itself has a role in the host then the C-terminal extension may act to keep the ubiquitin inactive, or aid in stability, until cleavage takes place within the host. The presence of the C-terminal extension was important for the suppression of host defences in some of the assays reported here (see below).

Western blotting showed that the UBI domain and C-terminal extension are cleaved from each other at the GG terminus of the UBI domain following expression of the Gp_UBI in *N. benthamiana.* It is possible that this cleavage is performed by DUBs and that DUBs are the target of the effector. An experiment using DUB silenced plants, in which cleavage is seen to be reduced, would provide direct evidence of the involvement of DUBs in the cleavage of the UBI-C-terminal fusion protein. Two *Arabidopsis dub* mutants, *12.1* and *13.1,* were found to be more susceptible to nematodes. This suggests DUBs have a role in defence against nematodes. Further to this, RNAi of the *GpUBI-EP* may determine if the Gp-UBI_EP is essential for

successful parasitism but such an experiment would prove technically difficult to achieve due to the similarity of the effector's UBI domain to other UBI genes, resulting in the possible down-regulation of off-target genes.

Confocal microscopy was used to determine the cellular localisation of the GpUBI-EP UBI domain and C-terminal extension following cleavage within the plant. In contrast to *H. glycines* and *H. schachtii* (Tytgat *et al.*, 2004), in which the cleaved UBI domain is cytoplasmically localised and the C-terminal extension is localised in the nucleus and nucleolus, the *G. pallida* UBI domain and C-terminal extension both remain in the cytoplasm. This is true for both the UBI_Del and UBI_WT isoforms. This suggests that the UBI domain has a functional role within the cytoplasm of the host, and that the C-terminal extension, which is highly variable and shares no similarity between cyst nematodes, may have different roles within the host cell. Alternatively, the C-terminal extension may simply be present in order to target the effector to a DUB, with the actual sequence of the extension being of limited relevance. It is also possible that different DUBs are targeted in different host plants and that these DUBs have different subcellular localisations.

The *GpUBI-EP* mRNA is expressed in the dorsal gland (Jones *et al.*, 2009b), qPCR and bioinformatics expression data show that the UBI_EP is highly up-regulated at the J2 stage which is a critical stage for the invading nematode, where suppressing host defences is essential for survival and successful parasitism. There is relatively low expression of GpUBI-EP at later life stages (7dpi, 14dpi, 21dpi and 28dpi) which suggests the presence of this effector may still be required. The J2 nematode invades the plant root and leaves a trail of destruction as it migrates through the root cells. However, once the nematode identifies a cell that it will attempt to modify into a feeding site, its behaviour changes. Now it delicately probes the cell with its stylet and introduces effectors into the cell that induce a feeding site. At this time it is essential for the nematode to suppress host defences. If the cell responds to the nematode by callose deposition (Hussey *et al.*, 1992) the nematode will select another cell. It is possibly significant that the expression of GpUBI-EP is highest at this phase of the nematode life cycle where it is essential to suppress host defences.

Infiltration and callose deposition assays were performed to determine if the various isoforms of GpUBI-EP suppressed host defences provoked by PAMPs including INF1 and FLG22. Both isoforms of GpUBI-EP did suppress INF1 induced PTI, whereas the UBI domain alone did not; however, GpUBI-EP did not suppress FLG22 induced PTI. The presence of the C-terminal extension may therefore be essential for the suppression of PTI induced by INF1. Further to this, both isoforms of GpUBI_EP suppressed PTI induced by co-expression of Cf4 and AVR4 whereas the UBI domain alone did not. Although Cf4/AVR4 is referred to in the literature as an effector/*R*-gene combination, Cf4 is an extracellular receptor and the defence responses induced by this receptor are considered to be PTI (Stergiopoulos and de Wit, 2009). In support of this, INF1 and Cf4/ AVR4 induced defence responses both signal through an E3 ligase called CMPG1. The *P. infestans* effector AVR3aKI suppresses INF1 and Cf4/AVR4 induced PTI by stabilising CMPG1 (Gilroy *et al.*, 2011). Although GpUBI-EP also suppresses PTI induced by INF1 and Cf4/AVR4, it did not stabilise a CMPG1-YFP fusion product. Therefore GpUBI-EP may have a different target within this pathway that results in the suppression of PTI.

Further assays were performed to investigate the potential role of GpUBI-EP in suppression of ETI. Surprisingly, all isoforms of the GpUBI-EP, including the UBI domain alone, suppressed ETI induced by the combination of R3a and Avr3aKI ($P<0.001$). No significant difference was found between the infiltration zones containing UBI_only, UBI_WT, and UBI_Del. This suggests that the main ubiquitin domain is important for the suppression of ETI induced by AVR3aKI and R3a. Further analysis showed that this suppression was specific to the R3a/Avr3aKI R gene/Avr gene combination, with no suppression of ETI induced by co-infiltration of Gpa2 and RBP1. In addition, other *G. pallida* effectors, infiltrated under precisely the same conditions, did not suppress R3a/Avr3a induced ETI and analysis using a GFP tagged protein showed that further *Agrobacterium* transformation was not inhibited by the presence of the GpUBI_EP.

The suppression of both PTI and ETI by GpUBI_EP was unexpected. This suggests that GpUBI-EP may interfere with more than one defence signalling pathway. The full length protein and/or the C-terminal extension may interfere with PTI while the

UBI domain may subsequently interfere with ETI. However, further analyses of the signalling pathways targeted are needed.

Although many effectors from other pathogens have been identified that suppress host defences, there are few published examples of nematode effectors involved in suppression of PTI or ETI. One such example from *H. glycines* and *H. schachtii* is an annexin-like effector which is expressed in their dorsal gland. Annexins are thought to belong to a family of proteins that are involved in calcium regulated activities at surface membranes. Transgenic *Arabidopsis* expressing *H. schachtii* annexin *Hs4F01* resulted in a greater infection of *H. schachtii* when compared to controls. Yeast two-hybrid data indicates that Hs4F01 interacts with an oxireductase of the 20GFe(II) oxygenase family, which is thought to be involved in plant defence and stress response. This data implicates annexin-like effectors in interfering with host defences (Patel *et al.*, 2010b). The data presented here suggest that GpUBI_EP may suppress both the PTI and ETI pathways. The increasing identification of effectors that interfere with or mimic the ubiquitin proteasome pathway strongly highlights its importance to both the host and to pathogens as a means of controlling cellular processes. In many cases effectors that act as avirulence proteins and the corresponding resistance genes have been identified, providing useful information regarding how plants resist a pathogen attack.

An effector gene from an economically important plant pathogen *P. infestans*, *Avr3a* encodes a protein that contains a conserved RxLR-EER motif that is found in many oomycete effectors. The RxLR-EER motif functions to deliver the effector to the plant cytosol. The corresponding R protein R3a that recognises Avr3a has been identified in potato (Birch *et al.*, 2006; Frei dit Frey and Robatzek, 2009). The identification of common motifs between effectors could therefore provide useful information for comparative genomics and the identification of effectors' mode of action may, in turn provide useful information about the plant defence system.

For most of the nematode effectors identified to date, no function has been ascribed. Given that cyst nematodes have a biotrophic phase that lasts several weeks, it is very likely that they produce effectors that have a role in suppressing the host defence system. The details of how feeding sites are induced are still not known. The

generation of high throughput genome sequence data will allow mining of genomes for candidate effectors and comparative genomics with other pathogens, to seek similarity between effectors of different species. In addition to generating novel scientific information this work has profound practical implications. Understanding which are the important effectors and their function will allow new control strategies based on biotechnological approaches to be developed. In addition, effectors that are recognised by resistance genes can be used in resistance breeding programmes and identifying essential effectors that are recognised by novel sources of resistance offers the prospect of durable resistance. The importance of such work will only increase due to the ever increasing list of banned pesticides and increasing demand for food across the world.

Summary

- A gene encoding a *G. pallida* ubiquitin extension protein, *GpUBI-EP*, is highly up-regulated at the J2 life stage as shown by qPCR and RNAseq.
- There are two isoforms of the effector in populations of nematodes.
- GpUBI-EP suppressed PTI induced by INF1 and Cf4/AVR4 in *Agro*-infiltration experiments.
- GpUBI-EP suppressed callose deposition induced by INF1 supernatant but did not suppress callose deposition induced by FLG-22.
- Suppression of PTI is correlated with the presence of the C-terminal extension.
- GpUBI-EP suppressed ETI induced by AVR3aKI/AVR3.
  DUB mutant *Arabidopsis* plants are more susceptible to nematode invasion.

6 General discussion

The potato cyst nematode *Globodera pallida* secretes a cocktail of proteins, which are synthesised in its oesophageal gland cells, through its stylet into the host cell. These proteins are termed effectors. Effectors are used to manipulate and control the host throughout the life cycle of the nematode (Smant and Jones, 2011; Sobczak and Golinowski, 2011). The aim of this project was to identify the full effector complement of *G. pallida* using several bioinformatic approaches, from newly available genome and transcriptome data. This is the first analysis of the full set of effectors from any PPN genome reported to date. A subset of the identified effectors were then characterised for their phenotypic effect when over-expressed in potato and *Arabidopsis.* One effector, a ubiquitin extension protein, was then characterised in more detail and was found to play a role in host defence suppression.

6.1 The *G. pallida* genome sequence

No robust mechanism for amplification of DNA from a single nematode is available. Therefore the sequencing project was conducted on DNA extracted from a substantial population of nematodes. Sequencing from a population of nematodes rather than an individual will inevitably result in sequence variation due to SNP and allelic variations, making the data much more difficult for the assembly software to work with. The resulting genome sequence obtained for *G. pallida* consists of 6872 scaffolds and 37% of these contain at least one "NNN" region. These regions represent clusters of sequences that are known to be linked from paired-end reads but for which the intervening sequence is not complete. In addition, analysis of the predicted protein set shows that over 7% of the predicted proteins do not start with a methionine, suggesting that gene predictions are not completely accurate. Indeed the subsequent manual annotation of effector sequences that was performed as part of this project required correction of many of these predicted gene models. The availability of RNAseq data was particularly useful in this process, providing information about transcribed regions that the gene prediction software had missed.

As genome sequencing technology advances, more advanced platforms are developed that are faster, cheaper and may require less starting material (Baker,

2010). The average read length produced by the latest sequencing technologies, termed third generation sequencing technologies (Schadt *et al.*, 2010), can be over 2000 bp, with read lengths of greater than 5000 bp claimed for the Nanopore system (Liu *et al.*, 2012). Re-sequencing of a genome and transcriptome with a technology that can produce long reads would be advantageous for assembly and may help to reduce the total number of contigs (Bao *et al.*, 2011), resulting in a better approximation of the genome itself. The re-sequencing of the *G. pallida* genome from a single worm, which may be possible with third generation technologies, would also provide information supporting the validity of the identified SPRY domain-containing family of genes reported in this thesis. An understanding of the extent and expression of the SPRY domain family in different individuals may also provide information about how this gene family functions in *G. pallida* and build on existing knowledge (Jones *et al.*, 2009b; Postma *et al.*, 2012; Rehman *et al.*, 2009b; Sacco *et al.*, 2009a; Sacco *et al.*, 2007). Refinement of the genome sequence could also provide the data required to allow promoter identification. The identification of promoters that control the expression of families of effectors may allow the transcription factors that bind to these regions to be used as targets for new control methods.

6.2 Effector identification

Bioinformatic analysis of genome data has proved to be a useful tool in identification of genes of interest as demonstrated for *Meloidogyne incognita* (Abad *et al.*, 2008), *M. hapla* (Opperman *et al*., 2008) and *Bursaphelenchus xylophilus* (Kikuchi *et al.*, 2011). Sequencing of mRNA, known as RNAseq, is a useful tool for identifying genes that are expressed (Haegeman *et al.*, 2011b; Nicol *et al.*, 2012) and can be used to examine quantitative expression profiles at different life stages (below) (Choi *et al.*, 2011). The analysis of the *G. pallida* genome revealed a large number of genes that were orthologues of effectors from closely related cyst nematodes. Some of these effectors were present in *G. pallida* in large gene families. Comparisons between cyst nematode effectors and those from *M. incognita* showed that there was little overlap between the two sets, with the exception of cell wall modifying enzymes. This is consistent with previously reported findings (Gao *et al.*, 2003a), although this earlier work analysed far smaller datasets than the whole genome

sequence described here. By contrast, comparisons of secreted proteins from *M. incognita* and *Brugia malayi* suggested that 26 proteins secreted by *M. incognita* were also secreted by *B. malayi* (Bellafiore *et al.*, 2008). However, it is important to note that this study compared total secreted proteins rather than effectors and that effectors will be a far more highly evolved subset of the total secreted proteins in a nematode.

