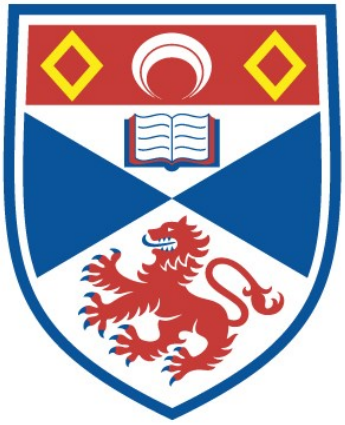
CHARACTERISATION OF THE POTATO *H2* RESISTANCE GENE AGAINST *GLOBODERA PALLIDA*

Shona Marie Strachan

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2019

Characterisation of the potato *H2* resistance gene against *Globodera pallida*

Shona Marie Strachan

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

September 2018

Candidate's declaration

I, Shona Marie Strachan, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 39,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2014.

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

Date Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a

copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Shona Marie Strachan, confirm that my thesis does not contain any third-party material that requires copyright clearance.

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

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date Signature of candidate

Date Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Shona Marie Strachan, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date Signature of candidate

General acknowledgements

I would like to thank Dr Vivian Blok and Dr Ingo Hein for their supervision and support throughout my PhD, and special thanks to Prof John Jones and Dr Glenn Bryan for their help and insight.

I would also like to thank the nematology lab, especially the other PhD students who helped immensely over the last four years.

Funding

This work was supported by Agriculture and Horticulture Development Board (AHDB) Potatoes; The Perry Foundation; and The Felix Cobbold Trust.

Abstract

Multiple pathotypes of the potato cyst nematode *Globodera pallida* are present in British potato fields; however, no single resistance gene has been identified which confers resistance to all three pathotypes which are present in British fields (Pa1, Pa2/3). Fortunately, the *H2* resistance gene has been identified from the wild potato diploid *S. multidissectum* which confers a high level of resistance to Pa1, and a lower level of resistance to Pa2/3.

A segregating F1 population was generated using susceptible cultivar Picasso and resistant breeding line P55/7 and using an initial population of 154 progeny clones the *H2* gene was verified to be a simplex dominant resistance gene. Bulked segregant analysis was undertaken with GenSeq and RenSeq enrichment to identify informative SNPs. Allele specific markers based on these informative SNPs were used to map the *H2* gene to a 4.7Mb region of potato chromosome 5, using the dihaploid DM genome as a reference. Using an expanded F1 population of 656 clones in combination with designed markers allowed the region of interest to be narrowed to 0.8Mb at the distal end of chromosome 5. Analysis of this region identified two NB-LRR structure resistance genes which became putative candidates of the *H2* gene.

Gene enrichment was also undertaken in pathotype Pa1 of *G. pallida* in order to identify putative candidate effectors which initiate resistance gene mediated cell death. A comparative analysis between virulent Pa2/3 and avirulent Pa1 revealed 19,873 allelic differences, which were stringently filtered to 23 variants in 10 candidate genes. One candidate gene was cloned and functionally tested via vacuum infiltration in resistant P55/7 potato leaves to observe a potential hypersensitive response. The data from the presented work will aid future research into cloning the functional *H2* gene.

Table of Contents

List of Figures

List of Tables

List of Supplementary Tables

List of Abbreviations

µg	Microgram
µl	Microlitre
AFLP	Amplified fragment length polymorphism
ATP	Adenosine triphosphate
Avr	Avirulence
BLAST	Basic local alignment search tool
bp	Base pair
CC	Coiled coil
cDNA	Complementary DNA
cm	Centimetre
CN	Cyst nematode
CPC	Commonwealth potato collection
CTAB	Cetrimonium bromide
CWDE	Cell wall degrading enzyme
DM	*S. tuberosum* clone Phureja DM1-3
DNA	Deoxyribonucleic acid
dpi	Days post infection
dNTP	Deoxynucleotide triphosphate
dsDNA	Double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ETI	Effector triggered immunity
F1	First generation progeny
GFP	Green fluorescent protein
GTP	Guanosine triphosphate
h	Hour
HF	Hatching factor
HI	Hatching inhibitor

HR	Hypersensitive response
HS	Hatching stimulant
Hz	Hertz
IAA	Isoamyl alcohol
IB	Infiltration buffer
INDEL	Insertion or deletion
ISC	Initial syncytial cell
J2	Second stage juvenile
KASP	Kompetitive Allele Specific PCR
l	Litre
LB	Luria-Bertani
LM	Light microscopy
LRR	Leucine rich repeat
min	Minute
M	Molar
MAMP	Microbe associated molecular pattern
Mb	Megabase
MES	2-ethanesulfonic acid
mg	Milligram
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
ml	Millilitre
mm	Millimetre
mM	Millimolar
MNP	Multiple nucleotide polymorphism
mRNA	Messenger RNA
$NaCl_2$	Sodium chloride
NB	Nucleotide binding
ng	Nanogram
NGS	Next generation sequencing

NLR	NOD-like receptor
OD	Optical density
PAMP	Pathogen associated molecular pattern
PCD	Programmed cell death
PCN	Potato cyst nematode
PCR	Polymerase chain reaction
PPN	Plant parasitic nematode
PRR	Pathogen recognition receptor
PTI	PAMP-triggered immunity/ Pattern-triggered immunity
PVX	Potato virus X
R	Resistance
RBP1	Ran-binding protein 1
RFLP	Restriction fragment length polymorphism
RKN	Root knot nematode
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
sec	Second
SMRT	Single molecule real-time
SNP	Single nucleotide polymorphism
SOC	Super optimal broth with catabolite repression
SPRYSEC	Secreted SP1a and Ryanodine receptor
STAND	Signal transduction ATPase with numerous domains
TIR	Toll and interleukin-1 receptor
TLR	Toll-like receptor
TRD	Tomato root diffusate

1. General Introduction

1.1 Origin of the Potato Cyst Nematode

1.1.1 Nematodes are everywhere

With the exception of insects, nematodes are the most widely distributed organisms on the planet – occupying every ecological niche. Although nematodes are all fundamentally similar, they differ dramatically in where and when they carry out their life cycle. Most identified species are non-parasitic and free-living, but there are others which parasitise animals, insects, and plants (Masler, 2013).

Nematodes have evolved to become diverse and effective in their ability to survive with or without a host. Their ability to adapt to a broad spectrum of environments has helped them become highly successful in their respective niches, for example the potato cyst nematode *Globodera pallida* which will be the focus of this research.

1.1.2 Plant Parasitic Nematodes (PPNs)

Plants are parasitised by a wide range of nematodes, with all agriculturally important crops being parasitised by at least one species of PPN (Ali et al., 2017). Nematodes have co-evolved and adapted to parasitise across multiple plant taxa which has led to their ability to expertly manipulate their host in order to fulfil their life cycle requirements, most often to the detriment of the host plant.

Nematode parasitism of plants has evolved several times in at least four clades of the phylum (Figure 1.1 (van Megen et al., 2009)). The phylum Nematoda comprises 12 clades with PPNs being spread across clades 1, 2, 10 and 12. Clade 1 (*Enoplida* and *Triplonchida*) and Clade 2 (subclass *Dorylaimia*) contain mostly migratory ectoparasites such as *Trichodorus* and *Xiphinema* spp. which are most commonly known to be economically important vectors of viral diseases. Clade 10 (*Aphelenchoidea* and *Panagrolaimomorpha*) mostly contains mycetophagous nematodes (*Aphelenchoides* spp.) but also contains the migratory endoparasites *Bursaphelenchus* spp.; exemplified by the pine

nematode *B. xylophilus* which causes wilt disease in pine trees (Nickle et al., 1981). Infection causes localised plant cell death by disrupting the flow of water through the plant (Korma and Sigareva, 2012) leading to widespread systemic damage and host death. Clade 12 (*Aphelenchidae* and *Tylenchida*) contains the most economically damaging species of crops, including the migratory endoparasites such as *Pratylenchus* and *Radopholus* spp. as well as the sedentary endoparasitic nematodes including root-knot (*Meloidogyne*) and cyst nematodes (*Heterodera* and *Globodera*) (van Megen et al., 2009, Jones et al., 2013).

1.1.3 Root-knot and Cyst Nematodes

The root-knot (*Meloidogyne* ssp.) and cyst (*Heterodera* and *Globodera* ssp.) nematodes from clade 12 (Figure 1.1) are the most economically important of the plant parasitic nematodes (Chitwood, 2003). Root-knot nematodes (RKNs) are characterised by the appearance of galls on infected root systems (Jones and Goto, 2011). These sedentary endoparasites have varying levels of host range with some, such as *Meloidogyne incognita* having a host range encompassing almost all vascular plants (Sasser, 1977). These nematodes parasitise many cash crops causing huge yield losses in both tropical and sub-tropical agriculture (Kiewnick and Sikora, 2006).

Solanaceous plants are targets for many root-knot nematodes with *M. incognita, M. arenaria* and *M. javanica* being found in Mediterranean, tropical and equatorial climates, while *M. hapla*, *M. fallax* and *M. chitwoodi* are found in temperate regions (Caromel and Gebhardt, 2011). During host infection RKN develop and maintain giant cells; a highly specialised feeding site created through the re-differentiation of between five and seven cells within the vascular cylinder to become multi-nucleate (Taylor and Sasser, 1978). Formation of the giant cells requires the nematode to induce karyokinesis (nuclear division) in the absence of cytokinesis in order to achieve the multinucleate state. Once these giant cells have developed the nematode becomes sedentary and completely dependent on

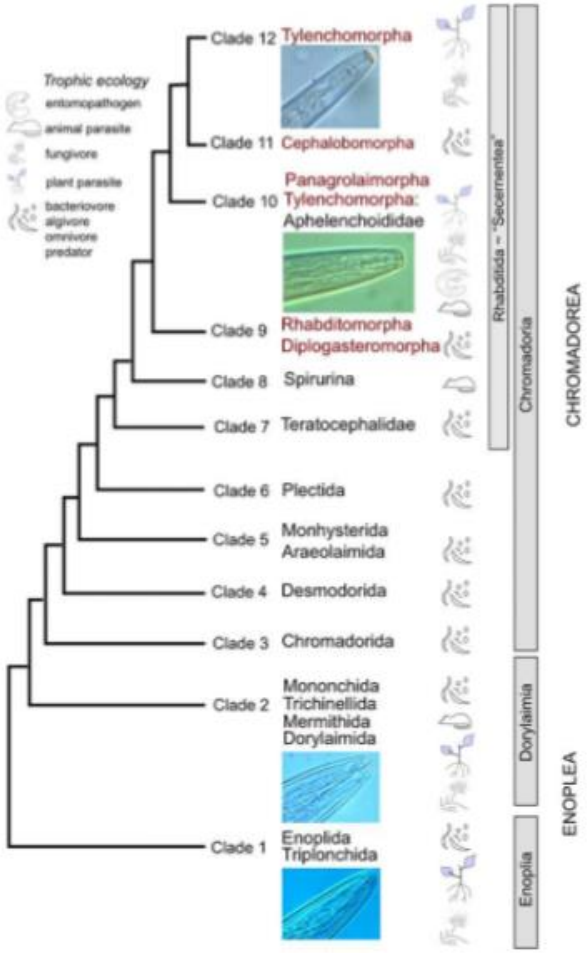


Figure 1.1 Phylogenetic position of nematode clades from (van Megen et al., 2009)

them for the nourishment required for all further development and growth. It is usually at this stage, one or two weeks after the initial infection, that the characteristic galls or knots begin to become visible on the outer surface of the root (Norton and Niblack, 1991, Ding et al., 1998).

In contrast to RKN, cyst nematodes, including the potato cyst nematodes (PCN) usually have a much more restricted host range. The most economically important cyst nematodes are those of the genera *Heterodera* and *Globodera*. From the genera *Heterodera*, *H. glycines* (soybean) and *H. schachtii* (sugar beet) are of most economic interest, however *H. avenae* is the most widespread, causing damage to wheat, barley and oat crops in more than 50% of cereal growing land in Europe (Lilley et al., 2005).

Globodera rostochiensis and *G. pallida* are the most prolific and economically important parasitic nematodes of potato crops, having the ability to reduce total crop yields by up to 70% (Stare et al., 2013). The genus *Globodera* is comprised of more than 10 species including *G. mexicana* and *G. tabacum*, and in 2012 a new species of cyst nematode was discovered in Oregon and Idaho in North America, later named *G. ellingtonae*; which is morphologically similar to other cyst nematodes apart from its distinctive J2 tail (Handoo et al., 2012). Both *G. rostochiensis* and *G. pallida* are comprised of several sub-populations, termed pathotypes, which can be differentiated based on their virulence and ability to multiply on a test set of *Solanum* ssp. clones, each containing resistance (Kort et al., 1977).

1.1.4 Introduction from South America

Potatoes were first brought to Europe from their native South America in the 16th century through two introductions; one to Spain in 1570 and a second to Britain in 1588. During this time the crop was not farmed, but instead treated as a curiosity, leading to the suggestion that any nematodes which were introduced with them would have been able to thrive and multiply uninhibited. Over time, multiple potato species were brought into Europe and despite their deep eyes and adaption to the short days of sub-tropical South America, became a staple food crop. By the end of the 17th century, potatoes were the main food source for many farming and crofting communities, and so the arrival of the late blight *Phytophthora infestans* in the cropping season of 1845-1846 resulted in the failure of crops, widespread famine, the death of nearly one million people, and mass emigration. In an attempt to stem losses and breed varieties which were late blight resistant, new wild species were imported from South America. Although this led to the development of potatoes with some resistance to late blight, it is also likely that this led to a further introduction of PCN due to imported, and most probably infected, potatoes from South America (Evans et al., 1975).

The close relationship which has formed between *Globodera* ssp. and *Solanum* ssp. indicates a long history of co-evolution and adaptation between the two (Picard et al., 2007). *G. pallida* is the primary nematode pest of potato crops in

South America, especially those found in the Andean Cordillera (Picard et al., 2007, Plantard et al., 2008). It has been suggested that *Globodera* ssp. originated 15-18 million years ago (Grenier et al., 2010) and the hypothesis that the Incas used 6-8 year crop rotations to avoid crop losses strengthens the argument that PCN infestation of potato is an ancient problem (Picard et al., 2007).

Natural resistance has been discovered in wild *Solanum* species, such as *Gro1* active against *G. rostochiensis* from *S. spegazzinii* (Barone et al., 1990) and *GroVI* from *S. vernei* (Jacobs et al., 1996), both of which originate from South America illustrating that resistances to be discovered would likely come from a wild relative of cultivated potato.

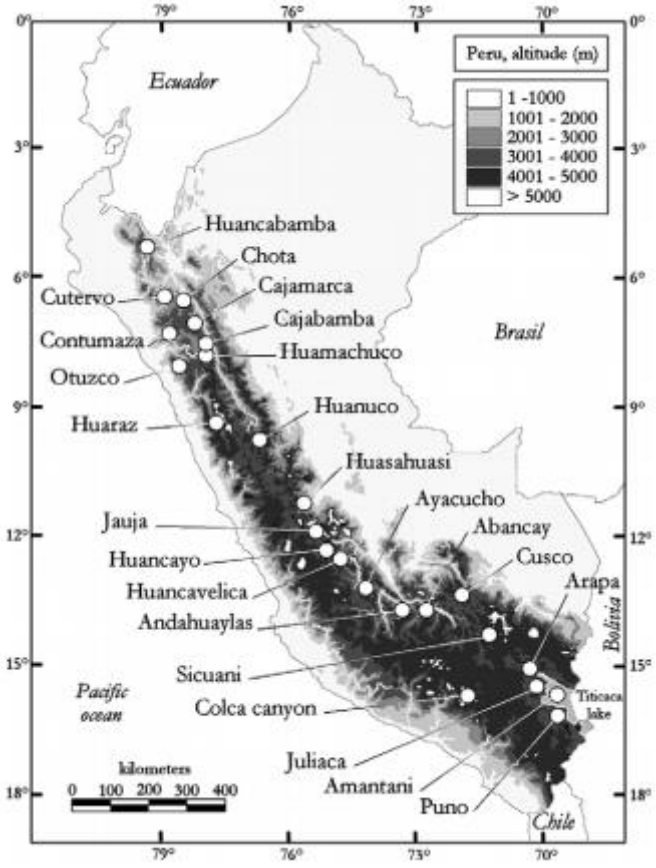


Figure 1.2 Map of Peru from (Picard et al., 2007). The Andean Cordillera, where *Globodera pallida* originates, is close to the Peruvian-Bolivian border and runs from Lake Titicaca to Cusco

All European populations of *G. pallida* studied to date originate from the Altiplano region of Peru; which extends from Lake Titicaca in the South towards Cusco in the North (Figure 1.2) (Plantard et al., 2008). Picard *et al* (2007) described five clades of *G. pallida* in South America with decreasing levels of genetic variability when moving south-to-north. The divergence in genetic makeup of the clades provides evidence that they diverged from one another long before human settlement and cultivation (Picard et al., 2007).

1.1.5 Theory of Multiple Introductions

European *G. pallida* is comprised of three pathotypes (Kort, 1974), with genetic differences between each of the pathotypes. Two explanations have been proposed to explain these differences; the first hypothesis is that most of the *G. pallida* populations in Europe come from one primary introduction, and that populations fragmented from the original larger primary population. Due to genetic drift and numerous founder effects, the three distinct pathotypes which are present today have arisen. The second hypothesis proposes that there were multiple introductions of pathotype Pa2/3, the most common European form. These multiple introductions may possibly have stemmed from two or more closely related areas which are too similar to distinguish between in South America (Blok et al., 1997). The resultant genetic differences would have arisen through founder effects acting on the original fragmentation of the species and the consequential selection pressures of the new location (Phillips and Trudgill, 1983).

The Pa1 pathotype is distinct from the Pa2/3 pathotype and seems likely that it was introduced separately into Britain (Phillips and Trudgill, 1983). It is hypothesised that Pa1 had one introduction into Northern Ireland where it has spread through the country with a small number of subsequent introductions to Scotland (Blok et al., 1997). Its limited geographical distribution, lack of virulence against the *H2* gene from *S. multidissectum,* as well as mitochondrial testing which differentiated its maternal lineage from that of Pa2/3 strengthens the argument that it was introduced separately from Pa2/3, as such a divergence is unlikely to have occurred in such a short time since the initial nematode

introduction (Phillips et al., 1992). The unique population “Luffness” was believed to be restricted to one area of Scotland; its genetic differences compared to other Pa2/3 populations poses the question as to whether it was introduced separately or whether it is a variant of another established population (Blok et al., 1997).

In comparison to *G. pallida*, British populations of *G. rostochiensis* are less genetically diverse. This lack of diversity is assumed to be due to a restricted number of introductions (possibly only one) into the UK (Evans et al., 1975, Bendezu et al., 1998).

1.1.6 Pathotypes

Before 1972 only one species of PCN was recognised, *Heterodera rostochiensis*, with several different pathotypes. Differences, including the body colour of young females, began to be recognised between the nematode populations, and the new species *H. pallida* was named (Canto Saenz and de Scurrah, 1977). Further study of the species led to the genus *Globodera* being created based on the differences in cyst morphology between *Heterodera* and *Globodera* cysts (Mulvey and Stone, 1976, Canto Saenz and de Scurrah, 1977).

As discussed earlier, resistance against European populations of PCN has been included in potato breeding programmes for decades. Once resistant cultivars were being developed it was noted that certain PCN populations could multiply on some genotypes but not others, suggesting that the species comprised more than one genotypic population (Kort et al., 1977). Research in The Netherlands and Britain resulted in the creation of two separate classification schemes. Observations in The Netherlands had noted six distinct populations, being characterised as A-E, while British research observed three distinct populations which were characterised as A, B and E to align with the Dutch scheme (Kort et al., 1977). Work undertaken by Canto-Sáénz and de Scurrah (1977) identified three new and distinct populations in Peru, Ecuador and Colombia, designated P_1B, P_3A and P_2A respectively. In 1977 Kort *et al.* proposed a scheme which numbered pathotypes and gave them the prefix Ro (for *G. rostochiensis*) or Pa (for *G. pallida*). This led to the recognition of five pathotypes of *G. rostochiensis* (Ro1-5) and three pathotypes of *G. pallida* (Pa1-3) (Kort et al., 1977). With this

scheme in place, a new population would be defined as a distinct pathotype if it differed in its ability to multiply on a resistant host (differential) compared to that of the known pathotypes (Kort et al., 1977).

This scheme was created in Europe but does not reflect the much wider PCN gene pool found within the Andes. Populations there have evolved with different selective pressures and European populations are a small subset of this diversity (Canto Saenz and de Scurrah, 1977). Further research based on Kort's scheme was undertaken on native Andean populations in order to relate them to the European populations (Table 1.1). The larger number of pathotypes found in South America reflects that this is the origin where the species evolved, was distributed and diverged over time. European populations represent a limited introduction(s) of this diversity and have been subjected to bottlenecking during secondary distribution potentially leading to lower levels of population diversity.

Based on the pathotype classification scheme put forward by Kort *et al* (1977), Pa1 can be differentiated from Pa2/3 populations based on whether P55/7 is a susceptible host. This tetraploid hybrid originating from the wild diploid *S. multidissectum,* was proposed to contain the single major resistance gene *H2* (Dunnett, 1961), and is the only major resistance gene to control Pa1 populations.

Table 1.1 Pathotype nomenclature devised to align Andean and European PCN populations to each other. Samples with (-) indicate that no orthologous pathotype is present in Europe.

Andean	R_1A	R_1B	R_2A	R_3A	P_1A	P_1B	P_2A	P_3A	P_4A	P_5A
European	Ro1	Ro4	Ro2	Ro3	Pa1	-	-	-	Pa2	Pa3

1.2 Impact of Potato Cyst Nematodes

1.2.1 Importance of the Potato

The global population currently stands at an estimated 7.4 billion (correct February 2018) with the net gain of one person every 15 seconds (Commerce, 2015). The population is projected to reach 8 billion by 2025 (Johnson, 2001), assuming the global population contiunes to rise annually as it has done as long

as records have been kept (Smith, 1966). The first noticeable spike in the human population came in the 17th century and since then birth rates have increased exponentially, culminating in a population explosion at the end of the 20th century (Smith, 1966). This increase is due to a multitude of factors including; better sanitation, as well as improved medicine, but importantly an increased awareness and improvement of nutrition, with potato playing a role in the bettering of nutrition levels (Smith, 1966).

In modern agriculture the most widely cultivated tuber-bearing crop is the potato (*Solanum tuberosum*), with annual production standing at 325 Mtonnes globally (FAO, 2012). Although potato ranks third behind the cereal crops rice and wheat as the most important food crop grown worldwide, potatoes produce more dry weight and protein per hectare than both rice and wheat (Burton, 1989). Approximately 50% of cropped potatoes are eaten fresh while the rest is either made into dried goods and animal feed or used to provide seed for the following season (Birch et al., 2012).

Each hectare of cropped land already feeds an average of four people, but this increases to 20 people per hectare in countries like Egypt where less arable land must feed larger populations. With the ever growing global population, it is clear that crop yields will also need to increase to deal with the intensification of demand. Any increase in crop yield would presumably indicate a requirment for intensification of productivity of the current cropping land as there is limited additional land available to be brought into cultivation (Johnson, 2001). Intensification can already be seen in potato crop production with a 21% increase in overall production between 1991-2007 (FAO, 2012), cementing the potato as a crop of global importance.

1.2.2 Economic Impact

More than one billion people eat potatoes globally, making it an important crop for food security during a rise in population and the threat of increased hunger rates (Birch et al., 2012). For the first time growth of potato in the developing world matched that of developed countries in 2005, with Asia being responsible

for 47.5% of global production in 2010 (FAO, 2012) proving once again the impact the potato crop has throughout the world.

Intensification of production can already be observed with total global production of potato in 2013 sitting at 376 million tonnes, an increase of just under 30% since 1963 (FAO, 2014) and this increase in food production has become heavily reliant on the implementation of monocultures. With this comes a greater risk of crops being attacked by opportunistic pests and pathogens. These pests can cause huge economic damage through yield losses and counterbalancing measures, with the estimated damage caused by *G. rostochiensis* and *G. pallida* in the UK valued at £45 million per annum, which equates to a 9% total yield loss (Nicol, 2010).

1.2.3 Control Strategies

There is a constant need to identify new ways to combat the threat of nematode infection of plants. Current control options available include crop rotation, deployment of resistant cultivars and nematicides.

Crop rotation is an ancient agricultural control strategy which has many benefits including a reduction in soil erosion as well as the maintenance of soil structure and nutrient levels (Peters et al., 2003). This strategy is used to combat PCN as it exploits the natural decline of populations over time due to spontaneous hatching of second stage juveniles (J2s) in the absence of a viable host, attrition through egg mortality and the destruction of unhatched cysts by other pests (Devine et al., 1999, Peters et al., 2003). The success of crop rotation hinges on the non-host which is planted between cycles of potato and the removal of volunteer potatoes (or ground keepers) (Emmond and Ledingham, 1972). Although rotations are highly effective in limiting increases in PCN population size, they are less effective in combating high population density infections (Peters et al., 2003). Hancock (1988) suggested that to return a field to a below detectable PCN threshold, a decline rate of 30%, for *G. rostochiensis*, per cropping season would be required in a 6-8 year rotation cycle (Devine et al., 1999). However, the soil climate greatly effects the rate of decline and studies in

the Republic of Ireland saw far greater levels of decline per year than those estimated by Hancock.

Nematicides are estimated to be used on approximately 1.7 million acres worldwide. Their commercial use began in 1945 after the development and introduction of DDT (dichlorodiphenyltrichloroethane) and ethylene dibromide in 1943 and 1951 respectively (Johnson, 1985). Nematicides are classified as either fumigants or non-fumigants. Fumigant compounds are most commonly liquids which are injected under the top layer of soil where they vaporise and begin to diffuse through the air pockets within the soil, dissolving into soil water and are taken up into the nematode through the cuticle (Spurr, 1985). The fumigants move deep within the soil and have a broad spectrum of biocidal activity against a wide range of organisms including other animals, fungi and bacteria (Van Gundy and McKenry, 1977). The widely used fumigant methyl bromide also has a wide environmental impact, contributing to the depletion of the ozone layer and so was removed from use, under the Montreal Protocol, in March 2010 (EU directive 2008/753/EC) (Karpouzas et al., 2004).

Non-fumigants were developed and commercially deployed in 1970. Most mixes are bound to granules but there are some which can be administered as liquids sprayed onto plant foliage (Whitehead et al., 1981). These chemicals move through the soil by percolation, and have narrow spectrums of biocidal activity; killing any target with a nervous system, and are non-phytotoxic at suggested usage rates allowing for immediate planting of crops (Spurr, 1985). The most commonly used chemical is aldicarb which inhibits acetylcholinesterases and cholinesterases. These esterase compounds are also present in humans, and so due to risks to human health as well as the enviroment, aldicarb was withdrawn from use in the EU in 2003 under directive EC91/414/EEC (Ruiz-Suárez et al., 2015).

The over use of chemical compounds has led to their reduced biological efficacy, leading to the requirement of higher dosages and concentrations. Biodegradation of the compounds has also been linked to their failure to control nematode infestations and with more stringent usage and policies enforced by the EU,

nematicide use is becoming less of a viable strategy in the control of PCN. Natural resistance can be a highly effective and environmentally friendly method of controlling many pests and diseases. This is discussed in detail in section *1.5 Resistance*.

1.3 Life Cycle of *Globodera pallida*

1.3.1 The Life Cycle

The life cycle of *G. pallida* both begins and ends in an encapsulated cyst. PCN can lay dormant for many years and hatch when the required signals are released. The ability of the nematode to hatch only in the presence of a suitable host is dictated by the release of root exudates from the host plant, allowing the nematode to co-ordinate its own life cycle with that of its host. Both *G. pallida* and *G. rostochiensis* are heavily regulated by root exudate induced hatching unlike *H. schachtii* and *H. avenae* which hatch in large quantities in the absence of an inducer (Jones et al., 1998).

The nematode hatches as a second stage juvenile (J2), having already undergone their first moult within the cyst (Perry, 1989). Once hatched the J2 begins a free-living phase whilst trying to locate a suitable root system. During this time the nematode does not feed but instead survives on internal lipid reserves. These energy stores are finite and if depleted to below 65% the juvenile will be unable to invade and parasitise a host plant (Curtis, 2007). Guided by root exudates the nematode invades a suitable root, most commonly at the zone of elongation, just behind the root tip. The nematode then uses its stylet, a hollow needle-like projection, to disrupt the plant cell wall and move intercellularly through the cortical cells toward the vascular cylinder, causing localised cell damage through the thrusting action of its stylet and the release of cell wall degrading enzymes (CWDE) (Davis et al., 2004). The damage produces tears in the cortical cell walls allowing the nematode to force its way through the root (Wyss and Grundler, 1992). A single cell, close to the vascular cylinder, is chosen to become the initial feeding cell. The nematode then secretes a suite of proteins into the cell that induce the formation of a syncytial feeding site (Goverse et al.,

2000, Gheysen and Mitchum, 2011). The most active molecules within the secretions are enzymes including cellulases and proteinases as well as CWDEs which degrade the pectic polysaccharides of the middle lamella (Smant et al., 1998). A secreted homolog of plant expansin is expressed in the subventral esophageal glands, the observed expansion of the plant cell wall suggests that this protein plays a role in the ability of the nematode to 'loosen' cell walls in order to facilitate easier intracellular migration. Esophageal secretions play a large role in successful parasitism of a root system, but secretions from the amphids (chemosensory organ) may also have a role (Hussey et al., 2002). A more in depth explanation of nematode secretions required for successful parasitism is given in chapter 6: *funtional testing of putative candidate H2 avirulence genes in resistant P55/7*. Cell wall dissolution, starts with a widening of plasmodesmata and fusion of adjacent protoplasts causes the expansion of the syncytium until is has incorporated up to 200 neighbouring cells (Gheysen and Fenoll, 2002). Endoreduplication of DNA creates enlarged nuclei and nucleoli, and the central vacuole fragments into smaller vacuoles. The syncytium develops invaginations that increase surface area to help facilitate water and nutrient uptake from the xylem (Golinowski et al., 1996).

Each time the nematode inserts its stylet into the syncytium to feed, a feeding tube forms. The tube-like structure formed ensures that the feeding site remains intact and is not damaged in any way as nematodes can only create a single feeding site, and damaging it would prove fatal (Eves-van den Akker et al., 2014a). Stylet secretions are produced throughout the lifetime of the feeding site to continually stimulate the up-keep of the sycytium as well as to suppress any host defence responses (Cotton et al., 2014). While feeding, the juvenile undergoes three moults to reach adulthood with each new stage lasting from between 3-4 days. Male nematodes become motile during the adult stage, at which point they leave the root and fertilise a nearby female. In order to attract males to their location the still sedentary female excretes sex pheromones, with *H. glycines* using vanillic acid to attract males to the correct root position (Jaffe et al., 1989). Females develop into adults, enlarging and swelling to a saccate sphere in preparation for fertilisation. Once fertilised, the cuticle of the female

darkens from creamy white (for *G. pallida*) to brown through polyphenol oxidation which creates a tough outer cyst coating. Each cyst can contain up to approximately 500 eggs awaiting the release of root exudates to stimulate hatch of the next generation (Brodie et al., 1993).

An important stage in the nematode life cycle is quiescence and diapause. Entering either of these stages allows the cyst to survive long periods of time without a host (Ebrahimi et al., 2014). Diapause is characterised by suspended development of nematodes immediately after J2 formation in the cyst, and does not end until predictable and unfavourable conditions have ended (cold temperatures) (Devine and Jones, 2001). This type of dormancy lasts for a set period of time and even if favourable conditions arise before sufficient time has elapsed, the eggs will not hatch. If the unfavourable conditions continue after the set time of diapause the cysts will then enter a state of quiescence (Perry, 1997). In contrast, quiescence is the delaying of development until conditions are optimal for the nematode to hatch. Exposure of cysts to root diffusates is the biggest factor influencing quiescence, its effect is discussed in detail below.

1.3.2 Factors controlling hatching and development

Nematodes are ectothermic organisms and so rely on their environment to maintain a constant body temperature. They have adapted to specific temperature ranges and often require varying optimal temperature for each phase of their life cycle. *G. pallida* has adapted to cooler conditions, requiring lower average temperature for hatch and development compared to *G. rostochiensis* (10°-18°C for *G. pallida*, 15°-23°C for *G. rostochiensis* (Franco, 1979, Kaczmarek et al., 2014).

Hatching of J2s is controlled by host plant root exudates, in particular hatching factors (HF) (Byrne et al., 2001). The HFs initiate a calcium-dependent change in permeability within the inner lipid membrane, resulting in trehalose flowing out from the perivitelline fluid, and an increased uptake of water, allowing for the rehydration of juveniles and their subsequent hatch. It was shown by Byrne *et al* (1998) that the exudate not only contains HFs but also hatching inhibitors (HIs) and hatching stimulants (HSs). The HIs inhibit HF-induced hatching, while HSs

alter the level of HFs which are produced. The inoculation of gravel pots with HF in the absence of viable host plants has been shown to generate mass hatching of nematodes, which could potentially lead to a novel control strategy in future (Byrne et al., 2001).

Both day length and light intensity have an indirect effect on hatching. Franco and Evans (1979) showed that more cysts hatch with 16 hours of daylight compared to 12 hours, while Hominick (1986) found that the amount and intensity of light that a plant receives has a direct correlation with the number of cysts which hatch. Increasing the day length increases the number of cysts which hatch, while short days give lower and more variable cyst hatching (Salazar and Ritter, 1993). Experiments carried out using *G. rostochiensis* found that cysts subjected to diffuse light hatched normally but cysts subjected to harsh, direct light stopped hatching completely (Franco and Evans, 1979).

1.4 Mechanism of Infection

The invasion of, and destruction caused by, the nematode does not go unnoticed by the plant and the host initiates several responses. Callose accumulates between the cell wall and plasma membrane around the site of stylet penetration in an attempt to halt feeding but has little effect as the tight interaction between plasma membrane and stylet excludes any callose from accumulating, allowing feeding to continue. This callose deposition may reflect a wound response as plant cells most commonly respond to mechanical damage by rapidly depositing callose along the inner surface of the infected wall to limit damage. Responses which are normally observed during stress or wounding are detected during nematode infection, including changes in gene expression. Experiments by Hammond-Kosack *et al* using potato roots infected with *G. rostochiensis* showed altered gene expression in the leaves and induced expression of pathogenesis-related proteins. Further research carried out by Niebel (1995) showed that a gene encoding for a catalase expressed during root bacterial infection is also expressed during root infection by *G. pallida* (Williamson and Hussey, 1996).

1.5 Plant Immune System

1.5.1 PAMP Triggered Immunity

Plants lack the mobile defence cells and adaptive immune system that is characteristic of mammals (Jones and Dangl, 2006). They instead rely on the innate immunity of individual cells and the systemic signalling of the infected cell. There are two methods by which inducible defence responses can be triggered within the plant cell.

The first uses pathogen associated molecule pattern (PAMP) recognition receptors (PRRs) and induces PAMP triggered immunity (PTI) (Zipfel, 2009). PAMPs are highly conserved host molecules, distributed across a diverse family of microbes, which are required for pathogen fitness (Schwessinger and Zipfel, 2008).

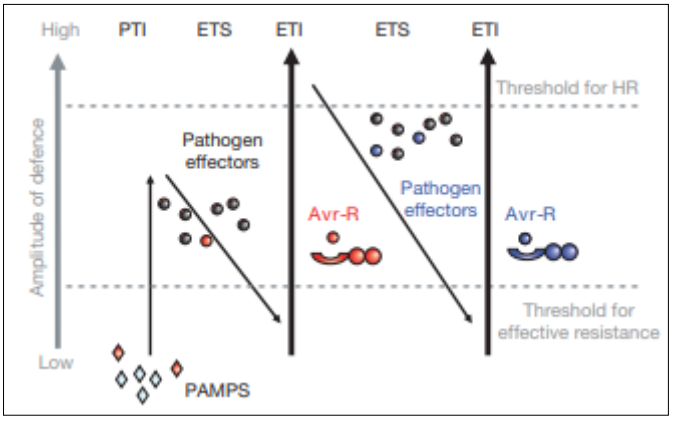


Figure 1.3 The 'zig-zag' model displaying host-pathogen interactions (Jones and Dangl, 2006)

When a pathogen infects a potential host plant it exposes itself to the surface localised PRRs and it is at that point that PTI is initiated (Figure 1.3). PAMP recognition activates many physiological pathways, including MAP kinase signalling, transcription of pathogen-associated genes, production of reactive oxygen species (ROS), and the deposition of callose on the inner membrane of the wounded cell wall (Chisholm et al., 2006). The quickest host response activates within minutes of recognition and causes an ion-flux across the plasma membrane, resulting in increased intracellular calcium, protein phosphorylation

and receptor endocytosis; by 30 minutes post recognition transcriptional changes have been induced. Experiments using the model host *Arabidopsis thaliana* show whole genome level changes due to infection; a total of 800 genes had expression changes, with 96 of them having prolonged changes over 12 hours pi (post infection) (Schwessinger and Zipfel, 2008).

There are two categories of PRR; transmembrane receptor kinases and transmembrane receptor-like proteins which act to recognise PAMP signals (Dodds and Rathjen, 2010). Little research has been carried out to characterise the nematode PAMPs which trigger PTI but one candidate, ascaroside 18 (ascr#18) has been identified. Ascarosides are signalling molecules found exclusively in nematodes and it has been found that they have the ability to induce defence and resistance responses *in planta* via effects analogous to bacterial flagellin signalling (Manosalva et al., 2015). Bacterial flagellin, a protein subunit of the flagellum, as well as chitin and ergosterol, major components of the cell wall, have all been identified as PAMPs.

1.5.2 Effector Triggered Immunity

Adapted pathogens suppress PTI through the release and activity of effectors (Zipfel, 2009). Effectors are molecules secreted by the pathogen which enable the alteration of host cell function and cellular structure. Some effector molecules allow the invading pathogen to suppress the PTI responses initiated by the host, but additionally assist in altering the plant's biology in order for the pathogen to carry out its life cycle (Win et al., 2012).

Effector molecules are secreted by the invading pathogen and can be secreted into host cells or into the apoplast. In the early stages of nematode infection and migration, effectors are secreted into the apoplasm to aid in cell wall degradation but effectors are also deposited into the cell cytoplasm (Hewezi and Baum, 2013). Several effector proteins have been found to have nuclear localisation signals (NLS) which facilitate the uptake of proteins into the nucleus, others are hypothesised to reach their target destination through interaction with host proteins. To date, no common consensus sequence has been identified which would act as an uptake signal, but experiments have shown that nematodes can

take up fluorescently tagged proteins from the feeding site proving that they have direct access to the host cell cytoplasm (Hewezi and Baum, 2013).

Resistance (R) proteins found within the host plant have the ability to detect effector proteins by either binding directly to them or through the detection of changes in activity of host proteins which themselves have been altered by effector proteins. R proteins are commonly encoded by NB-LRR (nucleotide binding leucine rich repeat) genes. If an effector is recognised by its corresponding R protein, ETI will ensue. ETI has a stronger response than PTI and often culminates in a programmed cell death event to limit the movement and feeding ability of the invading pathogen, such as observed during *Mi*- (RKN) and *Hero*-mediated (CN) resistance, which will be discussed in more detail in the proceeding section (Jones and Dangl, 2006, Bhattacharjee et al., 2013).

1.5.3 The Hypersensitive Response

The term ‘hypersensitive’ was first used by Stakman (1915) to describe the localised cell death seen in rust-resistant cereals. Further research revealed that this was a common response in pathogen-resistant plants and so was named the hypersensitive response (HR). A HR is characterised by rapid and localised cell death and a subsequent cell death lesion (Heath, 2000, Lam et al., 2001). Studies using mutants which spontaneously create hypersensitive-like lesions reveal that this cell death mechanism is under genetic control, and for a HR to be induced, the invading pathogen requires an avirulent (avr) effector protein which ‘matches’ a host R protein (Levine et al., 1996). The presence of genes in both resistant and susceptible host plants which encode pathways leading to the HR are also required and these RDR (required for disease resistance) genes may differ depending on the R protein triggered and the invading pathogen (Heath, 2000).

