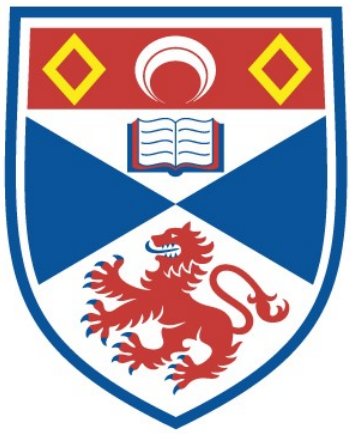
GENOME-BASED APPROACHES FOR IDENTIFICATION OF *AVIRULENCE* GENES IN POTATO CYST NEMATODES

Kyriakos Varypatakis

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



2019

Genome-based approaches for identification of *avirulence* genes in potato cyst nematodes

Kyriakos Varypatakis

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

February 2019

Candidate's declaration

I, Kyriakos Varypatakis, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 44,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2015.

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

Date 15/02/2019 Signature of candidate

Supervisor's declaration

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

Date 15/02/2019 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in

the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

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

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

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 15/02/2019 Signature of candidate

Date 15/02/2019 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

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

Date 15/02/2019 Signature of candidate

General acknowledgements

In September of 2015, I came to Scotland for starting this new, but exciting chapter of my professional and personal life. About 3.5 years after, it is time to mention and dedicate the following part of the book to those who contributed in a professional or personal manner to accomplish my PhD.

First of all, I would like to start by expressing my sincere gratitude to my supervisor Prof. John Jones for giving me the chance to work with his group these years. Thank you for being an excellent example of a supervisor, your guidance and of course the patience you showed to me throughout the whole project. Of course, I would also like to thank Dr. Vivian Blok too for her guidance, advices and positive energy. To both of you, many thanks for giving me this opportunity and I hope that my work met your initial expectations; I feel lucky being a student of both of you. Furthermore, I owe a big thank you to Dr. Peter Cock, who helped me a lot in my first steps in the complicated (for me) world of bioinformatics and coding.

To all the people from The James Hutton Institute (Dundee's site) who helped me in any stage of this project; Dr. Ingo Hein and Joanne Lim for their advices and contribution to the application of the target enrichment sequencing, as well as Dr. Sophie Mantelin who was always willing to help me and answering any of my questions. Special thanks to Dr. Peter Thorpe for his work on the new *G. pallida* genome assembly and Dr. Sebastian Eves-van den Akker for performing the flow cytometry of it. And of course, to all the members of the Nematology lab (Shona, Kerry, Jamie, James, Ulrike, Anne and Ailsa) who certainly helped me by making my days there better, as well as introducing me into the Scottish culture.

At this point, I should specially mention Dr. Benjamin Mimee and PY Véronneau from the Agriculture and Agri-Food Canada for accepting me working with them for several weeks and for the successful collaboration we had. Thank you for your help, advice and giving me this great opportunity to visit Quebec and having a very beautiful experience; I hope to see you again very soon.

To a more personal note, I should also mention and thank my friends Αντώνη, Jonny and Trisha for making my stay at JHI easier and for sure more interesting, as well as Jenny for making my life in the last years more beautiful and much easier. I couldn't forget all my friends back to Greece and the Netherlands for being always "next to me" all the years I am away; Κάτια, Ρίτα, Εύη, Βασίλη, Γιώργο, Πέτρο, Γιάννη Α, Δημήτρη, Γιάννη Σ, Χρύσα ευχαριστώ για όλα! And of course, last but not least, my parents, my sister and Ze for supporting me in every possible mean all these years.

Funding

This work was supported by The James Hutton Institute and the Agriculture and Horticulture Development Board (AHDB) Potatoes.

Author's publications

i. **Varypatakis, K.**, Jones, J. T. & Blok, V. C. (in press). Susceptibility of potato varieties to populations of *Globodera pallida* selected for increased virulence. *Nematology*. [Manuscript was accepted for publication in June 2019].

To my family and friends

«..Ἓν οἶδα ὅτι οὐδὲν οἶδα..», Σωκράτης (469-399 π.Χ.)

«..I know that I know nothing..», Socrates (469-399 BC)

Abstract

The use of natural resistance is the most effective way to control the potato cyst nematodes *Globodera pallida* and *G. rostochiensis*. Although pathotypes of the latter species can now be fully controlled, control of *G. pallida* is a real challenge because of the existence of multiple pathotypes with different virulence behaviour. Although the available resistance sources confer partial resistance only, such as that from *Solanum vernei* (*Gpa5*) and *S. tuberosum* ssp. *andigena* CPC2802 (*H3*), their use may drive strong selection towards virulence. Resistance operates through the detection of nematode-derived effectors by plant resistance (R) proteins. Upon recognition, immune responses are activated in plants and the recognised effectors are then characterised as avirulence (avr) proteins.

The study of the effectors and *avr* genes can provide important information on the mechanisms underlying selection. Fortunately, the development of novel sequencing technologies and genomic tools have enabled us to gain better insight towards complex genomes, such as that of potato cyst nematodes. Here, long-read sequencing was used for the generation of a new, more complete *G. pallida* genome. Based on this, analysis of captured, targeted effector-encoding genomic regions from virulent populations selected on the above resistance sources showed the presence of up to 54 candidate effectors that may determine virulence on these resistance sources. By cross-referencing those with the candidates identified from the re-sequencing of the same populations, 3 high-confidence candidate *avr* genes were listed. In parallel, the analysis of their allele frequencies and SNP distribution revealed that many potential effector genes are located within genomic "islands", which are all selected on a specific resistance. This organisation is believed to facilitate selection evolution and host adaptation.

Lastly, the *G. rostochiensis* candidate effector *g13394* was cloned from an avirulent and virulent to *H1* population, and subsequently was functionally validated through *Agrobacterium*-mediated transient expression using two different methods.

Table of Contents

List of Figures

List of Tables

List of Supplementary Data (CD)

List of Abbreviations

AF	allele frequency
AFLP	amplified fragment length polymorphism
ALT	alternate (allele)
Avr	avirulence
BAK1	BRI1-associated receptor kinase 1
BLAST	basic local alignment search tool
bp	base pair
BSA	bulked segregant analysis
BUSCO	benchmarking universal single-copy orthologs
CAF	change in AF
CC	coiled-coil
cDNA	complementary DNA
CEGMA	core eukaryotic genes mapping approach
CLE	CLAVATA elements
cM	centimorgan
cm	centimetre
CPC	Commonwealth Potato Collection
CWDE	cell wall-degrading enzymes
cytb	cytochrome b
DAMP	damage-associated molecular pattern
DNA	deoxyribonucleic acid
dpa	days post application
dpi	days post inoculation
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequenced tags
ETI	effector-triggered immunity
ETS	effector-triggered susceptibility
f-	Farcet
flg	flagellin
FLS2	flagellin sensing 2

GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GBS	genotyping by sequencing
gDNA	genomic DNA
GFP	green fluorescent protein
HGT	horizontal gene transfer
HMW	high-molecular weight
HR	hypersensitive response
hr	hour
J2	2nd-stage juvenile
kb	kilobase
LB	Luria-Bertani
LRR	leucine-rich repeat
MAMP	microbe-associated molecular pattern
MAPK	mitogen-activated protein kinase
Mb	mega bases
mg	milligram
min	minute
ml	millitre
MNP	multiple nucleotide polymorphisms
mRNA	messenger RNA
ms	milliseconds
n-	Newton
NB	nucleotide-binding
ng	nanogram
NGS	next generation sequencing
nM	nanomolar
OD	optical density
ORF	open reading frame
PacBio	Pacific Biosciences
PAMP	pathogen-associated molecular pattern
PCA	principal component analysis

PCN	potato cyst nematode
PCR	polymerase chain reaction
pDNA	plasmid DNA
PPN	plant-parasitic nematodes
PenSeq	pathogen enrichment sequencing
PF	pass-filter
PRR	pattern recognition receptor
PTI	PAMP-triggered immunity
PVX	Potato virus X
QC	quality control
qPCR	real-time (quantitative) PCR
QTL	quantitative trait *locus*
R	resistance (protein)
RAPD	random amplified polymorphic DNA
RBP-1	retinol-binding protein 1
REF	reference (allele)
RFLP	restriction fragment length polymorphism
RLK	receptor-like kinase
RLP	receptor-like protein
RNA	ribonucleic acid
ROS	reactive oxygen species
rpm	revolutions per minute
RQ	relative quantification
Sa_	*S. tuberosum* ssp. *andigena* CPC2802 (*H3*)
SDW	sterile distilled water
sec	second
SMRT	single-molecule real-time (sequencing)
SNP	single nucleotide polymorphism
SOC	super optimal broth with catabolite repression
SP	signal peptide
SPRYSEC	secreted SP1a and Ryanodine receptor

Sv_	*S. vernei*
TBE	Tris/Borate/EDTA buffer
T-DNA	transfer DNA
TLR	Toll-like receptor
TRD	tomato root diffusates
UV	ultra-violet
Vir	virulence
wpi	weeks post infection
µl	microlitre
µm	micrometre
µM	micromolar

1. General Introduction

1.1. Potato as a crop

1.1.1. Potato cultivation

The cultivated potato is one of the most important food sources worldwide; in 2016 more than 19 million hectares were cultivated with this crop yielding an average of 22 tonnes per hectare. In addition to making a significant contribution to global food security, potato cultivation provides an important economic input to the annual GDP of many countries (Xu et al., 2011) (FAOSTAT, 2016). In a global level, potato is the fifth most commonly cultivated crop after sugarcane and cereal crops, while in Europe comes fourth after cereals (e.g. wheat, barley, oats), maize and oil-seeded plants (e.g. rapeseed, sunflower) with 5.5 million hectares cultivated with this crop and 118 million tonnes production in total (Figure 1-1) (FAOSTAT, 2016).

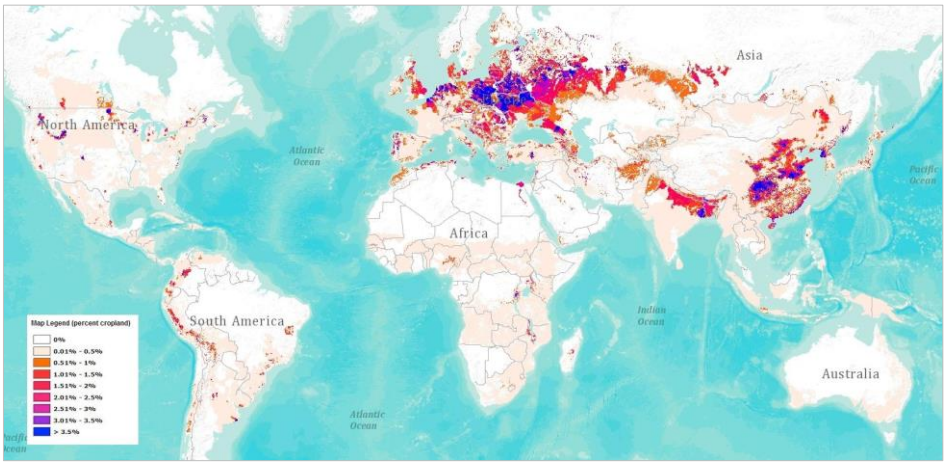


Figure 1-1 Geographical distribution of potato cultivation in the world in 2014 (CGIAR, 2014).