There were four effectors that were common between *G. pallida* and *M. incognita*, and one of these has also been identified in a migratory plant parasitic nematode (*Pratylenchus coffeae*) (Haegeman *et al.*, 2011b). The occurrence of this effector in cyst, root-knot and migratory nematodes could implicate the effector as having an important and conserved role in plant parasitism. If this effector proves to be essential for successful parasitism by plant parasitic nematodes, then this could be used as a target to control a wide range of these nematodes. Assuming a conserved function, this effector is likely to play a role in a process common to migratory and biotrophic nematodes. This may be migration or suppression of host defences during migration. A similar function may also be ascribed to chorismate mutase, which is also present in root-knot nematodes, cyst nematodes and migratory endoparasites. Previous work on this effector from *M. javanica* which suggested a role in manipulation of auxin metabolism (Lambert *et al.*, 1999b) may need to be reinvestigated.

The lack of overlap of effectors between RKN and cyst nematodes is consistent with the idea that biotrophic plant-parasitism evolved independently in these two genera of nematodes (van Megen *et al.*, 2009). RKN (*Meloidogyne* spp.) have a large host range (Hussey and Janssen, 2002) that includes more than 2000 plant species (Opperman *et al.*, 2009), whereas cyst nematode have a narrow host range (Gheysen *et al.*, 2006). Both RKN and cyst nematodes hatch as second stage juveniles (J2) from eggs and enter the plant root. RKN J2s migrate intercellularly until they reach the vascular cylinder. The nematode then induces plant cell mitosis in the absence of cytokinesis, leading to the formation of ‘giant cells’. Cyst nematodes, in-contrast migrate intracellularly through the zone of elongation to a site near the vascular tissue and induced a feeding site known as a syncytium (Gheysen *et al.*, 2006), The

amount of tissue destruction caused by the different migration strategies may be an important aspect in the need for host defence suppression at this stage. Therefore in spite of superficial similarities between these nematodes in terms of how they appear to infect plants, there are profound differences in their biology that are reflected in their effector profiles.

Cell-wall modifying enzymes have been identified from a wide range of plant parasitic nematodes and may be a prerequisite for plant parasitism by nematodes (Abad *et al.*, 2008; Kikuchi *et al.*, 2011; Opperman *et al.*, 2008). Nearly twice as many cell-wall modifying genes have been identified in *M. incognita* as in *G. pallida.* It is possible that the different migration methods used by RKN and cyst nematodes and differences in their host range could account for these differences in cell-wall modifying enzymes. The large number of cell-wall modifying genes identified in *M. incognita* may permit intercellular movement in a wide range of hosts, whereas those identified in *G. pallida* could be a specific set required for intracellular movement in *Solanaceous* species. Differences in cell wall degrading enzymes may also reflect varying gene family expansion rates in the two nematode groups or may simply be a reflection of the genome duplication that has taken place in *M. incognita* (Abad *et al.,* 2008).

Analysis of the transcriptome data generated as part of the *G. pallida* sequencing project suggest that *G. pallida* may use its own cell-wall modifying enzymes to metabolise the cell walls of the syncytium. This is a completely novel finding as previous studies have suggested that nematodes induce expression of the plant's cell wall degradation machinery in order to manipulate cell walls within the syncytium (e.g. Goellner *et al.,* 2001).

Differences in the feeding sites induced by RKN and cyst nematodes (discussed further below) could underlie the difference in the effector sets between these nematodes. A syncytium is formed by local cell wall degradation and subsequent fusion of the protoplasts of hundreds of cells. Nuclei within the syncytia undergo repeated S (synthesis) phases of the cell cycle (in which DNA is synthesised – also known as endoreplication) but without nuclear division (Gheysen *et al.*, 2006). RKN induce the formation of several binucleate cells that undergo mitoses without cell

division. This results in several large multinucleate cells, known as giant cells (Goverse *et al.*, 2000b). Cells that surround the giant cell undergo concurrent hyperplasia and hypertrophy resulting in galls on plants (Hussey and Janssen, 2002). However, both giant cells and syncytia act as metabolic sinks that deliver plant resources to the parasitic nematode. Therefore it would be logical to predict that a yet unknown effector(s) of similar function could be involved in redirecting plant resources to the feeding sites. One such candidate may be the invertases, these act as sink-source regulators by inducing a sink status in the cell where they are present (Roitsch, 1999).

The cell cycle is manipulated in both giant cells and syncytia although different modifications to this process are induced in the two structures. These differences could account for some of the differences in effector sets from RKN and cyst nematodes. Several secreted proteins identified from *M. incognita* have been hypothesised to be involved in giant cell formation and cell cycle regulation. These include a CDC48-like protein that could be involved in cell proliferation, an S-phase kinase-associated protein 1 with a nuclear localisation signal (SKP1 – see below), and translationally-controlled tumour proteins (TCTPs) that are thought to be involved in growth, cell cycle progression and protection of cells against stress and apoptosis. In addition NAC protein, Calcium-Dependent Protein Kinase (cell proliferation) and Nod-factors that may induce cytoskeletal changes were identified (Bellafiore *et al.*, 2008). However, it should be noted that there is still no evidence, other than their presence in secreted proteins collected from nematodes, that any of these proteins are actually effectors or that they have a function in the host-parasite interaction. Many lack signal peptides for secretion and may therefore simply have been present in collected secretions due to their release from dead worms within the sample. A SKP-1-like protein has also been identified as an effector in *H. glycines*. SKP1 is a component of the SCF complex that provides E3 ubiquitin-protein ligase activity (Gao *et al.*, 2003a). However, SKP-1 proteins are a large family and the proteins in cyst and root knot nematodes may have different functions.

The analysis of the *G. pallida* genome data revealed many novel putative effectors for which no significant homology could be found using standard BLAST searches.

PFam analysis of these novel sequences showed that some contain known domains. A ZF-RBK1 domain was identified in one novel effector, which is a sub-unit of SKP1 and these proteins may be involved in cell cycle control (Sasagawa *et al.*, 2003). In some of the other novel proteins, domains that are usually associated with transcription factors were identified. The induction of a feeding site is associated with large changes in gene expression in the host (Szakasits *et al.*, 2009b). It is possible that the nematode may directly induce these changes by introducing transcription factors into the host. In support of this, some of the effectors identified in this study have been shown to localise to the plant nucleolus in work underway at the JHI and Leeds groups. As part of this project, transgenic plants were produced that over-express some of the effectors. It would be possible to characterise the changes in plant gene expression in these plants using microarrays and to compare any changes in gene expression that occur as a result of the presence of the effector with changes known to occur in the feeding site (Ithal *et al.*, 2007a; Ithal *et al.*, 2007b; Puthoff *et al.*, 2003; Szakasits *et al.*, 2009a).

Novel effectors which had no sequence similarity to any known protein but contained calcium binding domains were identified. Calcium binding effectors may interfere with host cellular signalling pathways and may also suppress host defences, which rely on calcium signalling (Hogenhout and Bos, 2011). A calreticulin from *M. incognita* has been identified that suppresses host defences that may operate in this way, binding apoplastic calcium and preventing influx that occurs as part of normal defence signalling (Jaouannet *et al.,* 2012). Effectors containing calcium binding domains have also been identified from aphids (Nicholson *et al.*, 2012), bacteria (Aslam *et al.*, 2008), oomycetes (Xiong *et al.*, 2006) and fungi (Kloppholz *et al.*, 2011). This suggests calcium binding effectors play an important role in many pathosystems.

Novel effectors that have ubiquitin E2 and E3 domains were identified. The ubiquitin-proteasome pathway has been implicated in a number of plant pathogen interactions (Angot *et al.*, 2006b; Bellafiore *et al.*, 2008; Birch *et al.*, 2009a; Elling *et al.*, 2007; Ewan *et al.*, 2011; Gao *et al.*, 2003a; Goehre *et al.*, 2008b; Hewezi *et al.*, 2010b; Schrammeijer *et al.*, 2001b; Tytgat *et al.*, 2004). The increasing number of

putative effectors with similarity to components of the ubiquitin-proteasome pathway, including the novel effectors identified in this project, emphasises the importance of the ubiquitin-proteasome pathway in plant-pathogen interactions.

A family of cyst nematode effectors called SPRYSECs have previously been identified and some of these are able to supress host defences (SPRYSEC-19 *G. rostochiensis*). In addition, the SPRYSEC RBP1 is an AVR gene that is recognised by Gpa2 in potato (Jones *et al.*, 2009b; Postma *et al.*, 2012; Rehman *et al.*, 2009b; Sacco *et al.*, 2009a; Sacco *et al.*, 2007). However, until the availability of the genome sequence the size of the SPRYSEC gene family was not apparent. 299 SPRY domain containing proteins were identified in *G. pallida*. Effectors that are subject to strong selection pressure and thus rapid evolution due to the interaction with the host show signatures of high birth and death rate (Win *et al.*, 2012). An example of effectors that reflect this process is a class of effectors from oomycetes, called RXLR (Arg-X-Leu-Arg) effectors (Win *et al.*, 2012). The RXLR motif is diagnostic for this group and is required for translocation into the plant cell (Whisson *et al.*, 2007a). There are over 520 putative RXLR-effectors in *P. infestans* (Haas *et al.*, 2009). RXLR effectors show high divergence between species and are thought to be under high selection pressure due to their interactions with the host (Haas *et al.*, 2009). There are some intriguing parallels between the SPRY domain proteins in *G. pallida* and the RXLRs of *P. infestans.* Both are substantial gene families that have undergone rapid evolution within individual clades of nematodes/oomycetes. Both families seem to harbour *Avr* genes and both families include proteins that suppress host defences. In both cases a subset of the gene family is expressed in any given host. These data suggest that SPRY domain containing proteins may be excellent candidates for *AVR* gene identification and resistance screening.

Not surprisingly, there is no known overlap between effectors from nematodes and phylogenetically unrelated pathogens such as *P. infestans.* However, the effectors from each pathogen are likely to have some conserved functions in terms of making a host susceptible to infection. As science advances and structural biology is used more widely to reveal the structure of effectors, it may become possible to identify functionally conserved effectors based on structural as opposed to sequence

similarity. The structure of the effector is likely to be as important as the sequence, and there may be more than one way to encode a particular structure (Vens *et al.*, 2011). Once this kind of approach is available, it would be interesting to see if pathogens that are capable of successfully parasitizing the same host have evolved any structurally similar effectors. It would be logical to predict that effectors from different pathogens have evolved to target the same hubs (see below) within the host (Mukhtar *et al.*, 2011a).

A combination of modelling and yeast 2 hybrid studies has been used to analyse biochemical signalling networks in plants. Biological signalling networks are composed of many nodes, each of which can be thought of as a single point. Nodes that are connected together in a linear string form what is termed a 'bridge'. If one of these nodes is knocked-out in this bridge signalling pathway, then the whole system collapses. This is not a stable scenario for a host (Pritchard and Birch, 2011a) and it is therefore not surprising that bridge-style networks are rarely represented in biological networks (Fell and Wagner, 2000). Biological systems have instead evolved into large networks that are connected together by hubs. Hubs are central points that many distinct signalling pathways converge onto (Pritchard and Birch, 2011a). Single gene deletions of yeast genes show that just 17% of genes are essential for viability (Winzeler *et al.*, 1999). This suggests that biological networks are robust and are error tolerant. However, hubs provide an opportunity for pathogens to disrupt large parts of a biological network in order to parasitize a host. Hubs may be difficult for a plant to alter or lose through selection and therefore they are good targets for pathogens that need to interfere with defence signalling pathways (Brodsky and Medzhitov, 2009). The targeting of specific host hubs is counter-defended by guard proteins that monitor these hubs (Jones and Dangl, 2006). Bacterial pathogens that cause acute symptoms are shown to target hubs. An example of a hub signalling pathway is mitogen-associated protein kinase (MAPK). MAPK pathways are targeted by numerous effectors from bacteria, for example YopJ/P from *Yersinia* species, AvrA and SpvC from *Salmonella* species, OspF from *Shigella* species and lethal factors from *B. anthracis* (Brodsky and Medzhitov, 2009b). An extracellular protein called RCR3 is secreted by tomato. This is targeted by AVR2 from *Cladosporum fulvum* and also two effectors from *P. infestans* (EPIC1 and

EPIC2B) (Song *et al.*, 2009). A hub protein called RIN4 in *Arabidopsis* is targeted by three *P. syringae* effectors (AvrRmp1, AvrRpt2 and AvrB). RIN4 is guarded by two *R*-gene protein products (Mackey *et al.*, 2002).