An early step in HR is the production of superoxide (O_2^-) and a build up of hydrogen peroxide (H_2O_2) amassing in an oxidative burst resembling the mechanism seen in the activated macrophages of humans. This oxidative burst along with production of salicyclic acid and the flux of cytosolic calcium triggers the mechanism to launch the systemic immune response (Delledonne et al.,

2001). Commencement of HR also triggers the inducement of plant defence responses that help to limit the spread of the nematode (Lam et al., 2001). Physiological changes trigger the toughening of cell walls through the creation of oxidative cross-linking of structural proteins and changes in transcriptional regulation including the up-regulation of genes controlling the synthesis of phytoalexins (antimicrobial broad spectrum inhibitors) and the release of hydrolytic enzymes (Levine et al., 1996).

During nematode infection one of two types of HR is observed, either a rapid response best characterised by *Mi*-induced RKN resistance, or a more delayed response as observed in CN. The *Mi* based resistance from the wild tomato species *S. peruvianum* confers resistance against *M. incognita*, *M. arenaria*, and *M. javanica*, its rapid localised necrosis can be observed within 24 hours of infection, with some changes in physiology being visible after 8-12 hours. A loss of electron density during a HR occurs around the vacuoles, followed by membrane disruption, and an increase in electron density in the cytoplasm. This is followed by necrosis to the cells surrounding the nematode and its feeding site, which limits the ability of the nematode to feed, eventually starving the nematode to death (Paulson and Webster, 1972, Trudgill, 1991)

The second, slower HR seen in CN occurs after the establishment of the feeding site. Research undertaken on *Hero A*- induced *G. rostochiensis* resistance and *Gpa2*-mediated *G. pallida* resistance observed that nematodes established feeding sites and became sedentary as would be seen in a susceptible response (Sobczak et al., 2005). In time the cells surrounding the syncytium become necrotic, followed by detoriation of the feeding site, effectively cutting off the nutrient supply to the nematode which is therefore unable to complete its life cycle. This differing in response time has been theorised to be due to a weaker recognition interaction between R proteins and their corresponding nematode effector (Sacco et al., 2009).

1.5.4 NB-LRRs

Early research using the fungal flax rust pathogen (*Melampsora lini*) and its host flax (*Linum usitatissimum*) helped identified the gene-for-gene relationship. In

response to pathogen attack and the consequential effector release, plants have evolved NB-LRR proteins which act as resistance (R) genes. Either resistance or susceptibility is determined by whether the plant *R* gene recognises the pathogen avirulence gene (*avr* gene) (Eitas and Dangl, 2010).

R genes encoding NB-LRRs split into two groups based on the structure of their N-terminus. The first group shares a domain with homology to the *Drosophila* ssp. Toll and mammalian interleukin (IL)-1 receptor and are designated TIR-NB-LRRs (Dangl and Jones, 2001). The TIR-NB-LRR domain contains two monomers which form an asymmetric globular protein, each monomer consisting of five α-helices encapsulated in a five-stranded β-sheet (Takken and Goverse, 2012). Group two contains a putative coiled-coil domain (CC-NB-LRR), the structure of which was solved using the *Mla10 R* gene of barley (*Hordeum vulgare*). The domain is characterised by the linking of two coiled-coil dimers, each containing a 90Å helix-loop-helix, into a tightly folded rod-shaped homodimer with a larger surface area for interacting (Maekawa et al., 2011). In comparison, the LRR region of both groups is highly variable with unpredictable repeat lengths and polymorphisms, both aspects are thought to play a role in the specificity of the proteins (Dodds et al., 2001).

When in an inactive state, due to the absence of any pathogen effectors, the NB-LRR domains are suspended in a signalling-able but auto-inhibited state. This state is maintained by chaperone-assisted protein maturation; chaperones including Hsp90 and co-chaperones Rar1 and PP5 are required for the correct folding of the NB-LRR into its activated state (Azevedo et al., 2002). Activation is thought to involve the relief of NB domain inhibition, followed by multimerisation to recruit necessary proteins to further signalling (Dangl and Jones, 2001). Any reduction in the levels of (co-) chaperones risks a change in steady-state protein levels and the possible alterations in disease resistance (Takken and Goverse, 2012).

NB-LRRs function as *R* genes and are highly adapted for this function. However LRR domains are not exclusively found in R proteins but function in other proteins

as sites of protein-protein interaction, protein-carbohydrate interaction, and peptide-ligand binding (Dangl and Jones, 2001).

A direct effector interaction involves the NB-LRR protein recognising an effector protein and binding to it, similar to a ligand binding to its receptor. The first recorded instance of this interaction type was shown with the Pita NB-LRR immune receptor in rice (*Oryza sativa*) and the avr-Pita effector in the fungus *Magnaporthe grisea*. The Pita NB-LRR directly binds to avr-Pita and a single mutation in the NB-LRR can cause loss of recognition and consequently resistance (Caplan et al., 2008). During an indirect response the NB-LRR binds through an intermediate 'guardee' protein. The "Guard Hypothesis" theorises that R proteins monitor target proteins and activate when these are attacked. Within the "Guard Hypothesis" there are two theorectical mechanisms which may be implemented; the first being that a R protein is constitutively bound to its target protein, sensing changes directly, or alternatively the R protein only binds to its target protein if and when an effector binds to it (Caplan et al., 2008, Eitas and Dangl, 2010). Resistance against a single pathogen isolate can require recognition by multiple R proteins leading to the theory of quantitive trait loci (QTL) when trying to map resistance genes. If several proteins are required for defence responses, one hypothesis is that the R proteins form hetero-multimers which allow for successful detection (Eitas and Dangl, 2010). This is exemplified by the rice NLRs RGA4 and RGA5. In the absence of RGA5, RGA4 is in an autoactive state, while binding of RGA5 to form a heterodimer alters R protein expression and allows for successful effector recognition (Césari et al., 2014).

The presence of NB-LRR proteins or NB-LRR-like proteins across plants, animals and insects is not the result of a common ancestor but through convergent evolution. Plants have specific NB-LRRs which recognise specific effector proteins whereas mammalian TLRs (Toll-like receptors) and NLRs (NOD-like receptors) only recognise generic MAMPs (microbe associated molecular patterns) (Caplan et al., 2008).

There is great diversity in the cellular localisation of NB-LRRs, an increasing number including *Rx1* and *N* show nucleocytoplasmic localisation, while *RRS1-R* localises exclusively in the nucleus (Elmore et al., 2011).

1.5.5 Nematode *R* and *Avr* Genes

A potato cultivar can be referred to as resistant to nematodes when it can significantly inhibit the growth and development of a given nematode population/pathotype (EPPO, 2006). The resulting resistance can be defined as either broad, being effective against several species of pathotypes, or narrow, only having control over one specific pathotype or species (Williamson and Hussey, 1996).

Several *R* genes against PCN have been identified and their location, or linkage group, in the potato genome mapped (Williamson and Hussey, 1996). The genes *Gro1* on chromosome 7, derived from *S. spegazzinii*, *H1* from *S. tuberosum* ssp. *andigena* on chromosome 5, and *GroVI* from *S. vernei*, also from chromosome 5, all confer major dominant resistance against *G. rostochiensis* (Caromel and Gebhardt, 2011). Control of *G. pallida* is more complex than *G. rostochiensis* but several genes have been identified. *Gpa2*, from *S. tuberosum* ssp. *andigena* and *Hero A* from wild tomato species *Solanum pimpinellifolium* are the only major resistance genes isolated against *G. pallida*. High level resistances to *G. pallida* are usually determined by one major effect QTL and one or several minor effect QTLs (Caromel and Gebhardt, 2011).

The first breeding program to produce potato cultivars with nematode resistance was carried out in the 1950s when Ellenby (1952) discovered the *H1* gene in *S. tuberosum* ssp. *andigena* accession CPC 1673 (Janssen et al., 1991). The *H1* gene was found to confer almost complete resistance to *G. rostochiensis* pathotypes Ro1 and Ro4, and was found to be present in five accessions (out of a possible 1300 tested) of *S. tuberosum* ssp. *andigena* (Gebhardt et al., 1993). Using genetic analysis of selfed CPC1673 seeds, it was determined that *H1* was a single copy dominant gene, and was consequently used to breed resistance into *S. tuberosum* spp. *tuberosum* breeding lines, and producing several

European cultivars including Granola and Maris Piper (Gebhardt et al., 1993). The location of *H1* has subsequently been mapped using RFLP (restriction fragment length polymorphism) markers to the distal arm of chromosome 5 (Gebhardt and Valkonen, 2001, Bakker et al., 2004b). This is closely linked to the *GroVI* gene, originating from *S. vernei*, which also confers resistance to *G. rostochiensis* (Gebhardt and Valkonen, 2001). Use of *H1* along with several other minor effect genes originating from *S. vernei* and *S. spegazzinii*, the threat of *G. rostochiensis* pathotype Ro1 was diminished to almost zero (Gebhardt et al., 1993). The *H1* gene remains one of the most durable resistance genes known (Bakker et al., 2004b, Finkers-Tomczak et al., 2011).

Experiments carried out by Dunnett (1957) identified populations of PCN that were virulent against cultivars containing *H1* (Phillips et al., 1994). Although the existence of *G. pallida* was not suspected at this time, these populations were later designated as Pa1 (van der Voort et al., 1997). Further research carried out by Dunnett (1963) using wild diploid *S. multidissectum* found *H2* resistance which was effective against the virulent Pa1 population. The resistance gene discovered was found to be a major effect, dominant resistance gene (Phillips et al., 1994). Subsequent work carried out by Howard *et al* (1970), made efforts to identify sources of resistance from cultivated *S. tuberosum* ssp. *andigena*, leading to the discovery of the *H3* gene which also confers resistance to *G. pallida* (Phillips et al., 1994). Further work on *H3* by Franco and Evans (1978) identified that the *H3* resistance gene was only effective against European populations of *G. pallida*, its resistance toward South American populations was only partial (Phillips et al., 1994). Their work received criticism but they argued that the resistance which had been designated as *H3* was actually controlled by several genes (Phillips et al., 1994).

1.6 Scope of Thesis

Over the last two decades, the cost of whole genome sequencing has steadily decreased, enabling the analysis of large plant genomes. The potato reference genome of the doubled-monoploid *S. tuberosum* group Phureja clone DM1-3 516 R44 (DM) was published in 2011 (PGSC, 2011), and has enabled an in-depth

analysis of NLR gene diversity and organisation within the potato genome (Jupe et al., 2012). The development of NLR-specific enrichment sequencing (RenSeq), has facilitated a more comprehensive NLR gene annotation (Jupe et al., 2013). Furthermore, RenSeq which targets all 755 described NLRs in potato, has been successfully used to map and/or identify functional NLRs against late blight (Jupe et al., 2013, Witek et al., 2016, Chen et al., 2018). While GenSeq targeted enrichment sequencing of 1980 single or low-copy number genes that can be placed on the individual potato chromosomes with high confidence, has proven a versatile and effective tool for the mapping of new resistances when utilised in combination with RenSeq (Chen et al., 2018). The presented research focused on *H2* which has been identified as an ideal candidate for mapping through enrichment sequencing due to its simplex dominant nature. Although *G. pallida* Pa1 has a limited distribution, any major resistance gene which can be identified is a positive step forward in generating durable broad spectrum PCN resistance and has potential for use in pyramiding of resistances.

The specific aims of this project were to:

- Determine the genomic location of the functional *H2* resistance gene, using a combination of bulk segregant analysis and gene enrichment sequencing
- Identify putative Pa1 candidate avirulence genes which activate the *H2* resistance pathway, and functionally test them in an *H2*-resistant cultivar

2. General Materials and Methods

2.1 Biological Materials

2.1.1 PCN Populations

Cysts from the *Globodera pallida* pathotype Pa1 and Pa2/3 Lindley populations from The James Hutton Institute PCN collection were used for all experiments.

2.1.2 Plant Material

A cross between susceptible potato *S. tuberosum* ssp. *tuberosum* cultivar Picasso and resistant genotype P55/7 was carried out by breeders at The James Hutton Institute and yielded 1000 seeds. Initially 192 progeny plants were produced for resistance testing and then an additional 656 plants were produced for fine mapping. Individual Picasso and P55/7 plants were grown and used for parental controls.

2.1.3 Tomato Root Diffusate (TRD)

Tomato plants (cv. Moneymaker) were grown in compost for 6 weeks in glasshouse conditions (16 h light, 8 h dark, 16 h 20ºC, 8 h 16ºC). Excess soil was then removed from the roots of 2 plants and the roots were placed into a 1.5l glass beaker with 1l of sterile distilled water (SDW) and left overnight (a minimum of eight hours). Following this the plants were removed from the beaker and disposed of. The water in the beaker was filtered through Whatman paper in a glass funnel into 500ml glass bottles, and stored at 4°C until required (Blair et al., 1999).

2.1.3 Cyst viability testing and second stage juvenile hatching

Viability of the cysts was tested by adding single cysts to 2ml TRD in a well of a 12-well plate leaving the J2s to hatch for seven days at 20°C. Hatching was examined with a low power microscope. If at least 50 nematodes had hatched after 7 days the cysts were used for phenotyping assays.

To prepare second stage juvenile nematodes (J2s), cysts were placed in a 106µm sieve in a 20cm plastic Petri dish. Twenty ml of root diffusate was added to the dish, enough to soak the bottom of the sieve, wrapped in a layer of cling film and covered with aluminium foil. Dishes were left to incubate at 20°C for 7 days.

2.1.4 Infection Assays

2.1.4.1 Meristem Cutting Assays

In order to better visualise successful nematode infection and identification of females, all infection assays were undertaken in root trainers. Racks of eight root trainers (4 chambers/root trainer) (Haxnicks) were filled with compost (insecticide free). Meristem cuttings were taken from progeny plants with a scalpel, the cut end dipped into root growth hormone (Doff) with one cutting planted per chamber, and one root trainer containing 4 replicates of a single clone. After one week, to allow for the establishment of roots, a hole was made in the soil approximately half the depth of the root trainer and wells were infected with 15±2 Pa1 cysts. Infected plants were left to grow for a further eight weeks before root trainers were opened to count females present on the root systems (Figure 2.1).

Figure 2.1 Open root trainer displaying *G. pallida* infected potato roots. Root trainers allow for roots to grow down the outside of the soil in a relatively straight arrangement. Females become visible on the outer surface of the root.

2.1.4.2 Tuber Assays

Racks of root trainers were set-up as in Section 2.1.4.1 and infected with 15±2 Pa1 cysts. Progeny tubers were taken from the cold store one week prior to planting in order to allow for sprouting. One centimetre square pieces of tuber were cut around the sprout and were planted sprout down into infected wells. Plants were left to grow for eight weeks before root trainers were opened to count all females present on root systems.

2.2 Molecular Protocols

2.2.1 DNA Extraction

2.2.1.1 Single Cysts and Single Females

DNA was extracted from single cysts or females using a three day extraction protocol. A single cyst/female was placed into a 2ml microcentrifuge tube and frozen in liquid nitrogen. The contents of each tube were then crushed using a plastic micro pestle using a twisting action, then 600µl QIAGEN Cell Lysis Buffer was added before the pestle was carefully removed. Five microlitres of Proteinase K (20mg/ml) (Roche) was added to each tube before vortexing and incubating overnight at 56°C. Next; 4µl of RNase A (100mg/ml) (QIAGEN) was added to the incubated samples and mixed by inversion before being incubated for 10 min at room temperature (RT, ~20-22°C). After incubating, 200µl QIAGEN Protein Precipitation Buffer was added and samples were briefly vortexed and incubated on ice for 10 min. Samples were centrifuged for 10 min (11,000rpm, 4°C), the supernatant was transferred to a fresh 1.5ml microcentrifuge tube before adding 600µl cold isopropanol and incubating overnight at -20°C. On the final day samples were centrifuged for 10 min (12,000rpm, 4°C), the supernatant was discarded, 600µl 70% ethanol was added before centrifuging for a further 30 min (12,000rpm, 4°C), again the supernatant was discarded. The pellet was dried in a fume hood for 1 h, before 21µl elution buffer was added (QIAGEN, UK) and incubated for 1 h at room temperature. Extracted samples were stored at -20°C until required.

2.2.1.2 Plant Material

Three discs from 1 potato leaf were added to a well of a 96 deep-well plate containing a 4mm stainless steel ball bearing. To each well 200µl extraction buffer (1000µl RNase A (20mg/ml) (Thermofisher), 2.2ml Proteinase K, 19.8ml ATL Buffer (QIAGEN, UK)) was added. The plate was sealed with two foil lids and disrupted using a Retsch mill (1 min at 20Hz, change plate orientation, 1 min at 20Hz). The plate was briefly centrifuged at 3,000rpm and then incubated at 65°C for 60 min in a water bath. Once the incubation was complete the plate was centrifuged for 10 min at 5,000rpm. The lysate of each sample was pipetted into a QIAcube HT lysate plate (QIAGEN, UK). Using the QIAGEN QIAcube robotic workstation, a 96-well filter plate was loaded into the transfer carriage which was fitted into the channel adapter and finally fitted into the channel block holder. The buffer reservoirs were filled (64.6ml Buffer AW1, 64.6ml Buffer AW2, 62.6ml 90% Ethanol, 38.6ml Buffer ACB, 22.2ml Buffer AE) (QIAGEN), two boxes of 200µl filter tips were placed in the workstation, along with the tip bin, 96-well elution plate and finally the lysate plate. Following the safety instructions on screen, the QIAcube was started and proceded through the extraction protocol:

Load 350µl ACB into 96 well plate
Mix wells in 96 well plate
Incubate for 2:30 (min:sec)
Mix wells in 96 well plate
Incubate for 2:30 (min:sec)
Load 550µl lysate from 96 well plate into vacuum plate
Vacuum on (35kPa), 5 min
Vacuum off
Load 600µl AW1 into vacuum plate
Vacuum on (35kPa), 2 min
Vacuum off

Load 600µl AW2 into vacuum plate
Vacuum on (35kPa), 1 min
Vacuum off
Load 600µl 96% Ethanol into vacuum plate
Vacuum on (35kPa), 30 sec
Vacuum off
Vacuum on (55kPa), 1 min
Vacuum on (35kPa), 2 min
Load 200µl AE into vacuum plate

Once the protocol was complete, the capture plate was placed onto the elution plate, and both plates were centrifuged at 5,000rpm for 2 min to increase DNA yield. Each sample was quantified using a NanoDrop 2000 (ThermoScientific).

2.2.2 RNA Extraction

RNA was extracted from either J2 nematodes or young potato leaves using the TRIzol Reagent Kit (Life Technologies). For nematode extraction, hatched J2s were removed from the hatching petri dishes and pipetted into 50ml Falcon tubes. Each tube was centrifuged for 10 min at 2,500rpm to pellet the juveniles. Excess TRD was removed and the pelleted juveniles transferred to a fresh 2ml microcentrifuge tube and flash frozen in liquid nitrogen. For potato extraction, 1cm^2 pieces of young leaf material were flash frozen in liquid nitrogen. In both extractions the sample was then placed into a pre-cooled mortar. Material was crushed with a pre-cooled pestle before 1ml TRIzol Reagent was added. The tissue/TRIzol solution was transferred to a fresh 2ml microcentrifuge tube and left to incubate for 5 min at RT. After incubation 200µl chloroform was added and the tube was vigorously shaken by hand for 15 sec, before a 3 min incubation at RT. The samples were centrifuged for 15 min (10,000rpm, 4°C). The aqueous layer was removed and pipetted into a fresh tube to which 500µl 100% cold isopropanol

was added, and incubated at RT for 10 min. The sample was centrifuged for 10 min (10,000rpm, 4°C), before the supernatant was removed, leaving the RNA pellet. The pellet was washed with 1ml 75% ethanol and centrifuged for 5 min (8,000rpm, 4°C), before the wash was discarded, the pellet left to air dry and then re-suspended in 20µl RNAse-free H_2O.

Extracted RNA was DNAse treated using the RQ1 RNAse-Free DNAse Kit (Promega). Eight microlitres of RNA was mixed with 1µl each of RQ1 10x Reaction Buffer and RQ1 RNAse-Free DNAse before being incubated at 37°C for 30 min. After incubation, 1µl of RQ1 DNAse Stop Solution was added, and the sample incubated at 65°C for 10 min.

2.2.3 cDNA Synthesis

For both nematode and plant samples, DNAse-treated RNA was used with the SuperScript III Reverse Transcriptase kit (Invitrogen) to create cDNA libraries. The following was added to a nuclease-free microcentrifuge tube:

Reagent	Volume for 1 reaction
oligo$(dT)_{20}$ (50µM)	1µl
RNA	11µl
dNTP (10mM)	1µl
Total volume	13µl

The sample was incubated for 5 min at 65°C and incubated on ice for at least 1 min. The sample was briefly centrifuged before the following was added:

Reagent	Volume for 1 reaction
5x First-Strand Buffer	4µl
0.1M DTT	1µl
RNAseOUT	1µl
SuperScript Ⅲ RT	1µl
Total volume	20µl

Samples were incubated at 50°C for 60 min, before being heated to 70°C for 15 min to inactivate the reaction.

2.2.4 PCR and product purification

PCR experiments were carried out using Promega GoTaq G2 Flexi polymerase, or KOD Hot Start polymerase (Merck) if proofreading was required. For each reaction the following mix was used:

Reagent	Volume for 1 reaction
H_2O	6.9µl
5x GoTaq Buffer	3µl
MgCl2 (25mM)	0.75µl
dNTPs (2mM)	1.25µl
Primer (10mM)	0.5µl
Polymerase	0.1µl
Template	2µl
Total volume	15µl

The reaction mixture underwent a standard program, using a heated lid of 112°C:

94°C	3 min	
94°C	30 sec	
55-65°C	30 sec	40 cycles
72°C	3 min	
72°C	5 min	
10°C	hold	

To test for successful amplification, a 5µl aliquot of the PCR reaction was run on a 2% agarose gel (1g agarose, 50ml TBE). Five microlitres of PCR product was mixed with 3µl of 6x loading dye (Promega,UK), loaded into the gel well, and run at 75 volts for 15 min.

Positive samples were purified using the QIAquick PCR Purification Kit (QIAGEN, UK). Five volumes (50µl) of PB buffer was added to each reaction and mixed. Each sample was added to a QIAquick column and centrifuged for 60 sec at 13,000rpm, and the flow-through discarded. Samples were then washed in 750µl PE buffer and centrifuged for a further 60 sec. DNA was eluted through the column with a final volume of 30µl elution buffer before being quantified via Nanodrop.

2.2.5 PCR product purification from agarose gel

Positive amplified samples were excised from the agarose using a sharp scalpel under UV light. Three volumes of QG buffer was added to the gel sample and incubated at 50°C for 10 min, or until the gel dissolved. One volume of isopropanol was then added and the tube inverted several times to mix. An aliquot of 700µl was added to a QIAquick spin column and centrifuged at 13,000rpm for 1 min, the flow-through was discarded and the remaining sample was added to the column and centrifuged again. Each sample was washed by adding 750µl PE buffer and incubated at RT for 2 min before being centrifuged for 1 min. The column was subsequently placed into a fresh 1.5ml microcentrifuge tube and 30µl elution buffer was added to the membrane of the column. After a 4 min incubation at RT, the tubes were centrifuged for 1 min. Samples were then quantified by Nanodrop to determine their DNA concentration.

2.2.6 Primer Design

Target sequences were inputted into Primer3 primer designing software (http://primer3.ut.ee/) and the optimal primers were chosen based on: annealing temperature (forward and reverse primers must be <5°C apart), GC content (must be >50%), the absence of secondary structure and self-complementation, and the specificity in amplifying the target gene sequence. All primers were synthesised and quality checked by Eurofins Genomics (UK).

2.2.7 Gene Enrichment and Sequencing Library Preparation

2.2.7.1 Quantification of Purified DNA

Extracted DNA for either nematode or plant was quantified using the Qubit dsDNA HS Assay (ThermoFisher Scientific) to determine the exact concentration of the sample.

2.2.7.2 DNA Shearing

The Covaris Sonicator was prepared as detailed in the user manual. Fifty microlitres of sample was pipetted into a Covaris microtube and fragmented, to a target length of 500bp, under the following conditions;

target bp	500
peak incident power	50
duty factor	20%
cycles per burst	200
treatment time	32 sec
temperature	20°C

Samples were then analysed on a Bioanalyser Chip (Agilent) to check for correct fragment size and DNA integrity.

2.2.7.3 Purification of Sheared gDNA

Samples were purified using the protocol outlined in Jupe *et al* (2014). The bead drying time was altered to 3 min to stop DNA loss through over-drying. The DNA was eluted into 58µl RNAse-free H_2O, and 55.5µl of supernatant was transferred to a fresh 0.5ml tube.

2.2.7.4 Library Preparation – End Prep

Libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina. Fragmented DNA was mixed as follows;

End Prep Enzyme Mix	3.0µl
End Repair Reaction Buffer (10x)	6.5µl
Fragmented DNA	55.5µl
Total volume	65µl

The reaction was mixed by gentle pipetting and a very brief centrifugation at the lowest speed to collect all the liquid. Each tube was run in a thermal cycler (with heated lid) under the following conditions;

20°C	30 min
65°C	30 min
4°C	hold

2.2.7.5 Library Preparation – Adapter Ligation

The following components were added directly to the End Prep reaction mix immediately after heating;

Blunt/TA Ligase Master Mix	15µl
NEBNext Adaptor for Illumina	2.5µl
Ligation Enhancer	1µl
Total volume	83.5µl

The mix was incubated at 20°C for 15 min in a thermal cycler, before 3µl of USER enzyme was added, and the mix was incubated at 37°C for 15 min.

2.2.7.6 Sample Clean Up

To the adapter ligated samples 86.5µl AMPure XP beads was added and incubated at RT for 5 min. Tubes were placed on a magnetic stand to separate the beads from the supernatant. After the solution had become clear, the supernatant was removed and discarded. 200µl 80% ethanol was added to each tube while still placed in the magnetic stand, incubated for 30 sec at RT, before removing and discarding the supernatant. This wash step was repeated a total of three times. The beads were air dried for 3 min while the tubes were on the magnetic stand with their lids open. To elute the DNA, 28µl of elution buffer was added, the mix was pipetted well, placed on the magnetic stand and when the solution cleared, 23µl was transferred to a new PCR tube.

2.2.7.7 Library Preparation – PCR Amplification

Samples were amplified using the NEBNext Multiplex Oligos for Illumina Primer Set 1. The following components were mixed in a PCR tube;

Adaptor Ligated DNA Fragments	23µl
NEBNext High Fidelity 2X PCR Master Mix	25µl
Index Primer*	1µl
Universal Primer	1µl
Total volume	50µl

*Index primers were used based on their compatibility, a signal in both the red and green channel at each sequence base was required.

Samples were run using the following thermocycler conditions;

98°C	30 sec	1 cycle
98°C	10 sec	10 cycles
65°C	30 sec	
72°C	30 sec	
72°C	5 min	1 cycle
4°C	hold	

Samples were then cleaned as outlined in section 2.2.7.6.

2.2.7.8 Library Hybridisation

Samples were hybridised using the SureSelect^{XT} Target Enrichment System for the Illumina Paired-End Sequencing Library. The hybridisation reaction requires 750ng of DNA with a maximum volume of 3.4µl (221ng/µl) so library samples were vacuum concentrated at ≤45°C. Samples were completely lyophilised and reconstituted in 3.4µl nuclease-free H_2O. To prepare the hybridisation buffer the following were mixed:

Reagent	Volume for 1 capture (µl)
SureSelect Hyb #1	25µl
SureSelect Hyb #2	1µl
SureSelect Hyb #3	10µl
SureSelect Hyb #4	13µl
Total	49µl (40µl needed)

The SureSelect capture library mix for target enrichment was prepared in PCR tubes under the following conditions:

Capture Size	Volume of SureSelect Library	RNase Block dilution (block:water)	Volume of RNase Block Dilution
<3.0Mb	2µl	1:9 (10%)	5µl

To make the SureSelect Block mix the following was combined in a 0.5ml microcentrifuge tube:

Reagent	Volume for 1 reaction
SureSelect Indexing Block #1	5µl
Universal Nematode Block	1µl
SureSelect Indexing Block #3	0.6µl
Total	5.6µl

In a separate PCR plate, samples were prepared for target enrichment. To each well of row “B” 3.4µl of 221ng/µl library preparation was added, along with 5.6µl of SureSelect Block mix. The plate was placed in a thermocycler and the following programme was run:

95°C	5 min
65°C	hold

While the plate was maintained at 65°C, 40µl of hybridisation buffer was added to each well of row “A” of the plate and kept at 65°C for a minimum of 5 min. Next 7µl of the capture library was added to each well of row “C” and incubated at 65°C for 2 min. While maintaining the plate temperature at 65°C; 13µl of hybridisation buffer from row “A” was added to the SureSelect capture library in row “C” and mixed. Next the entire library from row “B” was added to the hybridisation solution in row “C”. The plate was sealed with strip caps and incubated for 24 h at 65°C with a heated lid at 105°C.

For each hybridisation, 50µl of re-suspended Dynabeads MyOne Streptavidin T1 was added to a 1.5ml LoBind tube (Eppendorf) and washed with 200µl SureSelect Binding Buffer, before being re-suspended in 200µl SureSelect Binding Buffer.

After the 24 h incubation the volume of sample was estimated, and the same volume of hybridisation bead solution was added. The hybrid capture/bead solution was incubated on a Nutator (40 rev/min) for 30 min at RT. The tube was briefly centrifuged before being placed on a magnetic stand to separate the beads from solution. The supernatant was removed and discarded, and the beads re-suspended in 500µl SureSelect Wash 1. Samples were incubated for 15 min at RT before being placed back into the magnetic stand to remove the supernatant. The beads were re-suspended in 500µl of 65°C pre-warmed SureSelect Wash 2 and vortexed to mix. The samples were incubated for 10 min at 65°C and placed in a magnetic stand to remove the supernatant. This washing was carried out for a total of 3 times. Finally, beads were re-suspended in 30µl nuclease-free water.

2.2.7.9 Post-Capture Processing for Multiplexed Sequencing

PCR mixture was prepared as follows:

Reagent	Volume for 1 reaction
Herculase Buffer	10µl
dNTP (2mM)	0.5µl
Primers	1.25µl
Polymerase	1µl
Beads	14µl
H_2O	22µl
Total	50µl

The sample was pipetted to resuspend the beads, placed in a thermal cycler and the following PCR amplification carried out:

98°C	2 min	1 cycle
98°C	30 sec	10 cycles
60°C	30 sec	
72°C	1 min	
72°C	10 min	1 cycle
4°C	hold	

The amplified capture library was purified as outlined in Section 2.2.7.6, except that 70% ethanol was used instead of 80%, and the capture library was eluted into a final volume of 30µl nuclease-free H_2O for sequencing.

2.2.8 *Agrobacterium*-mediated Transient Expression Assay

Five ml liquid cultures of *A. tumefaciens* clones (cognate avirulence genes in chapter 3 or candidate effector genes in chapter 5) (5ml YEB (0.5% beef extract, 0.1% yeast extract, 0.5% peptone, 0.5% sucrose, 10mM $MgSO_4$)) were incubated overnight at 28°C. Cultures were subsequently centrifuged at 3,500rpm, the YEB media removed, and the pellet resuspended in 5ml infiltration buffer (IB) (1M MES, 1M $MgCl_2$, 0.1M acetosyringone). Samples were centrifuged and resuspended a further two times. After the final centrifugation, samples were resuspended in 5ml IB and the optical density (OD_{600nm}) was measured. Samples

were then covered in aluminium foil and left to incubate on a shaking plate at RT for 3h (protocol based on (Kapila et al., 1997)).

After the incubation, samples were diluted to a final 0.5 OD_{600nm} in 100ml of IB (plus 0.002% Silwet L-77). Three leaflets of a single leaf cutting were dipped upside down into the IB and the beaker then placed into a vacuum manifold (Wei et al., 2007). The vacuum was switched on for 30s before the pressure was slowly released. Samples were held under vacuum twice before being removed from the beaker, patted dry and placed into damp paper towel lined plastic boxes (Figure 2.2). Boxes were wrapped in cling film and left in sunlight for a minimum of 6 days (maximum 10 days) before visualisation of a cell death response was recorded.

.

Figure 2.2 Layout of infiltrated leaf material in clear plastic box. Damp paper towels were concertinaed and leaves placed on top. Lids were replaced and the box covered in cling film.

3. Testing the novelty of *H2* using dRenSeq

3.1 Introduction

3.1.1 Resistance from wild potato species

Plants have developed control strategies which allow them to detect and prevent infection from a variety of pathogens, with resistance (*R*) genes playing a key role in this defence response (Jones and Dangl, 2006). However, the majority of established cultivated potato varieties (*Solanum tuberosum* ssp. *tuberosum*) are highly susceptible to many pathogens as they are grown, over many growing seasons, across vast areas of agricultural lands as genetically identical clones. Fortunately, wild relatives of cultivated potato are a rich source of pathogen resistance. Unfortunately, due to genetic barriers (such as differences in endosperm balance number), only a small number of species have the ability to be interbred with varieties directly (Jones et al., 2014).

Wild potato germplasms are maintained in genebanks all over the world; the International Potato Centre (CIP) in Peru, the USDA Potato Genebank in the USA and the Commonwealth Potato Collection (CPC) in Scotland are but a few (Bethke et al., 2017). The CPC contains over 1500 wild and locally adapted landraces comprising over 80 potato species which can be screened for potential new sources of pathogen resistance. One member; the diploid species *S. multidissectum* indigenous to Peru and Bolivia was screened for PCN resistance and was found to contain *H2*; which confers a high level of pathotype-specific *Globodera pallida* resistance (Dunnett, 1961).

Given that resistance to *G. pallida* is limited (*Gpa2* is currently the only major single resistance gene active against it) utilising and preserving major *R* genes is of utmost importance (Castagnone-Sereno, 2002). Furthermore, knowing which resistances are already present in a cultivar allows for more informed crop breeding. Tailoring combinations of effective resistance against nematodes contributes towards durability of resistances and ensures longevity of characterised genes.

3.1.2 The problem with tetraploids

An important point to consider when endeavouring to map resistance genes is the material used. In a simple diploid configuration, the use of parents which are both heterozygotic (Rr) for a given trait, leads to progeny with a segregation ratio of 1:2:1 (RR:Rr:rr), In this scenario, ¾ of the progeny shows the dominant phenotype, and therefore holds at least one copy of the dominant allele, while only ¼ of the progeny will contain the recessive allele(s) only and display a susceptible phenotype. Similarly, crossing a heterozygous resistant plant (Rr) with a homozygous susceptible parent (rr) will result in a simple 1:1 segregation ratio. In contrast, when carrying out similar crosses in a tetraploid environment, the genetics becomes more complex. For example, a cross between two heterozygotic tetraploid parents in a duplex configuration (RRrr), will yield a segregation ratio within the progeny of 1:8:18:8:1 (RRRR:RRRr:RRrr:Rrrr:rrrr) with 35:1 progeny being observed to have the resistant phenotype. When crossing a homozygotic susceptible (rrrr) parent with a heterozygotic simplex (Rrrr) parent to create an F1 progeny, the segregation ratio will conform to a a 1:1 (Rrrr:rrrr) resistant to susceptible ratio.

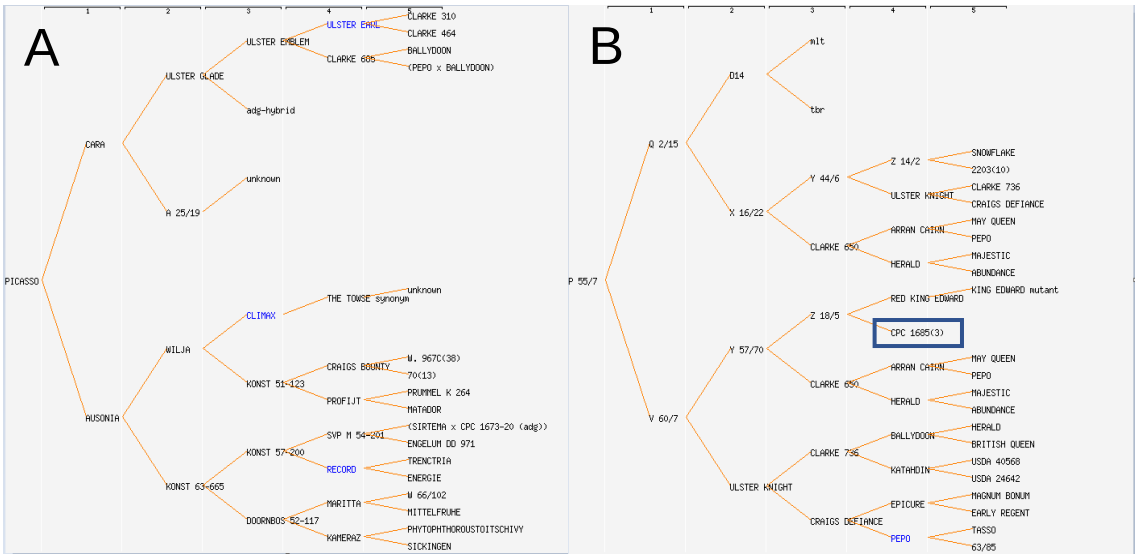


Figure 3.1 Pedigree of the homozygotic susceptible cultivar Picasso (tetraploid) and heterozygotic resistant clone P55/7 (tetraploid). Branches link a cultivar with its' parents. Highlighted within the blue box is *H2*-containing *S. multidissectum* accession CPC1685 (diploid) (SASA, 2018) (https://www.europotato.org/varieties/).

The work presented here focuses on mapping the *H2* resistance gene. In order to map *H2,* an F1 population segregating for *H2* was created from tetraploid parents. Picasso was chosen as the homozygotic susceptible parent. Picasso's direct parents are the Pa1 susceptible varieties Cara and Ausonia (Figure 3.1A). In addition, for at least the previous five generations which were used to breed Picasso, no cultivar was utilised that contains the *H2* gene. The resistant donor used was the breeding clone P55/7 (Figure 3.1B). P55/7 was bred from the wild diploid species *S. multidissectum* from which *H2* originates (denoted at mlt in Figure 3.1B) and is known to contain *H2* resistance. Initially, *S. multidissectum* accession CPC1685 was crossed with variety Red King Edward to yield breeding clone Z18/5 (Dunnett, 1961).

Parental pedigrees have been used by breeders to document the crosses carried out to produce new cultivars. The information which they contain gives insight into the material used to create cultivars. Until now little was known about the genetic make-up of each parent and cross and this proves important when breeding for specific traits. A question arising in this study was: how is it possible to identify new traits if it isn't known what is present to begin with?

3.1.3 Utilising new technologies

The increasing power of genomics means there are new approaches in how to use wild germplasm to our advantage. Breeders have previously focused solely on screening for phenotypic traits, limiting research to phenotypes which can be quantified. Embracing new technology allows for the genetic potential of both wild and cultivated species to be identified and exploited.

As many known resistance genes contain canonical domains associated with NB-LRR genes, it can be hypothesised that the *H2* gene will also follow this pattern allowing for targeted gene enrichment techniques to be utilised during mapping. Jupe *et al* (2013) designed and implemented a gene enrichment method specifically targeting NB-LRRs, termed RenSeq (Resistance gene enrichment Sequencing). This method allowed the annotation of 755 *R* genes within the potato genome, including sequence and loci position information. The method targets all NB-LRR-encoding genes which can be isolated through the use of

biotinylated probes, decreasing the number of genes sequenced from ~39,000 to 755, which in turn permits greater read-depth coverage of the target genes during re-sequencing (Jupe et al., 2013).

The sequence data produced by targeted enrichment can be mined to identify the presence of known pathogen *R* genes in a process known as dRenSeq (diagnostic RenSeq) (Van Weymers et al., 2016; Chen et al., 2018; Rui et al., 2018 and Armstrong et al., 2018). During dRenSeq, query sequences are screened for the presence of 29 known *R* genes which confer resistance to nematodes, *Phytophthora infestans*, bacteria, and viruses to test for their presence. By identifying the functional resistance genes which are already present in a given cultivar or population, it becomes easier to determine whether the resistance phenotype observed is controlled by a hitherto unknown *R* gene.

