Final Report

Pyramiding Resistances to Potato Cyst Nematodes to Produce Potato Cultivars with Durable and Broad-Spectrum Resistance

Ref: 11120021

Reporting Period: October 2014 – September 2018

Report Authors: Shona Strachan

Report No. 2019/9

AHDB Potatoes is a division of the Agriculture and Horticulture Development Board.

Additional copies of this report and a list of other publications can be obtained from:

Publications

AHDB Potatoes
Stoneleigh Park
Kenilworth
Warwickshire
CV8 2TL

Tel: 02476 692051
Fax: 02476 478902
E-mail: Potatoes.Publications@ahdb.org.uk

Our reports, and lists of publications, are also available at potatoes.ahdb.org.uk

CONTENTS

1. SUMMARY

1.1. Aim

The specific aim of this project was to determine the genomic location of the functional *H2* resistance gene against the potato cyst nematode *Globodera pallida*.

1.2. Methodology

PCN screen for resistant material – Progeny from a segregating susceptible : resistant (Picasso x P55/7) cross were screened using Pa1 cysts. A total of 154 progeny clones were tested, with each clone being screened six times across two technical replicates. Based on screening score, the twenty most resistant and most susceptible plants were taken forward for further analysis.

DNA extraction and targeted enrichment – DNA was extracted from young leaf material of the very susceptible, very resistant, Picasso (susceptible), and P55/7 (resistant) cultivars, with DNA from the twenty resistant and susceptible progeny was pooled in order to create four samples for analysis and sequencing. DNA samples underwent targeting in order to sequence only the resistance genes and a select number of conserved genes across the genome. Samples were then sequenced using Illumina MiSeq.

Sequence analysis – To identify differences which potentially control *H2*-mediated resistance variances between resistant and susceptible samples were compared, and those differences which occurred at a high enough frequency taken forward for further analysis as an area of interest.

Marker analysis – Single genetic differences were used to design markers which could discriminate between resistant and susceptible samples. Initially, the twenty most resistant and susceptible plants were tested using these markers, before an expanded population of 656 plants from the same cross were also screened.

Breeding – *H2*-containing P55/7 was crossed with *H1*- and *H3*-containing 10.Z.3.8.a to produce a progeny which was subsequently screened for potato cyst nematode resistance. Progeny plants which showed a strong resistance phenotype were taken forward in a breeding program to endeavour to produce a cultivar with broad-spectrum resistance to both *G. pallida* and *G. rostochiensis*.

1.3. Key findings

***H2* is a single dominant gene** – The progeny of the resistant : susceptible cross segregated with a close 1:1 ratio which is what is expected for a single dominant gene in cultivated potato. A single gene is ideal for breeding purposes.

***H2* maps to potato chromosome 5** – Identifying genetic differences between resistant and susceptible samples allowed the location of the functional H2 resistance gene to be mapped to the top end of potato chromosome 5

Identified a marker which segregated 100% with the resistant allele – Markers designed based on the genetic differences between resistant and susceptible samples allowed an area which contains two resistance genes to be identified, one of the markers was based on a genetic difference within the gene, so this gene became the candidate *H2* gene

Breeding material is very resistant to PCN – in parallel with the mapping of the *H2* resistance gene, preliminary breeding was undertaken using resistant P55/7 with a breeding line (10.Z.3.8.a) which contains resistance to *G. rostochiensis* (*H1*) and other *G. pallida* pathotype Pa2/3 (*H3*). Initial testing of a sub-set of the progeny revealed that they were highly resistant to both *G. rostochiensis* and *G. pallida*.

2. INTRODUCTION

2.1. Background

2.1.1. Origin of the Potato Cyst Nematode

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.

Cyst nematodes, including the potato cyst nematodes (PCN) usually have a restricted host range when compared to the root knot nematodes of the genus *Meloidogyne*. 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 which was 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).

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 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.

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) (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 and that the nematode made the shift from wild to cultivated potato independently of the human intervention of cultivating potato (Picard et al., 2007).

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).

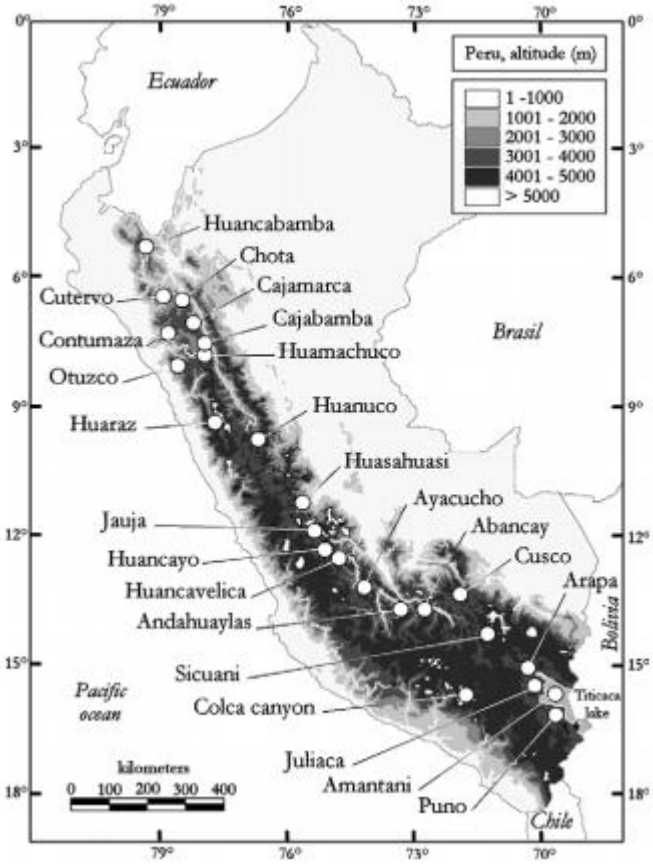

Figure 2 Map of Peru from (Picard et al., 2007). The Andean Cordillera, where *Globodera pallida* originates, is close to the Peruvian-Bolivian border and runs from Lake Titicaca to Cusco.