Oomycetes and bacteria separated over 2 billion years ago. An investigation into effectors from these pathogens shows that effectors converge on common sets of well-connected proteins called hubs, that are thought to be involved in defence responses and a hypothesis was made that 165 of these effector targets would also be targets for other pathogen effectors. This shows that independently evolved effectors converge on a common set of hubs in a plant defence system network (Mukhtar *et al.*, 2011b). Yeast-two hybrid screening of known hubs (Yu *et al.*, 2008) with effectors from *G. pallida* would provide interesting information about the effector targets, and similar experiments with the effectorome from other species could reveal effectors with similar functions that may not have sequence similarity.

RNAseq analysis of multiple life stages of a pathogen can reveal quantitative expression profiles of genes. For these studies, replication is essential and this may make RNAseq too expensive for use with all but the most important nematodes. RNAseq can provide information about the time during the life cycle that effectors (and other genes) are required by the nematode. This information can be used to focus in on certain subgroups of effectors that are important in certain processes. For example, effectors expressed at parasitic life stage of *G. pallida* would not be involved in migration, while studies that are focused on feeding would investigate effectors up-regulated at feeding stages. Relatively few large scale studies using RNAseq that compare expression profiles of genes expressed at different life stages within a pathogen have been published to date. Those that have been published are on helminths that are of high economic importance or of direct importance to humans. Examples include the human filarial parasite *B. malayi* (Choi *et al.*, 2011) and the human blood fluke *Schistosoma mansoni* (Protasio *et al.*, 2012). The generation of RNAseq data in this project has enabled the quantitative analysis of all gene models at several life stages. qPCR analysis has been used to confirm the RNAseq expression profiles. These studies showed that RNAseq is a robust tool for determining gene expression.

As the number of sequenced nematode genomes increases more data becomes available and comprehensive comparative genomics studies become feasible. Comparative RNAseq analysis on populations of virulent and avirulent *G. pallida* and *G. rostochiensis* could reveal variation in effectors that may be a reflection of strong selection pressure. Such effectors could be implicated as avirulence genes and therefore could be investigated further. Effectors similar to the implicated *AVR* genes could then be subjected to *R*-gene breeding programmes to develop resistant cultivars.

The best approach for control of plant pathogens, including nematodes, is the use of natural resistance traits. However, the identification of such genes and selectively breeding these from wild to commercial varieties can take several years and resistance is sometimes linked to other undesirable traits (Lilley *et al.*, 2011). Therefore, innovative transgenic approaches have been developed which are highly targeted to the organism of interest and can be faster to implement.

The transgenic introduction of an *R*-gene into a species of interest can by-pass the long, unpredictable process of conventional selective breeding. For example an *R*-gene against *H. schachtii* has been identified and is encoded by a single dominant gene from *Beta procumbens.* This gene has been transgenically introduced into sugar beet to confer resistance and avoid undesirable traits (Lilley *et al.*, 2011). *Mi-1.2* confers resistance against *Meloidogyne species* and has been successfully introduced into susceptible cultivars of tomato to confer resistance (Milligan *et al.*, 1998). However, it has proven extremely difficult to transfer nematode *R*-genes between species. For example, the *Hero* gene from tomato was introduced into potato but did not retain its function. It is thought that these difficulties may reflect changes in the host target that the *R*-gene is guarding between the donor and recipient species (Sobczak *et al.*, 2005; Ernst *et al.*, 2002). Transgenic plants expressing a double hairpin RNA construct have been used to induce RNAi in feeding nematodes. These plants showed reduced nematode infection rates (Huang *et al.*, 2006a). Data from the functional characterisation of effectors may be useful in allowing identification of an essential effector that could be used as a target for induced RNA silencing in the feeding site for the control of parasitic nematodes.

6.3 Functional characterisation of effectors

A number of effectors were transformed into *Arabidopsis* and potato to allow functional characterisation. The transgenic expression of effectors has been shown to be a useful tool in obtaining data regarding the mode of action of nematode effectors (Doyle and Lambert, 2003; Hamamouch *et al.*, 2012b; Hewezi *et al.*, 2010b; Huang *et al.*, 2006b; Patel *et al.*, 2010a). A number of growth phenotypes were observed in potato, including stunted growth and delayed flowering time. Some lines of transgenic plants showed altered susceptibility to pathogens such as *P. infestans* CS-12 and nematodes. Screening transgenic plants with pathogens to characterise altered susceptibility has been shown to be a useful method for identifying effectors that suppress host defences (Hewezi *et al.*, 2010b). Confirmation that transgenic plants were producing the expected protein proved to be difficult. A thorough preliminary investigation into protein-tags should be conducted to determine if these difficulties were due to the HA-tag. Transient expression of several effectors with a range of tags on their C-terminus should be subjected to western blot analysis to determine the best tag for identification. The best tag may be the longest tag, which could interfere with transgenic protein function, therefore this limitation should also be considered before choosing the tag for future experiments. If a suitable tag was identified then this could be used for pull-down assays which can complement yeast-two hybrid and split-YFP assay data.

The use of stable transformation has produced data implicating effectors in altering the phenotypes of plants which could not be investigated easily by other methods. Transient expression cannot be used to determine alterations in susceptibility to nematodes or identify growth phenotypes due to transgene expression. Therefore, the use of transgenics is an excellent tool for generating data to help decipher the mode of action of an effector. Transgenic plants produced in this project enabled some experiments to be employed that would otherwise have been difficult or impossible, such as screening a high number of transgenic lines with attenuated virulent pathogens as a high-throughput method of identifying interesting effectors. The main problem with the phenotypic analysis of transgenic plants is that data produced can show variability between assays (Queval *et al.*, 2008; Undurraga *et al.*, 2012; Wang *et al.*, 2011). This limitation therefore requires a large number of replicates to be

used to minimise the variability (Butaye *et al.*, 2005). However, this becomes labour intensive and resources such as space may not be available to satisfy the number of replicates required.

Full and detailed functional characterisation of effectors is the ideal outcome for researchers in the plant-pathology field (Gheysen and Jones, 2006). However, this is time consuming and can only be considered for effectors that have previously been shown to be important from other assays. A *G. pallida* effector which is similar to ubiquitin extension proteins (Jones *et al.*, 2009b) was subjected to detailed functional characterisation. Previously a similar effector from *Heterodera* species was reported (Tytgat *et al.*, 2004). Ubiquitin extension protein effectors are comprised of a signal peptide for secretion, a conserved ubiquitin domain and a short C-terminal extension. The absence of this type of effector in *Meloidogyne* species led previous researchers to speculate that the C-terminal extension of this effector may have a role in syncytium formation. However, it is not unusual to find species- or genus-specific effectors as these proteins are under strong host selection. The ubiquitin extension proteins from *Heterodera* species and *G. pallida* clearly have different roles. In *H. schachtii* the ubiquitin domain and C-terminal domain are cleaved and the C-terminal extension is targeted to the nucleus and nucleolus (Tytgat *et al.*, 2004) whereas in *G. pallida* no such localisation is seen after cleavage. *G. rostochiensis* has a nearly identical UBI-EP effector to *G. pallida*, except for two amino acid changes in the C-terminal extension, suggesting this effector may have conserved function within *Globodera* species.

Analysis of the *G. pallida* genome sequence identified another ubiquitin extension protein which has a signal peptide and a different C-terminal extension. The expression profile is different from the *GpUBI-EP* characterised in this thesis. The *GpUBI-EP* studied here is highly up-regulated at J2 while the other ubiquitin extension protein is up-regulated at 7 and 14 dpi. This suggests these effectors may have different functional roles. It would be interesting to characterise the UBI-EP that is up-regulated at 7 and 14 dpi to see what this effector does and compare this to the other UBI-EP characterised in this thesis. This would also shed light on the

importance and role of the ubiquitin domain which has previously been suggested to be just a chaperone (Tytgat *et al.*, 2004).

In eukaryotes all ubiquitin extension proteins are formed with a C-terminal extension to render the ubiquitin inactive until required (Hershko, 2005). Once the cellular machinery, known as de-ubiquitinating enzymes (DUBs), cleaves the C-terminal extension from the terminal Glycine-Glycine residues of the ubiquitin domain, the ubiquitin domain becomes active (Vierstra, 2009). Therefore, the nematode exploits the host machinery to cleave the C-terminal extension off the ubiquitin domain to release the C-terminal extension. This is a way of delivering a short peptide within the host by attaching it to a protein that already naturally occurs within the host cell. If further scientific evidence supports this theory, then the ubiquitin domain is merely acting as a chaperone, taking the peptide into the host cell. This could explain the variation in C-terminal extensions observed between *Heterodera* and *Globodera.* This method of delivering a functional peptide to a plant cell could be exploited. If there is a biotechnological requirement to deliver a peptide into a host cell then this could be expressed as a C-terminal extension to ubiquitin. Conserved lysine positions are required for ubiquitin function (Duncan *et al.*, 2006). The ubiquitin could have its lysine residues replaced with a non-functioning amino acid to render the domain inactive in the cell, if this was required.

GpUBI-EP expression is highly up-regulated at J2 with little or no expression in other life stages. The nematode invades the root as a J2. As it migrates through the root it induces damage-associated defence responses. Once the nematode is at a suitable site its behaviour changes and a cell is selected to become a feeding site (Smant and Jones, 2011). The nematode gently inserts its stylet into the host cell and injects effectors into the host cell in order to manipulate it (Sobczak and Golinowski, 2011). The nematode has to avoid activating defence responses, such as PTI (Smant and Jones, 2011). If PTI is activated the cell will respond by depositing callose around the nematode stylet (Hussey *et al.*, 1992) preventing infection. Our finding suggests that GpUBI-EP interferes with PTI responses and reduces callose deposition suggesting it may be important during this initial phase of the nematode life cycle.

Several attempts were made to clone the C-terminal extension of the UBI effector, but proved unsuccessful. RNAi has previously been used to study nematode effectors (Gleason *et al.*, 2008; Huang *et al.*, 2006a; Patel *et al.*, 2010a). RNAi could be used to determine how important the GpUBI-EP effector is in establishing successful parasitism. This could only be achieved by targeting the C-terminal extension for down-regulation, as targeting the ubiquitin domain would be lethal for the worm and result in off target genes down-regulated by RNAi. A synthetic clone of the C-terminal extension could also be included in the same infiltration and callose deposition assays already described, to determine whether the extension or the full protein are required for suppression of PTI. Yeast-two hybrid experiments have been successfully employed to study the host protein targets of effectors (Hamamouch *et al.*, 2012a; Hewezi *et al.*, 2010b; Lee *et al.*, 2011; Patel *et al.*, 2010a; Rehman *et al.*, 2009b). It would not be possible to include the GpUBI-EP in a yeast-two hybrid screen due to the nature of the ubiquitin domain, which would bind with many proteins, leading to false positives. However, if the C-terminal extension had been artificially cloned, it may be possible to subject this to yeast-two hybrid assays to determine which host protein(s) it may interact with.