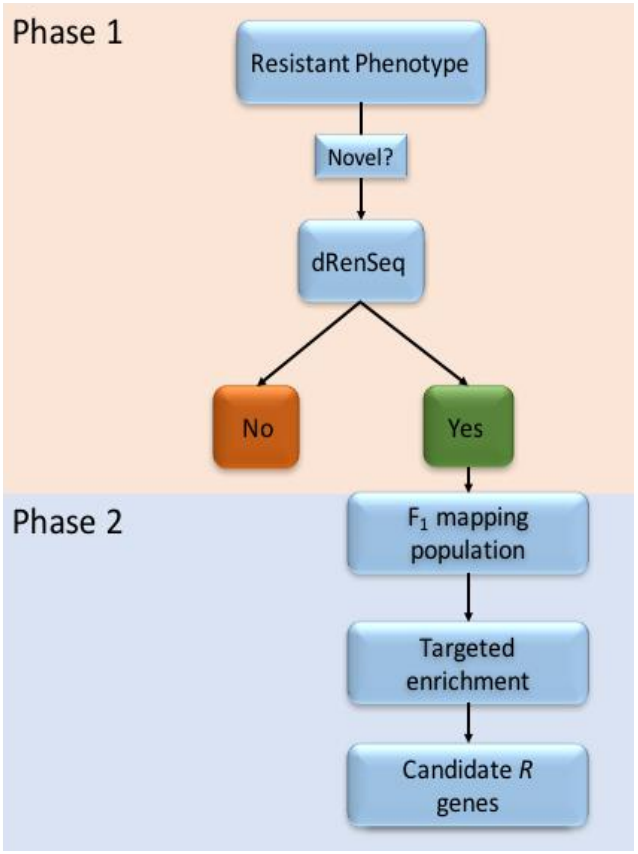


Figure 3.2 Graphical representation of the pipeline which takes known/unknown resistance traits through to a mapped genome position. Phase 1: A resistance phenotype is observed in a wild or cultivated species. The novelty of this resistance can then be verified through dRenSeq mapping. If the resistance is novel it can then be crossed with a susceptible cultivar to create a mapping population (Phase 2). The F1 progeny can then be used in GenSeq/RenSeq targeted enrichment sequencing to identify a chromosomal location as well as putative candidates within the region.

Since the development of dRenSeq, determining the novelty of a resistance is the first step in mapping and cloning new genes.

The pipeline outlined in Figure 3.2 details the steps required to successfully identify and map novel sources of resistance. Steps shown in phase 1 will be tested and discussed in this chapter, while phase 2 will be carried out in chapter 4: *genome mapping and fine mapping of H2 using an F1 segregating population*.

3.2 Aims

The aim of the chapter was to prove the novelty of *H2* as a major effective nematode resistance gene. The specific aims were:

a) To determine the known functional *R* gene(s) present in Picasso and/or P55/7 through the utilisation of dRenSeq technology
b) To validate the outcome of the dRenSeq experiment through challenging of parental plants in order to elicit an *R* gene mediated cell-death response through *Agrobacterium*-mediated transient expression assays

3.3 Materials and Methods

3.3.1 Pa1 infection assay

Infection assays were set-up as detailed in section 2.1.4. Tubers for cultivars Picasso and P55/7 were tested for their susceptibility/resistance to *G. pallida* pathotype Pa1.

3.3.2 Statistical analysis of infection data

Raw female count data was inputted into GraphPad (Prism) software to generate graphs and undertake statistical analysis. The standard deviation, standard error, average, and p-values were calculated for both cultivars.

3.3.3 dRenSeq

R genes enriched from Picasso and P55/7 gDNA (as section 2.2.7 and 4.3.6) were sequenced on Illumina MiSeq and reads were mapped as outlined in Armstrong *et al* (2018) (doi: https://doi.org/10.1101/360644) against a bespoke reference set of functionally validated *R* genes. Raw reads were trimmed using cutadapt (Martin, 2011) version 1.9.1 to a minimum length of 100bp. Trimmed reads were mapped to the reference *R* gene set (Table 3.1) using bowtie2 (Langmead and Salzberg, 2012) version 2.2.1 in very-sensitive end-to-end mode. The score-min parameter used was L, -0.01,-0.01 resulting in a mismatch penalty of 5 per 250bp, meaning that only reads which were identical to the reference set were mapped. Because of the high nucleotide similarity of NB-LRR sequences, reads were allowed to map up to 10 positions (-k 10). The produced bam file was aligned and indexed using samtools (Li et al., 2009) v1.3.1. Read depth and coverage was calculated using BEDtools (Quinlan and Hall, 2010). Data was subsequently transformed and plotted using a custom script in R Studio (Team, 2015).

3.3.4 *Agrobacterium*-mediated Transient Expression Assay

Samples were prepared and assayed as in section 2.2.8. *A. tumefaciens*-transformed RBP1, Avr2, Avr3a^{KI}, Avr3b, AvrRx, CRN2, and empty vector tagged eGFP were grown in liquid cultures (5ml + Rifamycin (25mg/ml)) overnight at 28°C. Young leaf material for both Picasso and P55/7 was harvested immediately prior to being vacuum infiltrated. Each *R*::Avr pair was tested in three independent replicates, with three leaflets being tested each time.

Table 3.1 List of reference genes which were used during the dRenSeq mapping (Armstrong et al., 2018). A GenBank identifier is given if one exists

Gene	GenBank Identifier	Reference
R1	GenBank: AF447489.1	(Ballvora et al., 2002)
R2	GenBank: FJ536325.1	(Lokossou et al., 2009)
R2-like	GenBank: FJ536323.1	(Lokossou et al., 2009)
R3a	GenBank: AY849382.1	(Huang et al., 2005)
R3b	GenBank: JF900492.1	(Li et al., 2011)
R3b$^{G1696/G3111}$		Armstrong et al (2018)
Rpi-sto1	GenBank: EU884421.1	(Vleeshouwers et al., 2008)
Rpi-pta1	GenBank: EU884422.1	(Vleeshouwers et al., 2008)
Rpi- blb1	GenBank: AY426259.1	(Van Der Vossen et al., 2003)
Rpi-blb2	GenBank: DQ122125.1	(Vossen et al., 2005)
Rpi-blb3	GenBank: FJ536346.1	(Lokossou et al., 2009)
Rpi-abpt	GenBank: FJ536324.1	(Lokossou et al., 2009)
Rpi-vnt1.1	GenBank: FJ423044.1	(Foster et al., 2009)
Rpi-vnt1.3	GenBank: FJ423046	(Foster et al., 2009)
Rpi-amr3	GenBank: KT373889	(Witek et al., 2016)
R8	GenBank: KU530153	(Vossen et al., 2016)
R9	https://www.google.com/patents/US20140041072	(Jo et al., 2015)
Rpi-ber		
Rpi-Ph3	GenBank: KJ563933.1	(Zhang et al., 2014)
Rx	GenBank: AJ011801.1	(Bendahmane et al., 1999)
Rpi-chc		
Rpi-Mcq1.1		
Rpi-tar1		
Mi1.1	GenBank: AF039681	(Milligan et al., 1998)
Mi1.2	GenBank: AF039682	(Milligan et al., 1998)
Gpa2	GenBank: AF195939.1	(Van Der Vossen et al., 2000)
Gpa2$^{\Delta C2992}$		Armstrong et al 2018
Gro1.4	GenBank: AY196151	(Paal et al., 2004)
Hero	GenBank: AJ457051	(Ganal et al., 1995)

3.4 Results

3.4.1 Cultivar P55/7 shows resistance to Pa1 while Picasso does not

Roots of Picasso and P55/7 plants were inoculated with Pa1 cysts and left for 8 weeks before the total number of females present on the root systems were counted (Supplementary Table 1). Four plants of each cultivar had three replicates and two independent screens to test their resistance/susceptibility to *G. pallida* Pa1 (Picasso had 21 out of 24 successful replicates, while P55/7 had 20). Picasso had a large spread of infection scores with one replicate having 54 females present, while another replicate had only 2 females, however the mean infection score for the cultivar was 17. P55/7 had a much narrower range on infection scores with a mean infection score of 0.65 females per replicate (Figure 3.3). The error bars were calculated based on the standard deviation for Picasso (SD=15.11) and P55/7 (SD=1.04) with a p-value <0.001 making the scores statistically significant, Picasso is susceptible to Pa1 while P55/7 is resistant.

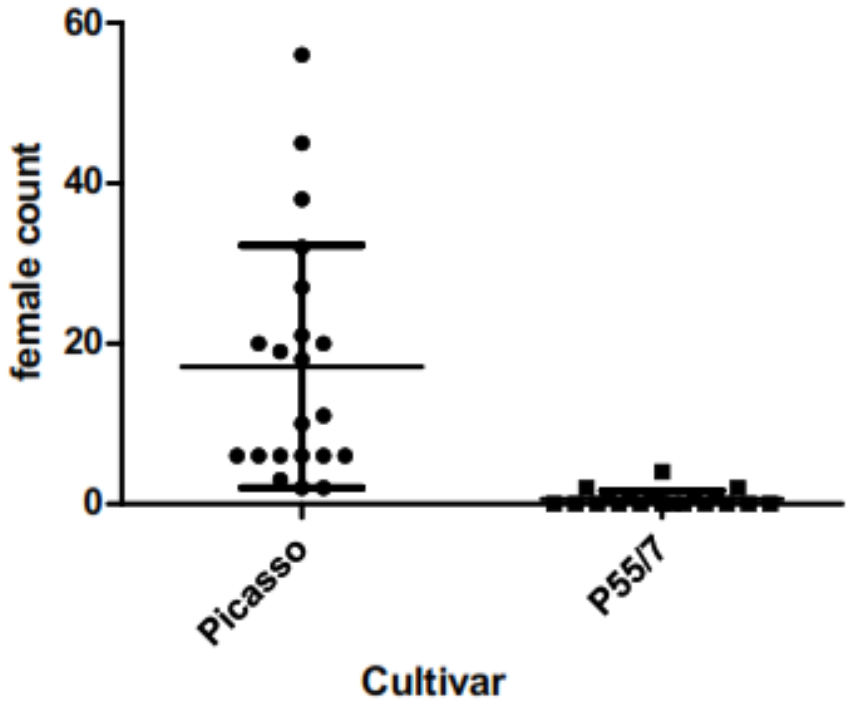


Figure 3.3 Graph of Pa1 scores from meristem infection for Picasso and P55/7. Error bars were calculated using the standard deviation, Picasso (SD=15.11) and P55/7 (SD=1.04), and a p-value of <0.001.

3.4.2 dRenSeq identifies *R* genes present in the susceptible parent Picasso

To evaluate if the resistance phenotype in the resistant clone P55/7 and segregating within the F1 progeny could be associated with an already characterised *R* gene, diagnostic RenSeq (dRenSeq) analysis was conducted. RenSeq enriched paired-end reads were mapped against a panel of known functional NB-LRRs including the nematode resistance genes *Gpa2, Gro1.4, Hero* and *Mi1.1/1.2* to assess their presence within Picasso and P55/7. For a gene to be designated as 'present' within a cultivar the sequence had to match 100% with the reference gene coding sequence. DRenSeq analysis revealed the presence of *G. pallida R* gene *Gpa2*, *P. infestans R* genes *R1, R3a* and *R3b*, and the viral PVX *R* gene *Rx* in Picasso; all whose presence was previously unknown. In comparison, no previously published functional *R* genes were revealed to be present in P55/7.

Results in Figure 3.4 give a visual representation as to how the reads mapped to each of the reference genes. As an example, in the case of *Gpa2* which is present in Picasso, read depth ranges between 10-100 across the whole length of the gene, with 100% of the gene being covered. By comparison, P55/7 which doesn't have a functional copy has variable levels of read depth with less than 25% of the gene being covered against the reference sequence. Some genes have sections of sequence similarity (*Rpi-blb1, Rpi-R8, Rpi-R9*) but have been designated as not present; this is due to the repetitive nature of NB-LRR genes, and the allowance of sequences to locate to 10 separate positions during mapping which can cause peaks to appear where small areas of sequence similarity are identified. Accepting only genes with 100% sequence coverage as present and functional revealed the presence of five functional *R* genes in susceptible Picasso, while no previously known functional *R* genes were identified in P55/7, giving a strong indication that *H2* is a discrete *R* gene which has not been previously described and is controlling the observed resistance phenotype.

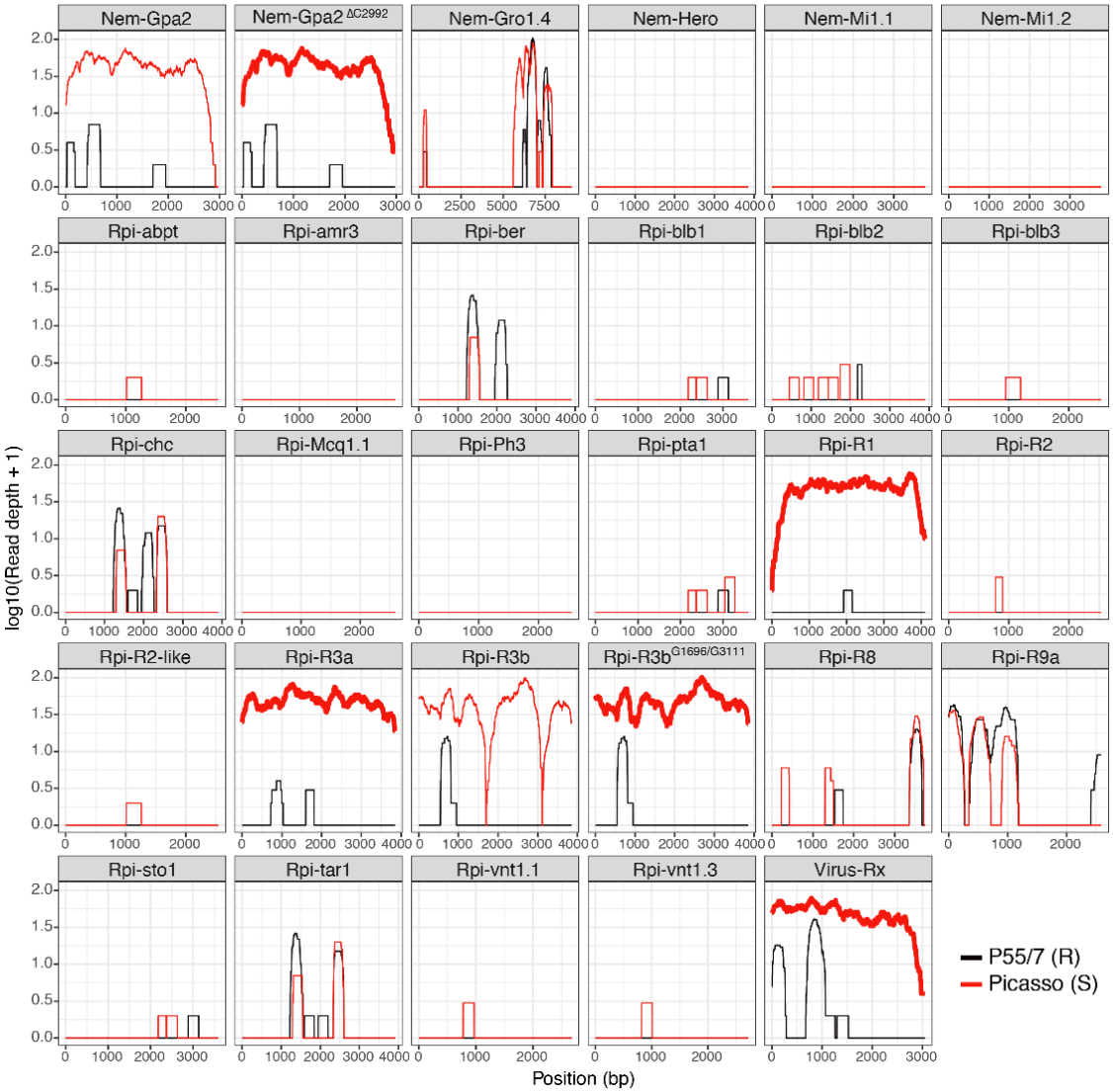


Figure 3.4 dRenSeq analysis of parent cultivars Picasso (red) and P55/7 (black). Each trace is representative of a known NB-LRR in a given sample. The x-axis displays the coding sequence of the gene and the y-axis is a $\log_{10}$ scale of read coverage. NLRs with full coding sequence representation are in bold.

Interestingly, the form of *R3b* which is present in Picasso, and has the ability to recognise *Avr3b* is not the published form of the gene. The identified copy of *R3b* contains within it

two non-synonymous mutations *R3b*$^{G1696/G3111}$ (Armstrong *et al,* 2018). These mutations have been verified as correct and present through sequencing and functional testing work undertaken by another student (Strachan *et al* (submitted)). This type of allele mining which dRenSeq allows, highlights its suitability as a tool for *R* gene identification and validation.

3.4.2 Transient expression proves presence of functional *R* genes

dRenSeq analysis identified the presence of five functional *R* genes in Picasso, while no already characterised *R* genes were identified in P55/7. To verify these findings, transient expression via vacuum infiltration was undertaken to prove the functionality of the identified genes. Cognate effectors for each resistance (*R3a*::Avr3a^{KI} (Bos et al., 2009), *R3b*::Avr3b (Li et al., 2011), *Rx*::Cp (Bendahmane et al., 1995) and *Gpa2*::RBP1 (Sacco et al., 2009)) were used as well as negative controls eGFP and Avr2 which should not give a response as *R2* is not present (Gilroy et al., 2011), and positive control CRN2 (an effector known to induce cell death) (Haas et al., 2009). Unfortunately *R1*::Avr1 could not be tested during this experiment, and so the functionality of *R1* could not be verified.

R gene-mediated cell death symptoms started to become visible after six days, however results became more pronounced 10 days post infection and so this time point was used to analyse the experiment. Picasso leaves showed varying symptoms of cell death with the *Rx*-mediated cell death showing complete necrosis across the entire leaf, while the eGFP control showed no signs of cell death or necrosis. Cognate effectors Avr3a, Avr3b and RBP1 showed cell death, while Avr2 did not show signs of necrosis (Figure 3.5). Elicitation of the target *R3a*, *R3b*, *Rx* and *Gpa2 R* genes proves both the presence and functionality of these genes within susceptible Picasso.

In comparison, transient expression of the same cognate effectors in P55/7 did not elicit any visible cell death response (Figure 3.5). After 10 days of incubation no *R* gene-mediated symptoms could be visualised and so it can be accepted that the dRenSeq analysis which identified that no currently characterised *R* genes are present in resistant P55/7 is correct. It could be argued that a lack of a cell death response in P55/7 is due to insufficient or no bacteria being successfully infiltrated into the leaf. However, the positive cell death response which is elicited when CRN2 is infiltrated (Torto et al., 2003) displays that it is a lack of the cognate *R* gene rather than an inability to infiltrate P55/7 which is causing no cell death symptoms to be observed.

Using the results from the transient expression experiment confirms the dRenSeq analysis that five characterised *R* genes are present in Picasso, while none of these are present in P55/7. These results also confirm that *Gpa2* (present in variety Picasso) does not control *G. pallida* pathotype Pa1 in P55/7 and the resistance phenotype is indeed mediated by the *H2 R* gene rather than a previously characterised *R* gene.

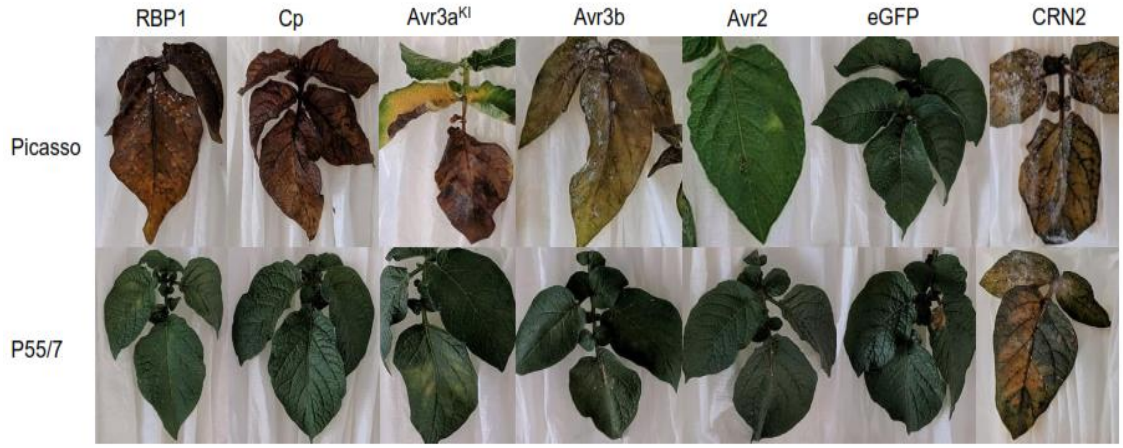


Figure 3.5 Vacuum infiltrated Picasso (A) and P55/7 (B) potato leaves. Cognate effectors Avr3a, Avr3b, Cp, and RBP1 were infiltrated in order to visualise a cell death response if the corresponding *R* gene was present. eGFP was used as a negative control, while Avr2 was used as a positive control, and CRN2 was used to demonstrate that P55/7 can be infiltrated. Samples were infiltrated and left for 10 days before results were analysed.

3.5 Discussion

3.5.1 Utilisation of dRenSeq reveals the presence of *R* genes which were previously unknown

Using dRenSeq mapping technology allowed for the identification of known *R* genes present in susceptible Picasso and resistant P55/7 potato cultivars. The mapping revealed the presence of 5 full length *R* genes in Picasso and no full-length *R* genes in resistant P55/7. The utilisation of dRenSeq confirmed that *H2* resistance phenotype cannot be explained by another previously identified resistance gene including those associated with viruses, nematodes, or oomycetes.

One of the genes identified in susceptible Picasso was the nematode resistance gene *Gpa2*, known to confer resistance to pathotype Pa2/3 populations. The presence of this gene in the susceptible background could go some way to explaining the varying levels of Picasso's susceptibility observed during the Pa1 infection assay (Figure 3.3). A proportion of the plants exhibited a Pa1 susceptible response, as was expected, while several plants showed low infection scores of only 6-8 females. It could be argued that this result is due to cysts not hatching, or poor root growth which could not sustain a successful Pa1 infection. However, as this lower level of infection happened over several plants it could be hypothesised lower infection scores were observed due to *Gpa2* having a minor resistance effect on Picasso.

3.5.2 Functional testing confirms the presence of functional *R* genes

Transient expression in potato is limited due to the structure of the leaf. Syringe infiltration of transformed bacterial suspensions is routinely undertaken in *N. benthamiana* by infiltrating via syringe into the intercellular space between veins (Yang et al., 2000). However, the small intercellular space present in potato leaves, as well as the presence of leaf surface hairs makes it difficult to infiltrate any volume of bacterial suspension. To overcome some of the problems faced during syringe infiltration, PVX-mediated transient expression was developed (Takken et al., 2000). Using a viral PVX vector was hypothesised to help spread a target construct through the leaf cells in order to help with expression. However, this technique also created problems during initial testing of the method, and so was abandoned for use during the functional testing experiment. The presence of the *Rx R* gene, which confers resistance to PVX, in Picasso means the technique was rendered useless as any cell death symptoms which may have been observed could have been due to the presence of the PVX vector, rather than the target construct which would have led to many false positive results.

Problems encountered during syringe agroinfiltration, and the inability to use PVX-mediated toothpick inoculation due to the presence of *Rx* led to the adaption of the Wei *et al* (2007) vacuum infiltration protocol for *Pseudomonas syringae* infection of *N. benthamiana*. Results in Figure 3.5 display the success of the

technique in delivering the target construct to the cell without causing any off-target damage, as can be seen in the healthy eGFP construct plants. The lack of cell death symptoms in the eGFP leaves highlights that the necrosis observed is due to the cognate *R*::Avr interactions rather than the leaves being subjected to a vacuum.

Utilising dRenSeq and transient expression techniques allowed for *R* genes to be identified in Picasso which were not previously known to be present. It also revealed that P55/7 did not have any of the currently characterised *R* genes. Knowing which forms of resistance are present in both Picasso and P55/7 ensures that the gene which controls the resistance phenotype and is being mapped is indeed a new and uncharacterised *R* gene and hasn't been mistaken for an existing resistance source. The lack of P55/7 RenSeq reads which map to the reference gene set is a strong indication that *H2* is indeed the gene conferring resistance to *G. pallida* Pa1.

4. Genome mapping and fine mapping of *H2* using an F1 segregating population

4.1 Introduction

Potatoes (*Solanum tuberosum* ssp. *tuberosum*) originate from South America and were brought into Europe through a limited number of introductions. Because of this, cultivated potatoes grown outside South America have a narrow genetic base (Bryan et al., 2004) and do not contain the genes required to mount a resistance response to most pathogens which are encountered. However, resistance to many pathogens, including nematodes, viruses, oomycetes and fungi, is present in wild relatives of potato from where it can be introgressed into commercial cultivars. Most cloned and functional resistance genes are all members of the STAND (Signal Transduction ATPase with Numerous Domains) protein family, and so typically contain NB-LRR domains (Lukasik and Takken, 2009) (see chapter *1.5.4 NB-LRRs*). As almost all known resistance genes are characterised by these domains, efforts are focused on this gene family to identify novel pathogen resistances as well as in the mapping and cloning of already known resistances.

Many sources of resistance to PCN have been identified in *Solanaceous* species. The *Gro1* resistance gene, on chromosome 7, which confers resistance to *G. rostochiensis* was the first gene to be mapped in potato using RFLPs (Restriction Fragment Length Polymorphism) (Gebhardt and Valkonen, 2001). Since then, 13 other PCN resistance gene loci have been mapped to eight linkage groups on chromosomes 3, 4, 5, 7, 9, 10, 11 and 12 (Finkers-Tomczak et al., 2009, Finkers-Tomczak et al., 2011, Bakker et al., 2011). In addition, *Gpa2* from *S. tuberosum* ssp. *andigena* has been mapped to chromosome 12 and subsequently cloned. (van der Voort et al., 1999, Van Der Vossen et al., 2000). In addition to Chromosome 4, 9 and 11, chromosome 5 is known as an a 'hot-spot' for pathogen resistance, and contains at least five *R* gene loci from which functional genes have arisen; *R1* which confers resistance to *Phytophthora infestans, Nb,*

and *Rx2* conferring resistance to Potato Virus X, and *Grp1* mediating resistance to *G. rostochiensis* and *G. pallida* (van der Voort et al., 1999).

The first major PCN resistance gene identified was the *H1* gene from the wild species *S. tuberosum* ssp. *andigena* CPC1673, which was introgressed into cultivated potato, and since then has been bred into a wide range of cultivars (Ellenby, 1954, Gebhardt et al., 1993). Deployment of *H1*-containing cultivars has been highly successful in controlling *G. rostochiensi*s infestations, but in the UK, it has led to a shift in species so that *G. pallida* is now predominant. Since multiple *G. pallida* pathotypes are present in British fields, most potato varieties do not contain sufficiently high levels of resistance for sustained control and suppression.

During attempts to identify robust sources of *G. pallida* resistance, the *H2* resistance from the wild species *Solanum multidissectum* was found to provide very high levels of suppression of multiplication of the Pa1 pathotype (Dunnett, 1961, Kreike et al., 1994). This pathotype is not widely spread in European fields and the *H2* source has therefore not been utilised in breeding programmes. However, *H2* resistance does confer a certain degree of resistance to other *G. pallida* pathotypes (Blok and Phillips, 2012) and thus should be considered for use in resistance gene pyramiding efforts. In addition, recent work using a mitochondrial DNA marker has shown that the diversity of *G. pallida* populations in Scottish potato fields (Figure 4.1) may be broader than previously thought, and that Pa1 populations may be more abundant than previously anticipated (Eves-van den Akker et al., 2015).

Due to the time required to introgress resistance into commercial cultivars, the emergence of virulent pathogen populations that are able to overcome deployed resistances is a serious concern (Castagnone-Sereno, 2002). The advantage of pyramiding several resistance genes into a single cultivar is that resistances are likely to be more durable (Mundt, 2014).

Figure 4.1 Distribution of Scottish *G. pallida* populations based on mitochondrial data. Type 1 (blue) and Type 3 (red) depict Pa2/3 types, while Type 2 (green) corresponds to Pa1 (Eves-van den Akker et al., 2015).

As no single gene has been discovered that confers complete resistance in potato to all European pathotypes of *G. pallida* (Pa1, Pa2/3), quantitative resistance is being used to develop potato crops that are protected from the mixed populations which are present in British fields. Breeding with quantitative resistance, which may be comprised of a mixture of several (*R*) genes including those with a minor effect, is unfavourable due to the complexities in incorporating multiple genes simultaneously into commercial cultivars. However, relying on a major *R* gene, which triggers a cell death response when the corresponding nematode avirulence (*avr*) gene is present, may not be durable (van der Voort et al., 2000a). Thus, pyramiding multiple major *R* genes or components of quantitative resistances with major effects into a single commercial cultivar is the approach now used to ensure that broad-spectrum resistance is more durable. With regard to PCN resistance, the *H2* resistance is a promising candidate due to its strong resistance phenotype against *G. pallida* Pa1 and partial resistance to Pa2/3. In order to take full advantage of its potential; the position and genetic makeup of the *H2* resistance must first be understood. To this effect, I could demonstrate through dRenSeq in the previous chapter, that the *H2* resistance cannot be explained by any currently cloned *R* gene. This warrants a detailed genetic study to ascertain a) the map position of the gene and b) the segregation ratio in a population as this infers the genetic composition.

4.2 Aims

The aim of this chapter was to determine the location of the *H2* resistance gene using established progeny of a Picasso x P55/7 cross (rrrr x Rrrr). The specific aims were:

a) Identify individual progeny from an F1 mapping population which showed a defined resistant or susceptible phenotype

b) Create pooled sequencing libraries which are representative of the genetic background of resistant and susceptible F1 progeny plants

c) Identify informative SNPs which can successfully differentiate between resistant and susceptible alleles

d) Determine the chromosomal location of the *H2* gene

e) Analyse the *H2* candidate gene(s) and compare with related *Solanaceous* species

f) Identify the functional *H2* gene

4.3 Materials and Methods

4.3.1 Cross between Picasso x P55/7

A cross between the homozygous susceptible cultivar Picasso (rrrr) with the heterozygous resistant P55/7 (Rrrr) yielded seed from which an initial population of 192 progeny plants was used for the first infection assay screen. Seed from this cross was subsequently used to grow an expanded population of an additional 656 plants to identify recombinants in the *H2* locus and fine map the resistance. For each F1 progeny, a parent clone was maintained for the production of apical stem cuttings. The cuttings were used for PCN infection assays and DNA extractions for KASP marker analysis. Recombinant progeny clones which were taken for further analysis were left to tuberise, and the tubers collected and stored at 4ºC.

4.3.2 PCN Infection Assay

4.3.2.1 Stem Cuttings

Progeny clones and parental cultivars were screened as detailed in section *2.1.4.1 meristem cutting assays*.

4.3.2.2 Tuber Pieces

Racks of root trainers were set-up as in Section 4.3.2.1 and infected with 15±2 Pa1 cysts. Progeny tubers were taken from the cold store one week prior to planting to encourage sprouting. One-centimetre square pieces of tuber were cut around the sprout and planted into infected wells. Plants were left to grow for eight weeks before root trainers were opened to count all females present on root systems.

4.3.3 Scoring of Females

At 8 weeks, root trainers were opened to expose the root systems. All visible females were counted from all sides of the root system, and the total number of visible females was recorded. Average scores were taken from the three technical replicates, and this average female count was used to determine whether a plant was classed as either susceptible or resistant. Two independent replicates were scored, and the data combined.

4.3.4 Bulk Segregant Analysis

The twenty most resistant (average ≤1 female) and most susceptible (average ≥18) progeny plants from the second independent infection assay were taken forward for further genetic analysis. Definition of a plant as either resistant or susceptible was based on the scores obtained for the resistant (P55/7) and susceptible (Picasso) parent plants. Young leaves were collected from each plant and stored in individual, labelled plastic bags at -80°C until required.

4.3.5 DNA Extraction

4.3.5.1 CTAB Extraction of Parental and Progeny Bulks

DNA from parents and progeny was extracted as follows. All centrifuge steps were done at 4°C unless otherwise stated. One cm round samples of leaf material were flash frozen in liquid nitrogen and ground with a micro pestle in a 2ml microcentrifuge tube. Next, 1ml of CTAB buffer (H_2O, Tris 1M (pH 8), EDTA 0.5M (pH 8), $NaCl_2$ (1M), CTAB (1.5%), β-mercaptoethanol) was added and samples were incubated in a 50°C water bath for 1h. Samples were centrifuged at 14,000rpm for 2 min to pellet the leaf material, the supernatant was transferred to a fresh 2ml microcentrifuge tube and 625µl chloroform/IAA (24:1) was added. Samples were centrifuged for 10 min at 14,000rpm and then the upper phase was pipetted into a fresh 1.5ml tube. Next, 600µl ice cold isopropanol was added and the samples incubated on ice for 10 min. Samples were centrifuged for 10 min at 14,000rpm to pellet the DNA before the supernatant was discarded and the pellet was washed with 400µl 70% ethanol. Samples were centrifuged for 5 min, the supernatant was removed, and the pellet air dried. Finally, the DNA was re-suspended in 100µl TE elution buffer (Qiagen) and quantified via Nanodrop (ThermoFisher). After extraction, DNA from each individual sample was normalised to a total of 50ng, and pooled to yield four samples; resistant progeny, susceptible progeny, resistant parent, and susceptible parent with each sample containing 1µg of total DNA.

4.3.5.2 Automated Extraction of Expanded Population

For the second round of KASP Marker Assays an expanded population of 656 F1 progeny plants were grown and DNA was extracted using the Qiagen QIAamp QIAcube HT Kit. Samples were prepared as outlined in section *2.2.1.2 DNA extraction of plant material*.

4.3.6 Illumina Library Preparation and Probes

Samples were prepared as outlined in section *2.2.7 Gene Enrichment and Sequencing Library Preparation* based on (Jupe et al., 2014). Two separate enrichments were carried out on the pooled DNA; the first was a genome-wide

enrichment, hereafter termed as GenSeq, while the second enriched solely for *R* genes, hereafter termed as RenSeq (Jupe et al., 2013). Probe library information for both GenSeq and RenSeq can be found at http://solanum.hutton.ac.uk/.

4.3.7 Read Mapping and SNP Calling

Bioinformatic analysis was carried out by members of the ICS department at The James Hutton Institute. Paired-end Illumina MiSeq reads were checked with FastQC and then quality and adapter trimmed with fastq-mcf to a minimum base quality of 20. The trimmed reads were then mapped to the potato (DM) reference genome using Bowtie2. Mapping was run at mismatch rates of 2%, 3%, and 5%. The BAM files for both the parents and bulks were sorted, merged and indexed using SAMtools and run through VarScan for variant calling.

4.3.8 Read coverage and on target estimation

Percentage of on target reads was calculated as the proportion of reads mapping to annotated GenSeq and RenSeq target regions within the DM reference genome. Intersecting these regions (+/- 1000bp) against the mapped reads using BEDTools gave the number of on-target reads. These on target reads were then taken as a proportion of total mapped reads. Read coverage to target regions was calculated by dividing the total number of GenSeq gene mapped base pairs (+/- 1000bp) by the total length of the GenSeq gene (+2000bp per gene).

4.3.9 SNP Filtering and KASP Markers

SNPs were filtered using custom Java code to retain informative SNPs present in both parental and progeny bulks. SNPs were filtered based on expected allele ratios for susceptible/resistant (susceptible: rrrr, resistant: Rrrr). For a SNP to be retained it required a minimum read coverage of 50 and an alternate allele ratio reflective of the genotype (0-5% alternate allele for susceptible and 20-30% allele for resistant, or 95-100% alternate allele for susceptible and 70-80% alternate allele for the resistant). BEDTools intersect was used to extract SNPs present in both parental and progeny bulks. The number of parental, bulk and informative SNPs were plotted in 1Mb bins across each chromosome.

KASP markers were designed against informative chromosome 5 SNPs identified in the different mismatch data sets. Sequence 50bp upstream of every SNP was extracted and the 51bp sequences (with the SNP at the 3' end) were used in a MEGABLAST against the DM genome v4.03 via the BLAST+ command line application (Camacho et al., 2009) at default settings. In total, 11 selected SNPs had no off-target BLAST hits back to the DM genome (defined as >95% sequence identity over at least 28 bp)and were used for KASP Marker synthesis (Table 4.1).

SNP markers were designed using the parameters and protocol put forward by LGC Genomics. DNA was extracted from young leaf material from the individuals used for the MiSeq sequencing (as in section 4.3.5.1), and diluted to a concentration of 20ng/µl. DNA was mixed with the KASP reagent and primer mix and run on a StepOne Plus (ThermoFisher) using the following parameters; 2 min at 20°C, 10 cycles of 15 min at 94°C, 20 sec at 94°C, 1 min at 62°C (decreasing by 0.7°C per cycle), 32 cycles of 20 sec at 94°C, 1 min at 55°C, and 2 min at 20°C. To tighten clusters and to try and rectify outliers, a recycling step was carried out on certain samples as follows: 3 cycles of 20 sec at 94°C, 60 sec at 57°C, and 1 min at 37°.

4.3.10 Graphical Genotypes

Results from the KASP marker allelic distribution plots were transformed into a tabulated format based on the parent allelic calls. Any progeny plant which clustered with the resistant parent, P55/7, was labelled as containing a resistant genotype (1; green), and every plant that clustered with the susceptible parent, Picasso, denoted as susceptible (0; red).

Table 4.1 KASP Marker information : Shown is whether the marker is based on GenSeq or RenSeq analysis (column 1), which mismatch rate the SNP was identified (column 2), whether the marker could successfully discriminate between alleles (column 3) the marker name as it will be referenced throughout the chapter and which reflects the position of the SNP in relation to DM chromosome 5 (column 4,) the gene ID where the SNP is located (column 5), and the sequence information to generate each KASP marker (column 6-8). Markers are arranged in chromosomal order, with the final 7-8 digits at the end of the marker name carrying information as to where the SNP is located on chromosome 5 potato genotype DM.