The United Kingdom is ranked within the 10 European countries with the largest harvested areas and highest production of potato. However, in the last decades, the total area planted with it has steadily declined (AHDB, 2015; FAOSTAT, 2016). Scotland and Eastern England are the areas where potato is cultivated the most (Figure 1-2). In Scotland potato production is the third highest among all the crops with a total output value of £209 million in 2016 (Scottish Government, 2016). The majority of the annual production of potato in Scotland is intended for the seed industry (around 65,000 tonnes of seed potatoes production compared to 10,000 tonnes of ware potatoes production)

and in recent years the export of seed potatoes has doubled (Scottish Government, 2016; SASA, 2018).

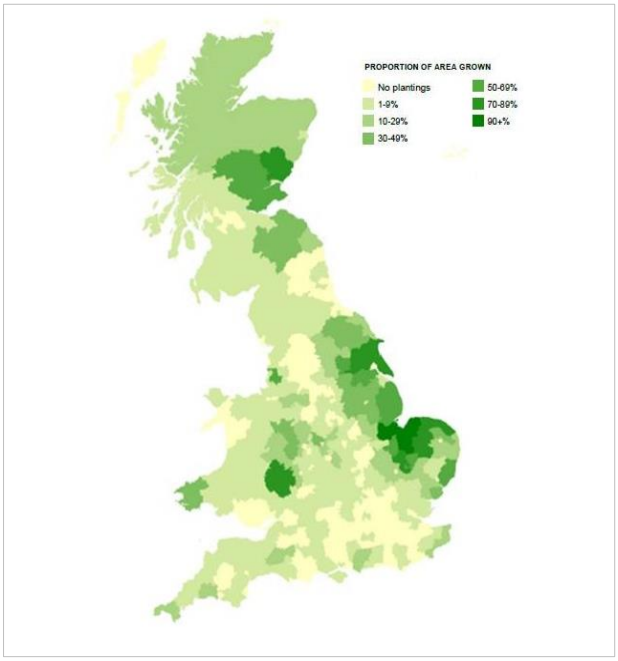


Figure 1-2 Distribution of areas in Britain grown with potato in 2015 (AHDB, 2015).

1.1.2. Botanics and introduction of potato into Europe

Taxonomically, potato (*Solanum tuberosum* L.) belongs to the *Solanaceae* family along with tomato (*S. lycopersicum*), eggplant (*S. melongena*) tobacco (*Nicotiana tabacum*) and peppers (*Capsicum* sp.) (Machida-Hirano, 2015). Potato has a large genetic pool containing many wild species that originate from the Andean highlands of Chile, Peru, Bolivia and northern Argentina (Xu et al., 2011). The polyploidy of its species, the high heterozygosity, the sexual compatibility with many other Solanaceous species and the ability to use asexual reproduction makes potato taxonomy complicated (Machida-Hirano, 2015; Xu et al., 2011). Cultivated potato species are morphologically diverse and can be diploid (2n = 2x = 24), triploid (2n = 3x = 36), tetraploid (2n = 4x = 48) or pentaploid (2n=5x = 60). Usually, triploids are genetically very unstable with very low fertility. Many *Solanum* species can be further sub-divided into different sub-species groups based on their morphological and molecular characteristics (Hijmans et al., 2007; Machida-Hirano, 2015).

The wild species are native to the areas along the American continent in two main diversity centres; in Central America (i.e. Mexico) and in South America (i.e. Andes) (Hijmans and Spooner, 2001; Machida-Hirano, 2015). Interestingly, the evidence shows a correlation between the ploidy level and the geographical distribution of the origin of the wild species, which allowed adaptation to different environmental conditions. As a general rule, potatoes with higher ploidy can be found in wetter and colder areas of the diversity centres (Hijmans et al., 2007; Machida-Hirano, 2015).

Potato, and more specifically the sub-species *andigena*, is believed to have been introduced into continental Europe in 1570 when Spanish sailors brought it from the Canary Islands, from where it had been introduced directly from South America a few years previously (Bradshaw and Ramsay, 2005; Hawkes and Francisco-Ortega, 1993). The 'common potato', i.e. Tuberosum form, was introduced in the middle of the 19th century into Europe (and specifically in Britain) from a narrow genetic base in the Chilean Andes as a breeding material against late blight (Bradshaw and Ramsay, 2005; Picard et al., 2007). Along with this species, a range of pathogens (e.g. fungi, bacteria, viruses) and pests (e.g. nematodes and insects) were also introduced and subsequently spread to the rest of the world. In parallel, the narrow genetic base of the 'common potato' compared to the wild species created favourable conditions for the spread of pathogenic diseases (Bradshaw and Ramsay, 2005).

1.2. Nematodes

1.2.1. A brief introduction to nematodes

The word for "nematode" comes from the Greek words "νήμα" [neéma] (meaning 'thread') and "είδος" [eédos] (meaning 'species' or '-like'). Thus, nematodes are non-segmented thread-like organisms and taxonomically they create the Phylum Nematoda in the Kingdom Animalia. Approximately 25,000 species of nematodes have been identified in total, but it is estimated that millions more species may exist in nature. Depending on their lifestyle, nematodes are categorised into four categories. The free-living species live in the soil, water or even in hot water springs and may feed on fungi,

bacteria and other microorganisms. Other species may be parasitic on insects (i.e. entomoparasitic), on plants or on animals (including that of humans). The species of the last two categories frequently can cause serious diseases on either plants or animals respectively (Lee, 2002; Perry and Moens, 2011a).

1.2.2. Nematodes as parasites of plants

More than 4,000 plant-parasitic species have been described to date (Decraemer and Hunt, 2013). In spite of their small body size, they have a major impact on crops causing large economic losses estimated at 80 billion U.S. dollars worldwide each year. The damage caused can be either direct (e.g. damage to the root system, fruits, depletion of water and nutrients), or indirect by vectoring viruses. While many species are endoparasites, some are ectoparasites or even semi-endoparasites (Nicol et al., 2011). The endoparasites are further sub-divided into sedentary species that spend their entire parasitic life feeding on a specific site of the plant, and migratory which migrate from one site to the other to secure their food sources. When they are fully developed, their body is less than 4mm long with a diameter between 15µm and 35µm and is protected by a hard cuticle (Figure 1-3) (Agrios, 2005).

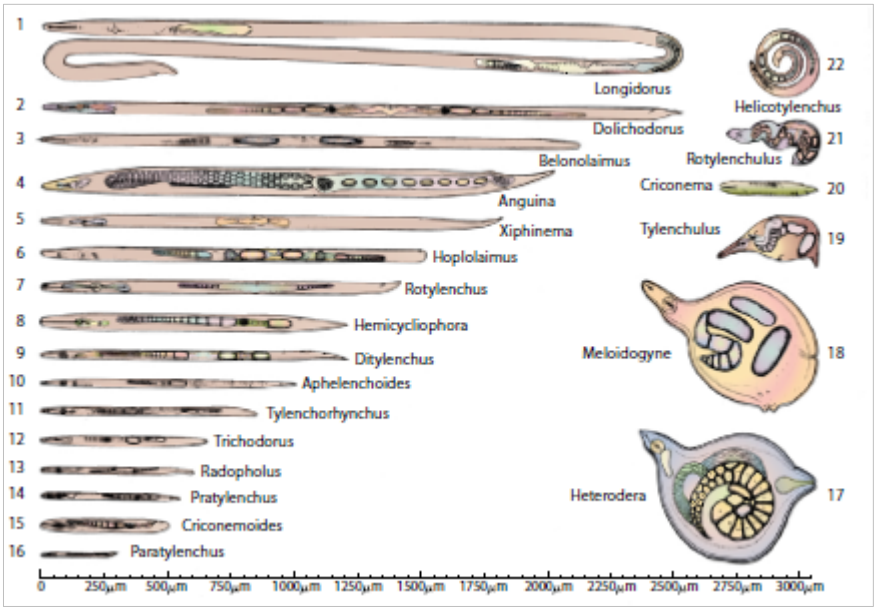


Figure 1-3 Important genera of plant-parasitic nematodes (reprinted from Agrios (2005)).

The nematode digestive system is relatively simple and consists of a tube that passes through the mouth, oesophagus and intestine ending up to the rectum. Plant-parasitic

nematodes are equipped with a protruded, hollow and rigid structure called the stylet, which besides breaking through plant cell walls, also delivers proteins and other molecules to facilitate host penetration, colonisation and modification of host cells into feeding structures (Agrios, 2005; Gheysen and Mitchum, 2011; Perry and Moens, 2011a). Root-knot (*Meloidogyne* spp.) and cyst nematodes (*Globodera* spp. and *Heterodera* spp.) are the most economically important plant-parasitic species with very similar lifecycles. Both are obligate sedentary endoparasites causing huge economic losses each year on crops in temperate and (sub)-tropical climates. Cyst nematodes tend to have a significantly smaller host species range compared to the polyphagous root-knot nematodes (Jones et al., 2013).

1.2.3. Plant-parasitism in nematodes

The ability to parasitize plants has evolved at least four times throughout evolutionary history. Phylogenetic studies based on a small subunit ribosomal RNA (SSU RNA) show that plant-parasitic species occupy 4 out the 12 clades that the phylum of Nematoda consists of (Figure 1-4) (Bird et al., 2015). Plant-parasitic nematodes show a number of morphological and genetic adaptations. For example, ectoparasitic species have evolved a long, strong and rigid structure called odontostylet. This structure, which is analogous to the stylet, help the ectoparasitic species to bypass the epidermis and reach the parenchymatic root cells. On the other hand, the endoparasitic species are equipped with a short but very robust stylet to allow physical strength to help breaching the plant cell walls (Hussey and Grundler, 1998). The stylet is essential for plant parasitism, although it can also be found in the non-parasitic species, which that may imply that it evolved to allow feeding on other organisms first (e.g. fungi) and subsequently became adapted for plant parasitism (Bird et al., 2015).

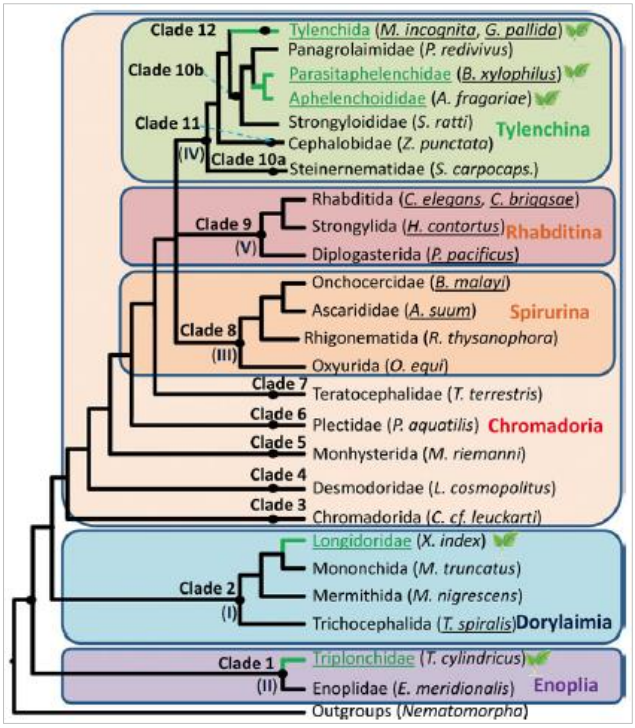


Figure 1-4 Phylogeny of the phylum Nematoda based on SSU RNA assays. Clades 1-12 are according to the classification proposed by van Megen et al. (2009). Latin numbers I-V correspond to clades according to the classification proposed by Blaxter et al. (1998). Green text shows the PPN species, underlined text the availability of genome assembly (reprinted from Bird et al. (2015)).