Although only a few examples of effectors involved in suppression of PTI or ETI have been identified in nematodes (Hewezi *et al.*, 2008; Hewezi *et al.*, 2010b; Postma *et al.*, 2012; Rehman *et al.*, 2009b), the functional analysis of GpUBI-EP suggests that this may interfere with both the PTI and ETI pathways. As mentioned above, many effectors that interfere with or exploit the ubiquitin proteasome pathway have been identified, highlighting its importance to both the host and to pathogens as a means of controlling cellular processes (Angot *et al.*, 2006b; Bellafiore *et al.*, 2008; Birch *et al.*, 2009a; Elling *et al.*, 2007; Ewan *et al.*, 2011; Gao *et al.*, 2003a; Goehre *et al.*, 2008b; Hewezi *et al.*, 2010b; Schrammeijer *et al.*, 2001b; Tytgat *et al.*, 2004). Moreover, bacteria genetics do not employ the same ubiquitin-proteasome pathway as that of eukaryotes. Eukaryotic ubiquitin possess a terminal Gly-Gly at the end of the ubiquitin molecule, whereas an identified prokaryotic ubiquitin-like molecule called Pup does not have the terminal Gly-Gly. Pup was shown to target proteins for proteolysis by the *Mycobacterium tuberculosis* proteasome (Darwin, 2009). Bacterial pathogens have evolved to use effectors that can manipulate or mimic components of

eukaryotic ubiquitin-proteasome pathways. For example, the *P. syringae* effector AvrPto (Nguyen *et al.*, 2010; Rosebrock *et al.*, 2007) is an E3-ligase and targets the FLS2/BAK-1 transmembrane receptor kinase for degradation by the ubiquitin-proteasome pathway (Nguyen *et al.*, 2010; Rosebrock *et al.*, 2007). FLS2/BAK-1 are transmembrane receptors that recognise the bacterial PAMP FLG-22 (Zipfel, 2009). Further to this, a human virus has also been shown to manipulate the ubiquitin-proteasome pathway (Howden and Huitema, 2012). This shows that prokaryotic pathogens have evolved to manipulate eukaryotic pathways and, as mentioned above eukaryotic pathogens (for example nematodes and *P. infestans*) also target the ubiquitin-proteasome pathways within their hosts.

For most of the nematode effectors identified to date, no function has been ascribed (Elling *et al.*, 2009a; Gao *et al.*, 2003a; Jones *et al.*, 2009b; Rosso and Grenier, 2011). Nematodes must produce effectors that have a role in suppressing the host defence system but little is known about this process in these organisms (Smant and Jones, 2011). The details of how feeding sites are induced are still not known. The generation of genomic sequence data has allowed the mining of genomes for candidate effectors and comparative genomics with other pathogens (Abad *et al.*, 2008; Elling *et al.*, 2009a; Gao *et al.*, 2003a; Haegeman *et al.*, 2011b; Jones *et al.*, 2009b; Kikuchi *et al.*, 2009; Kikuchi *et al.*, 2007; Opperman *et al.*, 2008), to seek similarity between effectors of different species. Effectors produced from different pathogens have a wide range of effects on the host. In all cases a successful pathogen will over-ride host signalling pathways for the benefit of the pathogen, and some evidence suggests they do this via the targeting of hubs (Mukhtar *et al.*, 2011b).

In addition to generating novel scientific information this work has profound practical implications. Understanding which effectors are important for the pathogen and how they function will allow new control strategies based on biotechnological approaches to be developed. In addition, effectors that are recognised by resistance genes can be used in resistance breeding programmes. Identifying essential effectors that are recognised by novel sources of resistance offers the prospect of durable resistance. The importance of such work will only increase due to the ever increasing list of banned pesticides and increasing demand for food across the world.

7 References

Abad, P., Gouzy, J., Aury, J.M., Castagnone-Sereno, P., Danchin, E., Deleury, E. *et al.* (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology* **26**: 909-915.

Abad, P. and McCarter, J.P. (2011) Genome analysis of plant parasitic nematodes genomics and molecular genetics of plant-nematode interactions. Jones, J., Gheysen, G. and Fenoll, C. (eds). Springer Netherlands, pp. 103-117.

Yeh, J. and Zimmerman, J. (2006) Introduction of the carrot HSP17. 7 into potato (*Solanum tuberosum* L.) enhances cellular membrane stability and tuberization in vitro. *Plant, Cell &Eenvironment* **29**: 95-104.

Alexander, L. and Grierson, D. (2002) Ethylene biosynthesis and action in tomato: A model for climacteric fruit ripening. *Journal of Experimental Botany* **53**: 2039-2055.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science Signalling* **301**: 653-657.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Altschul, S.F., Madden, T.L., Schañffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biology* **11**:10.R106

Angot, A., Peeters, N., Lechner, E., Vailleau, F., Baud, C., Gentzbittel, L. *et al.* (2006a) *Ralstonia solanacearum* requires F-box-like domain-containing type III effectors to promote disease on several host plants. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 14620-14625.

Angot, A., Peeters, N., Lechner, E., Vailleau, F., Baud, C., Gentzbittel, L. *et al.* (2006b) Ralstonia solanacearum requires F-box-like domain-containing type III effectors to promote disease on several host plants. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 14620-14625.

Arabidopsis, G.I. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796-815.

Artsaenko, O., Kettig, B., Fiedler, U., Conrad, U. and Doring, K. (1998) Potato tubers as a biofactory for recombinant antibodies. *Molecular Breeding* **4**: 313-319.

Arvidsson, S., Perez-Rodroguez, P. and Mueller-Roeber, B. (2011) A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytologist* **191**: 895-907.

Aslam, S.N., Newman, M.A., Erbs, G., Morrissey, K.L., Chinchilla, D., Boller, T. *et al.* (2008) Bacterial polysaccharides suppress induced innate immunity by calcium chelation. *Current Biology* **18**: 1078-1083.

Bailey, T.L., Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *Vectors*. **1:** 2

Baker, M. (2010) Next-generation sequencing: adjusting to data overload. *Nature Methods* **7**: 495-499.

Bakhetia, M., Urwin, P.E. and Atkinson, H.J. (2008) Characterisation by RNAi of pioneer genes expressed in the dorsal pharyngeal gland cell of *Heterodera glycines* and the effects of combinatorial RNAi. *International Journal for Parasitology* **38**: 1589-1597.

Baldwin, J.G., Nadler, S.A. and Adams, B.J. (2004) Evolution of plant parasitism among nematodes. *Annual Review of Phytopathology* **42**: 83-105.

Banerjee, A.K., Prat, S. and Hannapel, D.J. (2006) Efficient production of transgenic potato *S. tuberosum* L. ssp. *andigena* plants via *Agrobacterium tumefaciens*-mediated transformation. *Plant Science* **170**: 732-738.

Bao, S., Jiang, R., Kwan, W.K., Wang, B.B., Ma, X. and Song, Y.Q. (2011) Evaluation of next-generation sequencing software in mapping and assembly. *Journal of Human Genetics* **56**: 406-414.

Bar-Or, C., Kapulnik, Y. and Koltai, H. (2005) A broad characterization of the transcriptional profile of the compatible tomato response to the plant parasitic root knot nematode *Meloidogyne javanica*. *European Journal of Plant Pathology* **111**: 181-192.

Baumberger, N., Tsai, C.H., Lie, M., Havecker, E. and Baulcombe, D.C. (2007) The Polerovirus silencing suppressor P0 targets ARGONAUTE proteins for degradation. *Current Biology* **17**: 1609-1614.

Bellafiore, S., Shen, Z.X., Rosso, M.N., Abad, P., Shih, P. and Briggs, S.P. (2008) Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLoS Pathogens* **4:** e1000192**.**

Bendahmane, A., Farnham, G., Moffett, P. and Baulcombe, D.C. (2002) Constitutive gain of function mutants in a nucleotide binding site leucine rich repeat protein encoded at the Rx locus of potato. *The Plant Journal* **32**: 195-204.

Bennett, S. (2004) Solexa Ltd. *Pharmacogenomics* **5**: 433-438.

Berger, B., de Regt, B. and Tester, M. (2012) Applications of high-throughput plant phenotyping to study nutrient use efficiency. *Methods in Molecular Biology* **953**: 277.

Bethony, J., Brooker, S., Albonico, M., Geiger, S.M., Loukas, A., Diemert, D. and Hotez, P.J. (2006) Soil-transmitted helminth infections: *ascariasis*, *trichuriasis*, and hookworm. *The Lancet* **367**: 1521-1532.

Bhattarai, K.K., Xie, Q.G., Mantelin, S., Bishnoi, U., Girke, T., Navarre, D.A. and Kaloshian, I. (2008) Tomato susceptibility to root-knot nematodes requires an intact jasmonic acid signaling pathway. *Molecular Plant-Microbe Interactions* **21**: 1205-1214.

Birch, P., Armstrong, M., Bos, J., Boevink, P., Gilroy, E., Taylor, R. *et al.* (2009a) Towards understanding the virulence functions of RXLR effectors of the oomycete plant pathogen *Phytophthora infestans*. *Journal of Experimental Botany* **60**: 1133-1140.

Birch, P.R.J., Rehmany, A.P., Pritchard, L., Kamoun, S. and Beynon, J.L. (2006) Trafficking arms: oomycete effectors enter host plant cells. *Trends in Microbiology* **14**: 8-11.

Birch, P.R.J., Armstrong, M., Bos, J., Boevink, P., Gilroy, E.M., Taylor, R.M. *et al.* (2009b) Towards understanding the virulence functions of RXLR effectors of the oomycete plant pathogen *Phytophthora infestans*. *Journal of Experimental Botany* **60**: 1133-1140.

Bird, D.M.K. and Koltai, H. (2000) Plant parasitic nematodes: habitats, hormones, and horizontally-acquired genes. *Journal of Plant Growth Regulation* **19**: 183-194.

Blakeslee, J.J., Peer, W.A. and Murphy, A.S. (2005) Auxin transport. *Current Opinion in Plant Biology* **8**: 494-500.

Blanchard, A., Fouville, D., Esquibet, M., Mugniery, D. and Grenier, E. (2007) Sequence polymorphism of 2 pioneer genes expressed in phytoparasitic nematodes showing different host ranges. *Journal of Heredity* **98**: 611-619.

Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A. *et al.* (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* **392**: 71-75.

Bleve-Zacheo, T. and Zacheo, G. (1987) Cytological studies of the susceptible reaction of sugarbeet roots to *Heterodera schachtii*. *Physiological and Molecular Plant Pathology* **30**: 13-25.

Boller, T. and Felix, G. (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology* **60**: 379-406.

Bongers, T. and Ferris, H. (1999) Nematode community structure as a bioindicator in environmental monitoring. *Trends in Ecology & Evolution* **14**: 224-228.

Bos, J.I.B., Kanneganti, T.D., Young, C., Cakir, C., Huitema, E., Win, J. *et al.* (2006) The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. *Plant Journal* **48**: 165-176.

Bos, J.I., Armstrong, M.R., Gilroy, E.M., Boevink, P.C., Hein, I., Taylor, R.M. *et al.* (2010) *Phytophthora infestans* effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 9909-9914.

Bos, J.I.B., Chaparro-Garcia, A., Quesada-Ocampo, L.M., Gardener, B.B.M. and Kamoun, S. (2009) Distinct Amino Acids of the *Phytophthora infestans* Effector AVR3a Condition Activation of R3a Hypersensitivity and Suppression of Cell Death. *Molecular Plant-Microbe Interactions* **22**: 269-281.

Bostock, R.M. (2005) Signal crosstalk and induced resistance: Straddling the line between cost and benefit. *Annual Review of Phytopathology* **43**: 545-580.

Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R. and Garlach, J. (2001) Growth stage based phenotypic analysis of *Arabidopsis* a model for high throughput functional genomics in plants. *Plant Cell* **13**: 1499-1510.

Brodsky, I.E. and Medzhitov, R. (2009b) Targeting of immune signalling networks by bacterial pathogens. *Nature Cell Biology* **11**: 521-526.

Butaye, K.M.J., Cammue, B.P.A., Delaure, S.L. and De Bolle, M.F.C. (2005) Approaches to minimize variation of transgene expression in plants. *Molecular Breeding* **16**: 79-91.

C.elegans sequencing consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science (Washington)* **282**: 2012-2018.

Caillaud, M.C., Dubreuil, G., Quentin, M., Perfus-Barbeoch, L., Lecomte, P., de Almeida Engler, J. *et al.* (2008) Root-knot nematodes manipulate plant cell functions during a compatible interaction. *Journal of Plant Physiology* **165**: 104-113.

Carpentier, J., Esquibet, M., Fouvilled, D., Manzanares-Dauleux, M., Kerlan, M. and Grenier, E. (2012) The evolution of the Gp-Rbp-1 gene in *Globodera pallida* includes multiple selective replacements. *Molecular Plant Pathology* **13**: 546-555.

Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C., Menard, M. *et al.* (2004) Whole genome sequence comparisons and full-length cDNA sequences: a combined approach to evaluate and improve *Arabidopsis* genome annotation. *Genome Research* **14**: 406-413.

Catic, A., Collins, C., Church, G.M. and Ploegh, H.L. (2004) Preferred in vivo ubiquitination sites. *Bioinformatics* **20**: 3302-3307.