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 simple 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).

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.

2.1.2. Impact of Potato Cyst Nematodes

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). 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).

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.

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 et al., 2010).

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). 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). 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 will be discussed in more detail in the following section.

2.1.3. Resistance

Plants lack the mobile defence cells and adaptive immune system that is characteristic of mammals (Jones and Dangl, 2006). One method to induce defence responses uses pathogen associated molecule pattern (PAMP) which are recognised through 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).

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. These 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).

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 (Jones and Dangl, 2006, Bhattacharjee et al., 2013).

During nematode infection one of two types of cell death hypersensitive response (HR) is observed, either a rapid response best characterised by *Mi*-induced Root Knot Nematode resistance, or a more delayed response as observed in Cyst Nematodes (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 leaf 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).

Early research using the fungal flax rust pathogen (*Melampsora lini*) and its host flax (*Linum usitatissimum*) identified a gene-for-gene relationship which characterised the plant-pathogen interaction. The outcome of any interaction is determined by whether the plant resistance gene (*R* gene) coincides and recognises the pathogen avirulence gene (*avr* gene). In response to pathogen attack and the consequential effector release, plants have evolved NB-LRR proteins which act as *R* genes (Eitas and Dangl, 2010).

R genes split into two groups based on the structure of their N-terminus. The first group share 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). While 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*). 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). Resistance to *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). Through 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).

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 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 effective against European populations of *G. pallida*, however, resistance towards most 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).

2.2. Aims

The role targeted gene enrichment can play in the identification and mapping of resistance genes has been demonstrated by Jupe et al (2013) and then in *P. infestans* (Van Weymers et al., 2016). *H2* 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 (GenSeq and RenSeq)
- Identify putative Pa1 candidate avirulence genes which activate the *H2* resistance pathway, and functionally test them in an *H2*-resistant cultivar

3. MATERIALS AND METHODS

3.1. Biological Materials

3.1.1. 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.

3.1.2. Plant Material

A cross between susceptible potato *S. tuberosum* cultivar Picasso and resistant genotype P55/7 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.

3.1.3. Cyst viability and second stage juvenile hatching

Viability of the cysts was tested by adding single cysts to 2ml potato root diffusate in a well of a 12-well plate and left 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.

3.1.4. Infection Assays

3.1.4.1. 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 these 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).

3.1.4.2. Tuber Assays

Racks of root trainers were set-up as in Section 3.1.4.1 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.

Figure 3 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.

3.2. Molecular Protocols

3.2.1. DNA Extraction

3.2.1.1. 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. On day 2; 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.

3.2.1.2. Plant Material

Three, 1cm 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 instruction on screen, the QIAcube was started and preceded 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).

3.2.2. Gene Enrichment and Sequencing Library Preparation

3.2.2.1. 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.

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

3.2.2.2. 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.

3.2.2.3. 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.

3.2.2.4. Library Preparation – End Prep

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

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

3.2.2.5. Library Preparation – Adaptor 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.

3.2.2.6. 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 for 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.

3.2.2.7. 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 3.2.2.6.

3.2.2.8. 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 was 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 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.

3.2.2.9. Post-Capture Processing of 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 3.2.2.6, except 70% ethanol was used instead of 80%, and the capture library was eluted into a final volume of 30µl nuclease-free H_2O.

3.2.3. 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 1).

Table 1 KASP Marker information : 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

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, 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.

3.2.4. 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).

3.2.5. Semi-Quantitive 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	

3.2.6. *Agrobacterium*-mediated Transient Expression Assay

Five ml liquid cultures of *A. tumefaciens* clones (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 3). 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 3 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.

4. RESULTS

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

4.1.1. 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). 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). The progeny segregated with a 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) F1 clones were selected for further genetic analysis.

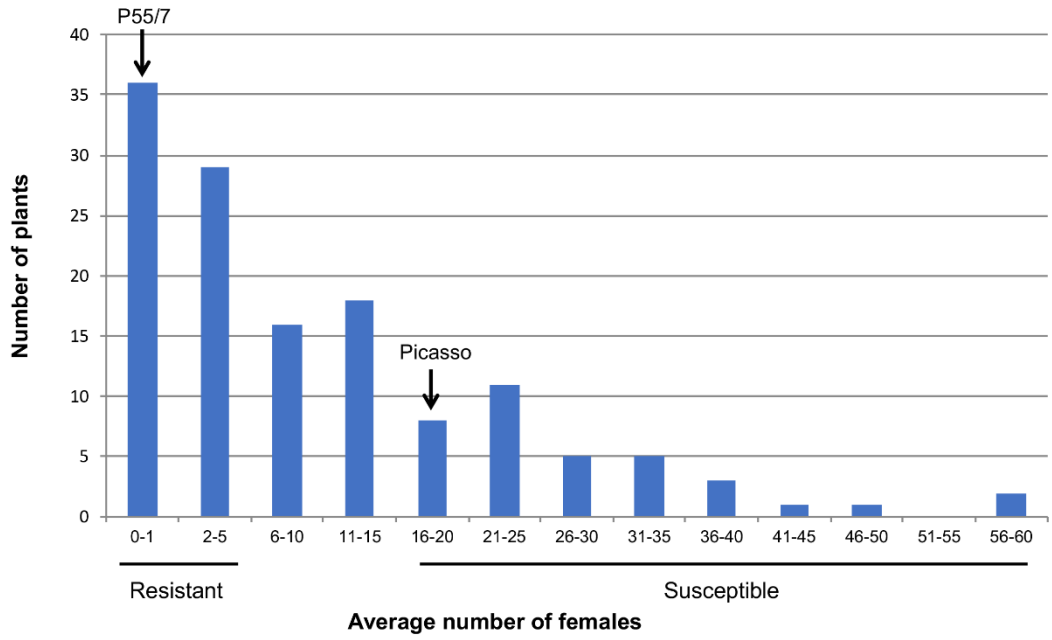

Figure 4 Histogram showing the distribution of infection levels of 154 Picasso x P55/7 progeny plants. The plants segregated in a ratio of 0.8:1 (resistant:susceptible) which is close to the 1:1 ratio which would be expected from a heterozygous simplex resistant : homozygous tetraploid cross. Resistant 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