In terms of genomic adaptations, plant-parasitic species contain genes that encode for proteins very similar in function and structure to proteins encoded in other microorganisms, mostly in fungi and bacteria. Nematodes have acquired these genes from the microorganisms through horizontal gene transfer (HGT) events and the genes involved are essential for their parasitic life inside the host. Cell wall-degrading enzymes (CWDE), such as xylanases, polygalacturanases, cellulases, glycosyl-hydrolases and expansin-like proteins are necessary for the invasion into the root system and the intracellular migration of the juvenile (Haegeman et al., 2012; Smant et al., 1998).

Other proteins, similar to the NodL proteins that the nitrogen-fixing soil bacteria encode for establishing root nodules, also contribute in the formation of the feeding sites in root-knot forming species. Furthermore, in the formation of feeding structures many proteins are also involved directly or indirectly by interfering plant hormonal pathways (e.g. the protein 19C07 secreted by cyst nematodes interacts with the auxin influx transporter LAX3). Species of the family Longidoridae (e.g. the grapevine parasite *Xiphinema index*) induce multinucleate feeding structures during their parasitic life similar to the giant

cells of the root-knot nematodes of the family Heteroderidae (i.e. *Meloidogyne* spp.). In the same family, other species (e.g. *Heterodera* spp., *Globodera* spp.) have also developed different strategies (e.g. syncytia) for inducing feeding structures. Therefore, this essential adaptation of plant parasitism has evolved independently on several occasions during the evolution of nematodes (Bird et al., 2015; Griffiths et al., 1982; Jones et al., 2013).

Plant-parasitic nematodes have also HGT-acquired genes for encoding proteins that suppress host immune responses (e.g. chorismate mutases) or biosynthesise nutrients (e.g. vitamins) (Bird et al., 2015; Haegeman et al., 2012; Jones et al., 2005). Most of the HGT-acquired genes are highly conserved among different genera even in species belonging to different clades (Danchin and Rosso, 2012; Haegeman et al., 2012).

1.3. Cyst nematodes

1.3.1. The lifecycle

All cyst nematode species have very similar lifecycle stages consisting of the eggs, four larval stages (alternatively, juveniles as they are known in the plant-parasitic species) and adults. Once the fertile females of the cyst nematodes die, the wall of their body is thickened and becomes darker due to polyphenol oxidation, forming a resistant structure known as cyst. Each cyst may contain approximately 400-500 eggs (Turner and Stone, 1984). Under favourable environmental conditions, the cysts break and the eggs are spread in the soil. Inside the eggshell an individual second-stage juvenile (J2) is protected. Its hatching is stimulated by host root exudates (Smant et al., 1997). Just before hatching, the permeability of the eggshell changes, the enclosed individual is hydrated and activated. The J2 escapes from the egg using the stylet to rupture the eggshell and, in some species, by secreting enzymes through its stylet (Jones et al., 1998; Perry, 2002). Eventually, the eggshell is ruptured and the newly-hatched J2 moves in the soil towards the host root system following host-derived chemical cues (Figure 1-5).

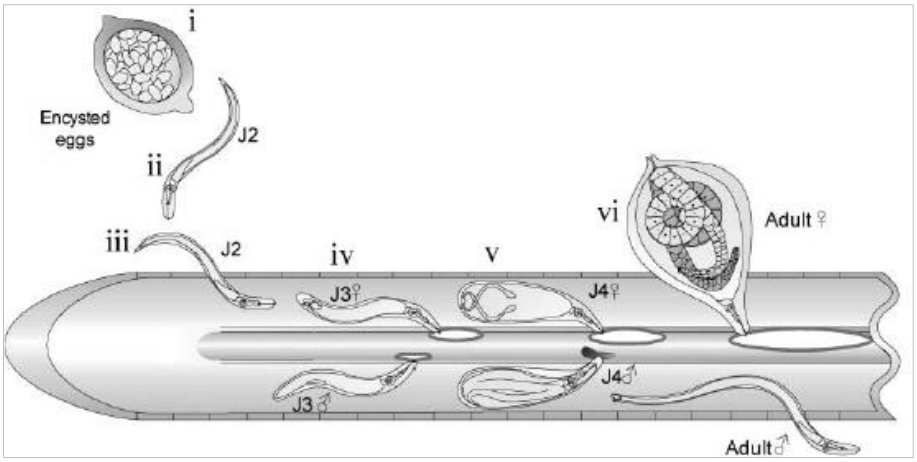


Figure 1-5 The lifecycle of cyst nematodes. (i) Each cyst contains hundreds of eggs, which in turn contains an individual dormant J2. (i) Under favourable conditions, the cyst breaks and the eggs hatch. (iii) The J2 enters the root and (iv-v) migrates intracellularly till it finds a suitable cell where it forms a syncytium-feeding site. Sex determination takes place between the J3 and J4. (vi) Usually the J2 is developed into a female. Once it dies, its cuticle hardens forming a cyst to protect the enclosing embryos. Depending on the environmental conditions, J2 can be developed into a male (bottom), which stops feeding and becomes mobile again (reprinted from Lilley et al. (2005)).

The invasion of the juvenile in the host occurs close to the root tip or in the lateral roots where the epidermis is thinner and easier to be penetrated. The infective J2 enters the root by exerting physical force with the stylet and secreting CWDEs through it. Once it enters, it migrates intracellularly until it reaches the cortical cells. Contrary to the cyst-forming species, the invading J2 of the root-knot species enter the root in the elongation zone and then they turn towards the root tip (in a U-shape) in order to evade the barrier of the Casparian band (Figure 1-6). When the J2 of the cyst-forming species arrives in the root cortex, it starts piercing several cells until it finds a suitable cell for syncytium induction (Gheysen and Mitchum, 2011; Jones et al., 2013; Wyss, 1992; Wyss and Zunke, 1986). The sedentary life of the nematode and a molecularly complex interaction with the host begins once it finds a cell that does not respond against it. At this point the J2 starts feeding using the cell as an initial food and water source. Feeding is also facilitated by formation of a tube-shaped structure extending the stylet till the host cell wall and being attached to the plasma membrane (Eves-van den Akker et al., 2015; Jones et al., 2013; Juvale and Baum, 2018; Sobczak et al., 1997).

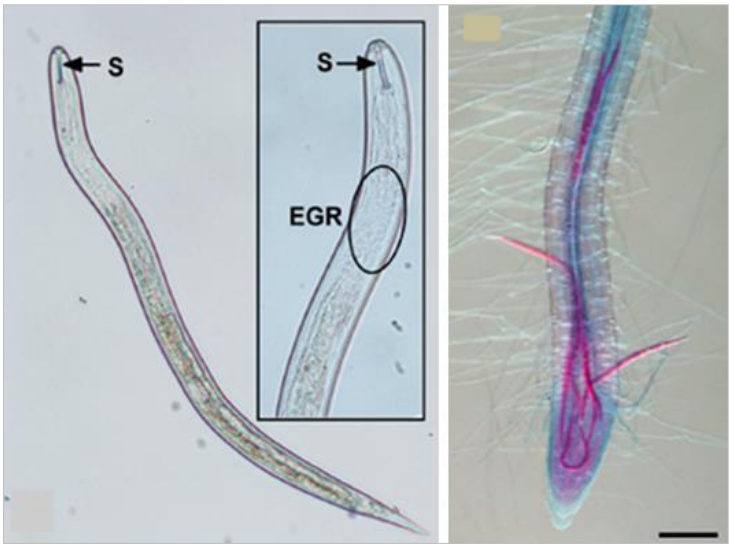


Figure 1-6 The stylet and the invasion of the J2s. *Left:* A plant-parasitic nematode juvenile and a close-up of the anterior region with the stylet (S) and the esophageal gland region (EGR) (reprinted from Juvale and Baum (2018)). *Right:* Infective juveniles (stage J2) (stained pink) enter from the root tip and migrate inside the root (bar size equals with 100μm) (reprinted from Gheysen and Mitchum (2011)).

The prolonged and complex interactions with the host start from the moment the nematode enters the root system till the formation of a feeding site. During this period, the nematode reprograms and manipulates host physiological and biochemical activities using a group of proteins or molecules that are crucial for its virulence activity, known as effectors. Effectors are expressed during infection and are delivered either to the host apoplast or cytoplasm targeting specific components of the host immune system (Jones and Dangl, 2006; Thomma et al., 2011). In nematodes, effectors are mainly produced in the dorsal and subventral gland cells that are based close to oesophagus (Davis et al., 2008; Gheysen and Mitchum, 2011). In the case of the cyst nematodes, the induced feeding cells are called syncytia, contrary to the giant cells of the root-knot forming species (Gheysen and Mitchum, 2011). Syncytium formation starts from a single initial cell located at the outer perimeter of the vascular cylinder. The cell wall of this initial cell is progressively partially degraded, and the protoplast is fused to those of the neighbouring joining cells.

At the same time, the subcellular organs (e.g. Golgi, ribosomes, mitochondria, vacuoles, nuclei) enlarge. The cytoskeleton of the feeding site cells is rearranged and their cell walls are transformed into finger-like structures aligning the syncytium with the vascular system in order to facilitate the flow of nutrients and water into the newly-

forming feeding site (Figure 1-7) (de Almeida Engler et al., 2004; Escobar et al., 2011; Wyss, 1992). Eventually, all the cells joining together show increased metabolic activity resulting from multiple duplications and transcriptional reprogramming (Jones and Northcote, 1972; Juvale and Baum, 2018; Lozano and Smant, 2011; Sobczak and Golinowski, 2011; Wyss and Zunke, 1986). In addition, plant hormonal pathways (e.g. auxin-induced pathway) that naturally regulate developmental processes in plants are manipulated in favour of the nematode (Gheysen and Mitchum, 2011; Goverse et al., 2000).

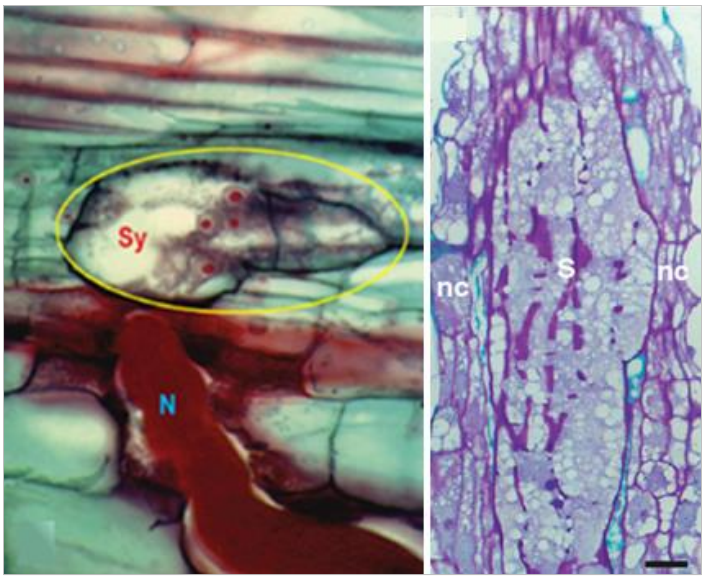


Figure 1-7 The syncytia of cyst nematodes. *Left*: Formation of a syncytium (Sy) by a juvenile (N) of the cyst nematode *Heterodera glycines* (reprinted from Juvale and Baum (2018). *Right*: Section of a syncytium (S) of the cyst nematode *H. schachtii* in *A. thaliana*. The surrounding neighbouring cells (nc) are still not part of the syncytium (bar equals to 25µm) (reprinted from Gheysen and Mitchum (2011)).