Chalfie, M. (1995) Green fluorescent protein. *Photochemistry and Photobiology* **62**: 651-656.

Chitwood, D.J. (2003) Research on plant-parasitic nematode biology conducted by the United States Department of Agriculture - Agricultural Research Service. *Pest Management Science* **59**: 748-753.

Chiu, M., Chen, I., Baulcombe, D. and Tsai, C. (2010) The silencing suppressor P25 of Potato virus X interacts with Argonaute1 and mediates its degradation through the proteasome pathway. *Molecular Plant Pathology* **11**: 641-649.

Choi, Y.J., Ghedin, E., Berriman, M., McQuillan, J., Holroyd, N., Mayhew, G.F. *et al.* (2011) A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, *Brugia malayi*. *PLoS Neglected Tropical Diseases* **5**: e1409.

Cock, P., Antao, T., Chang, J., Chapman, B., Cox, C., Dalke, A. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422-1423.

Craig, J., Bekal, S., Niblack, T., Domier, L. and Lambert, K. (2009) Evidence for horizontally transferred genes involved in the biosynthesis of vitamin B1, B5, and B7 in *Heterodera glycines*. *Journal of Nematology* **41**: 281-290.

Craig, J.P., Bekal, S., Hudson, M., Domier, L., Niblack, T. and Lambert, K.N. (2008) Analysis of a Horizontally Transferred Pathway Involved in Vitamin B6 Biosynthesis from the Soybean Cyst Nematode *Heterodera glycines*. *Molecular Biology and Evolution* **25**: 2085-2098.

Danchin, E., Rosso, M., Vieira, P., de Almeida-Engler, J., Coutinho, P., Henrissat, B. and Abad, P. (2010) Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences* **107**: 17651-17656.

Dangl, J.L. and McDowell, J.M. (2006) Two modes of pathogen recognition by plants. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 8575-8576.

Darwin, K.H. (2009) Prokaryotic ubiquitin-like protein (Pup), proteasomes and pathogenesis. *Nature Review Microbiology* **7**: 485-491.

de Almeida Engler, J., Van Poucke, K., Karimi, M., De Groodt, R., Gheysen, G., Engler, G. and Gheysen, G. (2004) Dynamic cytoskeleton rearrangements in giant cells and syncytia of nematode-infected roots. *The Plant Journal* **38**: 12-26.

De Boer, J.M., Mcdermott, J.P., Wang, X., Maier, T., Qui, F., Hussey, R.S. *et al.* (2002) The use of DNA microarrays for the developmental expression analysis of cDNAs from the oesophageal gland cell region of *Heterodera glycines*. *Molecular Plant Pathology* **3**: 261-270.

De Boer, J., Yan, Y., Wang, X., Smant, G., Hussey, R., Davis, E. and Baum, T. (1999) Developmental expression of secretory beta-1, 4-endoglucanases in the subventral esophageal glands of *Heterodera glycines*. *Molecular Plant-Microbe Interactions* **12**: 663-669.

de Jonge, R. and Thomma, B.P.H.J. (2009) Fungal LysM effectors: extinguishers of host immunity? *Trends in Microbiology* **17**: 151-157.

De Meutter, J., Tytgat, T., Witters, E., Gheysen, G., Van Onckelen, H. and Gheysen, G. (2003) Identification of cytokinins produced by the plant parasitic nematodes *Heterodera schachtii* and *Meloidogyne incognita*. *Molecular Plant Pathology* **4**: 271-277.

den Nijs, L. (2007) The Canon of Potato Science: 14. Cyst Nematodes. *Potato Research* **50**: 259-262.

Deshaies, R.J. and Joazeiro, C.A.P. (2009) RING Domain E3 Ubiquitin Ligases. *Annual Review of Biochemistry* **78**: 399-434.

Devine, K.J., Dunne, C., O'Gara, F. and Jones, P.W. (1999) The influence of in-egg mortality and spontaneous hatching on the decline of *Globodera rostochiensis* during crop rotation in the absence of the host potato crop in the field. Nematology **1**(**6**), 637-645.

Ding, X., Shields, J., Allen, R. and Hussey, R.S. (1998) A secretory cellulose-binding protein cDNA cloned from the root-knot nematode (*Meloidogyne incognita*). *Molecular Plant-Microbe Interactions* **11**: 952-959.

Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J. *et al.* (2011) Metabolic priming by a secreted fungal effector. *Nature* **478**: 395-398.

Dodds, P.N. and Rathjen, J.P. (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics* **11**: 539-548.

Doyle, E.A. and 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.

Doyle, E.A. and Lambert, K.N. (2003) *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. *Molecular Plant-Microbe Interactions* **16**: 123-131.

Duncan, L.M., Piper, S., Dodd, R.B., Saville, M.K., Sanderson, C.M., Luzio, J.P. and Lehner, P.J. (2006) Lysine-63-linked ubiquitination is required for endolysosomal degradation of class I molecules. *The EMBO Journal* **25**: 1635-1645.

Dyrl Bendtsen, J., Nielsen, H., von Heijne, G. and Brunak, S.+. (2004) Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology* **340**: 783-795.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Elling, A.A., Davis, E.L., Hussey, R.S. and Baum, T.J. (2007) Active uptake of cyst nematode parasitism proteins into the plant cell nucleus. *International Journal for parasitology* **37**: 1269-1279.

Elling, A.A., Mitreva, M., Gai, X.W., Martin, J., Recknor, J., Davis, E.L. *et al.* (2009a) Sequence mining and transcript profiling to explore cyst nematode parasitism. *BMC Genomics* **10**: 58

Ernst, K., Kumar, A., Kriseleit, D., Kloos, D.U., Phillips, M.S. and Ganal, M.W. (2002) The broad spectrum potato cyst nematode resistance gene (Hero) from tomato is the only member of a large gene family of NBS–LRR genes with an unusual amino acid repeat in the LRR region. *The Plant Journal* **31**: 127-136.

Ewan, R., Pangestuti, R., Thornber, S., Craig, A., Carr, C., O'Donnell, L. *et al.* (2011) Deubiquitinating enzymes AtUBP12 and AtUBP13 and their tobacco homologue NtUBP12 are negative regulators of plant immunity. *New Phytologist* **191**: 92-106.

Faik, A. (2010) Xylan Biosynthesis: News from the grass. *Plant Physiology* **153**: 396-402.

Felix, G., Duran, J.D., Volko, S. and Boller, T. (1999) Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant Journal* **18**: 265-276.

Fell, D.A. and Wagner, A. (2000) 12 Structural properties of metabolic networks: implications for evolution and modelling of metabolism. *Structural Properties of Metabolic Network.* **1**: 79-85

Fiers, M., Golemiec, E., Xu, J., van der Geest, L., Heidstra, R., Stiekema, W. and Liu, C.M. (2005) The 14 Amino Acid CLV3, CLE19, and CLE40 peptides trigger consumption of the root meristem in *Arabidopsis* through a CLAVATA2-dependent pathway. *Plant Cell* **17**: 2542-2553.

Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**: 29-37.

Fletcher, J.C. Shoot and floral meristem maintenance in *Arabidopsis* (2002). *Annual Review of Plant Biology* 53.1: 45-66

Franco, J. and Evans, K. (1978) Multiplication of some south American and European populations of potato cyst nematodes on potatoes possessing the resistance genes H1, H2 and H3. *Plant Pathology* **27**: 1-6.

Frei dit Frey, N. and Robatzek, S. (2009) Trafficking vesicles: pro or contra pathogens? *Current Opinion in Plant Biology* **12**: 437-443.

Friml, J. (2010) Subcellular trafficking of PIN auxin efflux carriers in auxin transport. *European Journal of Cell Biology* **89**: 231-235.

Gao, B.L., Allen, R., Maier, T., Davis, E.L., Baum, T.J. and Hussey, R.S. (2003) The parasitome of the phytonematode *Heterodera glycines*. *Molecular Plant-Microbe Interactions* **16**: 720-726.

Gelvin, S.B. (2009) *Agrobacterium* in the genomics age. *Plant Physiology* **150**: 1665-1676.

Genaev, M.A., Doroshkov, A.V., Pshenichnikova, T.A., Kolchanov, N.A. and Afonnikov, D.A. (2012) Extraction of quantitative characteristics describing wheat leaf pubescence with a novel image-processing technique. *Planta*: **5**:1943-1054.

Gheysen, G. and Fenoll, C. (2002) Gene expression in nematode feeding sites. *Annual Review of Phytopathology* **40**: 191-219.

Gheysen, G. and Jones, J.T. (2006) Molecular aspects of plant-nematode interactions. *Plant Nematology Wallingford, UK, CABI Publishing*: 234-254.

Gheysen, G. and Vanholme, B. (2007) RNAi from plants to nematodes. *Trends in Biotechnology* **25**: 89-92.

Gheysen, G., Jones, J.T., Perry, R.N. and Moens, M. (2006) Molecular aspects of plant-nematode interactions. *Plant Nematology*: 234-254.

Gilroy, E.M., Taylor, R.M., Hein, I., Boevink, P., Sadanandom, A. and Birch, P.R.J. (2011) CMPG1-dependent cell death follows perception of diverse pathogen elicitors at the host plasma membrane and is suppressed by *Phytophthora infestans* RXLR effector AVR3a. *New Phytologist* **190**: 653-666.

Gleason, C.A., Liu, Q.L. and Williamson, V.M. (2008) Silencing a candidate nematode effector gene corresponding to the tomato resistance gene *Mi-1* leads to acquisition of virulence. *Molecular Plant Microbe Interactions* **21**: 576-585.

Goecks, J., Nekrutenko, A., Taylor, J. and Galaxy Team. Galaxy: (2010) A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* 10:86

Goehre, V., Spallek, T., Haeweker, H., Mersmann, S., Mentzel, T., Boller, T. *et al.* (2008a) Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB. *Current Biology* **18**: 1824-1832.

Goehre, V. and Robatzek, S. (2008) Breaking the barriers: Microbial effector molecules subvert plant immunity. *Annual Review of Phytopathology* **46**: 189-215.

Goellner, M., Wang, X. and Davis, E.L. (2001) Endo beta-1, 4-glucanase expression in compatible plant–nematode interactions. *Plant Cell* **13**: 2241-2255.

Golzarian, M.R., Frick, R.A., Rajendran, K., Berger, B., Roy, S., Tester, M. and Lun, D.S. (2011) Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods* **7**: 2.

Goverse, A., Overmars, H., Engelbertink, J., Schots, A., Bakker, J. and Helder, J. (2000a) Both induction and morphogenesis of cyst nematode feeding cells are mediated by auxin. *Molecular Plant-Microbe Interactions* **13**: 1121-1129.

Goverse, A., van der Voort, J.R., van der Voort, C.R., Kavelaars, A., Smant, G., Schots, A. *et al.* (1999) Naturally induced secretions of the potato cyst nematode co-stimulate the proliferation of both tobacco leaf protoplasts and human peripheral blood mononuclear cells. *Molecular Plant-Microbe Interactions* **12**: 872-881.

Goverse, A., de Almeida Engler, J., Verhees, J., van der Krol, S., Helder, J. and Gheysen, G. (2000b) Cell cycle activation by plant parasitic nematodes. *Plant Molecular Biology* **43**: 747-761.

Green, J., Wang, D., Lilley, C.J., Urwin, P.E. and Atkinson, H.J. (2012) Transgenic potatoes for potato cyst nematode control can replace pesticide use without impact on soil quality. *Plos One* **7**.e30793

Grundler, F.M.W., Sobczak, M. and Lange, S. (1997) Defence responses of *Arabidopsis thaliana* during invasion and feeding site induction by the plant-parasitic nematode *Heterodera glycines*. *Physiological and Molecular Plant Pathology* **50**: 419-429.

Grundler, F.M.W. and Hofmann, J. (2011) Water and nutrient transport in nematode feeding sites. Genomics and Molecular Genetics of Plant-Nematode Interactions. Jones, J., Gheysen, G. and Fenoll, C. (eds). *Springer Netherlands*, pp. 423-439.

Grunewald, W., Cannoot, B., Friml, J. and Gheysen, G. (2009) Parasitic nematodes modulate PIN-mediated auxin transport to facilitate infection. *PLoS Pathogens* **5**(1):e1000266

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* **59**: 307-321.

Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H.Y., Handsaker, R.E., Cano, L.M. *et al.* (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**: 393-398.