4.1.2. 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 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 5A). 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 5B). Of those, 28 SNPs were identified at the expected frequency in both parent and bulk (Table 2), 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 5C). 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 2 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

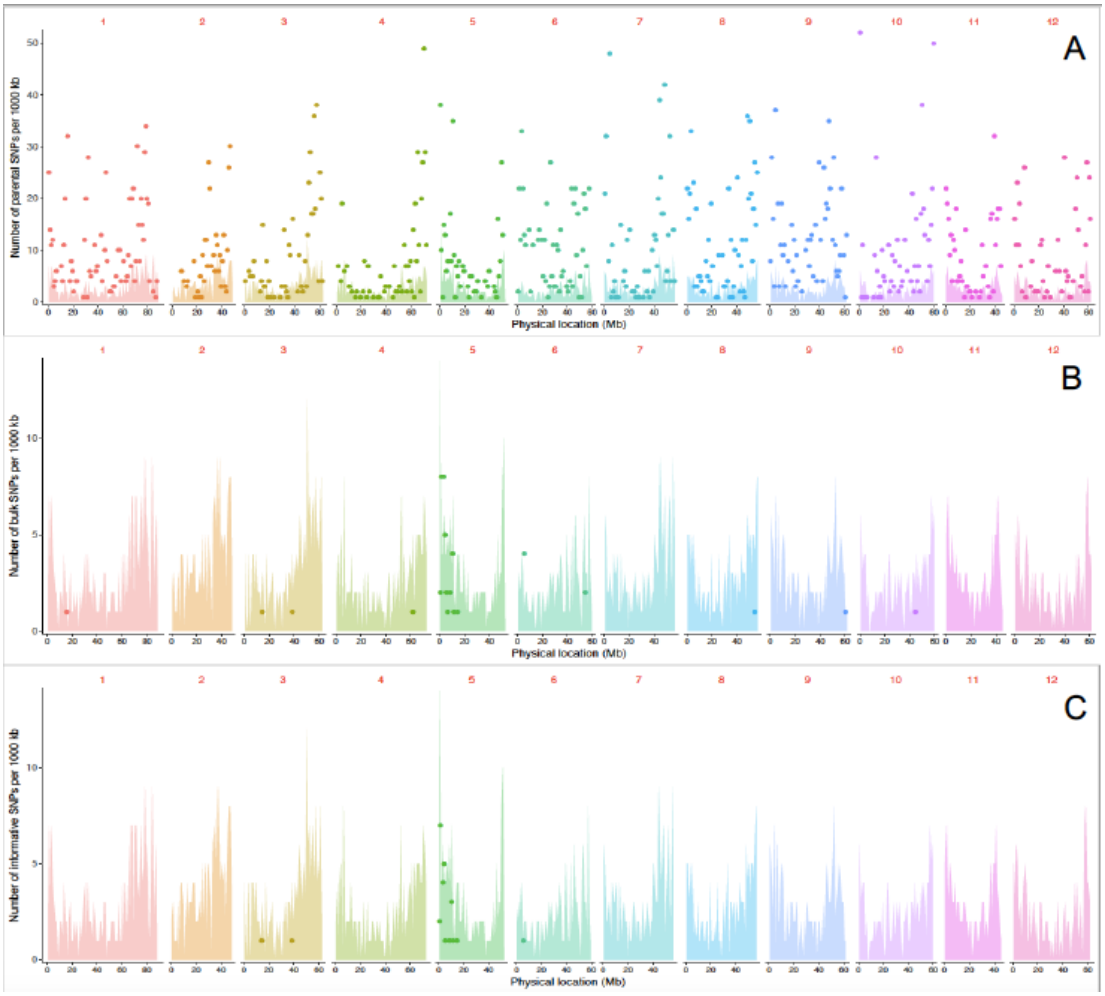

Figure 5 Graphical representation of the informative SNPs identified on the 12 potato chromosomes. Each coloured set of 'spikes' represents a chromosome, with the height of each spike representing the number of genes targeted by probes . Each dot represents a position were SNPs are located with its' 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%.

4.1.3. 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 6A). In the bulks, 106 SNPs passed the filtering conditions expected for susceptible progeny (rrrr) as well as resistant progeny (Rrrr) (Figure 6B). Of those SNPs 36 were found at the expected allele frequency in the parents and the bulks. (Table 3). 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 6C), while the remaining two SNPs correspond to two NB-LRRs on chromosome 9 (Table 3).

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. At a 2% mismatch rate 70% of SNPs are associated with the same interval. 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.