Marker designed from data set	Mismatch rate	Successful	Marker Name	Gene Name	Primer_AlleleFAM	Primer_AlleleHEX	Primer_Common	AlleleFAM	AlleleHEX	CG%_FAM	CG%_HEX	CG%_Common
GenSeq	3%	Yes	ST04_03ch05_1416331	PGSC0003DMG400025119	GAGGAGATGGAGGACAAAAGTTG	AGGAGGAGATGGAGGACAAAAGTTA	CCTTCAAGGTTCTTTTTGGAAGATCTGAA	G	A	47.8	44	37.9
GenSeq	3%	No	ST4_03ch05_1437439	PGSC0003DMG400025121	ACTTTATTCAATTAGGTCAATGACAAGAAC	ACTTTATTCAATTAGGTCAATGACAAGAAG	GTCGAAGATCTGAGCTCTGCTGTTT	C	G	30	30	48
GenSeq	3%	No	ST04_03ch05_1437827	PGSC0003DMG400025121	GAGAAGGGATGGGACTTGAC	GCTGAGAAGGGATGGGACTTGAA	CAACCACCTTTCCAAATCCTGGCAA	G	T	54.2	50	48
GenSeq	3%	Yes	ST04_03ch05_1438531	PGSC0003DMG400025121	ACCACCTGAAACTCCCATCCCT	ACCACCTGAAACTCCCATCCCA	AAGAAATGGCTGAGTTAGGTCTTCGTTAT	T	A	54.5	54.5	37.9
RenSeq	3%	Yes	ST04_03ch05_1503657	PGSC0003DMG400025099	CGAGATAACATCTAGATGAGGAGG	CGAGATAACATCTAGATGAGGAGA	GAGGCAATGGACAAAATAAAGAAGCAGAT	C	T	44	37	37.9
RenSeq	3%	Yes	ST04_03ch05_2202842	RDC0001NLR0076	CAAGTAGCTCCCAACTTTCATCTTC	CAAGTAGCTCCCAACTTTCATCTTT	GTGGCTAAGTGTTGTAATGATAAACCTCAT	G	A	44.4	38.5	36.7
GenSeq	3%	Yes	ST04_03ch05_3000757	PGSC0003DMG400014571	TATAAAAATTGATCCTAATAGCTTCTGCG	AAATTATAAAAATTGATCCTAATAGCTTCTGCA	CATCAAATCGTTAGGATTAGAACATAGCAA	C	T	30	22.9	33.3
GenSeq	3%	Yes	ST04_03ch05_4491040	PGSC0003DMG400018405	AGCCTTCAGAACTCAGACGAAAACT	CCTTCAGAACTCAGACGAAAACC	CGAGCTTTATGTTGTGGCGATTAATACAA	A	G	44	47.8	37.9
GenSeq	5%	Yes	ST04_03ch05_4737653	PGSC0003DMG400018411	CCCTTTGATTGATCTTGAGTTATATGTAT	CCCTTTGATTGATCTTGAGTTATATGTAC	GAGTAAAGAAGCATACTTGGCTTCTCAAA	T	C	30	32.3	37.9
RenSeq	5%	No	ST04_03ch05_4909072	RDC0001NLR0078	CAGTAAACTTCCTGCATCAATGTCAG	CAGTAAACTTCCTGCATCAATGTCAT	CTGTCCATATGAATAACTAGARTTTGCAAA	G	T	42.3	38.5	31.7
RenSeq	3%	Yes	ST4_03ch05_5727224	PGSC0003DMG400025611	CTCGATCATATCGTTCAAAGGAGC	CTCGATCATATCGTTCAAAGGAGA	ATTGAGCTAACAAAACATGAACACTTGGTA	G	T	44	37	33.3
GenSeq	2%	Yes	ST04_03ch05_6079232	PGSC0003DMG400017618	ACATCTGGCTTGTTTTGATCAATACTT	CTACATCTGGCTTGTTTTGATCAATACTA	GCTACTCTTGCTAGGTTGAGTACAAA	T	A	34.5	34.5	42.3
RenSeq	3%	Yes	ST4_03ch05_6533705	PGSC0003DMG401022603	CCTTCTCCCCTGCTAATGCCAT	CCTTCTCCCCTGCTAATGCCAA	CGAAGCATTCGCCTGAACTTTTCTCTT	T	A	54.2	52.2	44.4
GenSeq	3%	No	ST4_03ch05_8386459	PGSC0003DMG400030998	AGTGAACACTACAGTCATGAAGTGAAA	GTGAACACTACAGTCATGAAGTGAAG	GATTTCAGCACAGAAAATGTTCAAACTGTT	A	G	35.7	42.3	33.3
RenSeq	3%	Yes	ST4_03ch05_8625384	PGSC0003DMG400013506	CAATGAGGATAAGATACCTCTGACC	ATCAATGAGGATAAGATACCTCTGACA	GAATTGGCAGATAAGTTGCGTAAACTTCTA	C	A	42.3	37	36.7
RenSeq	3%	Yes	ST4_03ch05_9638908	RDC0001NLR0098	GAGAAAATGGGACTTCCATCAGCT	GAGAAAATGGGACTTCCATCAGCA	GGCCTCCTTTATCTCAACAGACTCAA	T	A	46.2	44	46.2
GenSeq	3%	No	ST4_03ch05_10598867	PGSC0003DMG400018598	GGTTATAAGTTGTATTTACATTATCTGTGCA	GTTATAAGTTGTATTTACATTATCTGTGCG	CTGAGAAATAGCTTGCTATTATTGGTTCAA	T	C	28.1	30	33.3
GenSeq	3%	Yes	ST04_03ch05_10615824	PGSC0003DMG400011727	ACACACGGAGGAAACTTCGCC	GACACACGGAGGAAACTTCGCA	CATGTACCTACGATGTCCTATTATCAGTT	G	T	57.1	52.2	37.9
GenSeq	3%	Yes	ST04_03ch05_11253634	PGSC0003DMG400010739	GCAACAACTGCAGGCTGAAT	CTGCAACAACTGCAGGCTGAAC	CCCGAGGTTAATTGAACAATTTCAGACTT	T	C	52.2	54.5	37.9

4.3.11 Cloning candidate genes from resistant P55/7

Primers for candidate genes RDC0001NLR0075 and RDC0001NLR0076 were designed as outlined in section 2.2.6, and target sequences were amplified and purified from extracted P55/7 gDNA as outlined in section 2.2.4.

RDC0001NLR0075	
forward	GTGTCGTTCAATCTTGTGGTC
reverse	AGTGGTGAATGATGCTGAGGA

RDC0001NLR0076	
forward	TTACAATCCAAGATCATGAGGG
reverse	AGTGGTGAATGATGCTGAGGA

Amplified target sequences were then cloned using the Promega pGEM-T Easy system. The following was mixed in a 0.5ml microcentrifuge tube:

Reagent	Volume for 1 reaction
2X Rapid ligation buffer	5µl
pGEM-T Easy vector (50ng)	1µl
purified PCR product	3µl
T4 DNA ligase	1µl

PCR amplicons and the vector were briefly mixed then incubated overnight at 4°C. The next day, samples were transformed into *E. coli* chemically competent JM109 cells (Invitrogen). Two microlitres of each sample was transformed into 40µl of competent cells and placed on ice for 20 min before being heat-shocked at 42°C for 45-50 sec. Next, 850µl of SOC media was added and cells were left to recover for 90 min at 37°C with gentle rocking before 250µl of the transformation was plated onto LB + AIX (Ampicillin resistance with X-Gal to yield blue/white colonies) (100mg/ml) and left to incubate overnight at 37°C.

White colonies were picked and screened to verify the insert using the PCR cycling parameters outlined above. Amplification of recombinants used the gene-specific reverse primer with the M13 forward (5' – TGTAAAACGACGGCCAGT). Positive clones were grown overnight at 37°C in 5ml LB + Amp (100mg/ml) liquid cultures, and then purified using the ThermoScientific GeneJET Miniprep Purification Kit. To Sanger sequence the purified DNA, additional primer pairs were used to span the >3Kb length of sequence.

RDC0001NLR0075_internal	
forward	AGGCAATGTAGCAGAAGCGT
reverse	ACCCGGGAAAAGTATAGCCG
RDC0001NLR0076_internal	
forward	GCGCTTAGTGGAAGGCAATG
reverse	ATCGCCGGAGCTTTGATAGG
RDC0001NLR0076_gap	
forward	CGATCCCAAGCTTCCATCGT
reverse	AGGTCGCGTTGAGAATCTTGT

4.3.12 Prediction of putative low copy number genes

Putative low copy number genes in the target region of interest were identified in order to rule out any other genes present as the source of the *H2* resistance phenotype. All genes predicted in DM to reside within the 0.8Mb region of interest of chromosome 5 were used in a BLAST search against the entire DM reference genome to ascertain potential copy numbers. BLAST hits were filtered at a query coverage of >50%, the remaining sequences were imported into R and filtered to a "number of hits" cut-off of 6. The list of low copy number genes was then annotated using information stored by NCBI.

4.4 Results

4.4.1 Segregation of the Picasso x P55/7 F1 population suggests the presence of a single, dominant *R* gene

The *H2*-containing resistant clone P55/7 and susceptible cultivar Picasso were crossed, and progeny assessed for segregation. Out of the 192 initial F1 progeny, 154 had three successful replicates from the two independent screens. In screen 1, 23 progeny clones were scored as very susceptible (>17 females), 11 progeny clones were scored as very resistant (≤1 females), and 28 clones as moderately susceptible (>1,<17). For screen 2, 32 progeny clones were scored as very susceptible (>18 females), 69 were scored as moderately susceptible (>1,<17), and 27 were scored as very resistant (≤1 females) (Supplementary Table 1). Based on the mean number of females present across the three replicated plants

per experiments and two independent experimental repeats, the distribution of the plant phenotypes was plotted (Figure 4.2). The progeny segregated with a

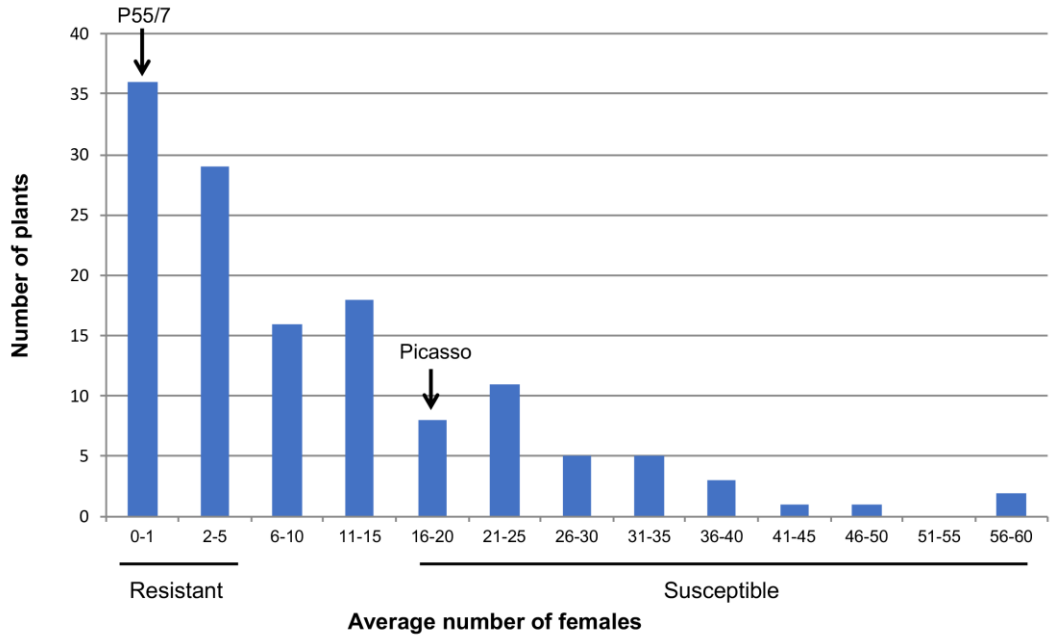


Figure 4.2 Histogram showing the distribution of infection levels of 154 Picasso x P55/7 progeny plants. The plants segregated in a ratio of 0.8:1 (resistant:susceptible) which is close to the 1:1 ratio which would be expected from a heterozygous simplex resistant : homozygous tetraploid cross. Resistant plants had the lowest number of females and cluster to the left of the graph, while susceptible plants showed the highest level of nematode infection and spread across the right of the graph

0.8:1 (resistant:susceptible) ratio (χ^2= 0.04, p>0.84) which is close to the 1:1 ratio expected for a simplex (Rrrr) dominant trait in a tetraploid rrrr x Rrrr cross. The twenty most consistently very resistant (scoring ≤1 female) and twenty most consistently very susceptible (scoring ≥18 females) (Supplementary Table 1) F1 clones were selected for further genetic analysis (Table 4.2).

4.4.2 GenSeq data suggests *H2* is located on chromosome 5

To genetically characterise the *H2* resistance, bulked segregant analysis using individually indexed parents, bulked susceptible and bulked resistant samples was conducted. The individually indexed genomic DNA samples were subjected to GenSeq-based enrichment which targets single or low-copy number genes that can be placed on the individually potato chromosomes with high confidence (Chen et al., 2018).

A total of 11,797,569 raw paired ends reads were obtained from the GenSeq analysis, with 11,634,150 passing read trimming. An on-target mapping rate to

Table 4.2 Pa1 screen scores for the twenty most resistant and twenty most susceptible F1 progeny plants which were taken forward for further analysis.

Resistant Bulk					Susceptible Bulk				
ID	Count 1	Count 2	Count 3	Average	ID	Count 1	Count 2	Count 3	Average
11	0	1	0	0.33	6	27	18	23	22.67
15	0	1	1	0.67	7	30	54	18	34.00
84	0	0	1	0.33	8	17	12	42	23.67
87	0	0	0	0.00	14	25	63	29	39.00
66	0	1	0	0.33	17	38	18	30	28.67
80	0	0	0	0.00	28	45	49	39	44.33
91	0	0	0	0.00	40	48	74	52	58.00
107	0	0	1	0.33	45	73	68	30	57.00
108	0	0	0	0.00	47	51	34	29	38.00
110	0	0	0	0.00	55	33	28	33	31.33
130	1	0	1	0.67	57	11	34	39	28.00
132	0	0	0	0.00	65	30	20	24	24.67
138	1	0	0	0.33	72	24	35	19	26.00
139	0	1	1	0.67	85	26	14	25	21.67
152	0	0	0	0.00	93	19	27	21	22.33
155	0	0	1	0.33	100	27	36	30	31.00
157	0	0	1	0.33	104	20	15	33	22.67
158	0	0	1	0.33	131	10	38	18	22.00
180	0	0	0	0.00	135	22	43	38	34.33
181	1	1	0	0.67	168	44	33	12	29.67

the DM reference was calculated to be between 44.85-75.01%. Based on the phenotypic segregation ratio of nearly 1:1, which suggests a single *R* gene in a simplex configuration, SNPs conforming to the expected ratio ([Rrrr] in P55/7 and [rrrr] in Picasso) were retained. SNP filtering was performed at a 2%, 3%, and 5% mapping mismatch rate to allow for sequence variation when compared to the DM reference genome. At a 3% mismatch rate, a total of 5,448 SNPs were identified between the parents Picasso and P55/7 that conformed to the expected ratio (Figure 4.3A). Relaxing the mapping mismatch rates to 5% or lowering the mismatch rate to 2% yielded 11,606 SNPs and 2,773 SNPs in the parents, respectively. In the bulks, 49 SNP passed filtering at the expected ratio of susceptible and resistant at a 3% mismatch rate (Figure 4.3B). Of those, 28 SNPs

were identified at the expected frequency in both parent and bulk (Table 4.3), with 25 (89%) of the SNPs corresponding to genes associated with the top-end of chromosome 5, while 2 SNPs corresponded to chromosome 3, and a single SNP to chromosome 6 (Figure 4.3C). Similarly, allowing for 2% and 5% mismatch rates, 47 and 699 SNPs at the respective mismatch rates passed the filtering criteria within the bulks, and of those that occurred at the expected frequency, the majority of informative SNPs (94% and 87%, respectively) mapped to a similar interval of potato chromosome 5 (Table 4.4 and 4.5).

Table 4.3 GenSeq informative SNPs identified at 3% mismatch. Column 1 denotes the chromosome where the SNP resides. Columns 2 and 3 give the start and end positions of the gene containing the SNP. Column 4 contains the gene name, and column 5 shows the number of SNPs present in each gene.

Chromosome	Start	Stop	Gene ID	Number of SNPs
3	14879240	14879866	ID=PGSC0003DMG400040532	1
3	38314819	38321395	ID=PGSC0003DMG400018852	1
5	644928	648054	ID=PGSC0003DMG401028313	1
5	668859	673110	ID=PGSC0003DMG400028364	1
5	1415273	1419957	ID=PGSC0003DMG400025119	1
5	1437168	1441274	ID=PGSC0003DMG400025121	6
5	2997356	3001120	ID=PGSC0003DMG400014571	2
5	3357219	3357723	ID=PGSC0003DMG400030589	1
5	3710910	3715061	ID=PGSC0003DMG400030518	1
5	4173679	4174911	ID=PGSC0003DMG400030500	4
5	4484319	4492247	ID=PGSC0003DMG400018405	1
5	5028894	5038966	ID=PGSC0003DMG400031261	1
5	8383814	8387263	ID=PGSC0003DMG400030998	1
5	10524338	10532794	ID=PGSC0003DMG400018598	1
5	10714245	10719910	ID=PGSC0003DMG400011723	2
5	11252622	11256056	ID=PGSC0003DMG400010739	1
5	14418005	14425294	ID=PGSC0003DMG400034313	1
6	5041666	5044538	ID=PGSC0003DMG402004406	1

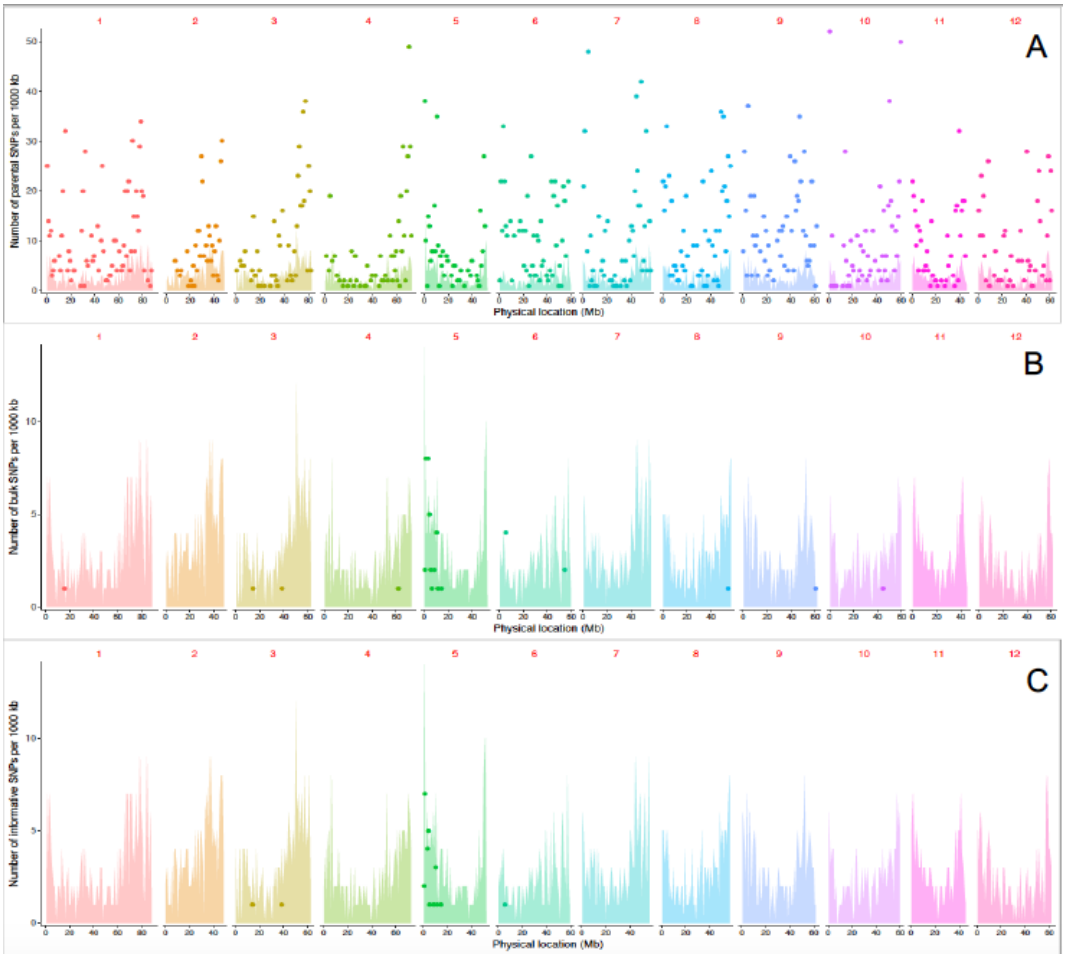


Figure 4.3 Graphical representation of the informative SNPs identified on the 12 potato chromosomes. Each coloured set of 'spikes' represents a chromosome, with the height of each spike representing the number of genes targeted by probes . Each dot represents a position were SNPs are located with its' position on the y-axis denoting how many SNPs are in a specific 1Mb bin. Panel A) displays the 5,448 parental SNPs identified, panel B) displays the 49 identified bulk SNPs, and panel C) displays the 28 identified informative SNPs in common between the parents and bulks. All panels were generated using a mismatch rate of 3%.

Table 4.4 Informative GenSeq SNPs identified at a 2% mismatch rate. Column 1 details the chromosome where the SNP resides. Columns 2 and 3 give the start and stop positions of the gene. Column 4 displays the gene ID, and column 5 details the number of SNPs identified within the gene

Chromosome	Start	Stop	Gene ID	Number of SNPs
5	11252622	11256056	ID=PGSC0003DMG400010739	1
5	1437168	1441274	ID=PGSC0003DMG400025121	5
5	2997356	3001120	ID=PGSC0003DMG400014571	2
5	3357219	3357723	ID=PGSC0003DMG400030589	1
5	3710910	3715061	ID=PGSC0003DMG400030518	1
5	4173679	4174911	ID=PGSC0003DMG400030500	2
5	4484319	4492247	ID=PGSC0003DMG400018405	1
5	5028894	5038966	ID=PGSC0003DMG400031261	1
5	6077929	6084820	ID=PGSC0003DMG400017618	2
5	7239163	7242611	ID=PGSC0003DMG400015575	1
8	40510845	40514964	ID=PGSC0003DMG400033649	1

Table 4.5 Identified GenSeq informative SNPs identified at a 5% mismatch rate. Column 1 details the chromosome where the SNP resides. Columns 2 and 3 give the start and stop positions of the gene. Column 4 displays the gene ID, and column 5 details the number of SNPs identified within the gene.

Chromosome	Start	Stop	Gene ID	Number of SNPs
3	352477	357676	ID=PGSC0003DMG400013442	1
3	3048502	3057254	ID=PGSC0003DMG400018667	3
3	3490272	3492775	ID=PGSC0003DMG400022531	1
3	14879240	14879866	ID=PGSC0003DMG400040532	1
3	32698196	32699073	ID=PGSC0003DMG400021795	1
5	644928	648054	ID=PGSC0003DMG401028313	1
5	1415273	1419957	ID=PGSC0003DMG400025119	1
5	1437168	1441274	ID=PGSC0003DMG400025121	8
5	1857415	1863020	ID=PGSC0003DMG400000805	3
5	2886031	2891012	ID=PGSC0003DMG400014546	3
5	2997356	3001120	ID=PGSC0003DMG400014571	1
5	3710910	3715061	ID=PGSC0003DMG400030518	2
5	3762886	3769061	ID=PGSC0003DMG400030563	2
5	3903471	3909624	ID=PGSC0003DMG400030510	1
5	4033954	4042260	ID=PGSC0003DMG400030549	2
5	4173679	4174911	ID=PGSC0003DMG400030500	4
5	4484319	4492247	ID=PGSC0003DMG400018405	1
5	5028894	5038966	ID=PGSC0003DMG400031261	1
5	6306477	6312721	ID=PGSC0003DMG400022618	2
5	6502663	6503756	ID=PGSC0003DMG400022601	1
5	7239163	7242611	ID=PGSC0003DMG400015575	1
5	8108580	8109092	ID=PGSC0003DMG400030975	1
5	8383814	8387263	ID=PGSC0003DMG400030998	2
5	9111929	9113418	ID=PGSC0003DMG400013488	2
5	9282913	9285557	ID=PGSC0003DMG400013494	1
5	10614999	10623108	ID=PGSC0003DMG400011727	1
5	10714245	10719910	ID=PGSC0003DMG400011723	4
5	11252622	11256056	ID=PGSC0003DMG400010739	1
5	14418005	14425294	ID=PGSC0003DMG400034313	1

4.4.3 Addition of RenSeq data confirms the location of *H2* to chromosome 5

To independently validate the GenSeq-inferred mapping position of *H2* to chromosome 5, the indexed samples were also subjected to RenSeq-based enrichment which specifically targets NB-LRR genes (Jupe et al., 2013). From a total of 8,511,314 paired-ends reads obtained from RenSeq, 8,477,489 passed the read trimming. The on-target mapping rate to the NB-LRR in the DM reference ranged from 30.37% to 61.86% at a 2% and 5% mismatch rate, respectively. At a 3% mismatch rate 3,314 SNPs were identified between the parent's Picasso (rrrr) and P55/7 (Rrrr) that conformed to the expected allele frequency (Figure 4.4A). In the bulks, 106 SNPs passed the filtering conditions expected for susceptible progeny (rrrr) as well as resistant progeny (Rrrr) (Figure 4.4B). Of those SNPs 36 were found at the expected allele frequency in the parents and the bulks. (Table 4.4). The 36 SNPs correspond to 15 NB-LRRs in the DM genome. More than 94% of these SNPs (34/36) reside in 13 NB-LRRs in an 8.1Mb interval on potato chromosome 5 (Figure 4.4C), while the remaining two SNPs correspond to two NB-LRRs on chromosome 9 (Table 4.6).

Table 4.6 RenSeq informative SNPs identified at a 3% mismatch rate. Column 1 denotes the chromosome where the SNP(s) was identified, column 2 and 3 give the start and stop positions of the gene, column 4 gives the gene ID, and column 5 displays the number of informative SNPs found within the gene.

Chromosome	Start	Stop	Gene ID	Number of SNPs
5	1500545	1506500	ID=PGSC0003DMG400025099	3
5	2063328	2066456	ID=PGSC0003DMG400000813	1
5	2075262	2079628	ID=RDC0001NLR0074	1
5	2185980	2190589	ID=RDC0001NLR0075	2
5	2201139	2204777	ID=RDC0001NLR0076	8
5	4227604	4230353	ID=PGSC0003DMG400030497	1
5	4589149	4595717	ID=PGSC0003DMG400018428	1
5	5469503	5473373	ID=PGSC0003DMG400023062	2
5	5723483	5731577	ID=PGSC0003DMG400025611	3
5	6506321	6508868	ID=RDC0001NLR0090	1
5	6528097	6537250	ID=PGSC0003DMG401022603	2
5	8619648	8627296	ID=PGSC0003DMG400013506	1
5	9635954	9642604	ID=RDC0001NLR0098	8
9	35461259	35467442	ID=RDC0001NLR0212	1
9	59518316	59519194	ID=PGSC0003DMG400024366	1

Relaxing the mapping mismatch rates to 5% or lowering the mismatch rate to 2% yielded 6192 SNPs and 1602 SNPs in the parents, respectively. In the bulks, 66 and 10 SNPs at the respective mismatch rates passed the filtering criteria.

In agreement with the 3% mismatch rate, using 2% and 5% mismatch rates predominantly yielded SNPs associated with NLRs on chromosome 5. At a 5% mismatch rate, 55/66 SNPs (>83%) can be attributed to 16 NB-LRRs that reside in the same interval (Table 4.7). At a 2% mismatch rate 70% of SNPs are associated with the same interval (Table 4.8). Combining the results from both the GenSeq and RenSeq analyses independently corroborated the mapping position of the *H2* resistance to a 11 MB interval on potato chromosome 5.

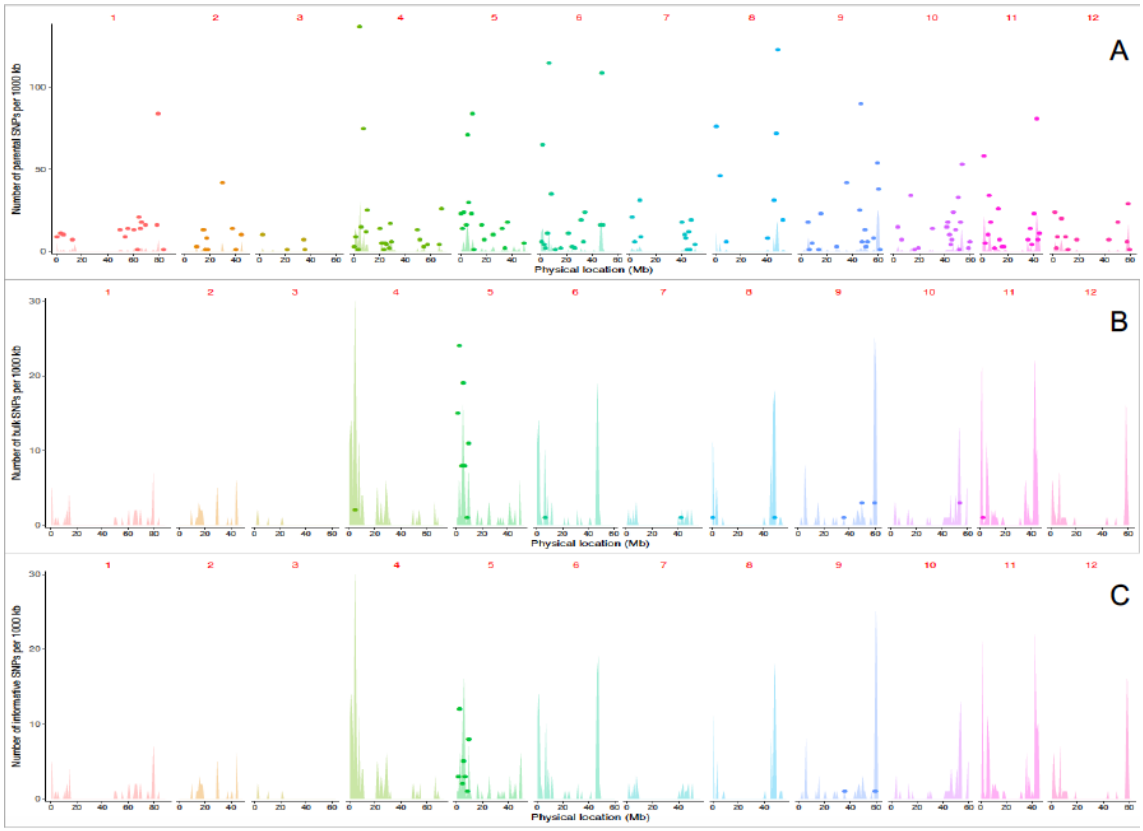


Figure 4.4 Graphical representation of the SNPs identified at a 3% mismatch rate. Each set of coloured data represents a specific chromosome. Coloured 'spikes' represent the number of NB-LRR genes targeted by probes across the chromosome. Each dot represents informative SNPs and its placement on the y-axis determines the number of SNPs identified in a given 1Mb region. Panel A) displays the 3,314 SNPs identified in the parental samples. Panel B) displays the 106 SNPs identified in the bulks, and Panel C) displays the 36 informative SNPs identified in bulks and parents.

Table 4.7 Informative RenSeq SNPs identified at a 5% mismatch rate. Identified informative SNPs discovered at a 5% mismatch rate. Column 1 details the chromosome where the SNP resides. Columns 2 and 3 give the start and stop positions of the gene. Column 4 displays the gene ID, and column 5 details the number of SNPs identified within the gene. A total of 55 SNPs were identified on chromosome 5, 1 SNP on chromosome 6, 7 SNPs on chromosome 9, and 3 SNPs on chromosome 11.

Chromosome	Start	Stop	Gene ID	Number of SNPs
5	1500545	1506500	ID=PGSC0003DMG400025099	2
5	2063328	2066456	ID=PGSC0003DMG400000813	5
5	2075262	2079628	ID=RDC0001NLR0074	5
5	2185980	2190589	ID=RDC0001NLR0075	2
5	2201139	2204777	ID=RDC0001NLR0076	8
5	4227604	4230353	ID=PGSC0003DMG400030497	1
5	4232244	4235519	ID=RDC0001NLR0077	1
5	4611906	4618405	ID=PGSC0003DMG400018429	2
5	4908113	4910054	ID=RDC0001NLR0078	2
5	5469503	5473373	ID=PGSC0003DMG400023062	3
5	5723483	5731577	ID=PGSC0003DMG400025611	5
5	6506321	6508868	ID=RDC0001NLR0090	2
5	6938661	6944482	ID=RDC0001NLR0093	1
5	8619648	8627296	ID=PGSC0003DMG400013506	6
5	9287263	9291478	ID=PGSC0003DMG401013522	1
5	9635954	9642604	ID=RDC0001NLR0098	9
6	6553332	6555332	ID=PGSC0003DMG400009686	1
9	35461259	35467442	ID=RDC0001NLR0212	1
9	59541619	59544402	ID=PGSC0003DMG403020585	1
9	59673213	59678764	ID=RDC0001NLR0223	1
9	60086974	60090937	ID=PGSC0003DMG400031521	1
9	60622644	60631553	ID=PGSC0003DMG400016601a	1
9	60622644	60631553	ID=PGSC0003DMG400016601b	1
9	60956586	60959599	ID=RDC0001NLR0230	1
11	5878059	5881256	ID=RDC0001NLR0264	3

Table 4.8 Informative RenSeq SNPs identified at a 2% mismatch rate. Identified informative SNPs discovered at a 5% mismatch rate. Column 1 details the chromosome where the SNP resides. Columns 2 and 3 give the start and stop positions of the gene. Column 4 displays the gene ID, and column 5 details the number of SNPs identified within the gene. A total of 3 SNPs were identified on chromosome 3 and 7 SNPs were identified on chromosome 5.

Chromosome	Start	Stop	Gene ID	Number of SNPs
3	2502618	2512246	ID=PGSC0003DMG400005052	3
5	2075262	2079628	ID=RDC0001NLR0074	1
5	5469503	5473373	ID=PGSC0003DMG400023062	1
5	6506321	6508868	ID=RDC0001NLR0090	2
5	8619648	8627296	ID=PGSC0003DMG400013506	2
5	9635954	9642604	ID=RDC0001NLR0098	1

4.4.4 SNP-based KASP Markers have a high success rate in discriminating between a susceptible and resistant allele

KASP (Competitive Allele Specific Primer) markers (LGC Genomics) contain two competitive forward primers; each with a tail sequence which interacts with one of the FRET molecules in the KASP Mastermix, and a single common reverse primer (Figure 4.5). The specific of the approach is based on the nucleotide at the 3' end of the forward primers which represents one or the other allele, respectively. Depending on the target DNA and the frequency or presence/absence of both alleles, the KASP assay will reflect this through the use of the complimentary primer(s) .

Figure 4.5 Graphic explanation of the mechanism of KASP marker assays. The green and red boxes represent the competitive forward primers, as well as demarking the direction of elongation, the blue box represents the universal reverse primer which anneals regardless of which forward primer anneals. The star represents the SNP present in the sequence which the marker is based upon. During the annealing and elongation phase either the green or red primer will anneal based on its sequence specificity of the SNP.

Generating markers for all 28 GenSeq informative SNPs and 34 informative RenSeq SNPs from chromosome 5 would have been too expensive. Of the above informative SNPs 19 successful markers were designed based on the GenSeq and RenSeq SNP data (Table 4.1) and the requirements put forward by LGC (desired SNP must no other SNPs 50bp upstream or downstream of it).

Each marker was named based on its location within the genome, for example; marker ST04_03ch05_1416331 is located on chromosome 5 at position 1,416,331bp within the DM reference genome. All marker names were generated using the same format. A KASP marker is successful if it can distinguish between the resistant and susceptible allele within the sample pool. Using marker ST04_03ch05_6079232 as an example (Figure 4.6), it was observed that out of all the polymorphisms identified within the gene the highlighted SNP is present at the correct allele frequency in the resistant reads whilst the susceptible alleles encoded the alternative polymorphism. However, although all the chosen SNPs were present at the correct allele frequency in the samples I also encountered one KASP (in addition to the 11 successful markers) that did not differentiate

between resistant and susceptible and is therefore not linked to the phenotype (Figure 4.7) shows this non-specific marker ST04_03ch05_1437439 (A) and

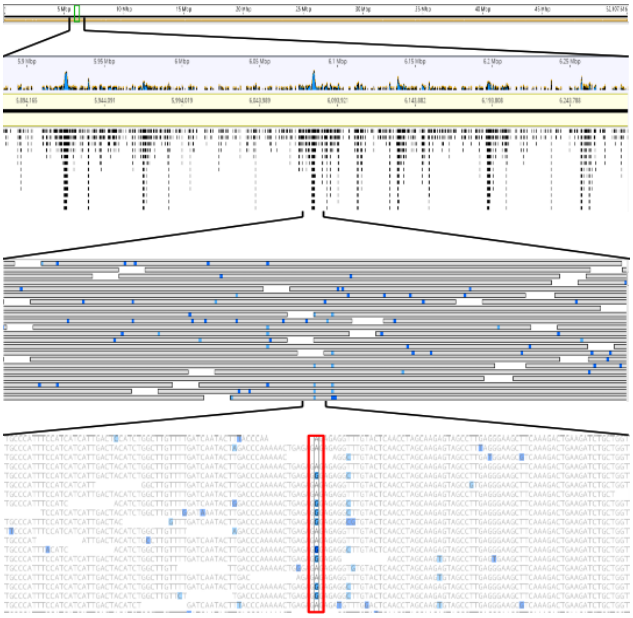


Figure 4.6 Graphical representation of informative SNP identification. An area of chromosome 5 is highlighted, and the GenSeq gene coverage is shown. All the Picasso and P55/7 reads which map to this region are highlighted, and nucleotides which do not resemble the reference genome are coloured blue. Viewing the sequence in this region reveals an allele which is present only in the resistant reads. The nucleotide highlighted within the red box was chosen to be generated into a KASP marker.

contrasts this with a successful markers ST04_03ch05_3000757 (B).

Each KASP marker was tested using the DNA from the individual progeny which were pooled during the enrichment sequencing and confirmed in the parents. Of the 12 markers designed based on the GenSeq data, 9 could competently differentiate between resistant and susceptible samples giving a success rate of 75%, while 6 out of 7 were competent for the RenSeq KASP markers, having a success rate of 85% (Table 4.1). This apparent SNP linkage with the *H2* resistance trait is further evidence that the correct genomic region has been found for the *H2* gene.

4.4.5 KASP Markers allow the *H2* gene to be mapped to a 4.7Mb region of chromosome 5

The individual progenies used to generate the pooled resistant and susceptible bulks were analysed using GenSeq and RenSeq-derived KASP markers (Table

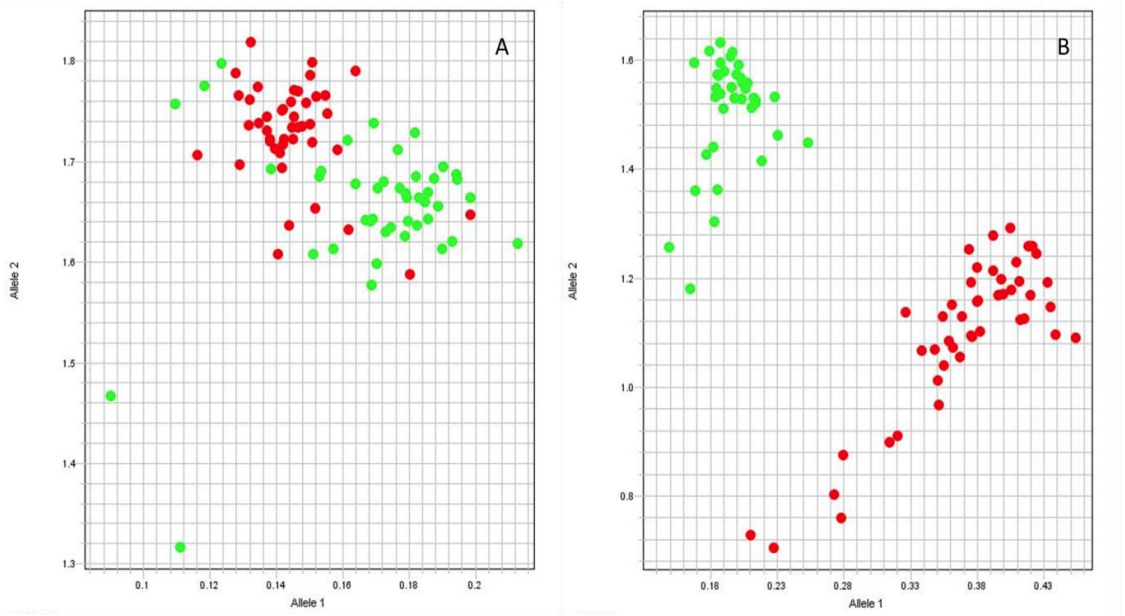


Figure 4.7 Allelic discrimination plots which demonstrate unsuccessful and successful KASP Markers. Each dot represents either a parent or progeny sample. Red dots are homozygotic (susceptible), and green dots are heterozygotic (resistant) for the candidate SNP. Panel A shows almost complete overlap between susceptible and resistant samples, while panel B shows successful clustering and separation of samples based on the candidate SNP.

4.1). Using the allele discrimination plot outputs from each KASP assay, a graphical genotype could be assigned that corresponds to the 'R' or 'r' alleles, respectively. Each dot was assigned a (1; green) if it contained the same allele as the resistant parent (P55/7), while those designated (0; red) showed the same allele as the susceptible parent (Picasso) (Table 4.9).

Progeny plant alleles were called on the basis of which parental allele they resembled; results were then ordered based on the phenotype which was observed during the infection assays. Arranging the phenotypes of the individual plants and the KASP marker-derived genotyped revealed three recombination events in resistant F1 progeny (clones 108, 110 and 152) as well as five recombinants in susceptible progeny (clones 8, 72, 93, 104 and 168).