Haegeman, A., Jones, J.T. and Danchin, E.G.J. (2011a) Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Molecular Plant-Microbe Interactions* **24**: 879-887.

Haegeman, A., Joseph, S. and Gheysen, G. (2011b) Analysis of the transcriptome of the root lesion nematode Pratylenchus coffeae generated by 454 sequencing technology. *Molecular and Biochemical Parasitology* **178**: 7-14.

Haegeman, A., Mantelin, S., Jones, J.T. and Gheysen, G. (2012) Functional roles of effectors of plant-parasitic nematodes. *Gene* **492**: 19-31.

Hamamouch, N., Li, C., Seo, P., Park, C. and Davis, E.L. (2011) Expression of *Arabidopsis* pathogenesis-related genes during nematode infection. *Molecular Plant Pathology* **12**: 355-364.

Hamamouch, N., Li, C., Hewezi, T., Baum, T.J., Mitchum, M.G., Hussey, R.S. *et al.* (2012) The interaction of the novel 30C02 cyst nematode effector protein with a plant beta-1, 3-endoglucanase may suppress host defence to promote parasitism. *Journal of Experimental Botany* **63**: 3683-3695.

Hartmann-Petersen, R., Seeger, M. and Gordon, C. (2003) Transferring substrates to the 26S proteasome. *Trends in Biochemical Sciences* **28**: 26-31.

Hauck, P., Thilmony, R. and He, S.Y. (2003a) A Pseudomonas syringae type III effector suppresses cell wall-based extracellular defense in susceptible *Arabidopsis* plants. *Proceedings of the National Academy of Sciences* **100**: 8577-8582.

Hauck, P., Thilmony, R. and He, S.Y. (2003b) A Pseudomonas syringae type III effector suppresses cell wall-based extracellular defense in susceptible *Arabidopsis* plants. *Proceedings of the National Academy of Sciences* **100**: 8577-8582.

Hermsmeier, D., Mazarei, M. and Baum, T.J. (1998) Differential display analysis of the early compatible interaction between soybean and the soybean cyst nematode. *Molecular Plant-Microbe Interactions* **11**: 1258-1263.

Hershko, A. (2005) The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. *Cell Death and Differentiation* **12**: 1191-1197.

Hewezi, T., Howe, P.J., Maier, T.R., Hussey, R.S., Mitchum, M.G., Davis, E.L. and Baum, T.J. (2010a) *Arabidopsis* Spermidine Synthase is targeted by an effector protein of the cyst nematode *Heterodera schachtii*. *Plant Physiology* **152**: 968-984.

Hewezi, T., Howe, P., Maier, T.R., Hussey, R.S., Mitchum, M.G., Davis, E.L. and Baum, T.J. (2008) Cellulose binding protein from the parasitic nematode *Heterodera schachtii* interacts with *Arabidopsis* pectin methylesterase: cooperative cell wall modification during parasitism. *Plant Cell* **20**: 3080-3093.

Hofmann, J. and Grundler, F.M.W. (2007) How do nematodes get their sweets? Solute supply to sedentary plant-parasitic nematodes. *Nematology* **9**: 451-458.

Hogenhout, S.A. and Bos, J.I.B. (2011) Effector proteins that modulate plant insect interactions. *Current Opinion in Plant Biology* **14**: 422-428.

Howden, A.J.M. and Huitema, E. (2012) Effector-triggered post-translational modifications and their role in suppression of plant immunity. *Frontiers in Plant Science* **3**: 160

Huang, G., Allen, R., Davis, E.L., Baum, T.J. and Hussey, R.S. (2006a) Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 14302-14306.

Huang, G., Dong, R., Allen, R., Davis, E.L., Baum, T.J. and Hussey, R.S. (2006b) A root-knot nematode secretory peptide functions as a ligand for a plant transcription factor. *Molecular Plant Microbe Interactions* **19**: 463-470.

Huang, G.Z., Gao, B.L., Maier, T., Allen, R., Davis, E.L., Baum, T.J. and Hussey, R.S. (2003) A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode Meloidogyne incognita. *Molecular Plant-Microbe Interactions* **16**: 376-381.

Hunter, T. (2007) The age of crosstalk: Phosphorylation, ubiquitination, and beyond. *Molecular Cell* **28**: 730-738.

Hurkman, W.J. and Tanaka, C.K. (1986) Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. *Plant Physiology* **81**: 802-806.

Hussey, R.S. (1989) Disease-inducing secretions of plant-parasitic nematodes. *Annual Review of Phytopathology* **27**: 123-141.

Hussey, R.S. and Janssen, G.J.W. (2002) Root-knot nematodes: *Meloidogyne* species. *Plant Resistance to Parasitic Nematodes CAB International, London, England*: 43-70.

Hussey, R.S., Mims, C.W. and Westcott, S.W. (1992) Immunocytochemical localization of callose in root cortical cells parasitized by the ring nematode; *Criconemella xenoplax*. *Protoplasma* **171**: 1-6.

Ithal, N., Recknor, J., Nettleton, D., Hearne, L., Maier, T., Baum, T.J. and Mitchum, M.G. (2007a) Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. *Molecular Plant-Microbe Interactions* **20**: 293-305.

Ithal, N., Recknor, J., Nettleton, D., Maier, T., Baum, T.J. and Mitchum, M.G. (2007b) Developmental transcript profiling of cyst nematode feeding cells in soybean roots. *Molecular Plant-Microbe Interactions* **20**: 510-525.

Jablonska, B., Ammiraju, J.S.S., Bhattarai, K.K., Mantelin, S., de Ilarduya, O.M., Roberts, P.A. and Kaloshian, I. (2007) The *Mi-9* gene from *Solanum arcanum* conferring heat-stable resistance to root-knot nematodes is a homolog of *Mi-1*. *Plant Physiology* **143**: 1044-1054.

Jacquet, M., Bongiovanni, M., Martinez, M., Verschave, P., Wajnberg, E. and Castagnone-Sereno, P. (2005) Variation in resistance to the root-knot nematode *Meloidogyne incognita* in tomato genotypes bearing the *Mi* gene. *Plant Pathology* **54**: 93-99.

Jammes, F., Lecomte, P., meida-Engler, J., Bitton, F., Martin-Magniette, M.L., Renou, J.P. *et al.* (2005) Genome-wide expression profiling of the host response to root-knot nematode infection in *Arabidopsis*. *Plant Journal* **44**: 447-458.

Jang, I.C., Niu, Q.W., Deng, S., Zhao, P. and Chua, N.H. (2012) Enhancing protein stability with retained biological function in transgenic plants. *The Plant Journal*. **72**(2): 345-354.

Janssen, R., Bakker, J. and Gommers, F.J. (1991) Mendelian proof for a gene-for-gene relationship between virulence of *Globodera rostochiensis* and the H-1 resistance gene in *Solanum tuberosum* Ssp *Andigena* Cpc 1673. *Revue de Nematologie* **14**: 207-212.

Jaouannet, M., Magliano, M., Arguel, M.J., Gourgues, M., Evangelisti, E., Abad, P. and Rosso, M.N.l. (2012) The root-knot nematode calreticulin Mi-CRT is a key effector in plant defense suppression. *Molecular Plant-Microbe Interactions*. **26**: 97-105.

Jaubert, S., Laffaire, J.B., Abad, P. and Rosso, M.N. (2002) A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. *Febs Letters* **522**: 109-112.

Jentsch, S. and Pyrowolakis, G. (2000) Ubiquitin and its kin: how close are the family ties? *Trends in Cell Biology* **10**: 335-342.

Jones, J., Amar, K., Pylypenko, L., Amarnath, T., Castelli, L., Chapman, S. *et al.* (2009) Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology* **10**: 815-828.

Jones, J.D.G. and Dangl, J.L. (2006) The plant immune system. *Nature.* **444**: 323-329.

Jones, J.T., Furlanetto, C., Bakker, E., Banks, B., Blok, V., Chen, Q. *et al.* (2003) Characterization of a chorismate mutase from the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology* **4**: 43-50.

Jones, J.T., Blok, V.C., Prior, A., Phillips, M.S., Popeijus, H., Bakker, E. *et al.* (2000) ESTs from potato cyst nematodes: From sequences to functional analysis. *Nematology* **2**: 765-774.

Jones, J.T., Furlanetto, C. and Phillips, M.S. (2007) The role of flavonoids produced in response to cyst nematode infection of *Arabidopsis thaliana*. Nematology **9**(5), 671-677.

Jones, J.T., Moens, M., Mota, M., Li, H. and Kikuchu, T. (2008) *Bursaphelenchus xylophilus*: opportunities in comparative genomics and molecular host–parasite interactions. *Molecular Plant Pathology* **9**: 357-368.

Jones, L.M., Giorgi, C.D. and Urwin, P.E. (2011) *C. elegans*; as a resource for studies on plant parasitic nematodes. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Jones, J., Gheysen, G. and Fenoll, C. (eds). *Springer Netherlands*, pp. 175-220.

Jones, M. and Goto, D.B. (2011) Root-knot Nematodes and Giant Cells. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Jones, J., Gheysen, G. and Fenoll, C. (eds). *Springer Netherlands*, pp. 83-100.

Kaiser, P. and Huang, L. (2005) Global approaches to understanding ubiquitination. *Genome Biology* **6**.233-249

Kaloshian, I., Desmond, O.J. and Atamian, H.S. (2011) Disease Resistance-Genes and Defense Responses During Incompatible Interactions. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Jones, J., Gheysen, G. and Fenoll, C. (eds). Springer Netherlands, pp. 309-324.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M. *et al.* (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**: 231-237.

Karimi, M., Inze, D. and Depicker, A. (2002a) GATEWAY vectors for *Agrobacterium-mediated* plant transformation. *Trends Plant Science* **7**: 193-195.

Keselman, H.J. (1976) A power investigation of the Tukey multiple comparison statistic. *Educational and Psychological Measurement* **36**: 97-104.

Kikuchi, T., Cotton, J.A., Dalzell, J.J., Hasegawa, K., Kanzaki, N., McVeigh, P. *et al.* (2011) Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathogens* **7**(9): e1002219

Kikuchi, T., Li, H.M., Karim, N., Kennedy, M.W., Moens, M. and Jones, J.T. (2009) Identification of putative expansin-like genes from the pine wood nematode, Bursaphelenchus xylophilus, and evolution of the expansin gene family within the Nematoda. *Nematology* **11**: 355-364.

Kikuchi, T., Aikawa, T., Kosaka, H., Pritchard, L., Ogura, N. and Jones, J.T. (2007) Expressed sequence tag (EST) analysis of the pine wood nematode Bursaphelenchus xylophilus and B. mucronatus. *Molecular and Biochemical Parasitology* **155**: 9-17.

Kim, G.T., Tsukaya, H. and Uchimiya, H. (1998) The CURLY LEAF gene controls both division and elongation of cells during the expansion of the leaf blade in *Arabidopsis thaliana*. *Planta* **206**: 175-183.

Kloppholz, S., Kuhn, H. and Requena, N. (2011) A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Current Biology* **21**: 1204-1209.

Koch, E.W. and Durako, M.J. (1991) In vitro studies of the submerged angiosperm *Ruppia maritima*: Auxin and cytokinin effects on plant growth and development. *Marine biology* **110**: 1-6.

Komatsu, S. (2007) Extraction of nuclear proteins. *Methods Molecular Biology* **355**: 73-77.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **305**: 567-580.

Kunkel, B.N. and Brooks, D.M. (2002) Cross talk between signaling pathways in pathogen defense. *Current Opinion in Plant Biology* **5**: 325-331.

Kurien, B.T. and Scofield, R.H. (2006) Western blotting. *Methods* **38**: 283-293.

Lambert, K.N., Allen, K.D. and Sussex, I.M. (1999a) Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*. *Molecular Plant-Microbe Interactions* **12**: 328-336.

Lee, C., Chronis, D., Kenning, C., Peret, B., Hewezi, T., Davis, E.L. *et al.* (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.

Leibfried, A., To, J.P.C., Busch, W., Stehling, S., Kehle, A., Demar, M. *et al.* (2005) WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators. *Nature* **438**: 1172-1175.

Libault, M., Wan, J., Czechowski, T., Udvardi, M. and Stacey, G. (2007) Identification of 118 *Arabidopsis* transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor. *Molecular Plant-Microbe Interactions* **20**: 900-911.