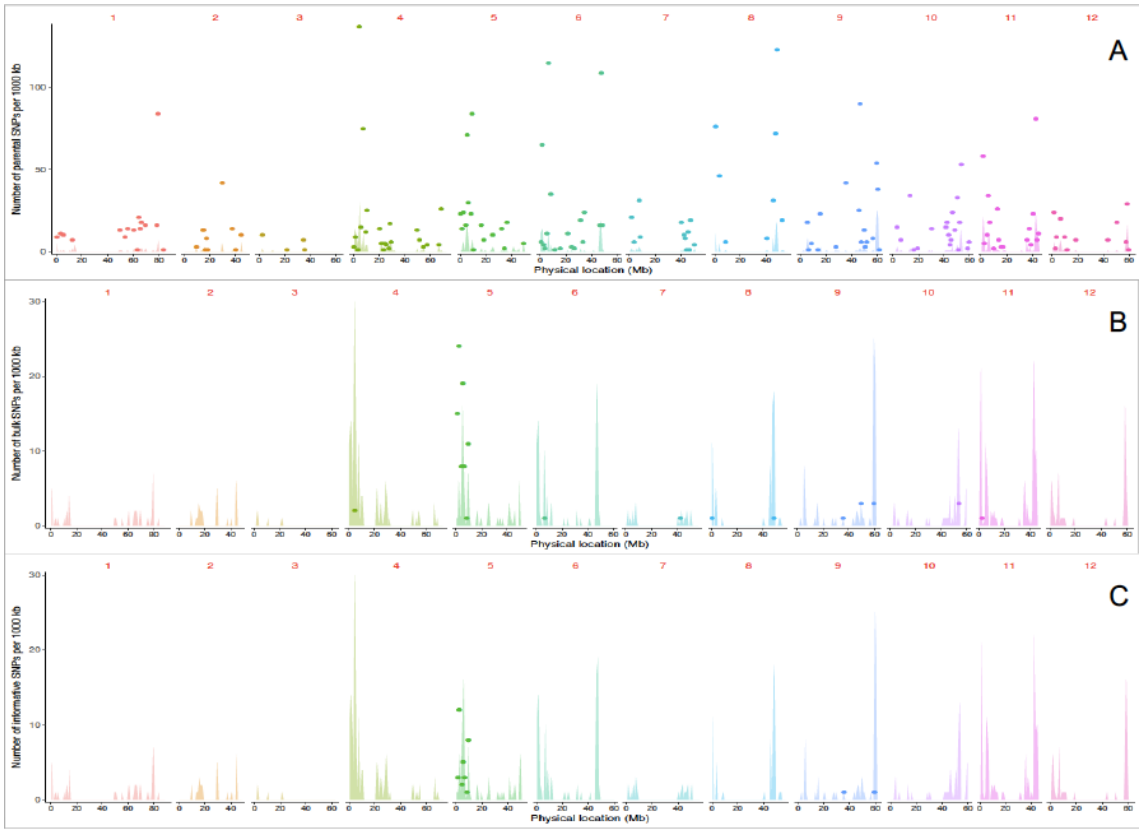

Figure 6 Graphical representation of the SNPs identified at a 3% mismatch rate. Each set of coloured data represents a specific chromosome. Coloured 'spikes' represent the number of NB-LRR genes targeted by probes across the chromosome. Each dot represents informative SNPs and its placement on the y-axis 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.

Chromosome	Start	Stop	Gene ID	Number of SNPs

Table 3 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.

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

4.1.4. 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 7). 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).

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 1).

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 8), 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

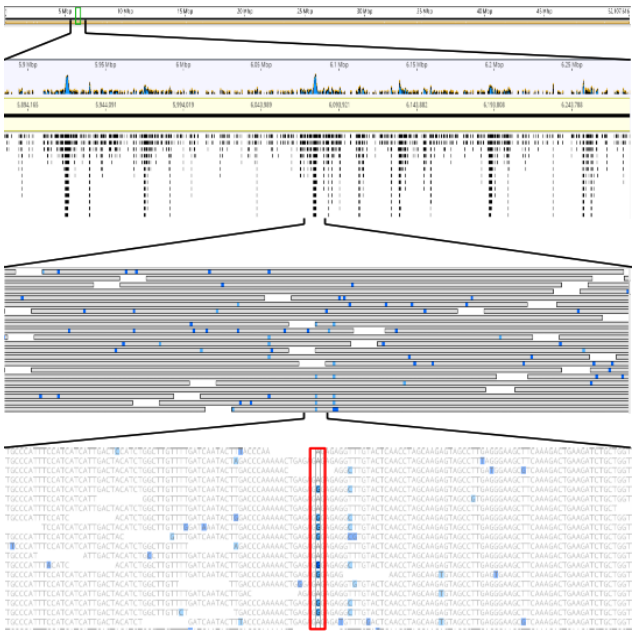

Figure 8 Graphical representation of informative SNP identification. An area of chromosome 5 is highlighted, and the GenSeq gene coverage is shown. All the Picasso and P55/7 reads which map to this region are highlighted, and nucleotides which do not resemble the reference genome are coloured blue. Viewing 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.

Figure 7 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.

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 phenotyped (Figure 9) shows this non-specific marker ST04_03ch05_1437439 (A) and 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 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.

4.1.5. 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 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).

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 genotypes 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).