Table 4.9 Graphical genotype of GenSeq-based KASP Marker Assay results of pooled progeny. Results are ordered by chromosome position; row 1 (ST04_03ch05_1416331) is at the top of the chromosome, moving down to row 12 (ST04_03ch05_11253634). Samples designated with a 1 contained the same allele as the resistant parent (P55/7), similarly; samples designated 0 contained the same allele as the susceptible parent (Picasso).

					Resistant Bulk															Susceptible Bulk														
KASP ID	Gene ID		Picasso	P55/7	11	66	84	87	91	107	108	110	130	132	138	139	152	155	157	158	180	6	7	8	17	40	45	47	55	57	65	72	93	100	104	131	135	168
ST04_03ch05_1416331	PGSC0003DMG400025119	GenSeq	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
ST04_03ch05_1437827	PGSC0003DMG400025121	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
ST04_03ch05_1438531	PGSC0003DMG400025121	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
ST04_03ch05_1503657	PGSC0003DMG400025099	RenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
ST04_03ch05_2202842	RDC0001NLR0076	RenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_3000757	PGSC0003DMG400014571	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_4491040	PGSC0003DMG400018405	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_4737653	PGSC0003DMG400018411	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_6079232	PGSC0003DMG400017618	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
ST04_03ch05_10615824	PGSC0003DMG400011727	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1
ST04_03ch05_11253634	PGSC0003DMG400010739	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1

The most informative markers for delimiting the *H2* interval were RenSeq-derived KASP marker ST04_03ch05_1503657 which is based on NB-LRR PGSC0003DMG400025099 and GenSeq marker ST04_03ch05_6079232 based on PGSC0003DMG400017618. These markers reduced the *H2* locus to a 4.7Mb region. Importantly, the marker order as inferred by the DM reference genome is conserved in this interval as no double recombination events were observed in the F1 progeny.

4.4.6 An expanded population allows the area of interest to be further defined to a 0.8Mb region on chromosome 5

Using the above mentioned *H2* locus as a guide, an expanded population of 656 plants from the original Picasso x P55/7 cross was genotyped with the two flanking markers of ST04_03ch05_1438531 and ST04_03ch05_6079232) to identify recombinants in this interval.

As with the pooled progeny results (Table 4.9), the expanded population allelic discrimination plots were transformed into graphical genotypes for ease of analysis. Out of the tested population, 65 (~10%) progeny plants were found to be recombinant in this interval. Phenotyping was only undertaken on these recombinants and a selected set of 25 controls (10 resistant progeny and 15 susceptible progeny), as well as the parents (Supplementary Table 2). Following the identification of recombinants through flanking markers, all recombinant plants were genotyped with the 5 KASP markers in the interval. Table 4.10 details the graphical genotype information alongside the observed phenotype of the recombinant and selected control plants. Based on the five recombination events present in the F1 resistant progeny (5, 34, 278, 374 and 604) as well as the three recombination events in the susceptible progeny (9, 175 and 331) the area of interest was diminished to a 0.8Mb interval between flanking markers ST04_03ch05_2202842 and ST04_03ch05_3000757. Most importantly it was the combination of both the GenSeq and RenSeq markers which allowed for the successful decrease in the mapping interval.

Table 4.10 Graphical genotype of the expanded recombinant progeny plants and selected control progeny from the initial mapping. Five GenSeq markers (ST04_03ch05_1438531, ST04_03ch05_3000757, ST04_03ch05_4491040, ST04_03ch05_4737653, and ST04_03ch05_6079232) and two RenSeq markers (ST04_03ch05_1503657 and ST04_03ch05_2202842) were combined to give an in-depth analysis of this target region. Susceptible alleles are labelled (0), while resistant alleles are labelled (1), any allele which could not be called as either resistant or susceptible was designated (-). Based on these recombinants, the region of interest can be narrowed to the 0.8Mb region between RenSeq marker ST4_03ch05_2202842 and GenSeq marker ST4_03ch05_3000757.

				Resistant																					
KASP ID		Picasso	P55/7	5	34	61	74	84	113	133	137	176	264	278	372	374	312	392	445	466	473	485	522	559	604
ST04_03ch05_1438531	GenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ST04_03ch05_1503657	RenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ST04_03ch05_2202842	RenSeq	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ST04_03ch05_3000757	GenSeq	0	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	0
ST04_03ch05_4491040	GenSeq	0	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	0
ST04_03ch05_4737653	GenSeq	0	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	0
ST04_03ch05_6079232	GenSeq	0	1	0	0	0	1	0	1	1	1	0	0	0	1	0	1	1	0	1	0	1	1	0	0

	Susceptible																					
KASP ID	2	9	12	41	64	90	92	102	104	131	143	163	165	169	170	175	190	193	200	206	223	233
ST04_03ch05_1438531	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_1503657	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_2202842	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_3000757	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
ST04_03ch05_4491040	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0
ST04_03ch05_4737653	0	1	1	0	0	1	1	1	0	1	0	0	0	0	0	1	0	0	1	0	-	0
ST04_03ch05_6079232	0	1	1	0	0	1	1	1	0	1	0	0	1	0	1	1	1	-	1	0	1	1

	Susceptible																						
KASP ID	289	292	315	328	331	362	389	390	391	404	412	430	447	456	464	468	481	505	545	563	568	575	584
ST04_03ch05_1438531	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_1503657	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST04_03ch05_2202842	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	-
ST04_03ch05_3000757	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
ST04_03ch05_4491040	0	0	0	0	1	0	0	0	1	0	0	1	0	0	-	1	1	0	0	0	0	0	1
ST04_03ch05_4737653	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	1	1	0	0	0	0	0	1
ST04_03ch05_6079232	1	0	0	0	1	0	1	0	1	1	0	1	0	0	1	1	1	0	0	0	-	0	1

4.4.7 Searching the DM reference genome reveals two *R* genes within the 0.8Mb region of interest

Narrowing the potential target region to 0.8Mb is the smallest interval which could be achieved with the current F1 population and the KASP marker set. Based on this reduced interval, the DM reference database was searched between positions 2,202,842 and 3,000,757bp of chromosome 5 to identify all genes present. Chromosome 5 is a hot-spot for several functional *R* genes including *Rx2*, *R1* and *H1*. During the mapping of these genes RFLP-based markers were generated and utilised to fine map their gene positions, with both *Rx2* and *R1* clustering between markers Gp21 and Gp179. To rule out this gene cluster as the location of the *H2* gene, both these markers were tested on the P55/7 (resistant) material, however, the identified candidate *H2* locus lies above both of these markers and therefore this *R* gene cluster (Figure 4.8).

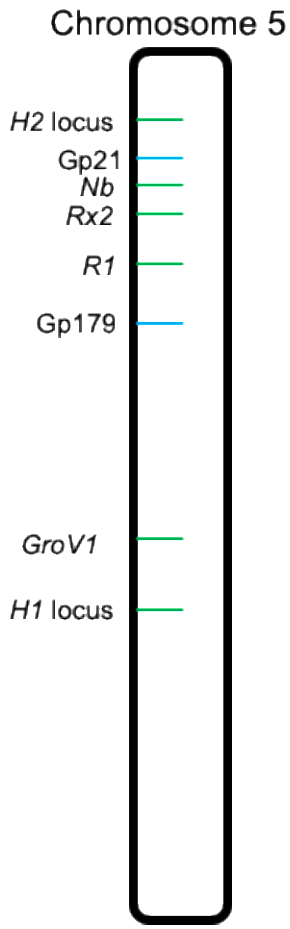


Figure 4.8 Schematic representation of potato chromosome 5 with loci of known pathogen *R* genes. Known *R* gene loci are highlighted in green, while RFLP marker loci are highlighted in blue.

A total of 49 genes reside in this 0.8Mb region (Table 4.11) and contain a wide variety of functions including coding for transporter proteins, transcription factors or containing a conserved domain or currently having unknown functions. Through the use of RenSeq, Jupe et al., (2013) identified two NB-LRR in this interval which can be viewed as a pre-set track available on the genome browser (solanum.hutton.ac.uk/gbrowse). Figure 4.9 details the *R* genes which are

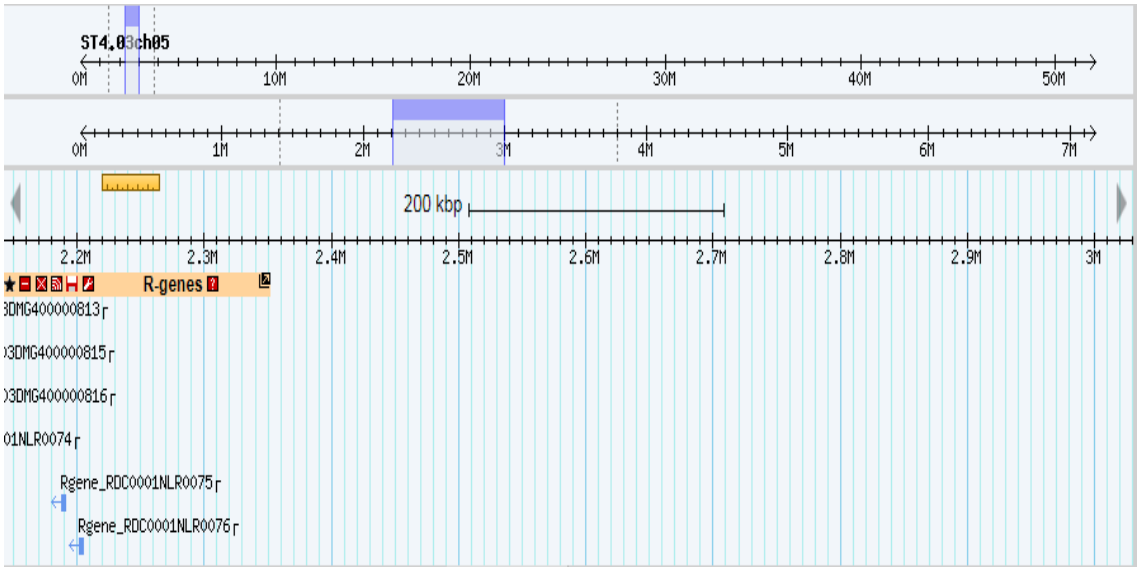


Figure 4.9 DM reference genome viewer showing the presence of two full length *R* genes within the 0.8Mb region between 2,202,842-3,000,757bp of the distal arm of chromosome 5.

present in the target region of chromosome 5 although they are not in the initial annotation of the potato genome (PGSC 2011). Two full length genes with the characteristic NB-LRR resistance gene structure are present within this region. It should be noted that this analysis was undertaken using the DM reference genome and not a genome from a resistant parent P55/7.Therefore the sequence of these genes represents a non-functional/susceptible allele for the *H2* gene, and the same genomic region may differ in the resistant clone P55/7. However, these two genes; RDC0001NLR0075 (positions 2,187,438-2,190,589bp) and RDC0001NLR0076 (positions 2,201,363-2,204,578bp) are nevertheless candidate genes and taken forward for further characterisation.

Table 4.11 Details of all 49 genes present between 2.2-3Mb of chromosome 5 in DM. Column 1 gives the gene identifier as it can be found within the DM reference, columns 2 and 3 give the start and stop base positions for each gene, and column 4 gives the gene annotation if one is known (as annotated on NCBI).

Gene	Start	End	Gene Annotation
PGSC0003DMG400012970	2205646	2207232	Plant Cadmium Resistance 9
PGSC0003DMG400012964	2213939	2215279	Catalytic
PGSC0003DMG400012971	2223501	2230860	Protein phosphatase 2A
PGSC0003DMG400012972	2232558	2233472	Aluminium-activated malate transporter 3
PGSC0003DMG400012973	2247566	2252022	Conserved Gene of unknown function
PGSC0003DMG401012965	2255814	2257171	Plastid-specific 30S ribosomal protein 3
PGSC0003DMG402012965	2258949	2260012	Homologous-pairing protein 2
PGSC0003DMG400012974	2260599	2262820	Conserved Gene of unknown function
PGSC0003DMG400012975	22700079	2276799	Prp4
PGSC0003DMG400012966	2289629	2293817	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase
PGSC0003DMG400012967	2310493	2311248	Conserved Gene of unknown function
PGSC0003DMG400012976	2313801	2314901	F-box family protein
PGSC0003DMG400010977	2390025	2390720	Gene of unknown function
PGSC0003DMG400010976	2393286	2394401	F-box family protein
PGSC0003DMG400019996	2457182	2458385	F-box family protein
PGSC0003DMG400019997	2484873	2485866	Rnase Phy3
PGSC0003DMG400019998	2489884	2490417	SFBB16-alpha
PGSC0003DMG400044653	2498172	2498510	Gene of unknown function
PGSC0003DMG400017423	2561579	2565344	Rnase Phy3
PGSC0003DMG400017422	2573742	2575142	Gene of unknown function
PGSC0003DMG400017392	2596168	2597196	S-locus F-box brothers
PGSC0003DMG400017391	2608825	2612970	GI10570
PGSC0003DMG400014539	2608825	2612970	Nitrate transporter
PGSC0003DMG400014558	2697954	2698566	Conserved Gene of unknown function
PGSC0003DMG400014559	2710633	2711038	Conserved Gene of unknown function
PGSC0003DMG400014540	2724264	2731649	Alpha-glucosidase
PGSC0003DMG400014541	2748542	2750911	DNA binding protein
PGSC0003DMG400014560	2761526	2764714	Bifunctional nuclease
PGSC0003DMG400014542	2781680	2782924	Conserved Gene of unknown function
PGSC0003DMG400014561	2788385	2789123	Isopentenyltransferase
PGSC0003DMG400014562	2792916	2793706	Gene of unknown function
PGSC0003DMG400014563	2800989	2801840	Isopentenyltransferase
PGSC0003DMG400014564	2816681	2818995	Binding protein
PGSC0003DMG400014543	2820134	2824308	Monoglyceride lipase
PGSC0003DMG400014544	2846037	2852520	50S ribosomal protein L15
PGSC0003DMG400014545	2864204	2867000	Conserved Gene of unknown function
PGSC0003DMG400014565	2873525	2880114	Zinc finger CCCH domain-containing protein 65
PGSC0003DMG400014580	2883189	2884281	Conserved Gene of unknown function
PGSC0003DMG400014546	2886031	2891012	Fiber protein Fb34
PGSC0003DMG400014581	2899124	2899899	Gene of unknown function
PGSC0003DMG400014547	2923715	2925346	LOB domain-containing protein
PGSC0003DMG400014566	2941260	2943570	Transcription factor
PGSC0003DMG400014567	2947216	2948013	Conserved Gene of unknown function
PGSC0003DMG400014548	2953645	2957592	Protein kinase PKN/PRK1
PGSC0003DMG400014568	2957727	2959209	3-ketoacyl-CoA synthase
PGSC0003DMG400014549	2963585	2969054	3-ketoacyl-CoA synthase
PGSC0003DMG400036385	2974083	2974595	AP2/ERF domain-containing transcription factor
PGSC0003DMG400014569	2989490	2989948	Gene of unknown function
PGSC0003DMG400014570	2993494	2994953	AP2/ERF domain-containing transcription factor

4.4.8 NLR-based KASP Marker segregates 100% with F1 progeny

Searching the DM reference genome identified RDC0001NLR0075 and RDC0001NLR0076 as putative candidate genes. A SNP in gene RDC0001NLR0076 was identified in the RenSeq mapping and used to design KASP marker ST04_03ch05_2202842. Based on the 65 recombinant plants from the expanded population tested, the marker ST04_03ch05_2202842 segregates 100% with all resistant progeny containing the resistant allele and all the

susceptible progeny containing the susceptible allele at this position (Table 4.6). Considering that the 65 recombinants identified originate from 650 additional progeny, suggests that this gene is either very close to the *H2* gene, or is in fact the functional gene itself. Based on this hypothesis, candidate RDC0001NLR0076 was taken forward for further sequence analysis to determine whether it could indeed be the functional *H2 R* gene.

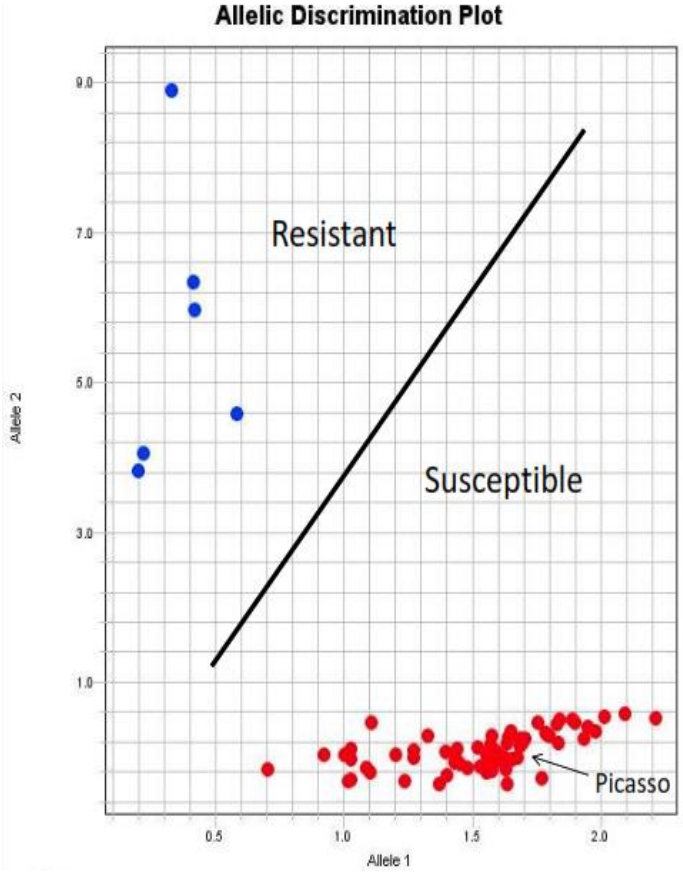


Figure 4.10 Allelic discrimination plot of KASP marker ST04.03ch05_2202842 tested on 20 RDC0001NLR0076 clones Susceptible parent Picasso was used to determine between resistant and susceptible alleles. Each clone was replicated in triplicate.

4.4.9 Identification of putative resistant *H2* allele

To further analyse candidate RDC0001NLR0076, the gene was cloned from resistant P55/7 by PCR amplification and transformation into pGEM-T. The cloning was required to discern putative haplotypes that were amplified during PCR. Following transformation of ligated amplicon and pGEM-T-easy vector into *E. coli* and plasmid preps, an initial 20 recombinant clones were tested with KASP

marker ST04_03ch05_2202842. This helped identify two clones (clone 1 and clone 8) which contained PCR amplicons in coupling with the resistant allele, clustering separately from the susceptible Picasso allele (e.g. clone 9) (Figure 4.10).

'Resistant' clones 1 and 8 as well as susceptible clone 9 were Sanger sequenced to cover almost the entire length of the candidate gene (78% coverage). The resistant clones (1 and 8) are 100% identical to each other, while the susceptible clone (9) contains 83 SNPs as well as an in-frame insertion of 18 nucleotides, and an out of frame insertion of 77 nucleotides compared to clone 1 and 8 (Figure 4.11). *In silico* translation of the sequence in all six frames 1 led to non-synonymous mutations and premature stop codons in the susceptible clone.

Translating the cloned sequences allowed some insight into the amino acid and potential protein structure of the candidate gene. Alignment of the cloned candidate fragments (resistant and susceptible clones) with the RDC0001NLR0076 gene from the DM reference gave a pairwise identity of 83.5%, with this drop in percentage identity accounted for by the missing 5' end

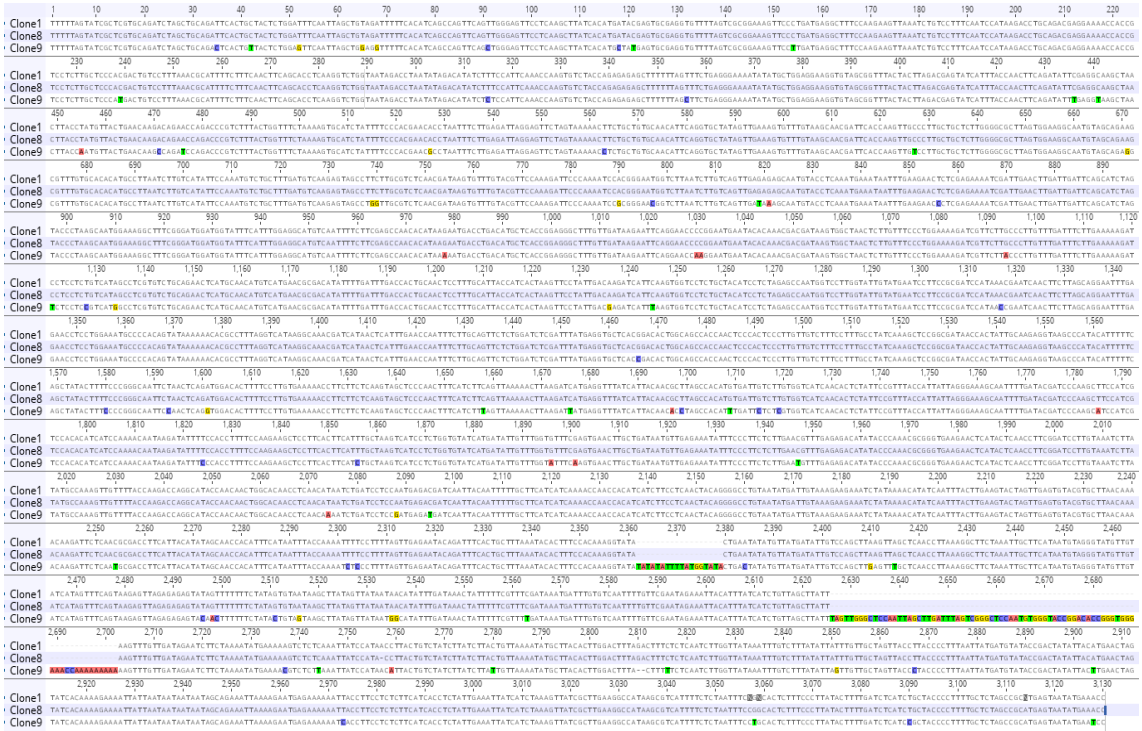


Figure 4.11 Aligned nucleotide sequence of candidate clones 1,8 (resistant) and 9 (susceptible) cloned from resistant P55/7. Sanger sequencing covered 78% of the length of the gene, with clones 1 and 8 being identical, and clone 9 containing sequencing differences. Polymorphisms are highlighted based on CLUSTAL colouring.

of the cloned candidate fragments (Figure 4.12). The sequence which could be aligned showed a high level of sequence identity, however SNPs are present in the resistant clones (clone 1 and 8) which aren't present in the susceptible clone (clone 9) or the DM reference; positions 510, 512, 516, and 548.

Utilising the NCBI Conserved Domain Database (NCBI CDD) allowed the theoretical structure of the DM reference gene RDC0001NLR0076 to be determined. Based on the conserved domains which are present within the amino acid sequence the protein was determined to be a member of the coiled-coil NB-ARC-LRR (CC-NB-ARC-LRR) family of *R* genes. Using sequence homology between the DM reference and the amplified clones it can be confirmed that they also contain NB-ARC-LRR domains (Figure 4.12). Although the 5' end of the amplified clones are missing, it can be hypothesised based on the sequence homology the fragments share with the reference copy of RDC0001NLR0076 that the candidate *H2* gene in P55/7 is also a member of the CC-NB-ARC-LRR family of *R* genes.

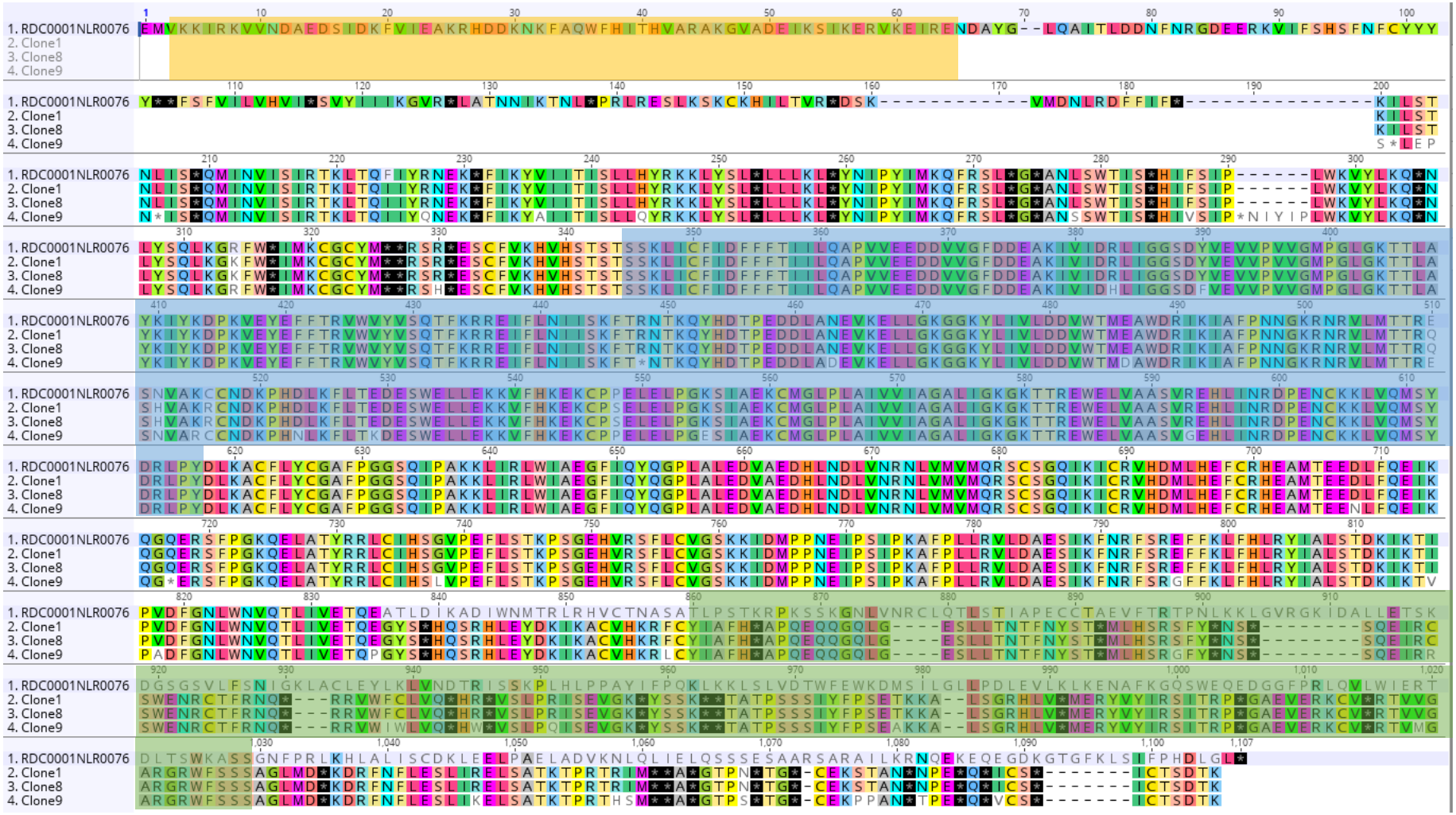


Figure 4.12 Aligned amino acid sequence of DM reference RDC0001NLR0076 with P55/7 cloned candidates. Clones 1 and 8 show the resistant allele, while Clone 9 is susceptible. Conserved domains are overlaid in coloured blocks; the coiled-coil (CC) domain (orange) at position 3-64, the NB-ARC domain (blue) at position 345-616, and the LRR domain (green) at position 860-1029. Non-synonymous mutations are present in resistant clones at positions 510, 512, 516, and 548, within the NB-ARC domain.

4.5 Discussion

4.5.1 Role of enrichment in gene mapping

The potato reference genome (DM) was generated through creating a doubled monoploid of the genotype *S. tuberosum* ssp. Phureja DM1-3 516 R44. Creating a doubled monoploid allowed the vast heterozygosity, which is present in tetraploid cultivated potato, to be simplified (PGSC, 2011). The reference genome contains over 844 million bases, coding for over 39,000 genes. This is a vast genome to search, even in its homozygotic form, for a single resistance gene. This highlights the need for reducing the complexity, with the drawback that a reduction in complexity leads to limitations in other areas such as R gene resistant allele sequence.

During an initial annotation of the DM reference genome, 438 NB-LRR genes were predicted within the 39,031 genome models provided based on 20 NB-LRR-specific motifs (PGSC, 2011, Jupe et al., 2012). The number of NB-LRR genes present within the DM reference increased to 755 loci when a re-annotation was done using the RenSeq workflow (Jupe et al., 2013). Together, the GenSeq and RenSeq (Jupe et al., 2013) enrichments target less than 3,000 genes and thereby approximately 1%, of the genome. This reduction in genome complexity, whilst at the same time targeting the areas of the genome (NB-LRRs) which would provide the greatest amount of useful information, has proven very powerful (Chen et al., 2018). Indeed, this reduction in genome complexity correlates with an increase in sequencing read depth, which resulted in higher confidence SNP calling during the subsequent filtering. Being able to identify SNPs which are 1) real, rather than sequencing artefacts, and 2) present through linkage to a trait of interest is fundamental when mapping genes. Furthermore, RenSeq reads for P55/7 and Picasso were used for a dRenSeq analysis to confirm that the *H2* resistance is not based on previously characterised resistance genes including *Gpa2* (Chapter 3).

4.5.2 Segregation of F1 progeny reveals that *H2* resistance in P55/7 is based on a simplex dominant gene

Classical Mendelian genetics states that a tetraploid cross between a homozygotic susceptible with a simplex heterozygotic resistant parent will give a 1:1 resistant : susceptible phenotype (Gebhardt and Valkonen, 2001, Bryan et al., 2002a). The susceptible parent Picasso is known to not contain the functional *H2* gene and to be completely susceptible to *G. pallida* Pa1 populations. The parent P55/7 is known to be resistant to Pa1 populations, but, until this study, the number of functional *H2* genes and the genetic makeup (e.g. being a single gene) remained elusive.

Screening the F1 progeny with *G. pallida* Pa1 gave an almost perfect 1:1 segregation of resistant:susceptible progeny. A chi-square (χ^2) test was done as part of the statistical analysis to measure whether the observed segregation ratio (0.8:1) fit with the expected segregation ratio (1:1). A p value of >0.84 was calculated, making the results not-significant, and proving that the progeny were segregating as if a single dominant gene was present. From this, we can infer that the resistant allele is present in a simplex (single copy) format in the resistant parent (P55/7) and that the resistance is most likely based on a single gene that segregates in the population. A single major gene in simplex is ideal for mapping and potential identification of the *H2* resistance gene.

4.5.3 RenSeq and GenSeq results agree that the *H2* gene is located on chromosome 5

The results outlined in Table 4.3 and 4.6 show an obvious skew in the number of SNPs present within chromosome 5 compared to the rest of the genome and across all three mismatch stringencies which were tested (2%, 3%, and 5%), with chromosome 5 being a known hotspot for pathogen resistances (Gebhardt and Valkonen, 2001, Bakker et al., 2004a) and the area of interest the distal short arm of the chromosome. Previous mapping of pathogen *R* genes have relied on AFLP- and RFLP-based markers and identified *R1* active against *P. infestans* (Meksem et al., 1995) as well as *Gpa5* and *Gpa6* active against *G. pallida* (van der Voort et al., 2000b). Interestingly, the area of interest for the location of *H2*

lies above these markers which have been used for these resistance trait mapping.

The combination of GenSeq and RenSeq targeted less than 3,000 genes, accounting for approximately 1% of the potato genome for re-sequencing. This reduction in genome complexity in parallel with the sequencing of low-copy genes and NB-LRRs helped inform the chromosomal location within the potato genome (GenSeq) as well putative candidate genes (RenSeq).

4.5.4 Flanking markers allow the area of interest to be decreased to 4.7Mb

The initial population of over 190 plants was phenotyped and used to identify the 20 most resistant and susceptible F1 plants, respectively. The bulking of these plants allowed a 4.7Mb area between ~1.4-6Mb of DM chromosome 5 to be identified as responsible for the resistance. The success of the markers meant they could be utilised to determine the genotype of any member of this mapping population. Increasing the tested population to include all 154 progeny clones from the initial mapping population did not yield any further information to narrow the area of interest to less than 4.7Mb.

4.5.5 Combining GenSeq and RenSeq markers on expanded F1 recombinant progeny population reduced area of interest to 0.8Mb

Expanding the initial progeny to include an additional 656 F1 clones allowed the target area to be greatly reduced. From the expanded population, an additional 65 recombinant plants were identified with flanking markers, reducing the area of interest to a 0.8Mb region between ~2.2-3Mb at the most distal end of chromosome 5. This is the smallest interval which could be defined using these GenSeq and RenSeq markers on this mapping population.

Within this interval, there are 49 low copy genes of varying roles based on the DM annotation. Of these 49 genes, 15 are of unknown function, while the rest have a diverse set of functions from mediating ubiquitination (F-box-related genes) to DNA binding, and transcription factors (Table 4.7). All genes in this region were identified in order to rule out the presence of any gene other than a NB-LRR controlling the resistance phenotype. This was done as work undertaken

on the *rhg1* locus of the soybean cyst nematodes *H. glycines,* discovered that the resistance was controlled by an α-SNAP (α soluble NSF attachment protein) family member (Matsye et al., 2012). While the *Rhg4*, also conferring resistance to *H. glycines* is controlled by a serine hydroxymethyltransferase (SHMT) (Liu et al., 2012). Interestingly both of these QTLs, *rhg1* on chromosome 18 and *Rhg4* on chromosome 8 of soybean, are required to elicit a resistance response toward *H. glycines* (Meksem et al., 2001). Observations of the genes within this interval did not highlight any which were hypothesised to control the *H2* resistance response apart from NB-LRR genes RDC0001NLR0075 and RDC0001NLR0076 which were only identified in this interval following bespoke, RenSeq-based reannotation of the DM genome (Jupe et al., 2013).

Based on the presence of these two *R* genes (RDC0001NLR0075 and RDC0001NLR0076) within this interval in the improved DM reference; it was not necessary to design further markers to decrease the area of interest further. These two genes became the focus of all further analysis due to their position within the genome.

Previous work carried out using the wild potato species *S. verrucosum* utilising both GenSeq and RenSeq allowed for the mapping of the *P. infestans* resistance gene *Rpi-ver1* in a 4.3Mb region of chromosome 9 (Chen et al., 2018). The success shown in mapping the *Rpi-ver1* gene increases the confidence in this study, and again highlights the power of this target enrichment-based sequencing approach for successful resistance gene mapping, as well as displaying the suitability of utilising enrichment approaches in tetraploid organisms.

4.5.6 Sequence analysis revealed candidate resistant clones are in coupling while susceptible clones are in repulsion

The KASP marker ST04_03ch05_2202842 was identified to co-segregate with the *H2* resistance allele based on the genetic map generated using the F1 progeny. This marker was based on a SNP present in candidate gene RDC0001NLR0076, and so this gene was taken forward for further analysis. All research up until this point was based on sequence from the DM reference. The DM reference genome can give insight into the sequence of a gene and where it

resides within the genome. However, it is only a reference and cannot be used to successfully identify functional genes in resistant cultivars. Because of this, all downstream analysis was done without the use of the reference genome.

The cloning of RDC0001NLR0076-like genes from P55/7 identified a resistant allele (clone 1 and 8) and a distinct susceptible allele. Within the PCR screen of 20 recombinants, only clone 1 and 8 were identified as resistance. This ratio (1:10) is slightly skewed as we could have expected a 1:4 ratio if all alleles amplified equally well. Nevertheless, the prediction of an ORF of over 2,000bp highlights the potential for this gene to be a functional candidate for *H2*. Future redesign of primers and re-sequencing of the locus aims to identify a start codon for this gene. This will be complimented by screening an existing BAC library from P55/7 and analysis long-read (PacBio) RenSeq enriched samples which are being generated for P55/7 by the Earlham Institute. The function of candidates will be assessed through transgenic assays using susceptible variety Desiree.

5. Functional testing of putative candidate *H2* avirulence genes in resistant P55/7

5.1 Introduction

Pathogen-caused disease is rare in natural environments as plants have evolved to recognise conserved pathogen associate molecular patterns (PAMPs) which results in PAMP-triggered immunity (PTI) (Zipfel, 2009). In order to overcome this defence, pathogens secrete effector proteins which suppress PTI. In response to this, host R proteins have developed which (in)directly recognise these proteins and initiate effector triggered immunity (ETI) (van der Hoorn and Kamoun, 2008). The presence of a large number of effectors allows the pathogen to continually try to overcome PTI and successfully infect the host. Effector proteins undergo continuous mutation in order to evade host defences; with this constant adaption causing pathotype and population-specific differences, some of which are advantageous allowing successful infection, while others are not. This pathotype-specific mutation of effector proteins to switch between avirulent and virulent will be the focus of the research within this chapter.

A wide range of definitions of what constitutes an effector have been proposed, ranging from very broad categorisations of any secreted molecule introduced into a host by a pathogen, to the very restricted definition of a protein that suppresses host defence responses. For the work carried out in this chapter we adhere to the definition of an effector put forward by (Hogenhout et al., 2009) which states that an effector is “a pathogen molecule that suppresses host defences or manipulates the host to allow provision of food to the pathogen”. Almost all effectors which have a role in PPN host parasitism are produced within the pharyngeal glands and are secreted into the host through the nematode’s hollow needle-like stylet.

In cyst nematode species, the pharyngeal glands are composed of two subventral and one dorsal gland cell. Morphological analysis of gland cells during development (Hussey and Mimms 1990) suggested that effectors from the separate gland types are expressed and secreted at different stages of host

parasitism. During early parasitism; that is, the stages of penetration of the root tip, migration toward the cortex, and the triggering of feeding site formation, the subventral gland cells show signs of being highly active as they are large and packed with secretory granules. This suggests that early parasitism is controlled through effectors produced within the subventral glands. By contrast, during the later sedentary stages of infection, where the nematode is mostly maintaining the syncytium and feeding, the dorsal glands become enlarged and active, suggesting that effectors produced here are required (Hussey and Mimms, 1990). Genomic analysis of changes in effector gene expression in PCN mirrored these findings with one group of effectors, including many known to be expressed in the subventral gland cells, upregulated at J2 and another group, including known dorsal gland cell genes, upregulated in later parasitic stages (Thorpe *et al.*, 2014).

Effectors that have a wide range of functional roles have been identified from plant-parasitic nematodes, including PCN. In the early stages of parasitism, cyst nematodes move intracellularly through the host root towards the cortex causing structural damage. In order to move, cyst nematodes secrete cell wall degrading enzymes which soften the cell wall (Smant et al., 1998, Popeijus et al., 2000). Together with other secreted proteins, such as expansin-like and cellulose-binding proteins that disrupt non-covalent bonds between cell wall components and allow enzymatic effectors more easy access to the cell wall components, the mixture allows for the dismantling of the plant cell wall (Hassan et al., 2010).

Cell wall degrading enzymes (CWDEs) were the first nematode secreted effector proteins to be identified and localised *in planta*. These proteins are produced in the subventral gland of the nematode (indicating that they are required for early parasitism) and through the use of immunolocalisation have been localised at the stylet orifice as well as along the migratory pathway of the cyst nematode *Heterodera glycines*, and the root-knot nematodes *Meloidogyne incognita* and *M. javanica* (Jaouannet and Rosso, 2013).

The first CWDE to be identified in plant parasitic nematodes was β-1,4-endoglucanase which breaks down cellulose, the major component of the plant cell wall. These cellulases have since been identified in many nematode genera

including *Meloidogyne, Heterodera, Globodera, Radopholus, Pratylenchus, Rotylenchulus, Ditylenchus* and *Aphelenchus* (Haegeman et al., 2012). Subsequent work identified a wide range of CWDEs present in plant-parasitic nematodes including pectate lyases, expansins, arabinogalactans, and arabinases (Cotton et al., 2014). The CWDEs present in plant-parasitic nematodes are thought to have been acquired from bacteria via horizontal gene transfer and analysis of the genomes of these nematodes has shown that other genes acquired in this manner are present, including other effectors. For example, both cyst and root-knot nematodes produce and secrete chorismate mutase, an enzyme that forms part of the shikimate pathway (Lambert et al., 1999). This enzyme (or pathway) is not present in animals and the substrate that it metabolises is not present, suggesting that it is likely to manipulate host biochemical processes. One possibility is that chorismate mutase depletes levels of chorismate in order to prevent conversion of this compound into the defence signalling molecule salicylic acid and thus suppresses host defences.