Liefhebber, J.M.P., Punt, S., Spaan, W.J.M. and Van Leeuwen, H.C. (2010) The human collagen beta (1-O) galactosyltransferase, GLT25D1, is a soluble endoplasmic reticulum localized protein. *BMC cell Biology* **11**: 33-50

Lilley, C.J., Kyndt, T. and Gheysen, G. (2011) Nematode Resistant GM Crops in Industrialised and Developing Countries. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Jones, J., Gheysen, G. and Fenoll, C. (eds). *Springer Netherlands*, pp. 517-541.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R. *et al.* (2012) Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology* Article ID 251364.

Liu, Y., Schiff, M. and Dinesh-Kumar, S.P. (2002) Virus-induced gene silencing in tomato. *The Plant Journal* **31**: 777-786.

Lotze, M.T., Zeh, H.J., Rubartelli, A., Sparvero, L.J., Amoscato, A.A., Washburn, N.R. *et al.* (2007) The grateful dead: damage-associated molecular pattern molecules and reduction/oxidation regulate immunity. *Immunological Reviews* **220**: 60-81.

Lu, S.W., Chen, S., Wang, J., Yu, H., Chronis, D., Mitchum, M.G. and Wang, X. (2009) Structural and functional diversity of CLAVATA3/ESR (CLE)-like genes from the potato cyst nematode *Globodera rostochiensis*. *Molecular Plant-Microbe Interactions* **22**: 1128-1142.

Mackey, D., Holt, F., Wiig, A. and Dangl, J.L. (2002) RIN4 Interacts with *Pseudomonas syringae* type III effector molecules and Is Required for RPM1-Mediated Resistance in *Arabidopsis*. *Cell* **108**: 743-754.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C. *et al.* (2011) CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* **39**: 225-229.

Mardis, E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**: 133-141.

Mazzucotelli, E., Belloni, S., Marone, D., De Leonardis, A.M., Guerra, D., Fonzo, N. *et al.* (2006) The E3 ubiquitin ligase gene family in plants: Regulation by degradation. *Current Genomics* **7:** 509-522.

Mersmann, S., Mentzel, T., Boller, T., de Torres, M., Mansfield, J.W. and Robatzek, S. (2008) Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB. *Current Biology* **18**: 1824-1832.

Milligan, S.B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P. and Williamson, V.M. (1998) The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell* **10**: 1307-1319.

Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F. and Wright, F. (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* **25**: 126-127.

Minnis, S.T., Haydock, P.P.J., Ibrahim, S.K., Grove, I.G., Evans, K. and Russell, M.D. (2002) Potato cyst nematodes in England and Wales-occurrence and distribution. *Annals of Applied Biology* **140**: 187-195.

Mitreva-Dautova, M., Roze, E., Overmars, H., de Graaff, L., Schots, A., Helder, J. *et al.* (2006) A symbiont-independent endo-1, 4-beta-xylanase from the plant-parasitic nematode *Meloidogyne incognita*. *Molecular Plant-Microbe Interactions* **19**: 521-529.

Monaghan, J. and Zipfel, C. (2012) Plant pattern recognition receptor complexes at the plasma membrane. *Current Opinion in Plant Biology*. **15**(4): 349–357

Mukhtar, M.S., Carvunis, A.R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J. *et al.* (2011a) Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science Signalling* **333**: 596-609

Muller, I.S. and Tester, M. (2007) Salinity tolerance of *Arabidopsis*: a good model for cereals? *Trends in Plant Science* **12**: 534-540.

Muller, R., Borghi, L., Kwiatkowska, D., Laufs, P. and Simon, R. (2006) Dynamic and compensatory responses of *Arabidopsis* shoot and floral meristems to CLV3 signaling. *Plant Cell* **18**: 1188-1198.

Newton, S.E. and Meeusen, E.N.T. (2003) Progress and new technologies for developing vaccines against gastrointestinal nematode parasites of sheep. *Parasite Immunology* **25**: 283-296.

Nguyen, H.P., Yeam, I., Angot, A. and Martin, G.B. (2010) Two virulence determinants of type III effector AvrPto are functionally conserved in diverse Pseudomonas syringae pathovars. *New Phytologist* **187**: 969-982.

Nicholson, S.J., Hartson, S.D. and Puterka, G.J. (2012) Proteomic analysis of secreted saliva from Russian Wheat Aphid (*Diuraphis noxia* Kurd.) biotypes that differ in virulence to wheat. *Journal of Proteomics* **75**: 2252-2268.

Nicol, J.M., Turner, S.J., Coyne, D.L., Nijs, L.d., Hockland, S. and Maafi, Z.T. (2011) Current nematode threats to world agriculture. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Jones, J., Gheysen, G. and Fenoll, C. (eds). *Springer Netherlands*, pp. 21-43.

Nicol, P., Gill, R., Fosu-Nyarko, J. and Jones, M.G.K. (2012) de novo analysis and functional classification of the transcriptome of the root lesion nematode, *Pratylenchus thornei*, after 454 GS FLX sequencing. *International Journal for Parasitology* **42**: 225-237.

O'Donnell, P.J., Calvert, C., Atzorn, R., Wasternack, C., Leyser, H.M.O. and Bowles, D.J. (1996) Ethylene as a signal mediating the wound response of tomato plants. *Science* **274**: 1914-1917.

Okkema, P. (2006) Transcriptional regulation. In WormBook., pp. 1-40.

Oliveira, A.R., de Moraes, G.J. and Ferraz, L.C.C.B. (2007) Consumption rate of phytonematodes by *Pergalumna* sp.(Acari: Oribatida: Galumnidae) under laboratory conditions determined by a new method. *Experimental and Applied Acarology* **41**: 183-189.

Opperman, C., Bird, D. and Schaff, J. (2009) Genomic analysis of the root-knot nematode genome. *Cell Biology of Plant Nematode Parasitism*: 221-237.

Opperman, C.H. and Bird, D.M. (1998) The soybean cyst nematode*, Heterodera glycines*: A genetic model system for the study of plant-parasitic nematodes. *Current Opinion in Plant Biology* **1**: 342-346.

Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, 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.

Painter, J.E. and Lambert, K.N. (2003) *Meloidogyne javanica* chorismate mutase transcript expression profile using real-time quantitative RT-PCR. *Journal of Nematology* **35**: 82-87.

Park, B.H., Karpinets, T.V., Syed, M.H., Leuze, M.R. and Uberbacher, E.C. (2010) CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20**: 1574-1584.

Patel, N., Hamamouch, N., Li, C.Y., Hewezi, T., Hussey, R.S., Baum, T.J. *et al.* (2010a) A nematode effector protein similar to annexins in host plants. *Journal of Experimental Botany* **61**: 235-248.

Pereyra-Irujo, G.A., Gasco, E.D., Peirone, L.S. and Aguirrezíbal, L.A.N. (2012) GlyPh: a low-cost platform for phenotyping plant growth and water use. *Functional Plant Biology*. 39(11) 905-913

Perry, R.N. and Moens, M. (2006) Plant nematology Wallingford: CABI.

Perry, R.N. and Wright, D.J. (1998) The physiology and biochemistry of free-living and plant-parasitic nematodes New York: C.A.B. International.

Plovie, E., De Buck, S., Goeleven, E., Tanghe, M., Vercauteren, I. and Gheysen, G. (2003) Hairy roots to test for transgenic nematode resistance: think twice. *Nematology* **5**: 831-841.

Popeijus, H., Blok, V.C., Cardle, L., Bakker, E., Phillips, M.S., Helder, J. *et al.* (2000) Analysis of genes expressed in second stage juveniles of the potato cyst nematodes *Globodera rostochiensis* and *G. pallida* using the expressed sequence tag approach. Nematology **2**(5): 567-574.

Postma, W.J., Slootweg, E.J., Rehman, S., Finkers-Tomczak, A., Tytgat, T.O.G., van Gelderen, K. *et al.* (2012) The effector SPRYSEC-19 of *Globodera rostochiensis* suppresses CC-NB-LRR-mediated disease resistance in plants. *Plant Physiology*.**5**: 112-125

Pozio, E., Varese, P., Morales, M.A.G., Croppo, G.P., Pelliccia, D. and Bruschi, F. (1993) Comparison of Human Trichinosis Caused by *Trichinella spiralis* and by *Trichinella britovi*. *American Journal of Tropical Medicine and Hygiene* **48**: 568-575.

Pritchard, L. and Birch, P. (2011) A systems biology perspective on plant microbe interactions: Biochemical and structural targets of pathogen effectors. *Plant Science* **180**: 584-603.

Protasio, A.V., Tsai, I.J., Babbage, A., Nichol, S., Hunt, M., Aslett, M.A. *et al.* (2012) A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* **6**: e1455.

Puthoff, D.P., Nettleton, D., Rodermel, S.R. and Baum, T.J. (2003) *Arabidopsis* gene expression changes during cyst nematode parasitism revealed by statistical analyses of microarray expression profiles. *The Plant Journal* **33**: 911-921.

Pylypenko, L., Uehara, T., Phillips, M., Sigareva, D. and Blok, V. (2005) Identification of *Globodera rostochiensis* and *G. pallida* in the Ukraine by PCR. *European journal of Plant Pathology* **111**: 39-46.

Qin, L., Kudla, U., Roze, E.H.A., Goverse, A., Popeijus, H., Nieuwland, J. *et al.* (2004) Plant degradation: A nematode expansin acting on plants. *Nature* **427**: 30.

Qin, L., Overmars, H., Helder, J., Popeijus, H., Rouppe Van Der Voort, J., Groenink, W. *et al.* (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.

Qin, L., Smant, G., Stokkermans, J., Bakker, J., Schots, A. and Helder, J. (1998) Cloning of a trans-spliced glyceraldehyde-3-phosphate-dehydrogenase gene from the potato cyst nematode *Globodera rostochiensis* and expression of its putative promoter region in *Caenorhabditis elegans*. *Molecular and Biochemical Parasitology* **96**: 59-67.

Queval, G., Hager, J., Gaki, B. and Noctor, G. (2008) Why are literature data for H2O2 contents so variable? A discussion of potential difficulties in the quantitative assay of leaf extracts. *Journal of Experimental Botany* **59**: 135-146.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005) InterProScan: Protein domains identifier. *Nucleic Acids Research* **33**: 116-120.

Rehman, S., Postma, W., Tytgat, T., Prins, P., Qin, L., Overmars, H. *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.

Rice, P., Longden, L. and Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**: 276-277.

Rico, A., Bennett, M.H., Forcat, S., Huang, W.E. and Preston, G.M. (2010) Agroinfiltration reduces ABA levels and suppresses *Pseudomonas syringae*-elicited salicylic acid production in *Nicotiana tabacum*. *Plos One* **5**: e8977.

Roberts, P.A. (1995) Conceptual and practical aspects of variability in root-knot nematodes related to host plant resistance. *Annual Review of Phytopathology* **33**: 199-221.

Roitsch, T. (1999) Source-sink regulation by sugar and stress. *Current Opinion in Plant Biology* **2**: 198-206.

Rosebrock, T.R., Zeng, L., Brady, J.J., Abramovitch, R.B., Xiao, F. and Martin, G.B. (2007) A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity. *Nature* **448**: 370-374.

Rosso, M.N. and Grenier, E. (2011) Other nematode effectors and evolutionary constraints. *Genomics and Molecular Genetics of Plant-Nematode Interactions*: 287-307.

Rothberg, J.M. and Leamon, J.H. (2008) The development and impact of 454 sequencing. *Nature Biotechnology* **26**: 1117-1124.

Roze, E., Hanse, B., Mitreva, M., Vanholme, B., Bakker, J. and Smant, G. (2008) Mining the secretome of the root-knot nematode *Meloidogyne chitwoodi* for candidate parasitism genes. *Molecular Plant Pathology* **9**: 1-10.

Sacco, M.A., Koropacka, K., Grenier, E., Jaubert, M.J., Blanchard, A., Goverse, A. *et al.* (2009a) The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2-and RanGAP2-dependent plant cell death. *PLoS Pathogens* **5**(8): e1000564.

Sacco, M.A., Mansoor, S. and Moffett, P. (2007) A RanGAP protein physically interacts with the NB-LRR protein Rx, and is required for Rx-mediated viral resistance. *The Plant Journal* **52**: 82-93.

Sacco, M.A., Koropacka, K., Grenier, E., Jaubert, M.J., Blanchard, A., Goverse, A. *et al.* (2009b) The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death. *PLoS Pathogens* **5**: e1000564.