As soon as the syncytium has been established, the J2 becomes bigger in size and develops into a third and fourth-staged juvenile (J3 and J4) with moulting phases at the end of each stage. Sex determination takes place in the moulting phase between the J2 and J3 in the cyst nematode *Heterodera schachtii*. From the last moult a male or a female is formed (Juvale and Baum, 2018; Sobczak and Golinowski, 2011; Wyss, 1992). The females remain sedentary and continue to feed from the syncytium. Their body gradually becomes bigger and spherical and eventually arises through the ruptured root epidermis. In the meantime, the males become mobile again, exit the root and travel towards females, attracted by sex pheromones, to fertilise their eggs. Immediately after

fertilisation, the female dies, the body hardens and is transformed into a cyst enclosing the eggs (Figure 1-8) (Jones et al., 2013; Williamson and Hussey, 1996).

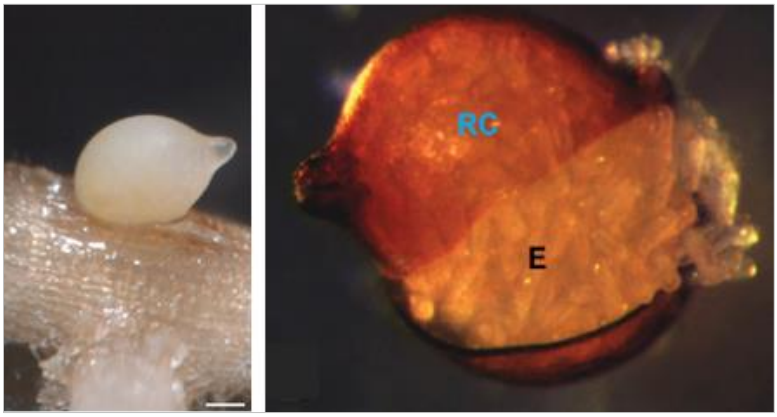


Figure 1-8 Cyst nematodes females and cysts. *Left*: Cyst nematode female developed on a soybean root at 30dpi (bar scale 100µm) (reprinted from Gheysen and Mitchum (2011). *Right*: Ruptured cyst (RC) of *H. glycines* with the contained eggs (E) (reprinted from Juvale and Baum (2018)).

1.3.2. Potato cyst nematodes

Cyst nematodes tend to be host specific (with exceptions), with this host specificity reflected in their common names. As such, the potato cyst nematodes (PCNs), *Globodera rostochiensis* (Woll.) and *G. pallida* (Stone) (family: Heteroderidae) are serious pests of potato. Both species are morphologically very similar, although the main difference can be found in the colour of maturing females which is yellow/gold in *G. rostochiensis* and white in *G. pallida* (Turner and Subbotin, 2013). They primarily infect plant roots consuming the water and nutrients that the plant uses; the root system and the tubers are therefore under-developed, and the above-ground parts grow very poorly. In the field, infected plants can appear as patches (Agrios, 2005).

It is thought that PCN were introduced in Europe in the middle of 19th century. After the catastrophic Irish famine caused by the late blight (*Phytophthora infestans*) outbreak in 1845, wild potato collections were brought into Europe from several sites of South America to be used as a breeding material against the oomycete, which in turn resulted in introduction of several PCN populations (Evans et al., 1975; Plantard et al., 2008). The first infestation of a potato field by PCN was reported in 1881 in Germany and a decade later in Scotland (Evans et al., 1975).

Today, both species are widespread, with Europe likely to have acted as a secondary source of PCN for many areas of the world, as a result of trade of infested seed potato. Every year 9% of global potato production is lost because of PCN (Turner and Subbotin, 2013). In Britain, PCNs were present in two-thirds of the potato fields (Minnis et al., 2002). However, they are still uncommon in Australia, north America, central Asia and India (Nicol et al., 2011). Recently, a significant and increasing problem was observed in Kenya (Mburu et al., 2018; Mwangi et al., 2015). Apart from these two species, other species belonging to the *Globodera* genus (e.g. *G. artemisiae*, *G. mexicana*, *G. ellingtonae*, *G. capensis*, *G. agulhasensis*) have been identified that parasitize *Solanaceae* sp. and *Compositae* sp. in places where *G. pallida* and *G. rostochiensis* are not as a significant problem, such as South Africa and New Zealand (Knoetze et al., 2017) (Figure 1-9).

In Europe, both *G. pallida* and *G. rostochiensis* constitute a threat to the potato industry, since the temperate European climate is favourable for them. It has been shown that *G. rostochiensis* eggs hatch at soil temperatures between 15°C and 27°C, while *G. pallida* is more adapted to lower temperature conditions (2 – 3°C degrees lower). This may be one reason why *G. pallida* populations are more prevalent in the northern European areas (e.g. Britain, Germany, the Netherlands) (Kaczmarek et al., 2014; Minnis et al., 2002). In addition, in fields where both populations are present, *G. pallida* shows greater multiplication level compared to that of *G. rostochiensis* (Hearne et al., 2017).

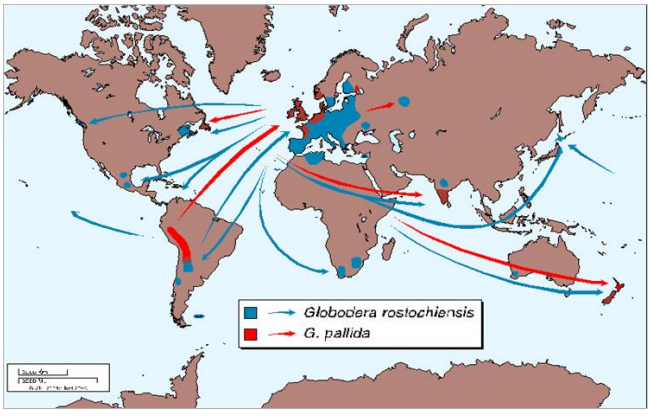


Figure 1-9 Reported worldwide distribution of PCN species and spread (USDA).

1.3.3. Controlling PCN

A number of measures have been developed for the control of PCN. One of the most usually used agricultural techniques is crop rotation. As previously described, potato

cyst nematodes (and other plant-parasitic species) need to find a suitable host to infect and parasitize. Continuous cultivation of a field with susceptible plants leads to a large multiplication rate of the pest. Switching to other non-hosts for some years may lead to a reduction of the number of eggs in the soil (Whitehead and Turner, 1998). However, the efficiency of rotation as a control strategy is hampered by the fact that cysts can remain viable in the soil in the absence of a suitable host or under unfavourable environmental conditions for many years, e.g. up to 20 years has been reported (Finkers-Tomczak et al., 2011; Lilley et al., 2005). However, crop rotation combined with other strategies (e.g. use of trap crops) can be still used to reduce nematode populations below the damage-threshold level (Phillips and Trudgill, 1998).

Nematicides have been used effectively in the past. The use of fumigants (e.g. methyl-bromide, 1, 3-dichloropropene) and non-fumigants (e.g. organophosphates) can be very expensive in large-scale applications, while their efficiency largely depends on the soil type and conditions (e.g. temperature, humidity, depth). Additionally, their use has a negative impact on the environment (soil and atmosphere) and on the health of non-target organisms (including humans). These considerations have led to the introduction of strict legislation by the EU and other global organisations on pesticides use (Directive 2009/128/EC) (European Union, 2009; Hughes et al., 2010; Trudgill et al., 2003; Trudgill et al., 2014; Whitehead and Turner, 1998). In addition to this, many countries categorise PCN species as quarantine pests to control their introduction to other areas.

In infected areas, early secondary potato varieties can be planted before the main one in order to stimulate the hatching of the eggs being in the soil. The hatched juveniles infect these trap plants, which are immediately removed and discarded. In countries with long periods of sunlight, infested plots can be covered with clear poly-ethylene plastics in order to create suffocating conditions by increasing the soil temperature. The main disadvantages of this control measure are the fact that the field should remain unexploited for a long period and that they may be effective only in the upper layers of soil (≈10cm), while cysts can be found in greater depths (Whitehead and Turner, 1998).

1.3.4. Resistance to PCN

Natural resistance is the most cost effective and environmentally-friendly way of controlling PCN (Bakker et al., 2011; Lilley et al., 2005). Fully resistant varieties allow no multiplication of nematodes. So far, fully resistant varieties have been developed only against *G. rostochiensis* (*H1*-containing varieties). Continuous use of these cultivars for 3 or 4 growing seasons can greatly reduce the number of viable eggs in the soil. Nevertheless this practice can lead to strong selection in favour of other species (e.g. *G. pallida*) or more virulent pathotypes (Whitehead and Turner, 1998). Resistance against *G. pallida* is more difficult to find, due to the complex nature of the *G. pallida* present in European soils. By ensuring their durability throughout time, newly developed varieties could be used along with other integrated pest management techniques ensuring nematode-free potato fields.

1.3.5. *Globodera pallida* populations

The existence of many different populations of *G. pallida* makes identifying major resistance sources complex. Phylogenetical studies analysing the mitochondrial genome of different *G. pallida* populations showed that these populations are genetically distant and can be categorised into different clades that may be related to their sites of origin. As mentioned in a previous section, PCN were introduced into Europe during the 19th century in different introduction events. These multiple introductions from different sites of the Andean region in South America possibly caused the emergence of multiple different *G. pallida* populations throughout Europe (Bryan et al., 2002; Evans et al., 1975; Hockland et al., 2012; Picard et al., 2007).

According to the Kort et al. (1977) classification system, three main pathotypes of *G. pallida* have been recognised: Pa1, Pa2 and Pa3. However, there is no clear distinction between some Pa2 and Pa3 populations in relation to their virulence against the two main resistance sources (*S. vernei* and *H3* derived from *S. tuberosum* ssp. *andigena* CPC2802) used to separate them (Blok et al., 1997; Phillips and Trudgill, 1998). In addition, the difference between Pa2 and Pa3 multiplication levels on these sources of resistance is quantitative and significantly depended on the populations origin, and there is therefore no reliable means by which to differentiate some populations into

either of these pathotypes (Phillips and Trudgill, 1998); hence they are often jointly referred as Pa2/3 pathotype. South American *G. pallida* populations tend to be more virulent on *H3* clones compared to the European populations, while both of them showed similar levels of virulence on *S. vernei* (Phillips and Trudgill, 1998). Random amplified polymorphic DNA (RAPD) analysis indicated significant genetic variation between the populations Pa1 and Pa2/3, though remarkably, the Scottish population Luffness, a Pa3 pathotype population, is genetically distant from both the rest of the European populations (e.g. British, Dutch and German populations) and Pa1. At the same time, Pa1 also seems to be distinct from the rest of the European *G. pallida* populations (Blok et al., 1997).