Some effectors are present as large multigene families, including the CLAVATA3/ESR (CLE)-like family and the secreted SP1a and Ryanodine receptor domain (SPRYSEC) family (Mitchum et al., 2013). The CLE proteins have multiple roles in plant growth and development. In plants the CLE proteins contain an N-terminal signal peptide, a variable domain, and a conserved 14 amino acid domain known as the CLE motif at or near the C-terminal. Research carried out in *G. rostochiensis* identified five CLE-like genes, with four out of the five encoding secreted effectors with multiple CLE motifs at the C-terminal. These effectors are produced and secreted from the dorsal gland (indicating their potential role in the later stages of host parasitism) and the presence of multiple CLE motifs demonstrates the nematode's ability to mimic host proteins to facilitate parasitism (Lu et al., 2009).

Perhaps the best example of an expanded multigene family is the SPRY domain family, which contains the SPRYSEC effector proteins. While all nematodes which have been studied have been found to contain SPRY domain proteins, *G. pallida* is predicted to contain over 300 SPRY proteins which is vastly expanded compared to other nematodes and to other PPNs; *M. incognita* contains 27, while

B. xylophilus contains 12 (Cotton et al., 2014). However most predicted SPRY domain coding proteins are not typically secreted (Rehman et al., 2009). Research undertaken in *G. pallida* revealed that only a fraction (approximately 10%) of SPRY domain containing proteins encode predicted signal peptides at their N-terminus, leading to the conclusion that the majority of SPRY proteins are not secreted into the host plant. Although there are cases where proteins lacking signal peptides are secreted into the host, there was a strong correlation between high expression at the J2 stage and the presence of a signal peptide. Conversely those SPRY domain proteins lacking signal peptides had much lower expression levels and less well defined expression profiles (Mei et al., 2015). Some, but not all, SPRYSEC proteins suppress host defence responses; several have been identified that suppress Effector Triggered Immunity induced by the recognition of RBP1 by *Gpa2* (Mei et al, 2015). Due to their hypervariable domain structure one role of the SPRYSEC family may be to target and interact with host proteins (Mitchum et al., 2013). One such role is to interact with cytoplasmic plant NB-LRR proteins; this interaction has been successfully demonstrated in *G. rostochiensis*. The SPRYSEC-19 effector protein of *G. rostochiensis* interacts with the LRR domain of SW5F, a member of the SW5 gene cluster containing several pathogen resistance genes, in tomato. Experiments to verify the interaction between SPRYSEC-19 and SW5F were carried out in both tobacco (*Nicotiana* ssp.) and tomato leaves, however the co-expression of the two proteins did not result in the classic programmed cell death response (Jaouannet and Rosso, 2013). This could possibly be explained by the resistance protein requiring further activation before a cell death response is triggered. Conversely the *G. pallida* SPRYSEC effector protein RBP1 does induce a cell death response when it is co-expressed with its cognate resistance gene *Gpa2*. The resistance protein Gpa2 from the wild potato species *S. tuberosum* ssp. *andigena* confers pathotype-specific resistance to a population of Pa2/3. Recognition of RBP1 by *Gpa2* is dependent on a single amino acid at position 187 within the SPRY domain. Mutation at this point can switch an avirulent nematode to virulent (Sacco et al., 2009). Effectors play an integral role in the host-pathogen interaction, and so a great deal of effort is put into research aimed at understanding their function.

A key aim is to identify the avirulence effectors recognised by host resistance genes. A better understanding of which resistances control nematode populations, or conversely which nematode effector mutations cause resistance breakdown, provides information that allows management of resistance to ensure durability.

The hypothesis for the presented work was that the differing responses produced between Pa1 (avirulent) and Pa2/3 (virulent) populations when exposed to *H2*-mediated resistance is controlled through a series of point mutations within an effector gene which alter its recognition rather than presence/absence of the effector.

5.2 Aims

In this study effector genes were enriched for and sequenced in *G. pallida* pathotypes Pa1 and Pa2/3 (Lindley population). Polymorphisms between the two groups were screened under stringent criteria to produce a list of testable *H2* avirulence gene candidates. The specific aims of this chapter were:

a) Identify putative candidate *H2 Avr* genes which showed polymorphic differences between the avirulent Pa1 and virulent Lindley populations
b) Clone effector gene candidates into gateway vectors in preparation for *Agrobacterium*-mediated transient expression in the *H2* resistant potato cultivar P55/7
c) Verify candidate genes are effector proteins through the use of *in situ* hybridisation to confirm gland cell expression
d) Visualise pathotype-specific programmed cell death through *Agrobacterium*-mediated transient expression in potato

5.3 Materials and Methods

5.3.1 DNA Extraction from *G. pallida* Females

Extractions of genomic DNA were carried out on female Pa1 and Pa2/3 (Lindley) nematodes in order to compare avirulent (Pa1) with virulent (Lindley) populations. Females of both populations were removed from susceptible potato plants with tweezers after 8 weeks of growth, divided into aliquots of 10 per 1.5ml microcentrifuge tube and were flash frozen in liquid nitrogen; Pa1 was harvested from susceptible Picasso x P55/7 progeny plants (Section 2.1.4), and Lindley was harvested from susceptible cv Desiree. In order to gain the correct target DNA concentration a total of 50 Lindley females and 30 Pa1 females were required for total DNA extraction.

Aliquots of frozen females in 1.5ml microcentrifuge tubes were crushed using a micropestle before 600µl of cell lysis buffer (Qiagen) and 5µl Proteinase K (20mg/ml) was added and the mixture vortexed. The samples were incubated overnight at 56°C. The next day 4µl of RNaseA (100mg/ml) was added to each tube and incubated for 10 min at room temperature (RT) (~20-22°C). Next, 200µl of protein precipitation buffer (Qiagen) was added before briefly vortexing the samples and incubating on ice for 10 min. Samples were centrifuged for 10 min (15,000rpm, 4°C), the supernatant was then transferred to a fresh tube to which 0.25µl glycogen (20mg/ml) and 600µl isopropanol was added before incubating the samples overnight at -20°C.

After the incubation, samples were centrifuged for 45 min (17,000rpm, 4°C) and the supernatant removed. Next, 600µl of 70% ethanol was added, and samples were centrifuged for 30 min (17,000rpm, 4°C). The supernatant was removed, and the pellet was left to dry in the fume hood for 1h. Finally, 20µl of elution buffer (Qiagen) was added and samples were left to resuspend for 1h at room temperature. Samples were initially quantified using a Nanodrop (Thermo Scientific) to determine the DNA concentration.

5.3.2 Enrichment Library Preparation

Libraries for Pa1 and Lindley were prepared as outlined in section 2.2.7 *Gene Enrichment and Sequencing Library Preparation.* Bait libraries were designed based on the effector complement detailed in Cotton *et al* (2014). Nematode libraries were multiplexed and sequenced by Illumina MiSeq.

5.3.3 Semi-Quantitative PCR

To test the success of the effector gene enrichment, qPCR was carried out on the pre-enriched vs post-enriched samples. Three effector genes (GpSPRY414-2, GpSPRY1719-1, and GpG16H02) which were part of the enrichment probe list were used as positive controls, while the cytochrome B gene was used as a control (not an effector and therefore should not have been enriched for).

GpSPRY414-2		GpG16H02	
Forward	GCTGTCTTCGCTGTTCAGTC	Forward	TATCCGAGTCCTTCACTACTG
Reverse	TTGCCGACACCATACCGT	Reverse	AAGATGATCATCCAGTCCAAG
GpSPRY1719-1		Cytochrome B	
Forward	AGAAAGGAGAGCACAACGGT	Forward	TGTAGGTGAACCTGCTGCTG
Reverse	CTCTTTGCCCAATCCACGC	Reverse	GTGTCCGTCAACAACAAACG

To reduce the introduction of bias into the qPCR, DNA concentrations for both pre- and post-enrichment libraries were standardised to 15.8ng/µl. The SYBR green Master Mix (ThermoFisher Scientific) was used for qPCR. Samples were prepared as follows:

Reagent	Volume
SYBR Green MasterMix	12.5µl
Forward primer (final conc. 300nm)	2µl
Reverse primer (final conc. 300nm)	2µl
DNA template	1µl
Water	7.5µl
	25µl

Samples were amplified under the following conditions (Applied Biosystems StepOnePlus):

95°C	20 seconds	
95°C	3 seconds	40 cycles
60°C	30 seconds	

5.3.4 Filtering SNPs and Primary Candidate Gene Analysis

Filtering of polymorphisms in the effector enrichment data culminated in 19,873 unique polymorphisms being identified within the Lindley and Pa1 libraries when compared to the *G. pallida* reference genome. In order to define a testable sub-set of effectors, polymorphisms underwent several rounds of filtering.

For a polymorphism (SNP, MNP, INDEL) to be kept it must: occur in an exon, cause a non-synonymous alteration in amino acid sequence, and the Lindley sequence must contain 100% of the reference allele, while the Pa1 sequence must contain 100% of the alternative allele. Although avirulence is dominant which would mean polymorphisms which occur at 50% in Pa1 could have been kept; the number of sequences needed to be reduced to a set that could feasibly be tested, and so only sequences with 100% alternative allele were kept.

Sequences that passed this pipeline underwent analysis using SignalP (http://www.cbs.dtu.dk/ services/SignalP/) to determine the presence of a signal peptide, as well as checking their expression profiles in the RNAseq dataset for *G. pallida* (Cotton et al., 2014).

5.3.5 Candidate Effector Gene Cloning

5.3.5.1 cDNA Synthesis

RNA was extracted from J2 nematodes of *G. pallida* Pa1 and Lindley populations using the TRIzol Reagent Kit (Life Technologies). Hatched J2s (from approximately 100 cysts) were removed from the hatching petri dishes and pipetted into 50ml tubes. Each tube was centrifuged for 10 min at 2,500rpm to pellet the juveniles. Excess TRD was removed and the pelleted juveniles

transferred to a fresh 2ml microcentrifuge tube. Tubes containing juveniles were flash frozen in liquid nitrogen before the pellet was placed in a pre-cooled mortar. Juveniles were crushed with a pre-cooled pestle before 1ml TRIzol reagent was added. The juvenile/TRIzol solution was transferred to a fresh 2ml microcentrifuge tube and left to incubate for 5 min at RT. After incubation 200µl chloroform was added and the tube was vigorously shaken by hand for 15 sec, before a 3 min incubation at RT. The samples were centrifuged for 15 min (10,000rpm at 4°C) before the aqueous layer was removed and pipetted into a fresh tube to which 500µl 100% isopropanol was added, and incubated at RT for 10 min. The sample was centrifuged for 10 min (10,000rpm at 4°C), before the supernatant was removed, leaving the RNA pellet. The pellet was washed with 1ml 75% ethanol and centrifuged for 5 min (8,000rpm at 4°C), before the supernatant was discarded and the pellet left to air dry and re-suspended in 20µl RNAse-free H_2O.

Extracted RNA was then DNase treated using the RQ1 RNase-Free DNase Kit (Promega). Eight microlitres of RNA was mixed with 1µl each of RQ1 10x Reaction Buffer and RQ1 RNase-Free DNase before being incubated at 37°C for 30 min. After incubation, 1µl of RQ1 DNase Stop Solution was added, and the sample incubated at 65°C for 10 min.

DNase-treated RNA was used with the SuperScript III Reverse Transcriptase (Invitrogen) to generate cDNA. The following was added to a nuclease-free microcentrifuge tube:

Reagent	Volume for 1 reaction
oligo$(dT)_{20}$ (50µM)	1µl
RNA	11µl
dNTP (10mM)	1µl
Total volume	13µl

The sample was incubated for 5 min at 65°C and then on ice for at least 1 min. The sample was briefly centrifuged before the following were added:

Reagent	Volume for 1 reaction
5x First-Strand Buffer	4µl
0.1M DTT	1µl
RNAseOUT	1µl
SuperScript Ⅲ RT	1µl
Total volume	20µl

Samples were incubated at 50°C for 60 min, before they were heated to 70°C for 15 min to inactivate the reaction.

5.3.5.2 Amplification of Candidate Effector Genes

Primers were designed using the full length of the effector gene (minus the signal peptide) using Primer3 (http://primer3.ut.ee/), as outlined in section 2.2.6. Primer sequences for each candidate were as follows:

GPLIN_000008800	
Forward	ACCATGACGGGGAAAAGTGGAGG
Reverse	TGATCAATATTCGATTCTTTGGTTTTG
GPLIN_000926600	
Forward	ACCATGACACCTAACGATAACCCGA
Reverse	TGATCAAGCACAGAAAGGCGAAA

The sequence in red is the artificial Kozak start (ACCATG) which was added to replace the ATG start codon which was removed with the signal peptide and stop (TGA) was included to ensure no additional sequence was added when expressed *in planta* which may affect function of each of the genes.

Genes were amplified from cDNA made from both Pa1 and Lindley using KOD hot start polymerase using the following reagents and conditions:

Reagent	Volume
Water	32µl
KOD Buffer (10X)	5µl
MgSO4 (25mM)	3µl
dNTPs (2mM)	5µl
Forward primer (10µM)	1.5µl
Reverse primer (10µM)	1.5µl
KOD (1u/µl)	1µl
J2 cDNA	1µl
	50µl

95°C	2 minutes	
95°C	20 seconds	x40 cycles
X°C	10 seconds	
70°C	X	
70°C	3 minutes	
12°C	hold	

The annealing temperature for GPLIN_000008800 was 63°C with a 1 min 45 sec extension, while GPLIN_000926600 has an annealing temperature of 58°C and an elongation time of 1 min 5 sec.

PCR products were visualised on a 1% agarose gel, and bands of the anticipated size were excised and purified using QIAquick Gel Extraction kit (QIAGEN) and eluted into 30µl of elution buffer.

5.3.5.3 Amplification of attB gene

Purified effector genes underwent a second amplification to incorporate the *attB* gene which is necessary for cloning into the pDONR vector.

The primers used for this were as follows:

GPLIN_000008800	
Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTCACCATGACGGGGAAAAG
Reverse	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATCAATATTCGATTC
GPLIN_000926600	
Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTCACCATGACACCTAAC
Reverse	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATCAAGCACAGAAAG

Samples were amplified as outlined above (see 5.3.5.2), but the purified candidate effector PCR product was used as template instead of cDNA. As above, samples were visualised on a 1% agarose gel and purified.

5.3.5.4 BP Clonase Ligation into pDONR207

Purified effector PCR product was mixed with vector and BP Clonase enzyme as follows:

Reagent	Volume
attB-PCR product	7µl
Vector pDONR207 (150ng/µl)	1µl
BP Clonase	2µl

The reaction was incubated at 25°C for 1h, before 1µl of Proteinase K was added and the sample incubated for 10 min at 37°C.

5.3.5.5 Transformation into JM109 Chemically Competent Cells

1µl of each BP reaction was transformed into 50µl of JM109 chemically competent cells (Invitrogen). Samples were placed on ice for 20 min before being heat-shocked for 47 sec at 42°C, 900µl of SOC was added and cells were left to recover for 90 min at 37°C before 100µl of the transformation was plated onto LB + Gentamycin (100µg/ml) plates.

Bacterial clones were PCR tested to ensure that they contained the correct insert using the PCR conditions detailed above (see 5.3.5.2), but using the following primer pair at an annealing temperature of 56°C:

pDONR207-FOR	CGGCGGATTTGTCCTAC
pDONR207-REV	AACACCCCTTGTATTACTGTTTAT

5.3.5.6 LR Reaction into pGR106GW Destination Vector

Cloned pDONR207 inserts were Sanger sequenced to ensure the target sequence was incorporated. The following steps to clone into the destination vector were only undertaken when the anticipated target sequence had been successfully cloned. An LR reaction was set-up as follows and was incubated overnight at room temperature:

Reagent	Volume
pENTRY clone	60ng
pGR106GW (25ng/µl)	6µl
LR clonase	1µl

After incubation, 1µl of Proteinase K was added and samples were incubated for 10 min at 37°C. Samples were then transformed into JM109 cells (see 5.3.5.5) and plated onto LB + Kanamycin (50µg/ml) plates. Colonies present on the plate were then screened (see 5.3.5.2) to check for the correct insert using the following primer set:

PVX-201-Seq-F	GCAGTCATTAGCACTTCCTTAGTGAGG
PVX-201-Seq-R	CCTGAAGCTGTGGCAGGAGTTGCGC

Positive samples were used to start liquid cultures (5ml LB + Kanamycin), left to incubate overnight at 37°C, then purified using GeneJET Plasmid Purification (ThermoFisher) with a final elution volume of 50µl.

5.3.5.7 Transformation of PVX vectors into electro-competent Agrobacterium

pGR106GW clones were diluted to a concentration of 10ng/µl and 2µl were transformed into GV3101 electro-competent *Agrobacterium*. Vector/bacteria mixtures were transferred to chilled electroporation cuvettes (Biorad) and shocked for 2 ms at 1.8kV. Cells were then re-suspended in 500µl SOC media (2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 20mM glucose, 10mM $MgCl_2$) and incubated for 2h at 28°C. Transformed cells were plated onto LB + Kanamycin (50µg/ml), Rifamycin (25µg/ml) and Chloramphenicol (30µg/ml) agar dishes and incubated at 28°C for 48h. Colonies grown on the plate were screened for the plasmid insertion. One positive clone for each construct was then re-propagated on low salt LB + mannitol plates (1% tryptone, 0.5% yeast extract, 0.25% $NaCl_2$, 1% mannitol) with the above antibiotics and incubated at 28°C for 48h.

5.3.5.8 Agrobacterium-mediated Transient Expression Assay

Samples were prepared as described in section 2.2.8. Liquid cultures of pGR106GW GPLIN_000926600 clones, as well as *Agrobacterium* transformed

CRN2 and GFP genes (clones received from The James Hutton Institute Nematology Lab) were used as positive and negative controls respectively, were grown overnight at 28°C. Resistant P55/7 leaves were harvested and used for this vacuum infiltration experiment. All steps were undertaken as outlined in section 2.2.8.

5.3.6 *In situ* Hybridisation

5.3.6.1 Amplification of in situ Probe

Primers were designed to amplify a unique ~200bp region of the candidate effector gene (GPLIN_000926600) and an effector gene with known localisation as a control (GPLIN_000892900). Primer sequences for each target sequence are:

GPLIN_000926600	
Forward	ACCCGATTGTGTACGAAGGT
Reverse	AACACCAACTCGGATACCGC
GPLIN_000892900	
Forward	ACCATGTCGGCAAAACCAAACAAAAAAC
Reverse	ACAGAACGCCACTCCCTTTT

Probes were generated using the purified Lindley candidate effector in the pDONR207 vector as template. Samples underwent two amplification steps to successfully produce the *in situ* probe:

PCR 1	
Water	9.8µl
GoTaq 5X Buffer	5µl
MgCl2 (25mM)	1.5µl
dNTPS (2mM)	2.5µl
Forward Primer	2.5µl
Reverse Primer	2.5µl
DNA Template	1µl
Taq Polymerase	0.2µl
	25µl

94°C	3 minutes	
94°C	30 seconds	x35 cycles
56°C	30 seconds	
72°C	45 seconds	
72°C	3 minutes	
10°C	hold	

Amplified samples were visualised on a 2% agarose gel, and positive samples were purified using the QIAquick Gel Extraction kit, then the DNA concentration quantified. The purified sample from PCR1 was then used as template for PCR2, where only the reverse primer was used to synthesise a labelled single strand, with the same cycle program:

PCR 2	
Water	7.8µl
GoTaq 5X Buffer	4µl
MgCl2 (25mM)	1.2µl
Digoxigenin labelled dNTPS	1.5µl
Reverse Primer	4µl
DNA Template (50ng)	0.5µl
Taq Polymerase	1µl
	20µl

Successful amplification of the probe was viewed as a smear when run on a 2% agarose gel.

5.3.6.2 Sucrose Flotation of Nematodes

Lindley juvenile nematodes were hatched as outlined in section *2.1.3 Second stage juvenile hatching.* Juveniles were harvested by pipetting the TRD (containing the J2s) into a 50ml tube and centrifuging for 10 min at 2,800rpm. The excess TRD was removed and the nematode pellet was re-suspended in 5ml 50% sucrose solution, 500µl of H_2O was layered on top of the sucrose and then centrifuged for 10 min at 2,800rpm. After centrifugation, the clean nematodes could be seen as a layer between the sucrose and water; they were removed and pipetted into a 1.5ml 'low-bind' tube (Eppendorf Protein LoBind). The nematodes were centrifuged for a further 10 min at 2,800rpm, and then left to settle for 30 min. Excess liquid was removed to leave only the pelleted nematodes.

5.3.6.3 Fixation and cutting of nematodes

Two ml of 2% paraformaldehyde was added to the pelleted nematodes to re-suspend them. The nematodes were left to fix in the paraformaldehyde for 18h at 4°C. Fixed nematodes were pelleted by centrifuging for 2 min at 8,000rpm, the fixative was removed, and the nematodes re-suspended in 10x diluted fixative.

Next, the nematode suspension was pipetted onto a glass microscope slide and cut with a single edge razor blade taped to a vibrating aquarium pump to cut the nematodes into small pieces. The chopped nematodes were washed off the microscope slide into a 2ml tube using 10x diluted fixative. Cut nematodes were pelleted in the tube by centrifuging for 2 min at 8,000rpm.

5.3.6.4 Permeabilisation of nematodes

The following steps were all carried out with centrifugation steps of 8,000 rpm for 2 min. Cut nematodes were washed twice with 1ml of M9 buffer (0.6% Na_2HPO_4, 0.3% KH_2PO_4, 0.5% NaCl, 0.025% $MgSO_4.7H_2O$) then incubated in 0.5ml proteinase K solution (26.4µl proteinase k (19mg/ml) in 1ml M9 buffer) for 30 min at 22°C while rotating. Samples were then washed with 1ml M9 buffer. The nematodes were frozen for 15 min before being re-suspended in 1ml cold (-20°C) methanol, and incubated for 30 sec. The nematodes were then re-pelleted at 13,000rpm for 30 sec, the supernatant was removed, and nematodes were re-suspended in 1ml cold (-20°C) acetone and incubated for 1 min. Nematodes were pelleted at 13,000rpm for 1 min; the acetone was removed until there was approximately 100µl left in the tube. The nematodes were then rehydrated in DEPC-treated H_2O (1% DEPC) to a volume of 200µl.

5.3.6.5 Hybridisation of nematodes

Nematodes were centrifuged (8,000rpm for 2 min) and the supernatant removed. Nematode sections were then washed with 500µl hybridisation buffer (5ml deionised formamide, 2ml 20x SSC (1M NaCl, 0.3 M Na_3Citrate), 1ml 10% blocking reagent (Merck) dissolved in maleic acid buffer, 1ml, 20% SDS, 100µl 100x Denhardts, 100 µl 0.1M EDTA, 200 µl herring sperm DNA (10 mg/ml), 62.5 µl yeast tRNA (500 units/ml) and 537.5 µl DEPC treated water) to remove the excess acetone. The nematodes were re-suspended in 150µl fresh hybridisation buffer per probe being tested. The nematodes were pre-hybridised for 15 min at 50°C in the hybridisation buffer (with rotating). DNA *in situ* probes (section 5.3.6.1) were heat denatured for 10 min at 100°C and snap cooled on ice. The denatured probe was then added to the nematode suspension and hybridised overnight at 50°C while rotating.

5.3.6.6 Washing hybridised nematodes

Hybridised nematode sections were washed three times for 15 min (rotating at 50°C) with 4x SSC (1ml 20x SSC in 4ml DEPC-H_2O) followed by three 20 min washes (rotating at 50°C) with 0.1xSSC/0.1% SDS (100µl 20x SSC and 100µl 20% SDS per 20ml DEPC-H_2O).

5.3.6.7 Staining nematodes

All centrifugations were carried out for 2 min at 8,000rpm. The nematodes were washed with 200µl maleic acid buffer for 1 min (rotating at room temperature). Next, nematodes were incubated for 30 min in 1x blocking reagent in maleic acid (1:10 10x blocking reagent in maleic acid). This solution was removed and nematodes were incubated for two hours in alkaline-phosphatase conjugated anti-digoxigenin antibody diluted 1:1000 in 1x blocking reagent in maleic acid.

Nematode samples were then washed three times, each for 15 min in maleic acid buffer (rotating at room temperature), before briefly being washed in alkaline phosphatase detection buffer (0.1M Tris-HCl, 0.1M NaCl, 50mM $MgCl_2.6H_2O$, dissolved in DEPC-H_2O).

Nematode sections were stained in Nitro Blue tetrazolium (NBT) (100mg/ml nitroblue tetrazolium salt in 70% (v/v) dimethylformamide)/X-phosphatase (50mg/ml 5-bromo-4-chloro-3-indolyl phosphate, toluidinium salt, in dimethyl formamide) overnight at 4°C. Stain was prepared fresh each time it was to be used by adding 3.4µl NBT with 3.5µl X-phosphate to 1ml alkaline phosphatase detection buffer. After an overnight staining the reaction was stopped by washing the nematodes in 0.01% Tween-20.

5.3.6.8 Visualisation of in situ hybridised nematodes

Nematodes were centrifuged for 2 min at 8,000rpm and all but 50µl of the solution was removed. Sixteen microlitres of the nematode solution was pipetted onto a microscope slide, a coverslip was placed over the sample, and the edges sealed with nail varnish. Samples were visualised using a Leica DMLFS light microscope at 100x, 200x, and 400x magnification.

5.4 Results

5.4.1 Semi-quantitive PCR confirms the success of the effector enrichment

Enrichment sequencing is a technique which has been used extensively to narrow the region of a genome being re-sequenced in order to answer a specific research question, but it is not a technique widely used in the study of cyst nematodes. The enrichment baits were designed to pull out all known effector genes, but it was not certain that they would be able to do this effectively and to a high enough efficiency.

In order to identify whether the selected genes had been enriched for, as well as identifying the risk of potential off-target enrichment; semi-quantitative PCR (qPCR) was undertaken on a select sub-set of effectors known to be in the enrichment bait panel, as well as a control gene which was absent. Effector genes Gp414-2, Gp1719-1 and Gp16H02 and a non-effector gene cytochrome b (Cyt B) were used to test for enrichment efficiency.

Samples were run in parallel and results shown in Figure 5.1 indicate the success of the enrichment library preparation. For all three effector genes tested, amplification was seen at an earlier cycle number; cycle 26 vs 30 for Gp414-2, cycle 20 vs 24 for Gp1719-1, and cycle 8 compared with cycle 30 for Gp16H02. As for the CytB control (Figure 5.1, panel D) no early amplification is present for post-enriched samples; both pre- and post-enriched libraries amplify at cycle 24.

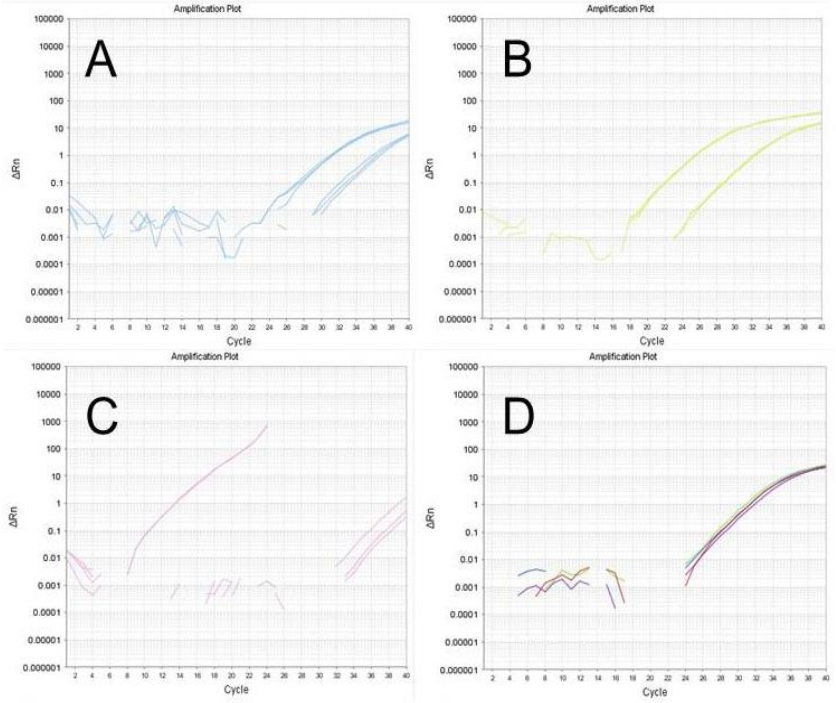


Figure 5.1 Quantitative PCR results for effector genes Gp414-2, Gp1719-1, Gp16H02 and control gene cytochrome B (CytB) in pre- and post-enriched library samples. Panel A) shows results for effector Gp414-2, B) effector Gp1719-1, C) effector Gp16H02, and D) control gene cytochrome b. In panels A-C there is a clear distinction between the cycle number where the post-enriched samples appear (A = cycle 26, B = cycle 20, C = cycle 8) compared to the pre-enriched samples (A = cycle 30, B = cycle 24, C = cycle 32). In panel D which shows CytB amplification there is no difference between the pre- and post-enriched libraries.

5.4.2 Filtering of polymorphisms allowed a shortlist of 10 genes to be identified

Research carried out into the effectors of *Phytophthora infestans* have identified that polymorphisms within avirulence genes can go some way to explaining the occurrence of virulent and more aggressive strains of the pathogen within Europe (Mantelin et al., 2017). With this in mind it can be hypothesised that instead of a presence/absence of avirulent genes between *G. pallida* pathotypes, mutations have taken place which have allowed the Pa2/3 pathotype to evade *H2*-mediated resistance while this mutation is not present in Pa1 and has rendered this pathotype avirulent.

Comparison of virulent Lindley (Pa2/3) populations with avirulent Pa1 at a nucleotide level allowed for the identification of potentially important allelic

changes which alter effector gene function, with the hope of revealing key avirulence genes required for *H2*-mediated resistance recognition.

Initial SNP findings revealed the presence of 19,873 allelic differences within Lindley and Pa1 enriched sequencing libraries compared to the *G. pallida* reference genome. This high number did not come as a surprise as *G. pallida* is known to be highly genetically diverse (Hoolahan et al., 2012).

Table 5.1 Details of putative candidate effector genes identified based on their adherence to filtering parameters. Genes chosen had differences: within protein coding regions, generated non-synonymous amino acid variants, Lindley samples contained the reference allele 100%, while Pa1 contained 100% of the alternate allele. Column heading information: scaffold – which scaffold the gene appears on in the *G. pallida* reference genome, position – the position of the relative start position of the gene, Ref allele – the allele called in the reference genome, Alternate allele – the allele called in the sequenced sample, Type – the sort of polymorphism seen (MNP – multiple nucleotide polymorphism, SNP – single nucleotide polymorphism, Del – deletion), Gene Function – the known/hypothetical role of the gene, Lindley/Pa1 – the allele observed in both copies of the gene; 0 indicates it is identical to the reference, 1 indicates it is different, Gene code – the gene identifier, Annotation – information indicating where the mutation is within the gene as well as the amino acid change.

Scaffold	Position	Ref allele	Alternate allele	Type	Gene Function	Lindley	Pa1	Gene Code	Annotation
pathogens_Gpal_scaffold_1	596862	GG	AA	mnp	HMG family member (hmg 5)	0/0	1/1	GPLIN_000008800	missense_variant|644_645GG>AA|Arg215Lys
pathogens_Gpal_scaffold_1	596877	A	G	snp	HMG family member (hmg 5)	0/0	1/1	GPLIN_000008800	missense_variant|659A>G|Lys220Arg
pathogens_Gpal_scaffold_27	265937	A	T	snp	RBP 1 protein	0/0	1/1	GPLIN_000157600	missense_variant|1076T>A|Ile359Asn
pathogens_Gpal_scaffold_27	266002	T	G	snp	RBP 1 protein	0/0	1/1	GPLIN_000157600	missense_variant|1011A>C|Leu337Phe
pathogens_Gpal_scaffold_27	266009	T	C	snp	RBP 1 protein	0/0	1/1	GPLIN_000157600	missense_variant|1004A>G|Asp335Gly
pathogens_Gpal_scaffold_27	266451	T	C	snp	RBP 1 protein	0/0	1/1	GPLIN_000157600	missense_variant|869A>G|Asp290Gly
pathogens_Gpal_scaffold_27	266515	A	G	snp	RBP 1 protein	0/0	1/1	GPLIN_000157600	missense_variant|805T>C|Phe269Leu
pathogens_Gpal_scaffold_131	74775	A	T	snp	rbp 1 protein	0/0	1/1	GPLIN_000509600	missense_variant|172A>T|Thr58Ser
pathogens_Gpal_scaffold_218	111330	C	A	snp	RBP 1 protein	0/0	1/1	GPLIN_000697500	missense_variant|420C>A|Asp140Glu
pathogens_Gpal_scaffold_263	7748	T	C	snp	rbp protein	0/0	1/1	GPLIN_000785600	missense_variant|491A>G|Glu164Gly
pathogens_Gpal_scaffold_263	7808	TCAAC	TC	del	rbp protein	0/0	1/1	GPLIN_000785600	inframe_deletion|427_429delGTT|Val143del
pathogens_Gpal_scaffold_263	7820	T	G	snp	rbp protein	0/0	1/1	GPLIN_000785600	missense_variant|419A>C|Asn140Thr
pathogens_Gpal_scaffold_275	60191	C	A	snp	RBP 1 protein	0/0	1/1	GPLIN_000803200	missense_variant|133C>A|Arg45Ser
pathogens_Gpal_scaffold_275	60201	T	C	snp	RBP 1 protein	0/0	1/1	GPLIN_000803200	missense_variant|143T>C|Met48Thr
pathogens_Gpal_scaffold_275	60224	TC	GG	mnp	RBP 1 protein	0/0	1/1	GPLIN_000803200	missense_variant|166_167TC>GG|Ser56Gly
pathogens_Gpal_scaffold_275	60414	G	A	snp	RBP 1 protein	0/0	1/1	GPLIN_000803200	missense_variant|284G>A|Arg95Lys
pathogens_Gpal_scaffold_275	60568	A	G	snp	RBP 1 protein	0/0	1/1	GPLIN_000803200	missense_variant|382A>G|Thr128Ala
pathogens_Gpal_scaffold_275	60602	C	A	snp	RBP 1 protein	0/0	1/1	GPLIN_000803200	missense_variant|416C>A|Ala139Glu
pathogens_Gpal_scaffold_361	11337	T	C	snp	gland protein	0/0	1/1	GPLIN_000926600	missense_variant|412A>G|Ser138Gly
pathogens_Gpal_scaffold_477	29449	C	T	snp	rbp 1 protein	0/0	1/1	GPLIN_001058700	missense_variant|164G>A|Arg55His
pathogens_Gpal_scaffold_651	41018	T	G	snp	transcribed hypothetical protein	0/0	1/1	GPLIN_001199500	missense_variant|953A>C|Asn318Thr
pathogens_Gpal_scaffold_651	41024	T	C	snp	transcribed hypothetical protein	0/0	1/1	GPLIN_001199500	missense_variant|947A>G|Asp316Gly
pathogens_Gpal_scaffold_1862	196	G	T	snp	rbp 1 protein	0/0	1/1	GPLIN_001446300	missense_variant|421C>A|His141Asn

Filtering was subsequently carried out based on the following criteria: presence of polymorphism in exon, alteration of amino acid encoded for, and the appearance of the reference or the alternate allele (100% reference in Lindley and 100% alternate in Pa1) and this allowed the list to be decreased to 23 polymorphisms in 10 genes (Table 5.1). The annotated function of these effector gene is heavily weighted toward members of the RBP1 gene family which is unsurprising as it is a member of the large SPRYSEC gene family which is known to rapidly mutate to evade recognition by the pathogen (Sacco et al., 2009). The other gene types pulled out are hypothesised to be effector genes based on similarity to known effectors from other plant-parasitic nematodes, but little work has been done to determine whether they are true effector genes.

5.4.3 Identification of signal peptides and downstream candidate effector analysis

As some genes are only hypothesised to be effectors based on conserved domain structure or amino acid sequence similarity, rather than through functional testing, further analysis to determine whether a gene could be a potential effector candidate was required. The amino acid sequence for each candidate was analysed using SignalP (http://www.cbs.dtu.dk/services/SignalP/) to determine whether a predicted signal peptide was present which would allow the effector protein to be secreted from the nematode into the plant (Table 5.2).

Table 5.2 Output of SignalP analysis of putative candidate effector genes.

Gene	Signal Peptide
GPLIN_000008800	No
GPLIN_000157600	No
GPLIN_000509600	No
GPLIN_000697500	No
GPLIN_000785600	No
GPLIN_000803200	No
GPLIN_000926600	Yes, cleavage between position 17 and 18
GPLIN_001058700	No
GPLIN_001199500	Yes, cleavage between position 20 and 21
GPLIN_001446300	No

Eight out of ten of the identified candidate proteins did not contain signal peptides which suggests that they may not be true effector proteins.

Problems during cloning of candidates meant GPLIN_001199500 was omitted and so from this list, two genes were taken forward for transient assays, GPLIN_000008800 and GPLIN_000926600. Both genes were searched for in the effector bait list and GPLIN_000008800 was not present. BLAST (Basic Local Align Search Tool) searching of GPLIN_000008800 against the effector bait list revealed that its closest match was known effector GPLIN_000372100. A second BLAST search of GPLIN_000372100 against the entire NCBI database revealed that it contains both an HMG box domain as well as a SPRY domain (analysis from SignalP), a domain which can be found in many non-effector proteins.

Gene GPLIN_000008800 contains an HMG box (Table 5.1) which is similar in sequence to the 5' end of GPLIN_000372100, and the presence of a 3' SPRY domain in GPLIN_000372100 is similar to that found in a 'real effector'. Due to the presence of these domains GPLIN_000008800 was captured during the enrichment process. However, subsequent analysis shows that this sequence does not have a signal peptide and does not have a similar domain structure as other known effectors. For these reasons, this sequence was also removed from the list.

Similar analysis was carried out on GPLIN_000926600 and it was found to be a putative effector with a signal peptide (Table 5.2). The final analysis to ensure that it is an effector was through expression analysis. Cyst nematodes penetrate and infect host plants during the parasitic J2 phase. Due to this it is expected that genes required for evasion of the host immune response, migration, and initiation of syncytial development would be highly expressed at this stage compared to at the egg or adult stages. Table 6.3 depicts the expression profile of GPLIN_000926600 compared to GPLIN_000008800 (as a putative non-effector control). This analysis shows that GPLIN_000926600 is highly expressed at the preparasitic J2 stage (highlighted in yellow); with a sharp decrease in expression to almost 0 by 7dpi.

Conserved domain analysis as well as expression data and the presence of a signal peptide leads us to the conclusion that GPLIN_000926600 is most probably a true effector gene and that GPLIN_000008800 is not.

Table 5.3 Candidate gene expression for GPLIN_000008800 and GPLIN_000926600. Highlighted in yellow is the parasitic J2 stage where effector genes are normally most highly expressed. Effector gene GPLIN_000926600 has an expression profile of between 97.70 and 170.74 compared to GPLIN_000008800 which is between 14.5 and 16.56.

	egg.4912_1	egg.6566_6	J2.5417_7	J2.6566_5	7dpi.6197_1	7dpi.6797_6_2	14dpi.5145_2	14dpi.6985_8	21dpi.3570_6	21dpi.6197_2	28dpi.3251_3	28dpi.6197_3	35dpi.3570_7	35dpi.6197_5	male.5145_1	male.6797_6_1
GPLIN_000926600	1.68	1.74	97.70	170.74	0.51	2.12	0.27	0.23	0.00	0.00	0.00	0.13	0.43	0.00	2.15	0.00
GPLIN_000008800	6.72	31.31	14.50	16.56	6.81	5.84	5.55	6.44	1.54	4.12	1.16	2.26	2.01	1.82	2.85	4.00

The final step in the verification of the remaining candidate gene (GPLIN_000926600) was ensuring that the SNP identified during enrichment (Table 5.1) was real and present in the Pa1 population. The effector gene was cloned in both Lindley and Pa1, sequenced, and translated (Figure 5.2), revealing that the mutation at position 412$^{A>G}$ which generated a non-synonymous amino acid change from serine to glycine was present.