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

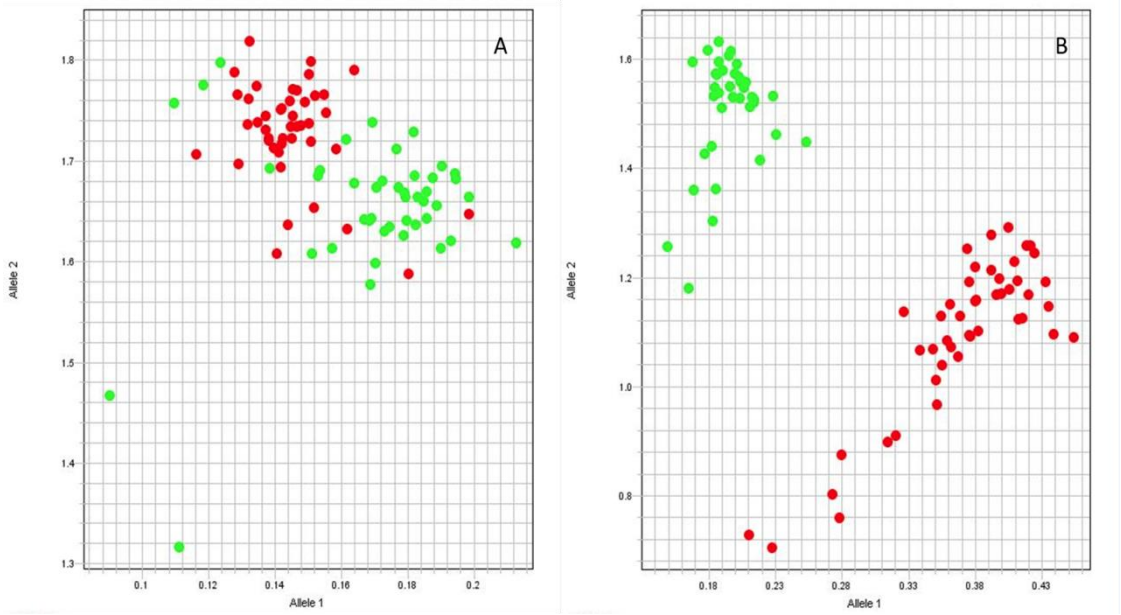

Figure 9 Allelic discrimination plots which demonstrate unsuccessful and successful KASP Markers. Each dot represents either a parent or progeny sample. Red dots are homozygotic (susceptible), and green dots are heterozygotic (resistant) for the candidate SNP. Panel A shows almost complete overlap between susceptible and resistant samples, while panel B shows successful clustering and separation of samples based on the candidate SNP.

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.

Table 4 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	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	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	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	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	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	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	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	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	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

4.1.6. 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), 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. Following the identification of recombinants through flanking markers, all recombinant plants were genotyped will the 5 KASP markers in the interval. Table 5 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 5 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	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	GenSeq	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	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	RenSeq	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	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	RenSeq	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	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_3000757	GenSeq	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	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	GenSeq	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	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	GenSeq	0	1	1	0	0	1	1	1	0	1	0	0	0	0	0	1	0	0	1	0	-	0	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	GenSeq	0	1	1	0	0	1	1	1	0	1	0	0	1	0	1	1	1	-	1	0	1	1	1	0	0	0	1	0	1	0	1	1	0	1	0	0	1	1	1	0	0	0	-	0	1

4.1.7. 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. A total of 49 genes reside in this 0.8Mb region (Table 6) 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).

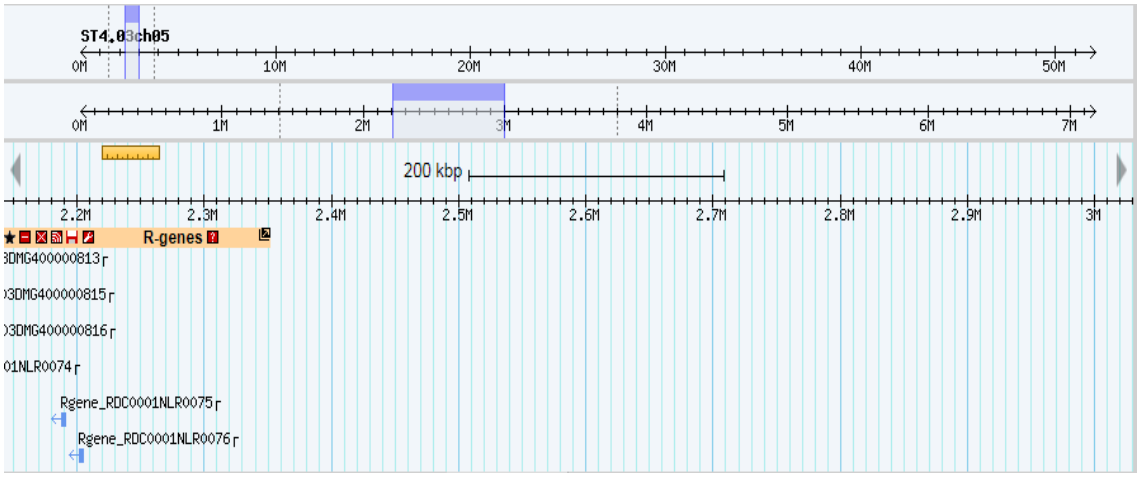

Figure 10 DM reference genome viewer showing the presence of two full length *R* genes within the 0.8Mb region between 2,202,842-3,000,757bp of the distal arm of chromosome 5.

Figure 10 details the *R* genes which are 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.

4.1.8. 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 5). 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.

Table 6 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 find 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

4.1.9. Identification of putative *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 to 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 11).