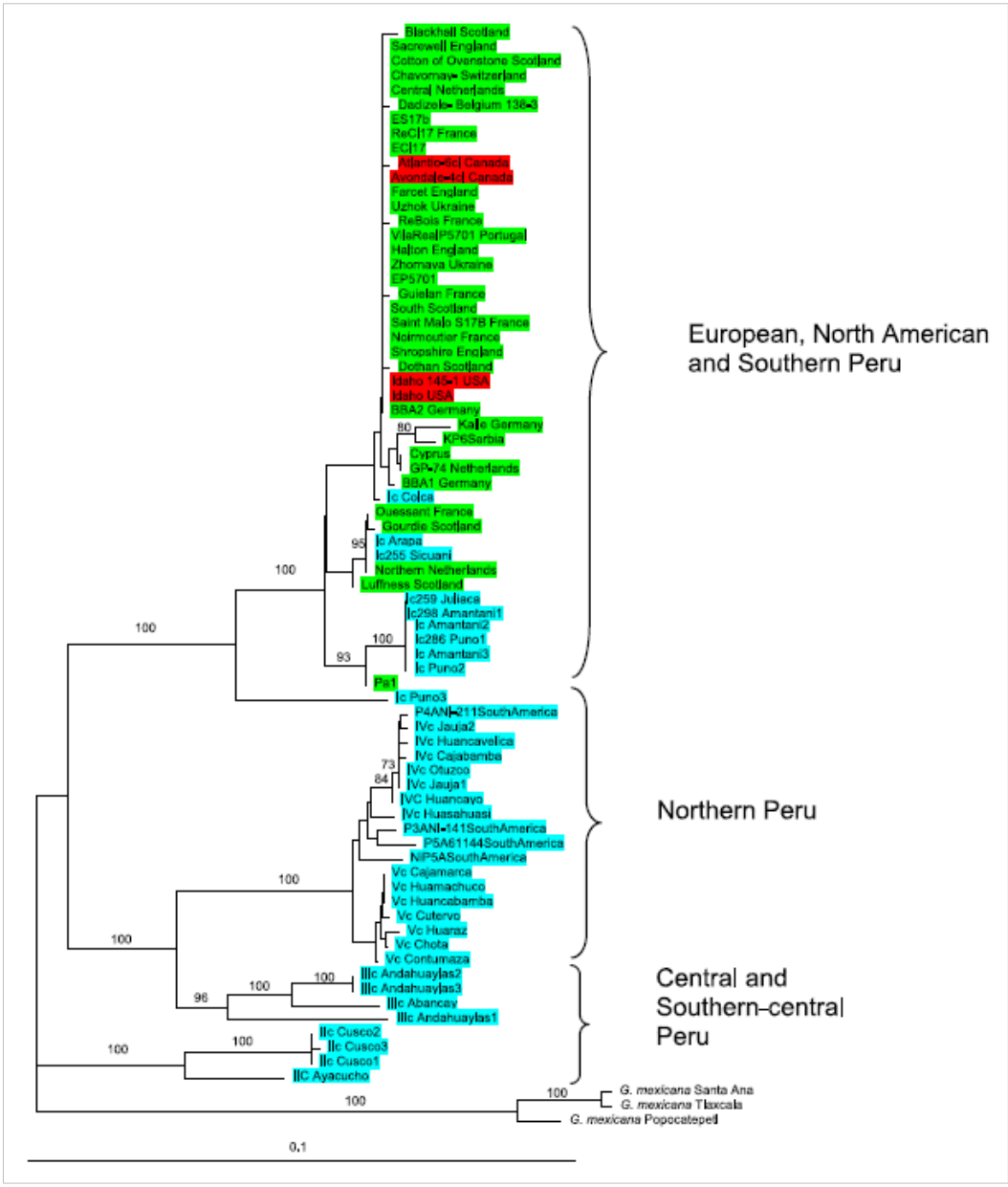


Figure 1-10 Phylogenetic analysis of *G. pallida* populations according to sequencing of the *cytb* gene. Populations highlighted with red colour indicates populations from North America, green from Europe and light blue from South America (reprinted and modified from Hockland et al. (2012)).

Sequencing of mitochondrial genes, such as that of *cytochrome b* (*cytb*) categorised most British populations in the same most common *cytb* haplotype with Western European populations (Plantard et al., 2008). Interestingly, Luffness and Pa1 showed a less common haplotype and this may be a result of the multiple introductions into the European continent as described before (Figure 1-10) (Eves-van den Akker et al., 2015; Hockland et al., 2012; Plantard et al., 2008). In line with this, *G. pallida* samples recently isolated from potato fields in Scotland showed the presence of three different *cytb*

sequences. Four-fifths of the fields sampled had a single type of DNA while the rest had mixed types, which also indicates the genetic variability of *G. pallida* populations (Eves-van den Akker et al., 2015).

1.3.6. *Globodera rostochiensis* pathotypes and resistance

Similar to *Globodera pallida*, *G. rostochiensis* has also been classified into several pathotypes on the basis of their ability to infect potato genotypes carrying a range of resistance sources. According to the international scheme of PCN species classification (Kort et al., 1977), five pathotypes have been recognised in *G. rostochiensis* (Ro1 to Ro5), with Ro1 being the most dominant in the United Kingdom (Eves-van den Akker et al., 2016).

Seven *loci* have been identified in wild-potato species related to resistance against *G. rostochiensis*. Only three of those confer almost full resistance i.e. the *loci H1* and *GroV1* in chromosome V and the *locus Gro1* in chromosome VII. On the other hand, the *loci Gro1.4* (chromosome III), *Grp1* (chromosome V), *Gro1.2* (chromosome X) and *Gro1.3* (chromosome XI) give only partial resistance to one or more pathotypes (Bakker et al., 2004; Gebhardt and Valkonen, 2001). Moreover, a single *locus* may confer resistance to more than one species such as *Grp1* that is linked with resistance to both *G. rostochiensis* and *G. pallida* (van der Voort et al., 1998).

The *H1* resistance gene is the only identified dominant gene conferring resistance to any PCN species (Huijsman, 1955). Since its discovery in 1952, *H1* has been exploited widely by the breeding industry and is still highly effective (Bradshaw and Ramsay, 2005; Evans, 1993). Although it confers resistance to a number of *G. rostochiensis* pathotypes such as Ro1 and Ro4, some pathotypes (i.e. Ro2, Ro3 and Ro5) are still able to overcome *H1*-derived resistance (Kort et al., 1977). It originates from a Commonwealth Potato Collection clone of *Solanum tuberosum* ssp. *andigena* CPC1673. Gebhardt et al. (1993) showed the presence of the restriction fragment length polymorphism (RFLP) marker CP113 linked with the *H1 locus.* A decade later, a bulked segregant analysis (BSA) confirmed the presence of the amplified fragment length polymorphism (AFLP) marker CM1 at a distance of less than 0.1cM from the *H1 locus*. This marker, which is very close to the previous CP113, co-segregates with the *H1 locus* in the F_1 progenies of the cross

between the *H1*-harbouring female parent and the male parent susceptible to *G. rostochiensis* (Bakker et al., 2004). By combining the above, it can be deduced that the *H1 locus* is located close to the edge of the long arm of the chromosome V of *S. tuberosum* ssp. *andigena* CPC1673. Mostly, in the same chromosomal region in *S. vernei*, the *locus GroV1* has also been mapped, which may indicate that the *loci H1* and *GroV1* may be allelic (Jacobs et al., 1996). Resistance to *G. pallida* is covered in detail in Chapter 3.

1.4. The plant immune system

1.4.1. Plant immunity and the gene-for-gene model

Pathogens and pests are engaged in a continuous battle with their host plants leading to the development of complex relationships. During these relations, the invading organisms try to settle on their hosts for survival and reproduction, whereas plants invest energy in defence to counteract them (Dangl et al., 2013; Jones and Dangl, 2006). Plants have an innate immune system organised in two layers. Most potential pathogens are unable to invade most plants because of the physical barrier that the waxy cuticular layer provides. Attackers that manage to break this barrier have to deal with the first level of plant defence. At this level, plants recognise widely conserved, pathogen/microbe-associated molecular patterns (PAMPs/MAMPs) with transmembrane pattern recognition receptors (PRRs). Based on their intracellular domain structure, i.e. presence or non-presence of an intracellular kinase domain, they are divided into receptor-like kinases (RLK) and receptor-like proteins (RLP). Although the kinase domain is responsible for the activation of the downstream signalling inside the cell, in the absence of it (e.g. in RLPs) activation requires the interaction with a co-receptor (Yang et al., 2012).

Both receptor types contain a single-pass transmembrane nucleotide-binding (NB) domain and their extracellular domains usually consist of a diverse leucine-rich repeat (LRR) domain, which is similar to the Toll-like receptor (TLR) in animals (e.g. in *Drosophila*) or the carbohydrate-binding LysM domain, with the first one being the most common in plants (Dangl et al., 2013; Ma et al., 2013; Monaghan and Zipfel, 2012). Once

the epitope of the PAMPs/MAMPs binds to the host LRR domain, basal defence responses (e.g. activation of ion fluxes and MAPKs, production of reactive oxygen species and local cell wall reinforcement) are activated; these are collectively known as PAMP-triggered immunity (PTI) (Jones and Dangl, 2006; Ma et al., 2013; Zipfel, 2008). PTI responses can also be induced by the detection and recognition of plant-derived molecules (such as plant cell wall-derived oligogalacturonides) produced by the fragmentation of their own cellular structure caused by microbial attack and infection. These molecules are also known as damage-associated molecular patterns (DAMPs) and they initiate basal defence responses similar to those induced by PAMPs/MAMPs (Boller and Felix, 2009; Monaghan and Zipfel, 2012; Yang et al., 2012).

One of the most well-studied examples of PAMP recognition is the recognition of the bacterial flagellin epitope flg22 by the receptor FLS2 (flagellin sensing 2) in *Arabidopsis thaliana*. The conserved 22-amino-acid epitope of the N-terminus of the flagellin, which is the main component of the most of the plant-parasitic bacterial flagella, is recognised by the LRR domain of FLS2 initiating immune responses, such as localised callose deposition and accumulation of defence proteins (Zipfel et al., 2004). The RLK receptor BAK1 (BRI1-associated receptor kinase 1) interacts with FLS2 during flg22-recognition (Chinchilla et al., 2007; Monaghan and Zipfel, 2012; Yang et al., 2012). Fungal chitin and xylanases, oomycete glucans and cell wall transglutaminases, and bacterial invertases, lipopolysaccharides and the elongation factor EF-Tu also elicit basal defence responses in plants (Nurnberger and Kemmerling, 2009). Only recently, a conserved family of pheromones, known as ascarosides, was found as the first nematode PAMP that triggers PTI responses (Manosalva et al., 2015); however, more research has to be done on this.