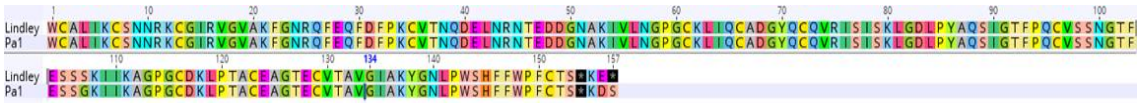


Figure 5.2 Protein translation of candidate effector gene GPLIN_000926600. Amino acid sequence is identical in both virulence Lindley and avirulent Pa1 except for the SNP (A>G) which mutates a serine (S) to a glycine (G) at amino acid position 107.

Analysis of putative candidate genes via SignalP, expression profile, and validation of polymorphisms reduced the number of *H2 avr* candidate genes to GPLIN_000926600. This gene was subsequently taken forward for *in situ* hybridisation and functional testing to try to elucidate its' role as an avirulence gene.

5.4.4 *In situ* hybridisation identifies a novel localisation of GPLIN_000926600

Any effector gene involved in nematode-host interactions is likely to be expressed and secreted from the pharyngeal glands. To test this, antisense cDNA probes

for the candidate effector gene GPLIN_000926600 and control GPLIN_000892900, a SPRYSEC with known localisation to the dorsal pharyngeal gland (Mei et al., 2015) were synthesised for *in situ* hybridisation of J2s. The control SPRYSEC probe specifically hybridised to the dorsal gland (Figure 5.3 A-D) as was expected.

Figure 5.3 *In situ* hybridisation of digoxigenin-labelled GPLIN_000892900 DNA probe in *G. pallida* J2s. The probe hybridised to the dorsal gland of second-stage juveniles.

During annotation of the *G. pallida* genome, the candidate avirulence effector GPLIN_000926600 was annotated as a 'gland protein' (Table 5.1) due to the similarity to an identified effector of another nematode (Table 5.2). However, *in situ* hybridisation demonstrated that instead of either dorsal or subventral pharyngeal gland hybridisation, candidate effector GPLIN_000926600 hybridises to the area beneath the metacorpal bulb, just below the stylet of the J2 (Figure 5.4 A-D).

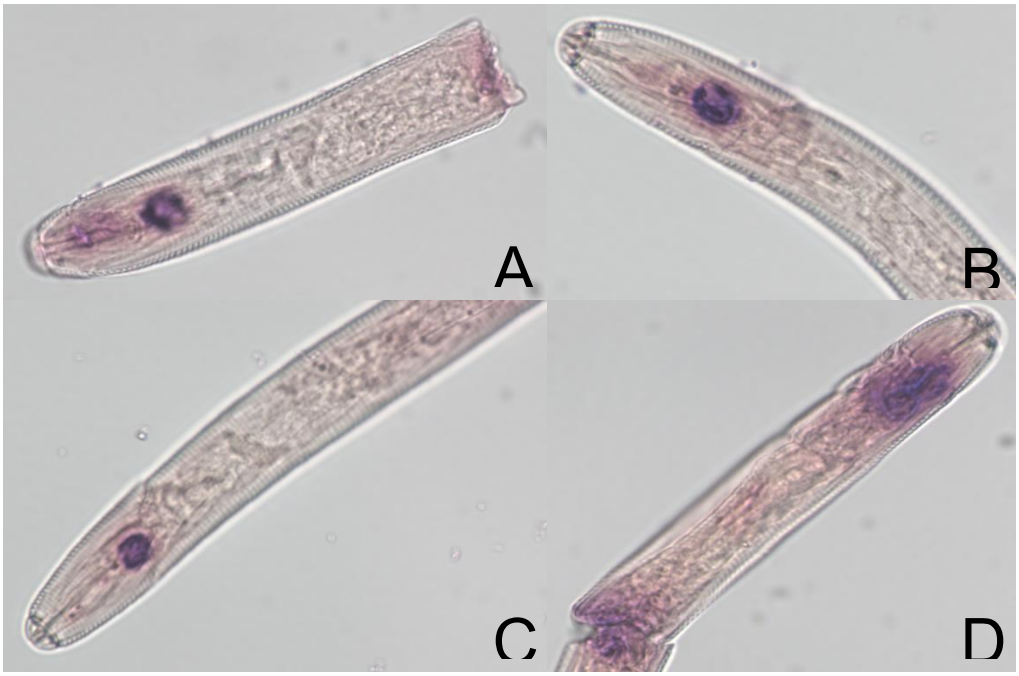


Figure 5.4 *In situ* hybridisation of digoxigenin-labelled GPLIN_000926600 DNA probe in *G. pallida* J2s. The probe hybridises to the metacorpal bulb at the base of the stylet of second-stage juveniles.

5.4.5 Transient Expression Assay

The functional test to investigate whether a candidate gene is indeed an avirulence gene is through its' co-expression with is potential cognate *R* gene. To test candidate effector GPLIN_000926600, the gene was cloned and transformed into *Agrobacterium,* before being vacuum infiltrated into *H2*-containing P55/7, alongside control samples CRN2 (a CRINKLER effector protein known to cause a cell death response), GFP, and the Lindley (Pa2/3) homolog of GPLIN_000926600. Each transient expression had 10 replicates carried out over two independent repeats with varying results (Figure 5.5). The positive control CRN2 showed a successful cell death response in five of the replicates (50%), while negative control GFP exhibited a death response in one replicate (10%). The results between Lindley and Pa1 were surprising, with Lindley cloned GPLIN_000926600, which

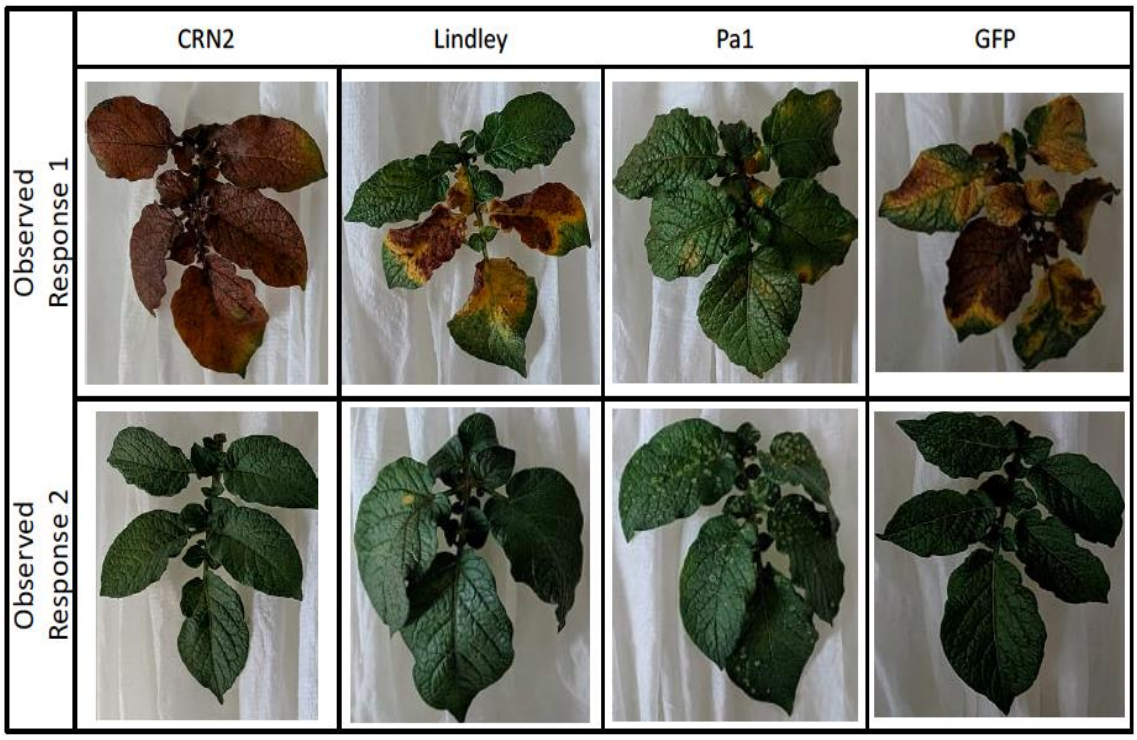


Figure 5.5 transient expression via vacuum infiltration of candidate effector and controls at 10dpi. Column 1 shows results of positive control CRN2, column 2 are the results of GPLIN_000926600 clones from Pa2/3 Lindley. Column 3 show results for candidate avirulence effector GPLIN_000926600 cloned from Pa1, and column 4 shows results for GFP. Each row shows a differing response observed during the experiment. All transient assays were undertaken using resistant cultivar P55/7, at an OD_{600nm} of 0.5.

hypothetically should cause no response as it should be the virulent form, causing a cell death response in 7/10 replicates, while Pa1 cloned GPLIN_000926600 caused cell death in 5/10 replicates. Interestingly, the Pa1 tested replicates which did not show a cell death response did show a level of chlorosis on their leaves, which could potentially be in response to the GPLIN_000926600 infection.

5.5 Discussion

5.5.1 Semi-quantitive PCR verified that effector genes were the target of the enrichment study

Non-synonymous polymorphisms allow for the diversification of effector genes allowing pathogens to evade recognition by plant host receptors (Ma and Guttman, 2008). Using this as a hypothesis, Pa1 and Lindley (Pa2/3) populations underwent effector gene enrichment prior to re-sequencing to identify

polymorphisms between the two which may control *H2*-mediated recognition in Pa1 but non-recognition in Lindley.

Analysis of the late blight oomycete *P. infestans* has shown that the characteristic RXLR and Crinkler motifs can be used to identify potential effector genes, some of which may be avirulence genes (Haas et al., 2009). To date, no single characteristic motif has been identified in cyst nematode effector genes. However, recent research has identified promoter motifs which are associated with expression in the pharyngeal gland cells and which may therefore indicate a gene as an effector. In cyst nematodes a 6bp dorsal gland box (DOG box) element was identified upstream of dorsal gland effectors, and genes which had multiple DOG boxes in their promoter region were more likely to encode signal peptides for secretion, a requirement for an effector protein (Eves-van den Akker et al., 2016). Similar work on the PPN *B. xylophilus* identified a STATAWAARS promoter motif which is associated with genes expressed in the pharyngeal glands, including effectors (Espada et al., 2018). Before this work into promoter regions, putative effector genes have been identified through other methods; for instance, the presence of a signal peptide and expression at parasitic stages, or similarity to other known effectors, coupled with *in situ* hybridisation to confirm expression in the gland cells. The enrichment bait library used for this experiment was compiled based on a list of candidate effectors identified following an analysis of the genome sequence of *G. pallida* (Thorpe *et al.,* 2014). The results shown in Figure 5.1 tested the enrichment of a select sub-set of effector genes, and showed the initial success of the experiment. However, the qPCR was only a baseline value of enrichment success as only a very small number of target genes were tested. The positive results obtained for the tested genes was taken as being indicative of success of the entire enrichment experiment.

5.5.2 Downstream analysis of sequence data revealed false positives

The enrichment allowed for the identification of almost 20,000 sequence variants within the gene set used. To decrease this vast quantity of variants to a testable sub-set required a set of stringent criteria to be established. Any sequence variant

type (SNP, MNP, INDEL) was accepted but it was required to occur within the CDS (coding DNA sequence), rather than in an intron which reduced the number of variants to examine by 2,170. Only choosing candidates which had 100% reference or alternate allele minimised the chance that the polymorphism was an error introduced through sequencing, and variants which did not change the amino acid sequence were discarded as they were considered unlikely to alter the protein function.

The 10 candidates which remained after this filtering underwent further analysis to determine their ability to be secreted from the nematode (through the presence of a signal peptide). Out of the 10 candidate genes only two (GPLIN_000926600 and GPLIN_001199500) contained a signal peptide. The other 8 sequences are likely to have been enriched as a result of their similarity to effectors that were represented in the bait list; for example, any sequence with a SPRY domain is likely to be enriched whether or not it is an effector. Since the discovery that the *G. pallida* resistance gene *Gpa2* was found to be activated by RBP1 (Ran binding protein 1), an effector gene with a SPRYSEC domain (Sacco et al., 2009), there has been a tendency for all SPRY-domain containing proteins to be labelled as effectors, whether they are true effectors or not (Kikuchi et al., 2017). Through this enrichment this fact has been highlighted, and future research should take extra care to analyse the dataset to ensure that any identified candidates are verified effector genes.

5.5.3 *In situ* hybridisation revealed non-gland expression of candidate effector

Effectors secreted through the stylet are the most widely studied in cyst and root-knot nematodes. Stylet-secreted effectors are produced in the dorsal and two subventral pharyngeal gland cells (Figure 5.6), are secreted into the metacarpal bulb chamber from the gland cells by exocytosis, and are delivered into the plant via the stylet (Jaouannet and Rosso, 2013). Secreted effectors fall into one of three categories: 1) CWDEs and cell wall loosening proteins, 2) feeding site inducing, developing, and maintaining proteins, or 3) plant resistance response

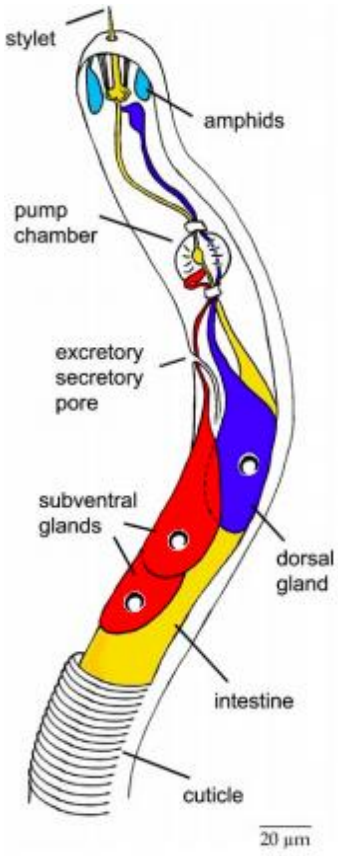


Figure 5.6 Diagram of a J2 endoparasitic nematode (Vanholme et al., 2004)

suppressors (Ali et al., 2017). The time at which the effector is required dictates which of the pharyngeal gland cells it is produced and secreted from. The subventral glands are highly metabolically active during initial root penetration, migration, and feeding site development, but their activity decreases during the maturation of the feeding site whereupon the dorsal gland activity increases (Hussey and Mimms, 1990). Immunolocalisation research carried out by Wang *et al* (1999) was able to verify this gland specific secretion pattern when they demonstrated that CWDEs originated from the two subventral glands.

It has long been thought that the pharyngeal glands were the lone site of effector protein synthesis and secretions, however the amphids may also produce effectors. The amphids are the primary sense organ within the nematode and it was previously thought that their role was limited to support during migration; however, recent studies have identified a family of hypervariable (HYP) effectors that are produced in the amphidial gland cells and secreted into the apoplast at the head of feeding nematodes (Eves-van den Akker *et al.,* 2014).

For this experiment digoxigenin-labelled anti-sense cDNA was synthesised from *G. pallida* tissue and utilised in an *in situ* hybridisation reaction. The control experiment using a known SPRYSEC effector GPLIN_000892900 hybridised to the dorsal gland which is in line with published data (Rehman et al., 2009, Mei et al., 2015). Hybridisation results for candidate effector GPLIN_000926600 unexpectedly did not localise to either pharyngeal gland, but instead hybridised near the metacorpal bulb at the base of the stylet. This area between the metacorpal bulb and stylet is the site of the esophagus and is densely packed with muscle fibres (Agrios, 2005), and to the best of knowledge, no effector proteins have ever been found to hybridise to this region of the nematode before. It is uncertain at this time exactly what structure in the nematode the candidate is

expressed in. It may be present in the muscle cells that are linked to the contraction and pumping of the stylet. Several neuronal cells that end in the anterior region of the nematode have substantial cell bodies in this region and it is possible that the gene is expressed in one of these. The gene may therefore not be an effector.

5.5.4 Transient expression of a candidate avirulence effector yielded varying levels of cell death response

Based on the work carried out on flax (*Linum usitatissimum*) and the flax rust fungus (*Melampsora lini*), the gene-for-gene concept put forward by Flor (1971) is widely accepted as the method of *R* gene-mediated resistance; that is, an avirulent effector is recognised by its cognate *R* gene and initiates a hypersensitive cell death response. The hypothesis during the transient assay experiments was that the polymorphisms present in Pa1 GPLIN_000926600 would cause it to become an avirulent form of the gene and be recognised by the *H2 R* gene, while expression of the virulent Lindley GPLIN_000926600 variant would show no response as this form would not be detected.

The *P. infestans* Crinkler (CRN) effector family has been widely studied. These CRN genes have the ability to alter host processes and cause necrosis (Haas et al., 2009). Because of their ability to cause host cell death they are an ideal positive control to monitor the appearance and timing of a cell death response. During the transient assay experiments, there was variance in leaves showing a cell death response when infiltrated with CRN2, although half of the replicates showed a strong response, the other half appeared healthy. A potential reason for the variance in response could be due to variation in successful infiltration of the effector, or the age of leaves infiltrated, with slightly older leaves taking longer to show symptoms. All experiments were carried out over a 6-10 day period, however CRN2 induced necrosis may take up to 12 days for symptoms to become visible (Torto et al., 2003). The variance in leaf age (between 4-6 weeks) and incubation time may have affected the cell death response exhibited across the biological replicates.

A variance in cell death symptoms was seen both within and between Lindley and Pa1 cloned GPLIN_000926600. Unexpectedly a cell death response was observed with both forms of the gene, leading to the hypothesis that the gene is avirulent in both pathotypes. If both forms of the gene cause recognition, then this clearly is not the cognate effector activating the *H2* resistance pathway. One explanation for a positive result from both pathotypes could be that the concentration of transformed bacteria may have caused off-target effects which gave the phenotype of a *R* gene-mediated HR, but in fact was due to stress because of over-expression of a foreign gene. One way to avoid the use of expression assays to test whether the effector is required for successful host infection would be to utilised RNAi (RNA interference), through nematode soaking, in order to silence the effector gene and monitor changes in the success/failure of infection (Elling and Jones, 2014)

The research in this chapter aimed to identify candidate avirulence effectors which have the potential to induce *H2*-mediated HR through the use of 'effectorome' enrichment. The most important issue faced during this research was the potential for the genomic sequence of the reference Lindley population and Pa1 being too dissimilar to successfully identify polymorphisms which underlie the avirulence phenotype. The intra- and inter-pathotype sequence diversity may have been a bigger issue than if the sequences which were pulled out were real effectors or not. If this experiment was to be repeated it may be advantageous to first generate a pathotype-specific reference genome in order to create a basis for whether a sequenced polymorphism is a true candidate for an avirulence phenotype, or population specific mutation. The enrichment was mostly a success, gaining a much greater read depth over the target effector sequences, however more caution will be required in future to ensure that genes determined as effectors are truly effectors. The one candidate effector which was identified during the analysis did not localise to either the dorsal or sub-ventral glands, but rather the cells directly below the stylet region. Transient expression revealed that it indeed causes a cell death response when infiltrated into *H2*-containing P55/7, but the thought-to-be virulent Lindley form also gives a similar response. The presented research shows that GPLIN_000926600 is probably not

the cognate avirulence gene which activates the *H2* resistance pathway; however, this first attempt at identifying putative candidate effector genes via enrichment has shown the wealth of information which can be gleaned from the technique, while also highlighting some potential issues.

6. General Discussion

Potato is the most important non-cereal food crop globally, and although its' growth and consumption has seen a decrease in the West recently, this crop has seen a consistent rise in growth in developing countries and Asia (FAO, 2012). Intensification in food production creates a greater need for pest and pathogen control strategies, and with the withdrawal of fumigant and non-fumigant nematicides, due to their detrimental health and environmental effects (Directive 91/414/EEC), the demand for resistant cultivars has intensified. Identifying and incorporating resistance sources from wild relatives of cultivated potato has been shown to be an effective method to introduce novel forms of resistance.

The discovery, breeding, and deployment of the *H1* gene in commercial cultivars from *S. tuberosum* ssp. *andigena* had a huge effect on controlling British populations of *G. rostochiensis* (Ross, 1986). Although populations of *G. rostochiensis* were reduced, the widespread growing of resistant cultivars allowed *G. pallida* populations to multiply unchallenged, as these cultivars were susceptible to the white potato cyst nematode. One single major gene in potato has not yet been found which has the ability to control all populations of *G. pallida* which are present in British fields therefore it is proposed that the best course of action for durable control of this pathogen is through the deployment of cultivars which contain several sources of resistance. As well as having cultivars which potentially control all populations of *G. pallida*, broad spectrum resistance would inhibit selection of virulent populations which could overcome the few major resistance sources that we currently have.

The research in this thesis aimed to map the genomic and chromosomal location of the functional *H2* resistance gene in tetraploid potato using the DM genome as a reference. This then led to the identification of a 0.8Mb region of chromosome 5 being the location of the *H2* gene. In addition to mapping the *H2* gene, this research aimed to identify a putative candidate Pa1 avirulence gene which could be functionally tested to observe whether it initiated a hypersensitive response.

6.1 dRenSeq Analysis

In order to determine for certain that *H2* is a novel resistance gene which is controlling the resistance phenotype seen in the Picasso x P55/7 cross, dRenSeq analysis was undertaken to identify the *R* genes already present in the susceptible and resistant parental background. During analysis it was discovered that Picasso contained resistance to *G. pallida* (*Gpa2*), *P. infestans* (*R1, R3a,* and *R3b*), and PVX (*Rx*). Conversely, mapping of P55/7 Illumina reads did not identify the presence of any known pathogen *R* genes. Normally when trying to identify or map an *R* gene, material which is known to be or suspected to contain the target gene is used. The *R*-gene-containing material is crossed with a susceptible variety to produce a segregating population which is subsequently used for mapping. DRenSeq allowed for the identification of orthologs of known *R* genes without the need for pathogen screening assays. Mapping Picasso Illumina reads to a known set of resistance genes revealed the presence of five *R* genes which were unknown to be present at the start of the experiment. Based on these data it can be argued that the resistance phenotype which is observed within the F1 progeny when infected with Pa1, as well as the phenotype observed in P55/7 is due to *H2*-mediated resistance. However, mapping of reads only reveals the hypothetical presence of a functional *R* gene, and *in planta* testing was still required to corroborate the findings of the dRenSeq mapping.

Mapped RenSeq reads determined the presence of the Potato Virus X resistance gene *Rx* in Picasso as well as resistances to *G. pallida* and *P. infestans* (Figure 3.4). In order to corroborate the results which identified the presence of *R* genes in Picasso and their absence in P55/7, functional testing of resistances was required. The preferred method of functional testing in potato plants is the utilisation of virus-mediated transient infection. This method uses PVX as a vector to transiently express avirulence genes through wound inoculation of *R. rhizobacter* onto plant leaves. The transformed bacteria allows the transfer and systemic spread of viral particles throughout the plant leaf (Kanneganti et al., 2007). However, due to the perceived presence of *Rx* in Picasso, carrying out PVX-mediated infection would cause the activation of the *Rx* gene pathway causing a HR and a false positive cell death response to be observed on plant

leaves. To verify that *Rx* is truly present and functional the avirulent coat protein (Cp) was expressed during vacuum infiltration, and as expected a cell death response was initiated on plant leaves, proving that Picasso contains PVX resistance, and that undertaking virus-mediated transient infection would have caused false positive results to be observed in all *R:Avr* interactions tested. Exploiting dRenSeq technologies allowed for the identification of *R* genes which were before now unknown to be present in the susceptible Picasso. Exploring the resistances already present and functional made sure the correct techniques could be employed to test the hypothesis.

6.2 Vacuum infiltration for transient expression in potato

Research by Flor (1971) identified a gene-for-gene relationship between *R* genes and the cognate avirulence protein. This relationship can be seen between the *G. pallida* resistance gene *Gpa2* and the avirulence protein RBP1, where the RBP1 protein triggers a HR in *Gpa2* containing potato cultivars (Sacco et al., 2009). Ideally, an avirulence factor and a *R* gene are co-expressed in the model plants *Arabidopsis* ssp. or *Nicotiana benthamiana* to test the relationship between the two, however this cannot be done when the *R* gene is unknown as in the case of *H2* resistance, or when the functionality of a gene is to be tested in potato. Utilisation of bacteria-mediated infiltration of plant material allows for the expression of genes *in planta* in order to observe the cell death response. Infiltration of proteins using a blunt end syringe is routinely used to transfect into plants, however the structure of potato leaves greatly reduced the efficiency of infiltration, making it almost impossible to 'push' the bacterial suspension into the plant. Vacuum infiltration allowed for the successful transient leaf expression of avirulence proteins *in planta*. Expression via vacuum infiltration was limited by the number of replicates which could be carried out due to the entire leaflet requiring to be infiltrated with a single transformation, which differs from syringe infiltration which allows for multiple transformations to be expressed on a single leaflet. A larger number of leaves are required to undertake the experiment; however, the results show a very definite susceptible or resistant phenotype on a whole leaf level. Implementing the use of vacuum infiltration allowed for the

expression of avirulence proteins across the entirety of the potato leaf, and effectively displayed the gene-for-gene cell death response expected between cognate *R:Avr* pairs. This infiltration method has shown an efficient way to transiently express proteins in potato leaves.

6.3 Mapping of the *H2* resistance gene

The first step in the mapping of any *R* gene is the production of a F1 segregating mapping population. Dunnett (1961) first explored *H2* resistance by carrying out crosses between diploid *S. multidissectum* and tetraploid *S. tuberosum*. His work confirmed that *H2* is indeed a dominant resistance gene; however the results were inconclusive as to whether it is a simplex or duplex gene. Crossing Picasso, which is known to be Pa1 susceptible, with P55/7 which is known to be resistant generated a F1 progeny which had a 0.8:1 resistant: susceptible ratio (chapter 4). This ratio is very close to the 1:1 ratio which is expected for a simplex x nulliplex cross of this type, confirming that *H2* is a simplex dominant resistance gene. Research undertaken on the tomato major resistance gene *Mi* identified that although *Mi* resistance is dominant, expressing the gene co-dominantly in a heterozygotic background reduced the effect of the resistance based on gene copy number (Tzortzakakis et al., 1998). The interaction between the two co-dominant *R* genes caused the formation of hetero-complexes which inhibited the action of the *R* genes, which when expressed separately conferred high levels of resistance to the nematode pathogen. Future breeding programmes could use this knowledge to research whether the stacking of multiple *R* genes with *H2* indeed creates a more resistant cultivar, or if off-target hetero-complexes causes an overall reduction in the resistance phenotype.

In order to identify the genomic and chromosomal location of the *H2* gene, the target enrichment technologies GenSeq and RenSeq were used in tandem to decrease genome complexity and focus on areas of genetic diversity. Combination of these two technologies was highly successful and allowed the area of interest to be reduced to a 4.7Mb region of chromosome 5 using 40 F1 progeny plants (20 resistant, 20 susceptible) as well as susceptible and resistant parents. The standard technique for the mapping of *R* genes and QTLs is through

the use of molecular markers, such as RFLPs and AFLPs, to construct linkage maps and localise the genetic region of the gene of interest (Kreike et al., 1993, Kreike et al., 1994, van der Voort et al., 2000a, Bryan et al., 2002b). Fortunately, through the sequencing of the potato genome (PGSC, 2011) and the knowledge that the majority of *R* genes are NB-LRR encoding (Van Der Biezen and Jones, 1998), advanced mapping techniques can be utilised. GenSeq allows for a reduction in genome complexity through the targeting of single or low copy genes across the entire genome, while RenSeq targets the 755 known pathogen *R* genes (Jupe et al., 2013). Combining the two allowed for the targeting of ~1% of the coding regions of the genome and the subsequent SNP filtering of phenotype-linked genes.

Both techniques require reads to be mapped to the DM reference genome, and although this can give information about which genic regions maybe used in further analysis, the DM reference is dihaploid and does not fully reveal the allelic diversity which is present in potato, meaning resistant alleles are not likely to be identified this way. Mapping to a reference genome also highlights the potential problems in filtering for SNPs; setting too relaxed a mismatch rate potentially allows for unlinked SNPs to be called as significant, conversely a too stringent mismatch threshold risks important SNPs being disregarded. A mismatch threshold of 3% was employed as this gave the most realistic number of SNPs (Table 4.3 and 4.4) for further analysis.

Identifying the genomic location of the *H2* gene begins with its chromosomal localisation, which was found to be chromosome 5 (chapter 4). In terms of nematode resistance, both *H1* (Gebhardt et al., 1993) and *GroVI* (Jacobs et al., 1996), genes active against *G. rostochiensis* have been mapped to chromosome 5 along with the major effect QTL *Gpa5* active against *G. pallida* (van der Voort et al., 2000a). In total 5 major resistance genes active against *P. infestans* (*R1* (Leonards-Schippers et al., 1992)), PVX (*Rx2* (Ritter et al., 1991) and *Nb* (De Jong et al., 1997)) have been mapped to chromosome 5, leading to it being labelled as a hot spot from pathogen resistance. It was therefore unsurprising that *H2* also mapped to chromosome 5.

In order to fine map the location of *H2* on chromosome 5, a larger number of progeny plants was required. Mapping efforts using the initial 40 progeny and parents was enough to refine the area of interest to 4.7Mb at the distal end of chromosome 5. Increasing the mapping population to over 600 plants allowed the area of interest to be refined to a 0.8Mb region which allowed for the identification of a candidate region. As with the first round of mapping it was from the recombinants that the most information was gleaned. Utilising both the GenSeq and RenSeq designed KASP markers on the expanded population gave a level of resolution that could not have been gained from the initial population.

The major outcome of the mapping was the identification of candidate gene RDC0001NLR0076 in P55/7 which is in coupling with the functional *H2* gene, meaning that it is either the *H2* gene or it is closely homologous to it. Using the DM genome as a reference is only a guide and the genetic interval in P55/7 may contain multiple tandemly duplicated genes which would all have the potential for being the *H2* gene.

A next step in the cloning of the functional *H2* gene would be the synthesis of a bacterial artificial chromosome (BAC) library based on resistant P55/7. The BAC library could be screened using RDC0001NLR0076 as a probe. Positive clones could have their ends sequenced to verify their location within the genome, and primer pairs could be used to sequence along the length of the chromosome to identify the gene. Screening of clones could then be done using KASP marker ST04_03ch05_2202842 to verify the resistant and susceptible alleles. Candidate clones could then be taken forward for complementation analysis through transformation into *Agrobacterium* to produce transgenic lines. Transformed potato lines could then be tested for their resistance to *G. pallida* pathotype Pa1. Observation of a resistant phenotype would confirm the functional *H2 R* gene.

In tandem with this, a bespoke reference genome of P55/7 could be generated utilising the long-read technology of PacBio sequencing (Pacific Biosciences Inc.). The creation of a bespoke reference would allow the region of interest to be studied in-depth, and would allow for the identification of true candidate genes, as well as allowing for any homologs or pseudogenes which are present that may

have arisen through gene duplication events to be localised. Having as much knowledge as possible from as many sources as possible will help to successfully clone the *H2* gene.

6.4 Identification of Pa1 candidate effectors

The gene-for-gene hypothesis explains how pathogen avirulent effectors interact with *R* genes to initiate host responses (Flor, 1971). Based on this, nematode effector gene enrichment was undertaken to identify potential avirulence candidates which activate the *H2* resistance pathway. Genetic variation in the form of a SNP is enough to switch a virulent gene to an avirulent gene (Carpentier et al., 2012); it was this knowledge which led to the comparison of Lindley (virulent) and Pa1 (avirulent) effectors to identify genetic mutations which may cause the cell death phenotype. This comparison led to the identification of 10 effector genes (table 5.1) which have the potential to trigger a host response.

Unlike *P. infestans* secreted effectors which carry the characteristic RXLR motif (Birch et al., 2006), nematode effectors do not contain a pre-determined sequence or domain structure which would identify them as such, although recent research has identified species-specific upstream promoter elements which may help to identify potential effectors (discussed in chapter 5) (Eves-van den Akker et al., 2014b, Espada et al., 2018). Sequence analysis to identify genes with similar domains to effectors already identified, such as the SPRYSEC or CLE-like gene families (described in chapter 5) or the presence of signal peptides, risks missing groups of effectors which do not fit into the “definition” of what an effector should be. Conversely, defining a gene solely on domain similarity can generate a list containing “pseudo” effectors. In this thesis the latter occurred as the designed baits had sufficient sequence homology to hybridise to effector genes (figure 5.2), but at the same time were hybridised to sequences which contained effector domains in unrelated non-effector or pseudogenes.

During allele mining, 19,873 allelic differences were identified between the Lindley and Pa1 populations. In an attempt to reduce the number of genes to be studied a set of stringent filtering parameters was adhered to which brought the

number down to 23 differences in 10 genes. Although this list had a feasibly testable number of effector candidates, perhaps the filtering parameters needed to be altered to give a better set of results. If effectors were first filtered based on the presence of a signal peptide, which is required for secretion into the host plant, a large number of false positives would have been removed, as even if genes had been defined as effectors due to domain homology their lack of signal peptide would have removed them from the list of potential effector genes from the outset.

The attempt to identify putative *H2* avirulence candidates through effector enrichment has the potential to be a very powerful tool in the search for effector genes, indeed candidates were identified which were taken forward for functional testing. However, the pros and cons of the criteria used in the selection need to be carefully considered.

In situ hybridisation analysis revealed the candidate effector gene GPLIN_000926600 to be localised to the amphid sheath cells. This was a surprising result as has been generally observed that most parasitism-related effector proteins are localised to the dorsal or subventral esophageal glands (Haegeman et al., 2012, Hewezi and Baum, 2013). An exception is the MAP-1 putative effector of the root knot nematode *M. incognita* which localises and is secreted from the amphids (Semblat et al., 2001). Also Eves-van den Akker (2014) showed that the HYP-family of effectors are localised to the amphidal sheath cells. The results presented in this thesis which showed the localisation of GPLIN_000926600 to the amphidal sheath cells, as well as the presence of a signal peptide, indicate that this gene may indeed have a role in host manipulation for successful infection for *G. pallida*. Based on the hybridisation experiment results, the gene annotation for GPLIN_000926600 can be changed from ‘gland protein’ (Table 5.1) to ‘putative effector’.

Transient expression of GPLIN_000926600 in *H2*-containing P55/7 endeavoured to determine whether the candidate gene had the ability to generate a cell death response due to it being the cognate effector of *H2*. A cell death response was observed in 40% of the replicates where Pa1 GPLIN_000926600 was expressed,

however necrosis was also observed in 70% of the Lindley GPLIN_000926600 replicates. From these results, it is unclear if the gene does codes for an avirulent effector which activates *H2*, however it may play some role in host manipulation. The intermediate effect observed during this experiment could be due to the need for more replicates of the experiment, increasing the sample size could have the potential to make the results seen more statistically significant. Alternatively, candidate GPLIN_000926600 may be a weak effect avirulence gene which interacts with many avirulence effectors to generate a *H2*-mediated HR, rather than a smaller number of strong acting avirulence genes creating the cell death response (Fabro et al., 2011). If this is true, a clear-cut HR would not be visualised as the effector expressed in isolation would be unable to initiate a successful cell death. Future research into this gene could include it in a yeast-two-hybrid screen to determine which, if any, proteins it interacts with.

All the work carried out on GPLIN_000926600 stemmed from the data gained from the effector enrichment. This experiment produced a vast amount of information, with only the smallest fraction being analysed during this research. Future work could utilise the wealth of sequence data which has yet to be examined. The data set could be re-analysed, firstly filtering to ensure that only genes which encode signal peptides, and lack transmembrane domains are kept. Next, candidate genes could have their protein structures resolved either computationally or using crystallography techniques, nematode effectors share structural homology which may help hypothesise their function. Candidate genes which pass these filtering steps and have potentially useful or interesting functions could then be taken forward for localisation analysis and functional testing.

Through the use of sequence enrichment techniques this work identified a candidate *H2* gene in addition to mapping the gene's location to a 0.8Mb region of potato chromosome 5. Furthermore, nematode sequence enrichment generated a list of *G. pallida* sequence polymorphisms which could potentially help identify pathotype-specific avirulence genes. Mapping *R* genes allows for markers to be generated which in turn can be used in breeding programmes. Identifying the avirulence genes which are recognised by *R* genes allows field

populations to be monitored to ensure that the correct resistance sources are deployed, which will only increase in importance as resistance sources are broken and potato crop production intensifies.

7. References

Agrios, G. 2005. Plant diseases caused by nematodes. *Plant Pathology,* 4**,** 565-597.

Ali, M. A., Azeem, F., Li, H. & Bohlmann, H. 2017. Smart parasitic nematodes use multifaceted strategies to parasitize plants. *Frontiers In Plant Science,* 8**,** 1699.

Armstrong, M. R., Vossen, J., Lim, T. Y., Hutten, R. C., Xu, J., Strachan, S. M., Harrower, B., Champouret, N., Gilroy, E. M. & Hein, I. 2018. Tracking disease resistance deployment in potato breeding by enrichment sequencing. *Plant Biotechnology Journal*.

Azevedo, C., Sadanandom, A., Kitagawa, K., Freialdenhoven, A., Shirasu, K. & Schulze-Lefert, P. 2002. The Rar1 interactor Sgt1, an essential component of *R* gene-triggered disease resistance. *Science,* 295**,** 2073-2076.

Bakker, E., Achenbach, U., Bakker, J., Van Vliet, J., Peleman, J., Segers, B., Van Der Heijden, S., Van Der Linde, P., Graveland, R. & Hutten, R. 2004a. A high-resolution map of the *H1* locus harbouring resistance to the potato cyst nematode *Globodera rostochiensis*. *Theoretical And Applied Genetics,* 109**,** 146-152.

Bakker, E., Borm, T., Prins, P., Van Der Vossen, E., Uenk, G., Arens, M., De Boer, J., Van Eck, H., Muskens, M. & Vossen, J. 2011. A genome-wide genetic map of NB-LRR disease resistance loci in potato. *Theoretical And Applied Genetics,* 123**,** 493-508.

Ballvora, A., Ercolano, M. R., Weis, J., Meksem, K., Bormann, C. A., Oberhagemann, P., Salamini, F. & Gebhardt, C. 2002. The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the Leucine Aipper/NBS/LRR class of plant resistance genes. *The Plant Journal,* 30**,** 361-371.

Barone, A., Ritter, E., Schachtschabel, U., Debener, T., Salamini, F. & Gebhardt, C. 1990. Localization by restriction fragment length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Molecular And General Genetics* 224**,** 177-182.

Bendahmane, A., Kanyuka, K. & Baulcombe, D. C. 1999. The *Rx* gene from potato controls separate virus resistance and cell death responses. *The Plant Cell,* 11**,** 781-791.

Bendahmane, A., Köhm, B. A., Dedi, C. & Baulcombe, D. C. 1995. The coat protein of Potato Virus X is a strain-specific elicitor of *Rx1*-mediated virus resistance in potato. *The Plant Journal,* 8**,** 933-941.

Bendezu, I. F., Evans, K., Burrows, P. R., De Pomerai, D. & Canto-Saenz, M. 1998. Inter and intra-specific genomic variability of the potato cyst nematodes *Globodera pallida* And *G. rostochiensis* from europe and South America using RAPD-PCR. *Nematologica,* 44**,** 49-61.

Bethke, P. C., Halterman, D. A. & Jansky, S. 2017. Are we getting better at using wild potato species in light of new tools? *Crop Science,* 57**,** 1241-1258.

Bhattacharjee, S., Garner, C. M. & Gassmann, W. 2013. New clues in the nucleus: transcriptional reprogramming in effector-triggered immunity. *Frontiers In Plant Science,* 4.

Birch, P. R. J., Bryan, G., Fenton, B., Gilroy, E. M., Hein, I., Jones, J. T., Prashar, A., Taylor, M. A., Torrance, L. & Toth, I. K. 2012. Crops that feed the world 8: potato: are the trends of increased global production sustainable? *Food Security,* 4**,** 477-508.

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

Blair, L., Perry, R. N., Oparka, K. & Jones, J. T. 1999. Activation of transcription during the hatching process of the potato cyst nematode *Globodera rostochiensis*. *Nematology,* 1**,** 103-111.