'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%

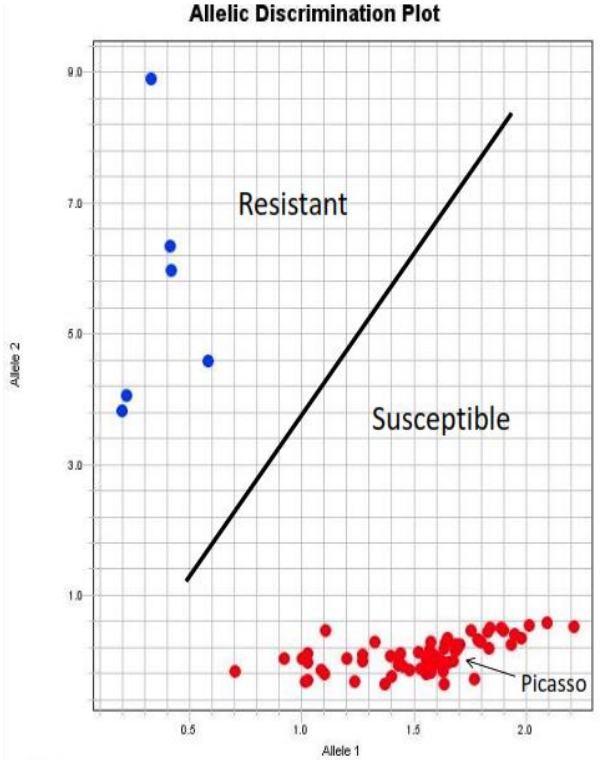

Figure 11 Allelic discrimination plot of KASP marker ST04.03ch05_2202842 tested on 20 RDC0001NLR0076 clones Susceptible parent Picasso was used to determine between resistant and susceptible alleles. Each clone was replicated in triplicates to.

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 12). *In silico* translation of the sequence in all six frames 1 led to non-synonymous mutations and premature stop codons in the susceptible clone.

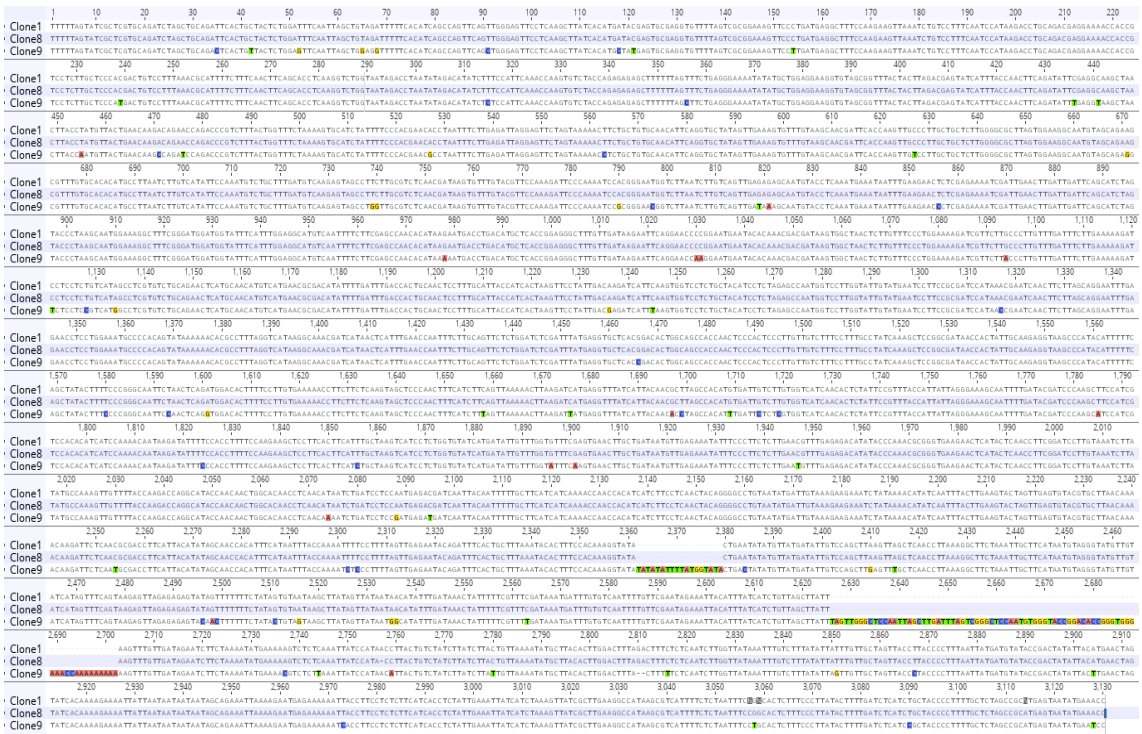

Figure 12 Aligned nucleotide sequence of candidate clones 1,8 (resistant) and 9 (susceptible) cloned from resistant P55/7. Sanger sequencing covered 78% of the length of the gene, with clones 1 and 8 being identical, and clone 9 containing sequencing differences. Polymorphisms are highlighted based on CLUSTAL colouring.

Focusing on the resistant clones, which under the hypothesis codes potentially for the functional *H2* gene, the translated sequence was screened for the presence of open reading frames (ORFs). A putative ORF of 2,200bp was identified within the coding sequence of the clone, however, no stop codon was identified (Figure 13). The lack of a stop codon could potentially be due to the way the raw sequence was trimmed, the sequence may have been of too poor quality and so was removed from further analysis. The missing stop codon could also be due to it being out with the primer designed amplified region, with the stop codon lying further downstream of what has been sequenced.

Figure 13 Nucleotide sequence of candidate clone 1 (resistant) showing an opening reading frame of 2,200bp. The sequence is missing a stop codon, this is hypothesised to have been deleted during read trimming due to poor sequence reads.

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

4.2.1. Semi-quantitative 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 14 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 14, panel D) no early amplification is present for post-enriched samples; both pre- and post-enriched libraries amplify at cycle 24.