1.4.2. Evolution of plant-pathogens/pest interactions and *avirulence* genes

During evolution, well adapted pathogens have evolved ways to evade or counteract PTI-responses by releasing highly specialised proteins, known as effectors. Effectors promote virulence of pathogens/pests by successfully suppressing PTI responses resulting in effector-triggered susceptibility (ETS) (Boller and Felix, 2009; Jones and Dangl, 2006). Specialised intracellular or transmembrane proteins (Resistance proteins –

R) are able to recognise these molecules, or their biochemical activity, activating a second defensive wall against the microbial attackers. These effector-triggered immunity (ETI) responses are more robust and faster compared to PTI and frequently result in local cell death (hypersensitive response - HR) (Dangl et al., 2013; Jones and Dangl, 2006; Ma et al., 2013). In the case of fungi and bacteria, an HR can prevent the further spread of the pathogens into the neighbouring healthy cells while, in the case of nematodes, it may lead to destruction of the developing feeding structure or may prevent spread of the developing feeding sites restricting food supply to the nematode (Lozano and Smant, 2011). The most common structure of the R proteins consists of a NB-LRR domain. Connected with the NB-LRR structure, some proteins have an additional domain that is a homologue to the Drosophila Toll and mammalian interleukin-1 receptor (i.e. TIR-NB-LRR) while others have a coiled-coil structure (i.e. CC-NB-LRR) (Coaker and Baker, 2013; Dangl and Jones, 2001; Ma et al., 2013).

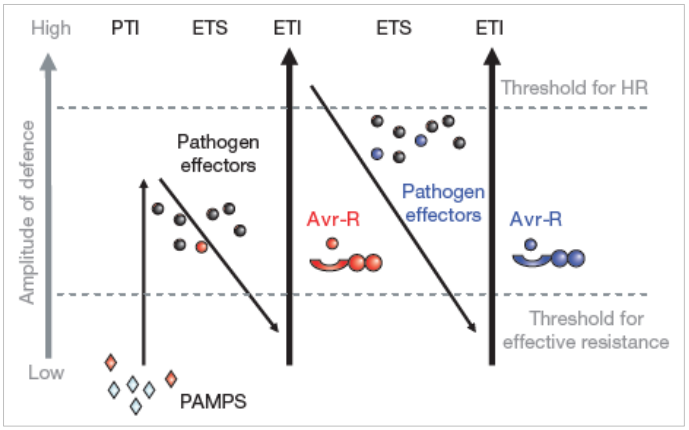


Figure 1-11 The 'zig-zag' model as proposed by Jones and Dangl (2006). PAMPs are recognised by plants during the attack and the first line of defence is activated (PTI). However, PTI can be bypassed by the secretion of effectors that promote virulence (ETS). Highly specific R proteins recognise specific effectors activating robust defence responses (e.g. HR); the recognised effector acquires then Avr activity. Sometimes, microbes may acquire or develop new effectors that can evade ETI and subsequently promote virulence. Selection over new resistant plants possessing new resistance proteins efficient to newly developed effectors can lead again to ETI activation. The energy cost for the host when activating ETI is higher to the PTI activation (reprinted by Jones and Dangl (2006)).

All the above mechanisms (Figure 1-11) can be summarised into a theoretical model known as the 'zig-zag' model. Secreted effectors that are recognised by R proteins leading to activation of immune responses are termed as *Avirulence* (*Avr*) genes and the interaction between them and the cognate *R* gene is highly specific (Bent and Mackey,

2007; Jones and Dangl, 2006; Sacco et al., 2009). In the absence of the corresponding *R* gene, *Avr* genes may contribute significantly to pathogen virulence (van der Hoorn and Kamoun, 2008).

Hence, ETI is a result of a continuous co-evolution of pathogens with their host plants. PTI defence was developed early triggered by the conserved PRR proteins that can interact with a wide number of pathogens-derived elicitors. ETI was evolved (and is constantly evolved) as a result of selection pressure, mutations, gain or loss of function of the elicitors. Resistance genes were developed later in evolution in order to recognise more specific effectors. Some of the ETI responses, such as cell death, can be very active and robust (Jones and Dangl, 2006; Thomma et al., 2011). Most of the strategies used by the plant breeding industry for obtaining resistant varieties is to identify resistance genes that recognise specific pathogen effectors and introgressing them into commercial varieties.

1.4.3. The 'guard' hypothesis

It was originally thought that a direct physical interaction between effectors with the LRR domain of the corresponding resistance protein was likely to underpin recognition of pathogens by resistance genes. However, this 'Avr-R' interaction is rarely observed in practice. It is widely believed that the recognition of pathogen effectors occurs mostly indirectly. In other words, receptors monitor changes brought about in the host by these effectors in order to initiate defence responses (Bozkurt et al., 2011; Lozano-Torres et al., 2012; van Esse et al., 2008). Plant R proteins monitor the activity of other host molecules (i.e. guards) that function as virulence targets. This model also explains why a single R protein can recognise multiple unrelated pathogens. A good example of this is the tomato-derived *R* gene *Mi-1.2* that confers resistance against the tomato psyllid *Bactericera cockerelli*, the potato aphid *Macrosiphum euphorbiae*, the whitefly *Bemisia tabaci* and root-knot nematodes *Meloidogyne* spp. (Casteel et al., 2006; Dangl and Jones, 2001; van der Hoorn and Kamoun, 2008). In this case, it is likely that each of the pathogens is targeting the same host process that is guarded by the gene *Mi1.2*.

Several "guarded" molecules that are targeted by diverse pathogens have been identified. The tomato cysteine protease Rcr3 (required for *Cladosporium fulvum*

resistance 3) contributes in the activation of immunity against several totally different phylogenetically pathogens. The receptor Cf-2 is able to sense perturbations of Rcr3 after its interaction with effectors secreted by the cyst nematode *Globodera rostochiensis* and the leaf mould fungus *C. fulvum* triggering immune responses (Lozano-Torres et al., 2012; Rooney et al., 2005).

1.5. Cyst nematode effectors and *avr* genes in their interaction with their hosts

As previously described, effectors promote virulence. Nevertheless, in the event of an incompatible interaction with a cognate R protein, downstream cascades are activated leading to defence responses (Figure 1-12). During this pathway, activation of transcription factors (e.g. WRKYs, basic-leucine zipper and myeloblast families) and upregulation of defence-related genes (such as peroxidases, lipoxygnases and ROS species) occur (Kaloshian et al., 2011). The analysis of expressed sequenced tags (EST) from cDNA fragments was until recently the only way to study and identify effectors in plant-parasitic nematodes. The publication of the genome assemblies of *G. rostochiensis* (Eves-van den Akker et al., 2016) and *G. pallida* (Cotton et al., 2014; Thorpe et al., 2014) provided the opportunity to study whole effector complements from these nematodes. The majority of the *G. rostochiensis* dorsal gland effector families contain a conserved motif in the promoter region (called the dorsal gland promoter element – DOG box) (Eves-van den Akker et al., 2016). This provides a new method for identifying effectors in genome sequences. These analyses suggest that several hundred effectors are deployed by PCN during the interaction with their hosts.

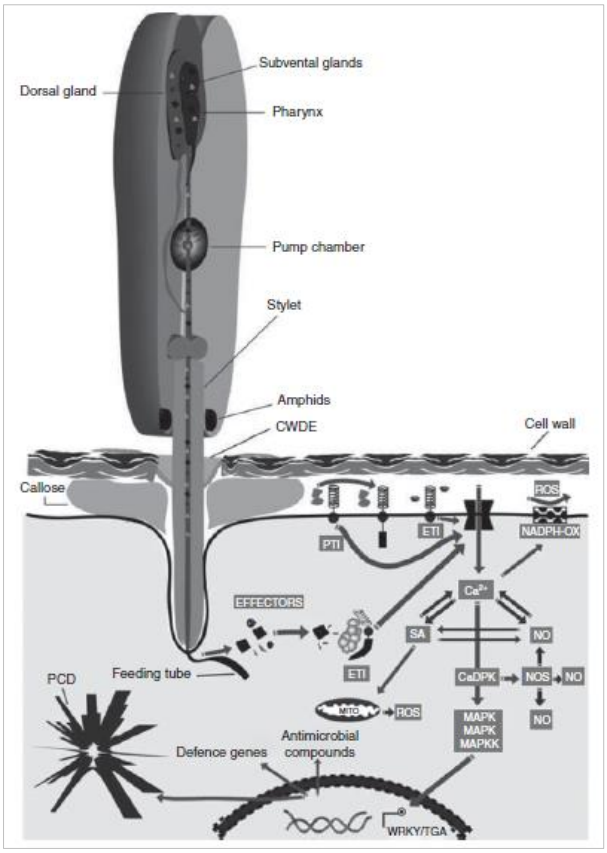


Figure 1-12 Secreting effectors through the stylet. A cocktail of effectors is produced in the two oesophageal gland cells (dorsal and subventral) and is secreted through the stylet into the host. CWDEs are secreted in order to degrade the cell wall and facilitate penetration and migration. Defence responses may be initiated upon the recognition of nematode-derived effectors. The activation of PTI and ETI leads to an activation of downstream signalling pathways where responses such as ion fluxes, hormonal signalling (SA) and production of reactive oxygen species (ROS) are triggered. Nematodes therefore produce effectors that suppress these responses (reprinted from Lozano and Smant (2011)).

One of the most diverse gene families encoding effectors is those containing the SRPY domain (named after the SP1a protein from an amoebea species and the mammalian Ryanodine Ca^{2+}-release channel receptor). The SPRY domain has a diverse N-terminus with no enzymatic activity attached to a signal peptide, which is necessary for secretion (Mei et al., 2015; Perfetto et al., 2013). Nematode SPRYSECs were firstly identified in the PCN species *G. rostochiensis* and *G. pallida* where they are expressed in the dorsal glands during the early stages of parasitism (Jones et al., 2009; Rehman et al., 2009). The effector SPRYSEC-19 from *G. rostochiensis* interacts physically with the LRR domain of the NB-LRR tomato protein SWF5 without causing defence-related programmed cell death.

Some SPRYSECs are known to suppress ETI. For example, SPRYSEC-19 seems that after its recognition by the resistance gene *Gpa2* and *Rx1* disturbs the activation of

programmed cell death (Mei et al., 2015; Postma et al., 2012). In contrast, the *G. pallida SPRYSEC Avr* gene *RBP-1* interacts with the LRR domain of the nematode R protein Gpa-2 triggering programmed cell death on *Nicotiana benthamiana* leaves. Variants of the *RBP-1* gene differing in only a single amino acid in the sequence of the variable N-terminus did not cause any defence response. Only a small change of the SPRY domain can determine avirulence or virulence activity. This also may prove that the great variability in the N-terminus is a result of the evolutionary attempt of nematodes to avoid NB-LRR-dependent immune responses (Diaz-Granados et al., 2016; Goverse and Smant, 2014; Sacco et al., 2009).

In addition to the SPRYSEC effectors, other proteins function as defence suppressors. In the secretome of some plant-parasitic species (e.g. *G. rostochiensis*, *M. incognita*, *H. schachtii*) effectors of the venom allergen-like proteins (VAPs) have been identified. Members of these families, which are conserved in both animal- and plant-parasitic nematode species, are produced in the gland cells. Juveniles secrete VAPs in the early stages of parasitism resulting in the suppression of immune responses mediated by surface-localised receptors in *Arabidopsis* (Lozano-Torres et al., 2012; Lozano-Torres et al., 2014). *G. rostochiensis* secretes the ubiquitin-like protein CEP12 that hampers flg22-associated PTI responses (Chen et al., 2013), while 30C02 from *Heterodera* spp. interacts with plant-derived endoglucanases to suppress immune responses (Hamamouch et al., 2012).