Blok, V. C. & Phillips, M. S. 2012. Biological characterisation of *Globodera pallida* From Idaho. *Nematology,* 14**,** 817-826.

Blok, V. C., Phillips, M. S. & Harrower, B. E. 1997. Comparison of british populations of potato cyst nematodes with populations from continental Europe and South America Using RAPDS. *Genome,* 40**,** 286-293.

Bos, J. I., Chaparro-Garcia, A., Quesada-Ocampo, L. M., Gardener, B. B. M. & 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.

Brodie, B., Evans, K. & Franco, J. 1993. Nematode parasites of potatoes. *Plant Parasitic Nematodes In Temperate Agriculture.***,** 87-132.

Bryan, G., Mclean, K., Bradshaw, J., De Jong, W., Phillips, M., Castelli, L. & Waugh, R. 2002a. Mapping QTLs for resistance to the cyst nematode *Globodera pallida* derived from the wild potato species *Solanum vernei. Theoretical And Applied Genetics,* 105**,** 68-77.

Bryan, G. J., Mclean, K., Pande, B., Purvis, A., Hackett, C. A., Bradshaw, J. E. & Waugh, R. 2004. Genetical dissection of *H3*-mediated polygenic PCN resistance in a heterozygous autotetraploid potato population. *Molecular Breeding,* 14**,** 105-116.

Burton, W. 1989. *The Potato Longman,* London.

Byrne, J. T., Maher, N. J. & Jones, P. W. 2001. Comparative responses of *Globodera rostochiensis* and *G. Pallida* to hatching chemicals. *Journal Of Nematology,* 33**,** 195-202.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. 2009. Blast+: Architecture and applications. *BMC Bioinformatics,* 10**,** 1.

Canto Saenz, M. & De Scurrah, M. 1977. Races of potato cyst nematode in Andean regions and a new system of classification. *Nematologica,* 23**,** 340-349.

Caplan, J., Padmanabhan, M. & Dinesh-Kumar, S. P. 2008. Many NB-LRR Immune receptors: from recognition to transcriptional reprogramming. *Cell Host & Microbe,* 3**,** 126-135.

Caromel, B. & Gebhardt, C. 2011. Breeding for nematode resistance: use of genomic information. *Genomics And Molecular Genetics Of Plant-Nematode Interactions.* Springer.

Carpentier, J., Esquibet, M., Fouville, D., Manzanares-Dauleux, M. J., Kerlan, M. C. & Grenier, E. 2012. The evolution Of The Gp-*Rbp*-1 gene in *Globodera pallida* includes multiple selective replacements. *Molecular Plant Pathology,* 13**,** 546-555.

Castagnone-Sereno, P. 2002. Genetic variability of nematodes: a threat to the durability of plant resistance genes? *Euphytica,* 124**,** 193-199.

Césari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R. & Kroj, T. 2014. The MB-LRR proteins *Rga4* and *Rga5* interact functionally and physically to confer disease resistance. *The Embo Journal,* 33**,** 1941-1959.

Chen, X., Lewandowska, D., Armstrong, M. R., Baker, K., Lim, T.-Y., Bayer, M., Harrower, B., Mclean, K., Jupe, F. & Witek, K. 2018. Identification and rapid mapping of a gene conferring broad-spectrum late blight resistance in the diploid potato species *Solanum verrucosum* through DNA capture technologies. *Theoretical And Applied Genetics***,** 1-11.

Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell,* 124**,** 803-814.

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.

Commerce, U. S. D. O. 2015. *U.S. And World Population Clock* [Online]. Available: Http://Www.Census.Gov/Popclock/ [Accessed 18/2/15 2015].

Cotton, J. A., Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., Tsai, I. J., Beasley, H., Blok, V., Cock, P. J. A., Eves-Van Den Akker, S., Holroyd, N., Hunt, M., Mantelin, S., Naghra, H., Pain, A., Palomares-Rius, J. E., Zarowiecki, M., Berriman, M., Jones, J. T. & Urwin, P. E. 2014. The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology,* 15**,** 17.

Curtis, R. H. 2007. Plant parasitic nematode proteins and the host–parasite interaction. *Briefings In Functional Genomics & Proteomics,* 6**,** 50-58.

Dangl, J. L. & Jones, J. D. G. 2001. Plant pathogens and integrated defence responses to infection. *Nature,* 411**,** 826-833.

Davis, E. L., Hussey, R. S. & Baum, T. J. 2004. Getting to the roots of parasitism by nematodes. *Trends In Parasitology,* 20**,** 134-141.

De Jong, W., Forsyth, A., Leister, D., Gebhardt, C. & Baulcombe, D. 1997. A potato hypersensitive resistance gene against potato virus x maps to a resistance gene cluster on chromosome 5. *Theoretical And Applied Genetics,* 95**,** 246-252.

Delledonne, M., Zeier, J., Marocco, A. & Lamb, C. 2001. Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response. *Proceedings Of The National Academy Of Sciences,* 98**,** 13454-13459.

Devine, K. J., Dunne, C., O'gara, F. & 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**,** 637-645.

Devine, K. J. & Jones, P. W. 2001. Effects of hatching factors on potato cyst nematode hatch and in-egg mortality in soil and in vitro. *Nematology,* 3**,** 65-74.

Ding, X., Shields, J., Allen, R. & 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.

Dodds, P. N., Lawrence, G. J. & Ellis, J. G. 2001. Six amino acid changes confined to the leucine-rich repeat β-strand/β-turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. *The Plant Cell,* 13**,** 163-178.

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

Dunnett, J. 1961. Inheritance of resistance to potato root eelworm in a breeding line stemming from *Solanum multidissectum* Hawkes. *Report Of The Scottish Plant Breeding Station***,** 39-46.

Ebrahimi, N., Viaene, N., Demeulemeester, K. & Moens, M. 2014. Observations on the life cycle of potato cyst nematodes, *Globodera rostochiensis* and *G. pallida*, on early potato cultivars. *Nematology,* 16**,** 937-952.

Eitas, T. K. & Dangl, J. L. 2010. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Current Opinion In Plant Biology,* 13**,** 472-477.

Ellenby, C. 1954. Tuber forming species and varieties of the genus *Solanum* tested for resistance to the potato root eelworm *Heterodera rostochiensis* Wollenweber. *Euphytica,* 3**,** 195-202.

Elling, A. A. & Jones, J. T. 2014. Functional characterization of nematode effectors in plants. *Plant-Pathogen Interactions.* Springer.

Elmore, J. M., Lin, Z.-J. D. & Coaker, G. 2011. Plant NB-LRR signaling: upstreams and downstreams. *Current Opinion In Plant Biology,* 14**,** 365-371.

Emmond, G. & Ledingham, R. 1972. Effects of crop rotation on some soil-borne pathogens of potato. *Canadian Journal Of Plant Science,* 52**,** 605-611.

Eppo 2006. Testing of potato varieties to assess resistance to *Globodera rostochiensis* and *Globodera pallida*. *Eppo Bulletin,* 36**,** 419-420.

Espada, M., Eves-Van Den Akker, S., Maier, T., Paramasivan, V., Baum, T., Mota, M. & Jones, J. T. 2018. STATAWAARS: a promoter motif associated with spatial expression in the major effector-producing tissues of the plant-parasitic nematode *Bursaphelenchus xylophilus*. *BMC Genomics,* 19**,** 553.

Evans, K., Franco, J. & De Scurrah, M. M. 1975. Distribution of species of potato cyst nematodes in South America. *Nematologica,* 21**,** 365-369.

Eves-Van Den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A. & Szitenberg, A. 2016. The genome of the yellow potato cyst nematode, *Globodera rostochiensis,* reveals insights into the basis of parasitism and virulence. *Genome Biology,* 17**,** 124.

Eves-Van Den Akker, S., Lilley, C. J., Ault, J. R., Ashcroft, A. E., Jones, J. T. & Urwin, P. E. 2014a. The feeding tube of cyst nematodes: characterisation of protein exclusion. *Plos One,* 9**,** E87289.

Eves-Van Den Akker, S., Lilley, C. J., Jones, J. T. & Urwin, P. E. 2014b. Identification and characterisation of a hyper-variable apoplastic effector gene family of the potato cyst nematodes. *Plos Pathogens,* 10.

Eves-Van Den Akker, S., Lilley, C. J., Reid, A., Pickup, J., Anderson, E., Cock, P. J., Blaxter, M., Urwin, P. E., Jones, J. T. & Blok, V. C. 2015. A metagenetic approach to determine the diversity and distribution of cyst nematodes at the level of the country, the field and the individual. *Molecular Ecology,* 24**,** 5842-5851.

Fabro, G., Steinbrenner, J., Coates, M., Ishaque, N., Baxter, L., Studholme, D. J., Körner, E., Allen, R. L., Piquerez, S. J. & Rougon-Cardoso, A. 2011. Multiple candidate effectors from the oomycete pathogen *Hyaloperonospora arabidopsidis* suppress host plant immunity. *Plos Pathogens,* 7**,** E1002348.

Fao. 2012. *Food And Agriculture Organisation Of The United Nations, Land Resources* [Online]. Available: Http://Www.Fao.Org/Nr/Land/Databasesinformationsystems/En/ [Accessed].

Finkers-Tomczak, A., Bakker, E., De Boer, J., Van Der Vossen, E., Achenbach, U., Golas, T., Suryaningrat, S., Smant, G., Bakker, J. & Goverse, A. 2011. Comparative sequence analysis of the potato cyst nematode resistance locus *H1* reveals a major lack of co-linearity between three haplotypes in potato (*Solanum tuberosum* ssp.). *Theoretical And Applied Genetics,* 122**,** 595-608.

Finkers-Tomczak, A., Danan, S., Van Dijk, T., Beyene, A., Bouwman, L., Overmars, H., Van Eck, H., Goverse, A., Bakker, J. & Bakker, E. 2009. A high-resolution map of the *Grp1* locus on chromosome V of potato harbouring broad-spectrum resistance to the cyst nematode species *Globodera pallida* and *Globodera rostochiensis*. *Theoretical And Applied Genetics,* 119**,** 165-173.

Flor, H. H. 1971. Current status of the gene-for-gene concept. *Annual Review Of Phytopathology,* 9**,** 275-296.

Foster, S. J., Park, T.-H., Pel, M., Brigneti, G., Śliwka, J., Jagger, L., Van Der Vossen, E. & Jones, J. D. 2009. Rpi-Vnt1. 1, A Tm-22 homolog from *Solanum venturii*, confers resistance to potato late blight. *Molecular Plant-Microbe Interactions,* 22**,** 589-600.

Franco, J. 1979. Effect of temperature on hatching and multiplication of potato cyst-nematodes. *Nematologica,* 25**,** 237-244.

Franco, J. & Evans, K. 1979. Effects of daylength on the multiplication of potato cyst-nematode (*Globodera* spp.) Populations. *Nematologica,* 25**,** 184-190.

Ganal, M., Simon, R., Brommonschenkel, S., Arndt, M., Phillips, M., Tanksley, S. & Kumar, A. 1995. Genetic mapping of a wide spectrum nematode resistance gene (*Hero*) against *Globodera rostochiensis* in tomato. *Molecular Plant-Microbe Interactions,* 8**,** 886-891.

Gebhardt, C., Mugniery, D., Ritter, E., Salamini, F. & Bonnel, E. 1993. Identification of RFLP markers closely linked to the *H1* gene conferring resistance to *Globodera rostochiensis* in potato. *Theoretical And Applied Genetics,* 85**,** 541-544.

Gebhardt, C. & Valkonen, J. P. 2001. Organization of genes controlling disease resistance in the potato genome. *Annual Review Of Phytopathology,* 39**,** 79-102.

Gheysen, G. & Fenoll, C. 2002. Gene expression in nematode feeding sites. *Annual Review Of Phytopathology,* 40**,** 191-219.

Gheysen, G. & Mitchum, M. G. 2011. How nematodes manipulate plant development pathways for infection. *Current Opinion In Plant Biology,* 14**,** 415-421.

Gilroy, E. M., Breen, S., Whisson, S. C., Squires, J., Hein, I., Kaczmarek, M., Turnbull, D., Boevink, P. C., Lokossou, A. & Cano, L. M. 2011. Presence/absence, differential expression and sequence polymorphisms between Piavr2 and Piavr2-like in *Phytophthora infestans* determine virulence on *R2* plants. *New Phytologist,* 191**,** 763-776.

Golinowski, W., Grundler, F. & Sobczak, M. 1996. Changes in the structure o f *Arabidopsis thaliana* during female development of the plant-parasitic nematode *Heterodera schachtii. Protoplasma,* 194**,** 103-116.

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

Grenier, E., Fournet, S., Petit, E. & Anthoine, G. 2010. A cyst nematode 'species factory' called the Andes. *Nematology,* 12**,** 163-169.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grünwald, N. J., Horn, K., Horner, N. R., Hu, C.-H., Huitema, E., Jeong, D.-H., Jones, A. M. E., Jones, J. D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., Maclean, D., Chibucos, M. C., Mcdonald, H., Mcwalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., Van De Vondervoort, P. J. I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., Van West, P., Ristaino, J., Govers, F., Birch, P. R. J., Whisson, S. C., Judelson, H. S. & Nusbaum, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature,* 461**,** 393.

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

Handoo, Z. A., Carta, L. K., Skantar, A. M. & Chitwood, D. J. 2012. Description of *Globodera ellingtonae* n. sp.(Nematoda: *Heteroderidae*) from Oregon. *Journal Of Nematology,* 44**,** 40.

Hassan, S., Behm, C. A. & Mathesius, U. 2010. Effectors of plant parasitic nematodes that re-program root cell development. *Functional Plant Biology,* 37**,** 933-942.

Heath, M. C. 2000. Hypersensitive response-related death. *Programmed Cell Death In Higher Plants.* Springer.

Hewezi, T. & Baum, T. J. 2013. Manipulation of plant cells by cyst and root-knot nematode effectors. *Molecular Plant-Microbe Interactions,* 26**,** 9-16.

Hogenhout, S. A., Van Der Hoorn, R. A., Terauchi, R. & Kamoun, S. 2009. Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant-Microbe Interactions,* 22**,** 115-122.

Hoolahan, A. H., Blok, V. C., Gibson, T. & Dowton, M. 2012. A comparison of three molecular markers for the identification of populations of *Globodera pallida*. *Journal Of Nematology,* 44**,** 7.

Huang, S., Van Der Vossen, E. A., Kuang, H., Vleeshouwers, V. G., Zhang, N., Borm, T. J., Van Eck, H. J., Baker, B., Jacobsen, E. & Visser, R. G. 2005. Comparative genomics enabled the isolation of the *R3a* late blight resistance gene in potato. *The Plant Journal,* 42**,** 251-261.

Hussey, R. S., Davis, E. L. & Baum, T. J. 2002. Secrets in secretions: genes that control nematode parasitism of plants. *Brazilian Journal Of Plant Physiology,* 14**,** 183-194.

Hussey, R. S. & Mimms, C. W. 1990. Ultrastructure of esophageal glands and their secretory granules in the root-knot nematode *Meloidogyne incognita*. *Protoplasma,* 156**,** 9-18.

Jacobs, J. M., Van Eck, H. J., Horsman, K., Arens, P. F., Verkerk-Bakker, B., Jacobsen, E., Pereira, A. & Stiekema, W. J. 1996. Mapping of resistance to the potato cyst nematode *Globodera rostochiensis* from the wild potato species *Solanum vernei. Molecular Breeding,* 2**,** 51-60.

Jaffe, H., Huettel, R. N., Demilo, A. B., Hayes, D. K. & Rebois, R. V. 1989. Isolation and identification of a compound from soybean cyst nematode, *Heterodera glycines,* with sex pheromone activity. *Journal Of Chemical Ecology,* 15**,** 2031-2043.

Janssen, R., Bakker, J. & Gommers, F. J. 1991. Mendelian proof for a gene-for-gene relationship between virulence of *Globodera rostochiensis* and the *H1* resistance gene in *Solanum tuberosum* ssp. *Andigena* CPC 1673. *Review Of Nematology,* 14**,** 213-219.

Jaouannet, M. & Rosso, M.-N. 2013. Effectors of root sedentary nematodes target diverse plant cell compartments to manipulate plant functions and promote infection. *Plant Signaling & Behavior,* 8**,** E25507.

Jo, K.-R., Visser, R. G., Jacobsen, E. & Vossen, J. H. 2015. Characterisation of the late blight resistance in potato differential Mar9 reveals a qualitative resistance gene, *R9a*, residing in a cluster of Tm-2 2 homologs on chromosome IX. *Theoretical And Applied Genetics,* 128**,** 931-941.

Johnson, A. 1985. The role of nematicides in nematode management. *An Advanced Treatise On Meloidogyne,* 1**,** 249-267.

Johnson, D. T. 2001. Will human food supply meet the need in 2025? A global and pacific regional analysis. *Journal Of South Pacific Agriculture,* 8**,** 6-17.

Jones, J. D. & Dangl, J. L. 2006. The plant immune system. *Nature,* 444**,** 323-329.

Jones, J. D., Witek, K., Verweij, W., Jupe, F., Cooke, D., Dorling, S., Tomlinson, L., Smoker, M., Perkins, S. & Foster, S. 2014. Elevating crop disease resistance with cloned genes. *Philosophical Transactions Of The Royal Society Of London B: Biological Sciences,* 369**,** 20130087.

Jones, J. T., Haegeman, A., Danchin, E. G., Gaur, H. S., Helder, J., Jones, M. G., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J. E. & Wesemael, W. M. 2013. Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology,* 14**,** 946-961.

Jones, J. T., Perry, R. N. & Wright, D. J. 1998. *The Physiology And Biochemistry Of Free-Living And Plant-Parasitic Nematodes*.

Jones, M. G. & Goto, D. B. 2011. Root-knot nematodes and giant cells. *Genomics And Molecular Genetics Of Plant-Nematode Interactions.* Springer.

Jupe, F., Chen, X., Verweij, W., Witek, K., Jones, J. D. & Hein, I. 2014. Genomic DNA library preparation for resistance gene enrichment and sequencing (RenSeq) in plants. *Plant-Pathogen Interactions: Methods And Protocols***,** 291-303.

Jupe, F., Pritchard, L., Etherington, G. J., Mackenzie, K., Cock, P. J., Wright, F., Sharma, S. K., Bolser, D., Bryan, G. J. & Jones, J. D. 2012. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics,* 13**,** 75.

Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G. J., Maclean, D., Cock, P. J., Leggett, R. M., Bryan, G. J., Cardle, L., Hein, I. & Jones, J. D. G. 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *The Plant Journal,* 76**,** 530-544.

Kaczmarek, A., Mackenzie, K., Kettle, H. & Blok, V. C. 2014. Increasing soil temperatures will likely benefit potato cyst nematodes. *The Dundee Conference.*

*Crop Protection In Northern Britain 2014, Dundee, Uk, 25-26 February 2014***,** 265-270.

Kanneganti, T.-D., Huitema, E. & Kamoun, S. 2007. *In planta* expression of oomycete and fungal genes. *Plant-Pathogen Interactions.* Springer.

Kapila, J., De Rycke, R., Van Montagu, M. & Angenon, G. 1997. An *Agrobacterium*-mediated transient gene expression system for intact leaves. *Plant Science,* 122**,** 101-108.

Karpouzas, D. G., Hatziapostolou, P., Papadopoulou-Mourkidou, E., Giannakou, I. O. & Georgiadou, A. 2004. The enhanced biodegradation of fenamiphos in soils from previously treated sites and the effect of soil fumigants. *Environmental Toxicology And Chemistry,* 2**3,** 2099-2107.

Kiewnick, S. & Sikora, R. 2006. Biological control of the root-knot nematode *Meloidogyne incognita* by *Paecilomyces lilacinus* strain 251. *Biological Control,* 38**,** 179-187.

Korma, A. & Sigareva, D. 2012. *Results of a survey on infestation by Bursaphelenchus spp. In pine forests in Ukraine*.

Kort, J. 1974. Identification of pathotypes of the potato cyst nematode. *Eppo Bulletin,* 4**,** 511-518.

Kort, J., Ross, H., Rumpenhorst, H. J. & Stone, A. R. 1977. International scheme for identifying and classifying pathotypes of potato cyst nematodes *Globodera rostochiensis* and *G. pallida. Nematologica,* 23**,** 333-339.

Kreike, C., De Koning, J., Vinke, J., Van Ooijen, J., Gebhardt, C. & Stiekema, W. 1993. Mapping of loci involved in quantitatively inherited resistance to the potato cyst-nematode *Globodera rostochiensis* pathotype Ro1. *Theoretical And Applied Genetics,* 87**,** 464-470.

Kreike, C., De Koning, J., Vinke, J., Van Ooijen, J. & Stiekema, W. 1994. Quantitatively-inherited resistance to *Globodera pallida* is dominated by one major locus in *Solanum spegazzinii*. *Theoretical And Applied Genetics,* 88**,** 764-769.

Lam, E., Kato, N. & Lawton, M. 2001. Programmed cell death, mitochondria and the plant hypersensitive response. *Nature,* 411**,** 848-853.

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

Langmead, B. & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9**,** 357-359.

Leonards-Schippers, C., Gieffers, W., Salamini, F. & Gebhardt, C. 1992. The *R1* gene conferring race-specific resistance to *Phytophthora infestans* in potato is located on potato chromosome V. *Molecular And General Genetics* 233**,** 278-283.

Leonards-Schippers, C., Gieffers, W., Schäfer-Pregl, R., Ritter, E., Knapp, S., Salamini, F. & Gebhardt, C. 1994. Quantitative resistance to *Phytophthora infestans* in potato: a case study for QTL mapping in an allogamous plant species. *Genetics,* 137**,** 67-77.

Levine, A., Pennell, R. I., Alvarez, M. E., Palmer, R. & Lamb, C. 1996. Calcium-mediated apoptosis in a plant hypersensitive disease resistance response. *Current Biology,* 6**,** 427-437.

Li, G., Huang, S., Guo, X., Li, Y., Yang, Y., Guo, Z., Kuang, H., Rietman, H., Bergervoet, M. & Vleeshouwers, V. G. 2011. Cloning and characterization of *R3b*; members of the *R3* superfamily of late blight resistance genes show sequence and functional divergence. *Molecular Plant-Microbe Interactions,* 24**,** 1132-1142.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. 2009. The sequence alignment/map format and Samtools. *Bioinformatics,* 25**,** 2078-2079.

Lilley, C. J., Atkinson, H. J. & Urwin, P. E. 2005. Molecular aspects of cyst nematodes. *Molecular Plant Pathology,* 6**,** 577-588.

Liu, S., Kandoth, P. K., Warren, S. D., Yeckel, G., Heinz, R., Alden, J., Yang, C., Jamai, A., El-Mellouki, T. & Juvale, P. S. 2012. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature,* 492**,** 256-260.

Lokossou, A. A., Park, T.-H., Van Arkel, G., Arens, M., Ruyter-Spira, C., Morales, J., Whisson, S. C., Birch, P. R., Visser, R. G. & Jacobsen, E. 2009. Exploiting knowledge of R/Avr genes to rapidly clone a new LZ-NBS-LRR family of late blight resistance genes from potato linkage group IV. *Molecular Plant-Microbe Interactions,* 22**,** 630-641.

Lu, S.-W., Chen, S., Wang, J., Yu, H., Chronis, D., Mitchum, M. G. & 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.

Lukasik, E. & Takken, F. L. 2009. Standing strong, resistance proteins instigators of plant defence. *Current Opinion In Plant Biology,* 12**,** 427-436.

Ma, W. & Guttman, D. S. 2008. Evolution of prokaryotic and eukaryotic virulence effectors. *Current Opinion In Plant Biology,* 11**,** 412-419.

Maekawa, T., Cheng, W., Spiridon, L. N., Töller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q.-H., Micluta, M. A. & Somssich, I. E. 2011. Coiled-coil domain-dependent

homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell Host & Microbe,* 9**,** 187-199.

Manosalva, P., Manohar, M., Von Reuss, S. H., Chen, S., Koch, A., Kaplan, F., Choe, A., Micikas, R. J., Wang, X. & Kogel, K.-H. 2015. Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. *Nature Communications,* 6**,** 7795.

Mantelin, S., Thorpe, P. & Jones, J. T. 2017. Translational biology of nematode effectors. or, to put it another way, functional analysis of effectors–what's the point? *Brill*.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet. Journal,* 17**,** Pp. 10-12.

Masler, E. P. 2013. Free-living nematodes. *In:* Kastin, A. J. (Ed.) *Handbook Of Biologically Active Peptides.* Second Ed. Amsterdam: Elsevier.

Matsye, P. D., Lawrence, G. W., Youssef, R. M., Kim, K.-H., Lawrence, K. S., Matthews, B. F. & Klink, V. P. 2012. The expression of a naturally occurring, truncated allele of an α-SNAP gene suppresses plant parasitic nematode infection. *Plant Molecular Biology,* 80**,** 131-155.

Mei, Y., Thorpe, P., Guzha, A., Haegeman, A., Blok, V. C., Mackenzie, K., Gheysen, G., Jones, J. T. & Mantelin, S. 2015. Only a small subset of the SPRY domain gene family in *Globodera pallida* is likely to encode effectors, two of which suppress host defences induced by the potato resistance gene *Gpa2*. *Nematology,* 17**,** 409-424.

Meksem, K., Leister, D., Peleman, J., Zabeau, M., Salamini, F. & Gebhardt, C. 1995. A high-resolution map of the vicinity of the *R1* locus on chromosome V of potato based on RFLP and AFLP markers. *Molecular And General Genetics,* 249**,** 74-81.

Meksem, K., Pantazopoulos, P., Njiti, V., Hyten, L., Arelli, P. & Lightfoot, D. 2001. 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *rhg1* and *Rhg4* loci. *Theoretical And Applied Genetics,* 103**,** 710-717.

Milligan, S. B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P. & 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. *The Plant Cell,* 10**,** 1307-1319.

Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M. & Davis, E. L. 2013. Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist,* 199**,** 879-894.

Mulvey, R. & Stone, A. 1976. Description of *Punctodera matadorensis* n. Gen., n. Sp.(Nematoda: *Heteroderidae*) from Saskatchewan with lists of species and

generic diagnoses of *Globodera* (n. Rank), *Heterodera*, and *Sarisodera*. *Canadian Journal Of Zoology,* 54**,** 772-785.

Mundt, C. C. 2014. Durable resistance: a key to sustainable management of pathogens and pests. *Infection, Genetics And Evolution,* 27**,** 446-455.

Nickle, W. R., Golden, A. M., Mamiya, Y. & Wergin, W. P. 1981. On the taxonomy and morphology of the pine wood nematode, *Bursaphelenchus xylophilus*. *Journal Of Nematology,* 13**,** 385-392.

Nicol, J. M. T., S. J.; Coyne, D. L.; Den Nijs, L.; Hockland, S.; Tahna Maafi, Z. 2010. Current nematode threats to world agriculture. *In:* Jones, J. G. G. F., C (Ed.) *Genomics And Molecular Genetics Of Plant-Nematode Interactions.* Dordrecht: Springer.

Norton, D. & Niblack, T. L. 1991. Biology and ecology of nematodes. *Manual Of Agricultural Nematology***,** 47-71.

Paal, J., Henselewski, H., Muth, J., Meksem, K., Menéndez, C. M., Salamini, F., Ballvora, A. & Gebhardt, C. 2004. Molecular cloning of the potato *Gro1-4* gene conferring resistance to pathotype *Ro1* of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *The Plant Journal,* 38**,** 285-297.

Paulson, R. E. & Webster, J. M. 1972. Ultrastructure of the hypersensitive reaction in roots of tomato, *Lycopersicon esculentum* l., to infection by the root-knot nematode, *Meloidogyne incognita. Physiological Plant Pathology,* 2**,** 227-234.

Perry, R. 1989. Dormancy and hatching of nematode eggs. *Parasitology Today,* 5**,** 377-383.

Perry, R. N. 1997. Plant signals in nematode hatching and attraction. *Cellular And Molecular Aspects Of Plant-Nematode Interactions.* Springer.

Peters, R., Sturz, A., Carter, M. & Sanderson, J. 2003. Developing disease-suppressive soils through crop rotation and tillage management practices. *Soil And Tillage Research,* 72**,** 181-192.

Pgsc 2011. Genome sequence and analysis of the tuber crop potato. *Nature,* 475**,** 189-195.

Phillips, M., Bradshaw, J. & Mackay, G. 1994. Inheritance of resistance to nematodes. *Potato Genetics.***,** 319-337.

Phillips, M. S., Harrower, B. E., Trudgill, D. L., Catley, M. A. & Waugh, R. 1992. Genetic-variation in british populations of *Globodera pallida* as revealed by isozyme and DNA analyses. *Nematologica,* 38**,** 304-319.

Phillips, M. S. & Trudgill, D. L. 1983. Variations in the ability of *Globodera pallida* to produce females on potato clones bred from *Solanum vernei* or *Solanum tuberosum* ssp *andigena* CPC 2802. *Nematologica,* 29**,** 217-226.

Picard, D., Sempere, T. & Plantard, O. 2007. A northward colonisation of the andes by the potato cyst nematode during geological times suggests multiple host-shifts from wild to cultivated potatoes. *Molecular Phylogenetics And Evolution,* 42**,** 308-316.

Plantard, O., Picard, D., Valette, S., Scurrah, M., Grenier, E. & Mugniery, D. 2008. Origin and genetic diversity of western european populations of the potato cyst nematode (*Globodera pallida*) inferred from mitochondrial sequences and microsatellite loci. *Molecular Ecology,* 17**,** 2208-2218.

Popeijus, H., Overmars, H., Jones, J., Blok, V., Goverse, A., Helder, J., Schots, A., Bakker, J. & Smant, G. 2000. Enzymology: degradation of plant cell walls by a nematode. *Nature,* 406**,** 36-37.

Quinlan, A. R. & Hall, I. M. 2010. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

Rehman, S., Postma, W., Tytgat, T., Prins, P., Qin, L., Overmars, H., Vossen, J., Spiridon, L.-N., Petrescu, A.-J. & Goverse, A. 2009. A secreted SPRY domain-containing protein (SPRYSEC) from the plant-parasitic nematode *Globodera rostochiensis* interacts with a CC-NB-LRR protein from a susceptible tomato. *Molecular Plant-Microbe Interactions,* 22**,** 330-340.

Ritter, E., Debener, T., Barone, A., Salamini, F. & Gebhardt, C. 1991. RFLP mapping on potato chromosomes of two genes controlling extreme resistance to Potato Virus X (PVX). *Molecular And General Genetics* 227**,** 81-85.

Ross, H. 1986. Potato breeding-problems and perspectives. *Fortschritte Der Pflanzenzuechtung (Germany)*.

Ruiz-Suárez, N., Boada, L. D., Henríquez-Hernández, L. A., González-Moreo, F., Suárez-Pérez, A., Camacho, M., Zumbado, M., Almeida-González, M., Del Mar Travieso-Aja, M. & Luzardo, O. P. 2015. Continued implication of the banned pesticides carbofuran and aldicarb in the poisoning of domestic and wild animals of the canary islands (Spain). *Science Of The Total Environment,* 505**,** 1093-1099.

Sacco, M. A., Koropacka, K., Grenier, E., Jaubert, M. J., Blanchard, A., Goverse, A., Smant, G. & Moffett, P. 2009. The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGap2-dependent plant cell death. *Plos Pathogens,* 5**,** E1000564.

Salazar, A. & Ritter, E. 1993. Effects of daylength during cyst formation, storage time and temperature of cysts on the in vitro hatching of *Globodera rostochiensis* and *G. pallida*. *Fundamental And Applied Nematology,* 16**,** 567-572.

SASA. 2018. *The European Cultivated Potato* [Online]. [Accessed].

Sasser, J. 1977. Worldwide dissemination and importance of the root-knot nematodes, Meloidogyne spp. *Journal Of Nematology,* 9**,** 26.

Schwessinger, B. & Zipfel, C. 2008. News from the frontline: recent insights into PAMP-triggered immunity in plants. *Current Opinion In Plant Biology,* 11**,** 389-395.

Semblat, J.-P., Rosso, M.-N., Hussey, R. S., Abad, P. & Castagnone-Sereno, P. 2001. Molecular cloning of a cDNA encoding an amphid-secreted putative avirulence protein from the root-knot nematode *Meloidogyne incognita*. *Molecular Plant-Microbe Interactions,* 14**,** 72-79.

Smant, G., Stokkermans, J. P., Yan, Y., De Boer, J. M., Baum, T. J., Wang, X., Hussey, R. S., Gommers, F. J., Henrissat, B. & Davis, E. L. 1998. Endogenous cellulases in animals: isolation of β-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings Of The National Academy Of Sciences,* 95**,** 4906-4911.

Smith, R. 1966. Family limitation. *Proceedings Of The Royal Society Of Medicine,* 59**,** 1149.

Sobczak, M., Avrova, A., Jupowicz, J., Phillips, M. S., Ernst, K. & Kumar, A. 2005. Characterization of susceptibility and resistance responses to potato cyst nematode (*Globodera* spp.) Infection of tomato lines in the absence and presence of the broad-spectrum nematode resistance *Hero* gene. *Molecular Plant-Microbe Interactions,* 18**,** 158-168.

Spurr, H. 1985. Mode of action of nematicides. *An Advanced Treatise On Meloidogyne,* 1**,** 269-275.

Stare, B. G., Sirca, S. & Urek, G. 2013. *Are We Ready For The New Nematode Species Of The Globodera Genus?*

Takken, F. L., Luderer, R., Gabriëls, S. H., Westerink, N., Lu, R., De Wit, P. J. & Joosten, M. H. 2000. A functional cloning strategy, based on a binary PVX-expression vector, to isolate HR-inducing cDNAs of plant pathogens. *The Plant Journal,* 24**,** 275-283.

Takken, F. L. W. & Goverse, A. 2012. How to build a pathogen detector: structural basis of NB-LRR function. *Current Opinion In Plant Biology,* 15**,** 375-384.

Taylor, A. & Sasser, J. 1978. Biology, identification and control of root-knot nematodes. *North Carolina State University Graphics*.

Team, R. 2015. Rstudio: integrated development for R. *Rstudio, Inc., Boston, Ma Url Http://Www. Rstudio. Com*.

Torto, T. A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N. A., Van West, P. & Kamoun, S. 2003. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Research,* 13**,** 1675-1685.

Trudgill, D. 1991. Resistance to and tolerance of plant parasitic nematodes in plants. *Annual Review Of Phytopathology,* 29**,** 167-192.

Tzortzakakis, E., Trudgill, D. & Phillips, M. 1998. Evidence for a dosage effect of the *Mi* gene on partially virulent isolates of *Meloidogyne javanica*. *Journal Of Nematology,* 30**,** 76.

Van Der Biezen, E. A. & Jones, J. D. 1998. Plant disease-resistance proteins and the gene-for-gene concept. *Trends In Biochemical Sciences,* 23**,** 454-456.

Van Der Hoorn, R. A. & Kamoun, S. 2008. From guard to decoy: a new model for perception of plant pathogen effectors. *The Plant Cell,* 20**,** 2009-2017.

Van Der Voort, J. R., Kanyuka, K., Van Der Vossen, E., Bendahmane, A., Mooijman, P., Klein-Lankhorst, R., Stiekema, W., Baulcombe, D. & Bakker, J. 1999. Tight physical linkage of the nematode resistance gene *Gpa2* and the virus resistance gene *Rx* on a single segment introgressed from the wild species *Solanum tuberosum* subsp. *Andigena* CPC 1673 into cultivated potato. *Molecular Plant-Microbe Interactions,* 12**,** 197-206.

Van Der Voort, J. R., Van Der Vossen, E., Bakker, E., Overmars, H., Van Zandvoort, P., Hutten, R., Lankhorst, R. K. & Bakker, J. 2000. Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. *Theoretical And Applied Genetics,* 101**,** 1122-1130.

Van Der Voort, J. R., Wolters, P., Folkertsma, R., Hutten, R., Van Zandvoort, P., Vinke, H., Kanyuka, K., Bendahmane, A., Jacobsen, E. & Janssen, R. 1997. Mapping of the cyst nematode resistance locus *Gpa2* in potato using a strategy based on comigrating AFLP markers. *Theoretical And Applied Genetics,* 95**,** 874-880.

Van Der Vossen, E., Sikkema, A., Hekkert, B. T. L., Gros, J., Stevens, P., Muskens, M., Wouters, D., Pereira, A., Stiekema, W. & Allefs, S. 2003. An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. *The Plant Journal,* 36**,** 867-882.

Van Der Vossen, E. A., Van Der Voort, J. N. R., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D. C., Bakker, J., Stiekema, W. J. & Klein-Lankhorst, R. M. 2000. Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. *The Plant Journal,* 23**,** 567-576.

Van Gundy, S. & Mckenry, M. V. 1977. Action of nematicides. *Plant Disease: An Advanced Treatise, 1st Edn. Elsevier***,** 263-283.

Van Megen, H., Van Den Elsen, S., Holterman, M., Karssen, G., Mooyman, P., Bongers, T., Holovachov, O., Bakker, J. & Helder, J. 2009. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology,* 11**,** 927-950.

Vanholme, B., De Meutter, J., Tytgat, T., Van Montagu, M., Coomans, A. & Gheysen, G. 2004. Secretions of plant-parasitic nematodes: a molecular update. *Gene,* 332**,** 13-27.

Vleeshouwers, V. G., Rietman, H., Krenek, P., Champouret, N., Young, C., Oh, S.-K., Wang, M., Bouwmeester, K., Vosman, B. & Visser, R. G. 2008. Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence genes. *Plos One,* 3**,** E2875.

Vossen, E. A., Gros, J., Sikkema, A., Muskens, M., Wouters, D., Wolters, P., Pereira, A. & Allefs, S. 2005. The *Rpi-blb2* gene from *Solanum bulbocastanum* is an *Mi-1* gene homolog conferring broad-spectrum late blight resistance in potato. *The Plant Journal,* 44**,** 208-222.

Vossen, J. H., Van Arkel, G., Bergervoet, M., Jo, K.-R., Jacobsen, E. & Visser, R. G. 2016. The *Solanum demissum R8* late blight resistance gene is an Sw-5 homologue that has been deployed worldwide in late blight resistant varieties. *Theoretical And Applied Genetics,* 129**,** 1785-1796.

Wei, C. F., Kvitko, B. H., Shimizu, R., Crabill, E., Alfano, J. R., Lin, N. C., Martin, G. B., Huang, H. C. & Collmer, A. 2007. A *Pseudomonas syringae* pv. tomato dc3000 mutant lacking the type III effector hopq1-1 is able to cause disease in the model plant *Nicotiana benthamiana*. *The Plant Journal,* 51**,** 32-46.

Whitehead, A., Tite, D. & Bromilow, R. 1981. Techniques for distributing non-fumigant nematicides in soil to control potato cyst-nematodes, *Globodera rostochiensis* and *G. pallida*. *Annals Of Applied Biology,* 97**,** 311-321.

Williamson, V. M. & Hussey, R. S. 1996. Nematode pathogenesis and resistance in plants. *Plant Cell,* 8**,** 1735-1745.

Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D., Yoshida, K., Dong, S., Schornack, S., Zipfel, C., Robatzek, S. & Hogenhout, S. Effector biology of plant-associated organisms: concepts and perspectives. Cold Spring Harbor Symposia on Quantitative Biology, 2012. Cold Spring Harbor Laboratory Press, 235-247.

Witek, K., Jupe, F., Witek, A. I., Baker, D., Clark, M. D. & Jones, J. D. 2016. Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nature Biotechnology,* 34**,** 656.

Wyss, U. & Grundler, F. 1992. Feeding behavior of sedentary plant parasitic nematodes. *Netherlands Journal Of Plant Pathology,* 98**,** 165-173.

Yang, Y., Li, R. & Qi, M. 2000. *In vivo* analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *The Plant Journal,* 22**,** 543-551.

Zhang, C., Liu, L., Wang, X., Vossen, J., Li, G., Li, T., Zheng, Z., Gao, J., Guo, Y. & Visser, R. G. 2014. The *Ph3* gene from *Solanum pimpinellifolium* encodes CC-NB-LRR protein conferring resistance to *Phytophthora infestans*. *Theoretical And Applied Genetics,* 127**,** 1353-1364.

Zipfel, C. 2009. Early molecular events in PAMP-triggered immunity. *Current Opinion In Plant Biology,* 12**,** 414-420.