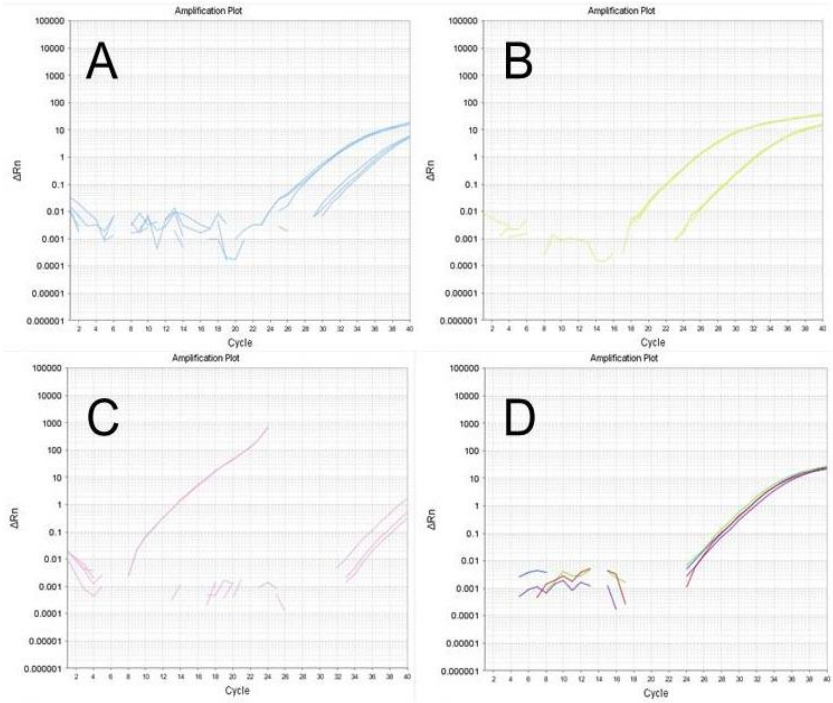

Figure 14 Quantitative PCR results for effector genes Gp414-2, Gp1719-1, Gp16H02 and control gene cytochrome B (CytB) in pre- and post-enriched library samples. Panel A) shows results for effector Gp414-2, B) effector Gp1719-1, C) effector Gp16H02, and D) control gene cytochrome b. In panels A-C there 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.

4.2.2. 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).

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 7). 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.

Table 7 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

4.2.3. Identification of signal peptides and downstream candidate effector genes

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

Table 8 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

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 7) 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 8). 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 9 depicts the expression profile of GPLIN_000926600 compared to GPLIN_000008800 (as a putative non-effector control). This analysis shows that GPLIN_000926600

Table 9 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

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.

The final step in the verification of the remaining candidate gene (GPLIN_000926600) was ensuring that the SNP identified during enrichment (Table 7) was real and present in the Pa1 population. The effector gene was cloned in both Lindley and Pa1, sequenced, and translated (Figure 15), revealing that the mutation at position $412^{A>G}$ which generated a non-synonymous amino acid change from serine to glycine was present.

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.

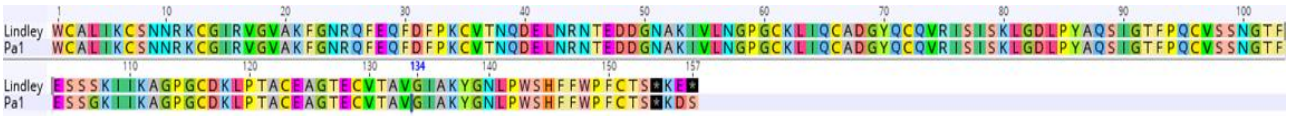

Figure 15 Protein translation of candidate effector gene GPLIN_000926600. Amino acid sequence is identical in both virulence Lindley and avirulent Pa1 except for the SNP (A>G) which mutates a serine (S) to a glycine (G) at amino acid position 107.

4.2.4. 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 18). 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 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.

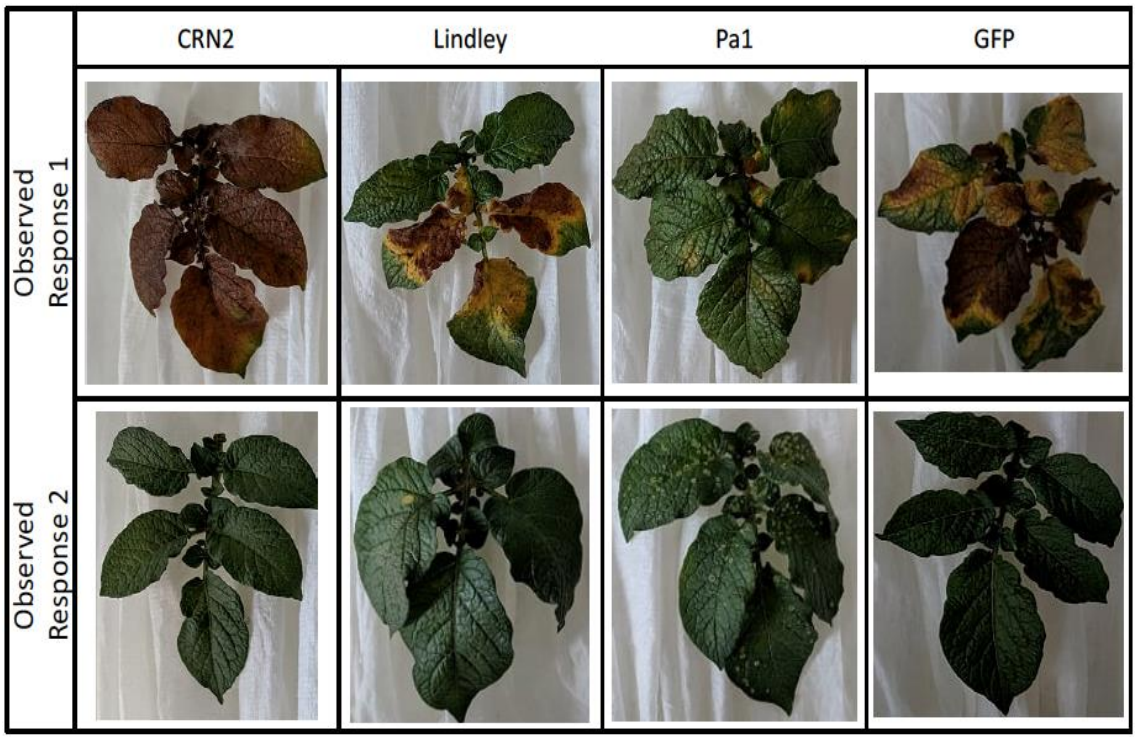

Figure 18 transient expression via vacuum infiltration of candidate effector and controls at 10dpi. Column 1 shows results of positive control CRN2, column 2 are the results of GPLIN_000926600 clones from Pa2/3 Lindley. Column 3 show results for candidate avirulence effector GPLIN_000926600 cloned from 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.