Other effectors have been identified that have other roles in the plant-nematode interaction. The best-studied effector group in cyst nematodes encode small peptides similar to Arabidopsis CLAVATA (CLV) elements (CLEs) (Guo et al., 2011; Lu et al., 2009). The CLE family in plants includes a variety of proteins that function as ligands for plant RLK/RLP receptors, such as CLV1 and CLV2, with an important role in several developmental pathways (e.g. cell fate, meristem development, balance between cell proliferation and differentiation etc.) (Rojo et al., 2002). Cyst nematodes, including *Heterodera glycines, H. schachtii* and *G. rostochiensis*, secrete CLE-like proteins to mimic plant-derived CLEs and therefore, facilitate their parasitism on the host (Guo et al., 2011; Haegeman et al., 2012; Lu et al., 2009; Wang et al., 2011). In many plant developmental

pathways, phytohormones (e.g. auxin) play an important role. Many plant-parasitic nematode species can interfere in the auxin-regulated signalling pathways by secreting chorismate mutases in favour of establishing feeding sites (Grunewald et al., 2009; Haegeman et al., 2012). Other proteins (such as the 19C07) target auxin transporter family members (e.g. AUXIN RESISTANT1/LIKE AUX1 – AUX1/LAX in *A. thaliana*) in order to redirect their cellular localisation towards the feeding sites (Lee et al., 2011).

The role of CWDEs in the lifecycle of plant-parasitic nematodes was mentioned before. Through the stylet, a cocktail of effectors (including CWDEs such as β-1,4-endoglucanase, cellulases, pectinase) is secreted to degrade the components of cell wall (i.e. cellulose, hemicellulose and pectin) accompanied with the exerted mechanical disruption (Haegeman et al., 2012; Rehman et al., 2009; Smant et al., 1998).

It can be understood that plant-parasitic nematodes and cyst nematodes specifically, have developed a large arsenal of effectors to assist their host invasion, migration, manipulations of host biochemical pathways and to allow them to suppress host defence responses. A high level of variability can also be found within the members of an effector family that may indicate evolutionary attempts of nematode species (or even populations) to avoid recognition by plant immune receptors according to the 'zig-zag' model described in a previous section of this chapter. Further analysis on evolution of effectors and more specifically on *Avr* genes will be done in the next chapters of this thesis.

1.6. Thesis' outline

Effectors play a key role in the invasion and establishment of pathogens inside the host plants. On the other side, plant-derived resistance proteins are able to recognise specific effectors activating immune responses against the invader; these recognised effectors are consequently termed as *avirulence*. Studying of this pathosystem is essential in the breeding industry. Towards this, in the last years, new sequencing technologies as well as the decrease of their service cost, have made the identification of effectors quicker and more accurate, especially for small and complicated organisms like PCNs. The genome

assembly of *G. pallida* was the first PCN assembly published a few years ago (Cotton et al., 2014) followed by that of *G. rostochiensis* (Eves-van den Akker et al., 2016). Apart from the identification of effector families, these studies have also revealed important aspects of parasitism. The present PhD project exploited up-to-date genome-based approaches for the identification of candidate *avr* genes in the PCNs.

As a first step, the diversity in virulence of *G. pallida* selected populations is examined. These populations, which had been generated by Phillips and Blok (2008) for increased virulence on two main resistance sources (*Gpa5* from *S. vernei* and *H3*), are screened on potato genotypes containing different levels of resistance from those sources (Chapter 3). These phenotypic results helped us to study differences in virulence between the unselected and selected populations along the different resistances, as well as providing us enough unique nematode material for the next steps. As a next step (Chapter 4), effector-encoding genomic regions from each defined selected population were sequenced using captured technology (target enrichment sequencing) (Jupe et al., 2013) and achieving large read coverage of the captured genes by performing short-read sequencing. Analysis of the captured sequences using computational biology tools reveals polymorphisms depending on the population selection background. Variant sequences represent candidate effector genes.

The same selected populations are also whole-genome re-sequenced (Chapter 5) to provide us with important information about host adaptation and selection. For both downstream analyses, the newly developed and improved *G. pallida* genome assembly (Chapter 4) is used from DNA extracted from the cyst of the founder, unselected population (Newton) and sequenced on a long-read sequencing platform.

In the last part of this study, functional validation of a candidate *G. rostochiensis avr* gene, which had been identified by Eves-van den Akker et al. (2016) is performed (Chapter 6) by expressing it *in planta*, using *Agrobacterium*-mediated transient expression system. All the findings are discussed and summarised in a more general context in chapter 7. The discussion also points out how the gained knowledge can contribute to the main goal of the breeding industry for dealing with the PCN problem. In addition to that, suggestions for future research based on these results are given.

2. Materials & Methods

2.1. Biological material

2.1.1. Tomato root diffusate

The roots of a 4-week old tomato plant (*S. lycopersicum* cv. Moneymaker) were carefully washed in order to remove soil and were placed in 500ml sterile distilled water overnight at room temperature. After this time, the tomato root diffusate (TRD) solution was filtered using a filter paper (Whatman) to remove any remaining soil or root tissues and stored at 4°C before being used. TRD was used within a few weeks of being produced.

2.1.2. Nematode material

Nematode cysts from two species were used in this study; *Globodera pallida* and *G. rostochiensis* (Table 2-1). All cysts were stored at 4°C organised in batches according to the species, population/pathotype and the harvest year. In Chapter 3, two English *G. pallida* field populations (Newton and Farcet) and generated sub-populations selected on various resistance sources were used. These "founder" populations had been reared on four partially resistant potato clones for 12 generations (Phillips and Blok, 2008); two of them derived from *Solanum vernei* (clones Sv_8906-or Guardian and Sv_11305-or Morag) and two from *S. tuberosum* ssp. *andigena* CPC2802 (clones Sa_11415 and Sa_12674). In each generation, a subset of 10 cysts developed during the previous generation development was used for the inoculation of a specific potato clone. At the end of each series, all the developed cysts from all the 4 biological replicates of each 'nematode population x potato clone' combination were pooled together. For the next generation, 10 cysts (of the previously developed pool) per biological replicate were randomly chosen to inoculate a potato plant of a specific genotype (Phillips and Blok, 2008).

Subsequently, the Newton (n) selected populations are respectively called here n-8906, n-11305, n-11415 and n-12674, while the Farcet (f) selected populations as f-8906, f-11305, f-11415 and f-12674 respectively (Figure 2-1 and Table 2-1). In chapter 6, two *G. rostochiensis* pathotypes were used; Ro1 and Ro5 (Table 2-1).

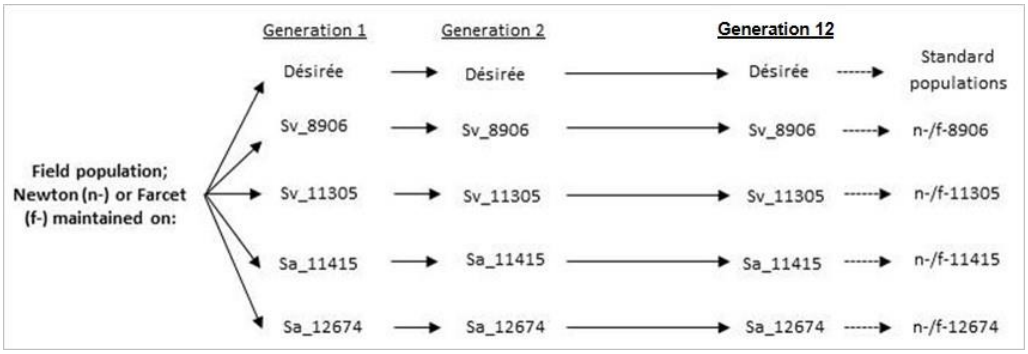


Figure 2-1 Diagram showing the development of the selected Newton or Farcet *G. pallida* populations. The prefixes 'n-' stand for Newton, 'f-' for Farcet, 'Sv_' for *S. vernei* and 'Sa_' for *S. tuberosum* spp. *andigena* CPC2802.

Table 2-1 Summary of starting nematode (species, pathotypes, populations) material used in this study.

Species	Pathotype	Population	Chapters used
G. pallida	Pa2/3	Newton standard	3, 4
G. pallida	Pa2/3	n-8906	3, 4, 5
G. pallida	Pa2/3	n-11305	3, 4, 5
G. pallida	Pa2/3	n-11415	3, 4, 5
G. pallida	Pa2/3	n-12674	3, 4, 5
G. pallida	Pa2/3	Farcet standard	3
G. pallida	Pa2/3	f-8906	3
G. pallida	Pa2/3	f-11305	3
G. pallida	Pa2/3	f-11415	3
G. pallida	Pa2/3	f-12674	3
G. rostochiensis	Ro1	-	6
G. rostochiensis	Ro5	-	6

2.1.3. Hatching

Prior to each experiment, the cysts were assessed for vitality and response to root diffusates. For the hatching test, 24-well tissue culture plates (Corning-Falcon) were used at room temperature in the dark. The wells were filled with 1ml of sterile distilled water and 2 cysts from each species/population were transferred into each well with forceps with four replicates for each. After 48hrs, the sterile water was replaced with the same volume of TRD solution. The cysts used in the control treatment were exposed to fresh sterile distilled water in place of TRD. Juveniles emerging from cysts were counted 3 days after the TRD application and at regular intervals for the following 7 days.

2.1.4. Harvesting J2s

To collect second-stage juveniles (J2) at least 200 cysts of a specific species and population were placed on a sieve with a 25µm aperture. Then they were placed in a 150mm petri dish filled with 100ml of fresh TRD solution, enough for the cysts to be in contact with it. The cysts were incubated at 18°C in the dark. The hatched J2s were able to move through the sieve openings while the cysts remained on the sieve. The TRD solution (containing the hatched J2) was placed in 50ml tubes (Corning-Falcon) and centrifuged for 10min at 2,500rpm to pellet the J2s. The pellet was then transferred into a fresh 15ml tube (Corning-Falcon) with 5ml of the supernatant. In the same tube, 5ml of 50% sucrose solution was added with 500µl of sterile distilled water on top of it. The tube was then centrifuged again for 10min at 2,500rpm. The J2s were gathered from the layer between the sucrose and the sterile distilled water and transferred into a 1.5ml low-bind Eppendorf tube by pipetting followed by a short spin to remove any excess of water and sucrose solution. The collected pellet (consisting of J2s) was frozen in liquid nitrogen and stored at -80°C till further use. The same procedure was repeated regularly for approximately a month until all the eggs hatched.

2.1.5. Potato material

Nine potato genotypes were screened in virulence tests (chapter 3). Five of these were commercial (tetraploid) varieties (Désirée, Vales Everest, Innovator, Royal and Arsenal) and 4 were (diploid) clones from the Commonwealth Potato Collection (CPC) of The James Hutton Institute (Sv_8906, Sv_11305, Sa_11415 and Sa_12674). The Sv_8906 and Sv_11305 lines have also been used for developing the varieties Guardian and Morag respectively as mentioned in a previous study (Phillips and Blok, 2008). The resistance sources in these potato lines were derived from either *Gpa5* from *S. vernei* or *S. tuberosum* ssp. *andigena* CPC2802 (*H3*), except for the susceptible cultivar Désirée and Royal. Royal (parentage Midas x 92-BUY-1) is a Danish variety highly resistant to *G. rostochiensis* Ro1 pathotype with low resistance to all the *G. pallida* pathotypes. Vales Everest (parentage 12674ab1 x Cara) is a variety developed and registered by the Scottish Crop Research Institute in 2005 and has *H3*-derived resistance. The most resistant varieties tested were Arsenal (parentage AR92-1146 x Silvester) and Innovator (pedigree of the Canadian Shepody and the Dutch clone RZ-84-2580) from the Netherlands that both contain

resistance from *S. vernei* (AHDB potato variety database; Scottish Agriculture Science Agency database) (Table 2-2 and Supplementary Data 1).