Sasagawa, Y., Urano, T., Kohara, Y., Takahashi, H. and Higashitani, A. (2003) *Caenorhabditis elegans* RBX1 is essential for meiosis, mitotic chromosomal condensation and segregation, and cytokinesis. *Genes to Cells* **8**: 857-872.

Schadt, E.E., Turner, S. and Kasarskis, A. (2010) A window into third-generation sequencing. *Human molecular genetics* **19**: 227-240.

Scholl, E.H., Thorne, J.L., McCarter, J.P. and Bird, D.M. (2003) Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biology* **4**: R39.

Schrammeijer, B., Risseeuw, E., Pansegrau, W., Regensburg-Tuink, T.J.G., Crosby, W.L. and Hooykaas, P.J.J. (2001) Interaction of the virulence protein VirF of *Agrobacterium tumefaciens* with plant homologs of the yeast Skp1 protein. *Current Biology* **11**: 258-262.

Semblat, J.P., Rosso, M.N., Hussey, R.S., Abad, P. and Castagnone-Sereno, P. (2001) Molecular cloning of a cDNA encoding an amphid-secreted putative avirulence protein from the root-knot nematode *Meloidogyne incognita*. *Molecular Plant Microbe Interactions* **14**: 72-79.

Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika* **52**: 591-611.

Sijmons, P.C., Grundler, F.M.W., Mende, N., Burrows, P.R. and Wyss, U. (2008) *Arabidopsis thaliana* as a new model host for plant parasitic nematodes. *The Plant Journal* **1**: 245-254.

Smant, G. and Jones, J. (2011) Suppression of Plant Defences by Nematodes. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Jones, J., Gheysen, G. and Fenoll, C. (eds). *Springer Netherlands*, pp. 273-286.

Smant, G., Stokkermans, J.P.W.G., Yan, Y., De Boer, J.M., Baum, T.J., Wang, X. *et al.* (1998) Endogenous cellulases in animals: Isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences* **95**: 4906-4911.

Sobczak, M., Avrova, A., Jupowicz, J., Phillips, M.S., Ernst, K. and 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* **18**: 158-168.

Sobczak, M. and Golinowski, W. (2008) Structure of cyst nematode feeding sites. *Plant Cell* **10**:153-187

Sobczak, M., Golinowski, W.A. and Grundler, F.M.W. (1999) Ultrastructure of feeding plugs and feeding tubes formed by *Heterodera schachtii*. *Nematology* **1**: 363-374.

Sobczak, M., Fudali, S. and Wieczorek, K. (2011) Cell Wall Modifications Induced by Nematodes. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Jones, J., Gheysen, G. and Fenoll, C. (eds). *Springer Netherlands*, pp. 395-422.

Sobczak, M. and Golinowski, W. (2011) Cyst Nematodes and Syncytia. *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Jones, J., Gheysen, G. and Fenoll, C. (eds). *Springer Netherlands*, pp. 61-82.

Sohn, K.H., Lei, R., Nemri, A. and Jones, J.D.G. (2007) The downy mildew effector proteins ATR1 and ATR13 promote disease susceptibility in *Arabidopsis thaliana*. *Plant Cell* **19**: 4077-4090.

Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M. *et al.* (2009) Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. *Proceedings of the National Academy of Sciences* **106**: 1654-1659.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N. *et al.* (2003) The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLoS Biology* **1**(2): 45-55

Stephenson, L.S., Holland, C.V. and Cooper, E.S. (2000) The public health significance of *Trichuris trichiura*. *Parasitology* **121**: 73-95.

Stergiopoulos, Ioannis, and Pierre JGM de Wit. "Fungal effector proteins." *Annual Review of Phytopathology* 47 (2009): 233-263.

Strange, Richard N., and Peter R. Scott. "Plant disease: a threat to global food security." *Annal. Review Phytopathology.* 43 (2005): 83-116.

Szakasits, D., Heinen, P., Wieczorek, K., Hofmann, J., Wagner, F., Kreil, D.P. *et al.* (2009) The transcriptome of syncytia induced by the cyst nematode *Heterodera schachtii* in *Arabidopsis* roots. *Plant Journal* **57**: 771-784.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L. *et al.* (1999) The rde-1 Gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**: 123-132.

Team, R.D.C. (2008) R: A language and environment for statistical computing. *R Foundation Statistical Computing*. **31**: 299-314

Thompson, W., Rouchka, E.C. and Lawrence, C.E. (2003) Gibbs Recursive Sampler: finding transcription factor binding sites. *Nucleic Acids Research* **31**: 3580-3585.

Tobin, J.D., Haydock, P.P.J., Hare, M.C., Woods, S.R. and Crump, D.H. (2008) Effect of the fungus Pochonia chlamydosporia and fosthiazate on the multiplication rate of potato cyst nematodes (*Globodera pallida* and *G. rostochiensis*) in potato crops grown under UK field conditions. *Biological Control* **46**: 194-201.

Turner, S.J., Stone, A.R. and Perry, J.N. (1983) Selection of potato cyst-nematodes on resistant Solanum vernei hybrids. *Euphytica* **32**: 911-917.

Turner, S.J. and Stone, A.R. (1984) Development of Potato Cyst-Nematodes in Roots of Resistant Solanum Tuberosum Ssp. Andigena and S. Vernei Hybrids. *Nematologica* **30**(3), 324-332.

Tytgat, T., Vanholme, B., De Meutter, J., Claeys, M., Couvreur, M., Vanhoutte, I. *et al.* (2004) A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes. *Molecular Plant-Microbe Interactions* **17**: 846-852.

Undurraga, S.F., Press, M.O., Legendre, M., Bujdoso, N., Bale, J., Wang, H. *et al.* (2012) Background-dependent effects of polyglutamine variation in the *Arabidopsis thaliana* gene ELF3. *Proceedings of the National Academy of Sciences* **109**: 19363-19367.

Urwin, P.E., Lilley, C.J., McPherson, M.J. and 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.

Valentine, T., Shaw, J., Blok, V.C., Phillips, M.S., Oparka, K.J. and Lacomme, C. (2004) Efficient virus-induced gene silencing in roots using a modified tobacco rattle virus vector. *Plant Physiology* **136**: 3999-4009.

Van der Hoorn, R.A.L., Laurent, F., Roth, R. and De Wit, P.J.G.M. (2000) Agroinfiltration is a versatile tool that facilitates comparative analyses of Avr9/Cf-9-induced and Avr4/Cf-4-induced necrosis. *Molecular Plant-Microbe Interactions* **13**: 439-446.

van Megen, H., van den Elsen, S., Holterman, M., Karssen, G., Mooyman, P., Bongers, T. *et al.* (2009) A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* **11**(6), 927-950.

Vanholme, B., De Meutter, J., Tytgat, T., Van Montagu, M., Coomans, A. and Gheysen, G. (2004) Secretions of plant-parasitic nematodes: a molecular update. *Gene* **332**: 13-27.

Vanholme, B., Haegeman, A., Jacob, J., Cannoot, B. and Gheysen, G. (2009) Arabinogalactan endo-1, 4-beta-galactosidase: a putative plant cell wall-degrading enzyme of plant-parasitic nematodes. *Nematology* **11**: 739-747.

Vens, C., Rosso, M.N. and Danchin, E.G.J. (2011) Identifying discriminative classification-based motifs in biological sequences. *Bioinformatics* **27**: 1231-1238.

Vierstra, R.D. (2009) The ubiquitin 26S proteasome system at the nexus of plant biology. *Nature Reviews Molecular Cell Biology* **10**: 385-397.

Wang, J., Replogle, A., Hussey, R., Baum, T., Wang, X., Davis, E.L. and Mitchum, M.G. (2011) Identification of potential host plant mimics of CLAVATA3/ESR (CLE) like peptides from the plant parasitic nematode *Heterodera schachtii*. *Molecular Plant Pathology* **12**: 177-186.

Wang, X.H., Mitchum, M.G., Gao, B.L., Li, C.Y., Diab, H., Baum, T.J. *et al.* (2005) A parasitism gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR (CLE) of *Arabidopsis thaliana*. *Molecular Plant Pathology* **6**: 187-191.

Wang, X., Allen, R., Ding, X., Goellner, M., Maier, T., De Boer, J.M. *et al.* (2001) Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. *Molecular Plant-Microbe Interactions* **14**: 536-544.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M.l. and Barton, G.J. (2009) Jalview Version 2 a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.

Weerasinghe, R.R., Bird, D.M. and Allen, N.S. (2005) Root-knot nematodes and bacterial Nod factors elicit common signal transduction events in *Lotus japonicus*.

Proceedings of the National Academy of Sciences of the United States of America **102**: 3147-3152.

Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M. *et al.* (2007a) A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **450**: 115-118.

Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M. *et al.* (2007b) A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **450**: 115-118.

Williamson, V.M. and Gleason, C.A. (2003) Plant nematode interactions. *Current Opinion in Plant Biology* **6**: 327-333.

Williamson, V.M. and Kumar, A. (2006) Nematode resistance in plants: the battle underground. *Trends in Genetics* **22**: 396-403.

Wilson, M.J. and Kakouli-Duarte, T. (2009) Nematodes as environmental indicators (2009) CABI.

Win, J., Krasileva, K.V., Kamoun, S., Shirasu, K., Staskawicz, B.J. and Banfield, M.J. (2012) Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. *PLoS Pathogens* **8**: e1002400.

Wing, S.S. (2003) Deubiquitinating enzymes the importance of driving in reverse along the ubiquitin–proteasome pathway. *The International Journal of Biochemistry & Cell Biology* **35**: 590-605.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B. *et al.* (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**: 901-906.

Woo, J.S., Imm, J.H., Min, C.K., Kim, K.J., Cha, S.S. and Oh, B.H. (2006) Structural and functional insights into the B30.2/SPRY domain. *EMBO* **25**: 1353-1363.

Wyss, U. (1992) Observations on the Feeding Behaviour of Heterodera-Schachtii Throughout Development Including Events During Moulting. *Fundamental and Applied Nematology* **15**: 75-89.

Wyss, U. and Grundler, F.M.W. (1992) Feeding behavior of sedentary plant parasitic nematodes. *European Journal of Plant Pathology* **98**: 165-173.

Wyss, U., Grundler, F.M.W. and Munch, A. (1992) The parasitic behavior of 2nd-stage juveniles of *Meloidogyne incognita* in roots of *Arabidopsis thaliana*. *Nematologica* **38**: 98-111.

Xiang, T., Zong, N., Zou, Y., Wu, Y., Zhang, J., Xing, W. *et al.* (2008) *Pseudomonas syringae* effector AvrPto blocks innate immunity by targeting receptor kinases. *Current Biology* **18**: 74-80.

Xiong, T.C., Bourque, S., Lecourieux, D., Amelot, N., Grat, S., Brire, C. *et al.* (2006) Calcium signaling in plant cell organelles delimited by a double membrane. *Biochim Biophys Acta* **1763**: 1209-1215.

Xu, R. and Li, Q.Q. (2008) Protocol: Streamline cloning of genes into binary vectors in Agrobacterium via the Gateway-TOPO vector system. *Plant Methods* **4**.21-32

Yadav, B.C., Veluthambi, K. and Subramaniam, K. (2006) Host-generated double stranded RNA induces RNAi in plant-parasitic nematodes and protects the host from infection. *Molecular and Biochemical Parasitology* **148**: 219-222.

Ye, Y. and Rape, M. (2009) Building ubiquitin chains: E2 enzymes at work. *Nature Review Moleclar Cell Biology* **10**: 755-764.

Yu, H., Braun, P., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T. *et al.* (2008) High-quality binary protein interaction map of the yeast interactome network. *Science* **322**: 104-110.

Zhang, K., Letham, D.S. and John, P.C.L. (1996) Cytokinin controls the cell cycle at mitosis by stimulating the tyrosine dephosphorylation and activation of H1 histone kinase. *Planta* **200**: 2-12.

Zhou, J.M. and Chai, J. (2008) Plant pathogenic bacterial type III effectors subdue host responses. *Current Opinion in Microbiology* **11**: 179-185.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiology* **136**: 2621-2632.

Zipfel, C. (2009) Early molecular events in PAMP-triggered immunity. *Current Opinion in Plant Biology* **12**: 414-420.

Zipfel, C. and Robatzek, S. (2010) Pathogen-associated molecular pattern-triggered immunity: Veni, vidi...? *Plant Physiology* **154**: 551-554.