5. DISCUSSION

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* 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 genome complexity.

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, while 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*.

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.

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 one 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).

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 plans 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.

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 DM. 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). 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.

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 great 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 ration (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 stop codon for this gene. This will be complimented by screening an existing BAC library from P55/7 and analysis of 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.

5.7. 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 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 gland, 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.

5.8. 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.

5.9. 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.

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. CONCLUSIONS

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. 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). Future breeding programmes can use this knowledge to determine the effect of dosage on the gene.

In order to identify the genomic and chromosomal location of the *H2* gene, 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)), and PCN, as well as 5 QTLs active against PCN and *P. infestans* (*Pi* (Leonards-Schippers et al., 1994)) 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.

7. REFERENCES

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.

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 rostocbiensis. *Molecular and General Genetics MGG,* 224**,** 177-182.

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

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., Bradshaw, J. E., De Jong, W. S., Phillips, M., Castelli, L. & Waugh, R. 2002b. 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.

Burton, W. 1989. The Potato Longman. *London 3rd edn 742pp*.

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. M. 1977. Races of potato cyst nematode in Andean region and a new system of classification. *Nematologica,* 23**,** 340-349.

Caromel, B. & Gebhardt, C. 2011. Breeding for Nematode Resistance: Use of Genomic Information. *Genomics and Molecular Genetics of Plant-Nematode Interactions.* Springer.

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.

Commerce, U. S. D. O. 2015. *U.S. and World Population Clock* [Online]. Available: http://www.census.gov/popclock/ [Accessed 18/2/15 2015].

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

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.

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.

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

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.

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.

FAO. 2012. *Food and Agriculture Organisation of the United Nations, Land Resources* [Online]. Available: http://www.fao.org/nr/land/databasesinformationsystems/en/ [Accessed].

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.

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.

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.

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.

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.
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. *Rev Nematol,* 14**,** 213-219.

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.

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.

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.

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 toGlobodera pallida is dominated by one major locus inSolanum spegazzinii. *Theoretical and Applied Genetics,* 88**,** 764-769.

Leonards-Schippers, C., Gieffers, W., Salamini, F. & Gebhardt, C. 1992. TheR1 gene conferring race-specific resistance toPhytophthora infestans in potato is located on potato chromosome V. *Molecular and General Genetics MGG,* 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.

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

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

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*.

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.

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 MGG,* 249**,** 74-81.

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.

Nicol, J. M., Turner, 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.

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.

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.

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.

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 MGG,* 227**,** 81-85.

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.

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.

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.

Stare, B. G., Sirca, S. & Urek, G. 2013. *Are we ready for the new nematode species of the Globodera genus?*
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 Voort, J. R., Van Der Vossen, E., Bakker, E., Overmars, H., Van Zandroort, P., Hutten, R., Lankhorst, R. K. & Bakker, J. 2000a. 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., Van Der Vossen, E., Bakker, E., Overmars, H., Van Zandvoort, P., Hutten, R., Lankhorst, R. K. & Bakker, J. 2000b. 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 Weymers, P. S., Baker, K., Chen, X., Harrower, B., Cooke, D. E., Gilroy, E. M., Birch, P. R., Thilliez, G. J., Lees, A. K. & Lynott, J. S. 2016. Utilizing “Omic” technologies to identify and prioritize novel sources of resistance to the oomycete pathogen Phytophthora infestans in potato germplasm collections. *Frontiers in plant science,* 7**,** 672.

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.

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.

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

8. KNOWLEDGE TRANSFER ACTIVITIES

Solanaceae 2017, Valencia (Spain), 3rd-6th September 2017, Oral Presentation

AAB Advances in Nematology, London (England), 12th December 2017, Oral Presentation

Dundee Effector Consortium, Birnam (Scotland), 18th-19th June 2018, Oral Presentation

Potato Association of America, Boise (Idaho, North America), 22nd-26th July 2018, Oral Presentation

European Society of Nematologists, Ghent (Belgium), 9th-13th September 2018, Oral Presentation

Strachan S.M., Armstrong M.R., Kaur A., Wright K.M., Lim TY., Baker K., Jones J.T., Bryan G., Blok V and Hein I. (2018) Mapping the *H2* resistance effective against *Globodera pallida* pathotype Pa1 in tetraploid potato. Theor App Gen (submitted)

Baker K, Stephen G, **Strachan S**, Armstrong M, Hein I. (2018) BLASTmap: A Shiny-Based Application to Visualize BLAST Results as Interactive Heat Maps and a Tool to Design Gene-Specific Baits for Bespoke Target Enrichment Sequencing. Methods Mol. Biol; 1848: 199-206

Armstrong M.R,M Vossen J., Lim T.Y., Hutten R.C.B., Xu J., **Strachan S.M**., Harrower B., Chamouret N., Gilroy E.M. and Hein I. (2018) Tracking disease resistance deployment in potato breeding by enrichment sequencing. Plant Biotechnol J. doi:10.1111/pbi.12997.

9. 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.