In chapter 6, two varieties were used for transient expression of *Agrobacterium tumefaciens in planta* with a candidate *G. rostochiensis* effector gene; Désirée (as a susceptible control) and Maris Piper as a *H1*-containing resistant variety) (Table 2-2).

Table 2-2 Summary of the starting potato material used in this study (information taken by The European Cultivated Potato Database and The AHDB Potato Variety Database).

Variety/Clone	Resistance to *G. pallida*	Resistance to *G. rostochiensis*	Chapter used
Désirée	none	none	3, 6
Vales Everest	medium	medium	3
Innovator	very high	low	3
Royal	low	high	3
Arsenal	very high	medium	3
Sv_8906	high	-	3
Sv_11305	high	-	3
Sa_11415	high	-	3
Sa_12674	high	-	3
Maris Piper	low	very high	6

2.1.6. Virulence tests

In Chapter 3, the reproductive ability of *G. pallida* nematode populations selected for high virulence on different potato lines was examined. Thus, screening tests were performed in the greenhouse facilities of The James Hutton Institute. The experiments took place between November and December 2015 and between May and September 2016. The average day temperature in the greenhouse was set to 18°C and average night temperature at 15°C. Eight hours of continuous dark conditions were controlled with additional artificial yellow light during the short-days period and with ambient lighting (combined with shading when needed) in long-days period.

The tests were carried out in root-trainer pots (12cm deep; Haxnicks, Bristol, UK). These were three-quarters filled with insecticide-free compost (prepared by greenhouse staff) containing 20 cysts. A tuber section (sized ≈ 1.5cm x 1.5cm), from each potato genotype, containing a single sprouted eye was planted on the surface of the compost with the sprout pointing downwards. Each experiment was repeated twice and each 'nematode

population x potato line' combination had 4 replicates in each complete randomised experiment (Figure 2-2). All the cysts had previously been tested for vitality and fungus contamination in hatching tests and in case of failure the virulence test was postponed until a new batch was available.

Figure 2-2 Set up of virulence tests. (a) Each group contains 36 root trainer compartments. All the root trainers in each group were inoculated with the same population to avoid cross-contamination. A single sprouted bud from a specific potato genotype was placed in each root trainer in a complete randomised design. (b) The root trainers allow the development of roots in parallel. (c) At 7wpi, the number of females feeding on roots were counted and collected. The females are visible as small white spheres developed along roots (scale bar equals with 1cm).

In each experiment, a Newton and Farcet field (standard) population along with their 4 selected sub-populations were tested on 9 potato genotypes (lines) in total. However, one line (Arsenal) was used in the Farcet screening tests only once, due to the lack of availability of tubers. Seven weeks post-inoculations, the root trainers were opened carefully without disrupting the developed root system and were placed under a magnifier lens. Females that were visible on the root surfaces were counted and collected using fine tweezers without damaging them and stored in sterile Eppendorf tubes. The collected nematode material was frozen in liquid nitrogen and subsequently stored at -20^{o}C for later use.

2.1.7. Statistical analysis

The number of females was calculated as a mean number of the counted developed females for each 'population x potato line' interaction. The mean number from both technical repeats per population was calculated. Prior to this, paired t-tests (p-value ≤ 0.05) were carried out in order to find whether each specific 'population x potato line' application differed significantly in each technical repeat. In general, in both experiments each 'population x potato line' showed a similar trend with regard to a number of developed females. Statistical analysis was performed using the software SigmaPlot Version 13.0 (Systat Software Inc., San Jose California USA, www.systatsoftware.com).

2.2. Molecular Biology

2.2.1. Primers

All the primers (Eurofins) used in the current project for cloning candidate effectors or diagnostic purposes are listed below (Table 2-3). The primers G12477WLF_sp, G12477WLR, G13394WLF_sp, G13394WLR and G13394WLF were designed with the online tool Primer 3 Plus Version 2.4.2. The signal peptide present on these sequences was predicted using the online tool SignalP 4.1 (Petersen et al., 2011). The primers 18S-UNI, PITSr3 and PITSp4 are universal primers used for distinguishing *G. rostochiensis* and *G. pallida* (Bulman and Marshall (1997)). The primers M13-FOR and M13-REV are universal vector primers. The primers PVX-201 were used specifically to verify the cloned insert in the vector pGR106GW.

Table 2-3 A summary of the primers used in this study. Their sequence, the annealing temperature of each and the chapter where they were used are also shown.

Gene	Sequence (5'-3')	Tm (°C)	Chapter used
18S-UNI	CGTAACAAGGTAGCTGTAG	54.5	6
PITSr3	AGCGCAGACATGCCGCAA	58.2	6
PITSp4	ACAACAGCAATCGTCGAG	53.7	6
G12477WLF_sp	ATGTCTGCTAATAATCTTACCGTTTTC	58.9	6
G12477WLR	TTAGCCTAACTTAGGATCGCGC	60.3	6
G13394WLF_sp	ATGAATGGACTGATCGGAATATTG	57.6	6
G13394WLR	TTAACCTGCAGAATCCGGGC	59.4	6
G13394WLF	ACCATGCAACCAAGCACAAGTC	60	6
M13-FOR	GTAAAACGACGGCCAG	57.1	6
M13-REV	CAGGAAACAGCTATGAC	52.6	6
PVX-201-Seq-F	GCAGTCATTAGCACTTCCTTAGTGAGG	63	6
PVX-201-Seq-R	CCTGAAGCTGTGGCAGGAGTTGCGC	63	6
GAPDH-FOR	GTGATTAGCAACGCTTCGTG	55.9	4
GAPDH-REV	GTCATCAGCCCTTCGATGAT	56.1	4
SPRY-414-2L	GCCAAGGTTACAGGAAAGAA	54	4
SPRY-414-2R	TTTGTTTGGTCGCAAGTCCA	56.5	4
SPRY-1719-1L	AGAGAAAGGAGAGCACAACG	56	4
SPRY-1719-1R	TTTGAGTATGCGTAAGTGCC	54.1	4
G16H02L	GTCGTTCTCCGTCATTTTGG	55.5	4
G16H02R	GGAAAGCGTGTGAAAGGCAC	59.2	4

2.2.2. Small-scale DNA extractions from cysts or females

DNA was extracted from ~40 either cysts of females of each nematode population. Firstly, they were frozen in liquid nitrogen in 1.5ml Eppendorf tubes and crushed using an autoclaved micropestle. Six hundred microlitres of Cell Lysis Buffer (Qiagen) was added and additional crushing was performed. Next, 5µl Proteinase K (20mg/ml) (Qiagen) was added and the sample was mixed by vortexing. The samples were incubated for about 18hrs (until the next morning) at 56°C on a heating block and mixed occasionally by inversion.

The next day, 4µl RNAse A (100mg/ml) (Qiagen) was added, mixed by inversion and the samples were incubated for 10min at room temperature. Into the tube, 200µl Protein Precipitation Buffer (Qiagen) was added, followed by brief vortexing and a 10min-incubation on ice. To remove precipitated proteins the samples were centrifuged at 12,500rpm for 10min at 4°C. The supernatant (containing DNA) was carefully transferred into a fresh 1.5ml Eppendorf tube and 0.25µl Glycogen solution was added (20mg/ml)

(ThermoFisher). Six hundred microlitres of Isopropanol (Sigma-Aldrich) was added followed by mixing by inversion. The samples were then incubated overnight at -20°C to precipitate DNA.

The next day, the tubes were centrifuged at 13,500rpm for 45min at 4°C and the supernatant was carefully removed without losing any of the pellet. The pellet was then washed with 600µl 70% Ethanol and mixed by inversion to re-suspend it. Subsequently, the tube was again centrifuged at 13,500rpm for 30min at 4°C and the supernatant was discarded. The pellet was air-dried for 1hr to allow excess ethanol to evaporate (without letting the pellet over-dry). Lastly, the pellet was re-suspended by adding 21µl Elution buffer AE (10mM Tris-Cl, 0.5mM EDTA) (Qiagen) and incubating for 1hr at room temperature. DNA samples were quantified on a Qubit™ 3.0 Fluorometer (Invitrogen).

2.2.3. RNA isolation

RNA was isolated from 2nd stage juveniles; previously harvested 2nd stage juveniles were pelleted, frozen in liquid nitrogen and immediately transferred into an autoclaved mortar. One millilitre of TRIzol® reagent (Invitrogen) was added to the mortar and the pellet was thoroughly homogenised. Then, the ground tissue was transferred by pipetting into a 1.5ml Eppendorf tube and was incubated for 5min at room temperature to allow complete dissociation of nucleoprotein complexes bound to RNA molecules.

After the incubation, 0.2ml of chloroform was added and the tube was shaken vigorously for about 15sec, followed by an additional incubation for 3 more minutes at room temperature. Subsequently, the sample was centrifuged at 12,000rpm for 15min at 4°C to separate the lower pink organic phase, an interphase and the top colourless (aqueous) phase that contained RNA; the top phase was then transferred into a fresh 1.5ml Eppendorf tube by pipetting (without picking any of the other phases). Into this tube, 0.5ml of 100% chilled isopropanol was added and incubated for 10min at room temperature. The sample was then centrifuged at 12,000rpm for 10min at 4°C to pellet the RNA from the supernatant, which was carefully removed by pipetting afterwards. The pelleted RNA was washed with 1ml of 75% Ethanol and vortexed briefly to dislodge the pellet. Another step of centrifugation was done at 8,000rpm for 5min at 4°C, followed by removal of the supernatant without disturbing the pelleted RNA.

As a next step, the lid of the tube was left open for about 10min to dry the RNA pellet. Once it was dried, the pellet was re-suspended by adding 30µl of RNAse-free water followed by mixing by pipetting and incubation on a heating block for 15min at 55°C. To treat the isolated RNA with DNAse, the following digestion reaction was set in a 0.5ml Eppendorf tube and then was incubated at 37°C for 30min:

- 8µl re-suspended RNA,
- 1µl RQ1 RNAse-free DNAse 10x reaction buffer (Promega),
- 1µl RQ1 RNAse-free DNAse (Promega)

To terminate the reaction, 1µl of RQ1 DNAse Stop Solution (Promega) was added followed by incubation at 65°C for 10min. The DNAse-treated RNA was then stored at -20°C until further use.

2.2.4. Synthesis of first-strand cDNA

For constructing cDNA, in a sterile 0.5ml Eppendorf tube, the following reaction was set up and incubated in a heating block at 65°C for 5min followed by incubation on-ice for at least 1min:

- 0.5µl of oligo$(dT)_{20}$ (100µM) (Eurofins Genomics),
- 1µl dNTPs (10mM each),
- Xµl RNA extracted from J2s,
- Xµl (filled up to 13µl) RNAse-free water

The reaction was briefly centrifuged to draw all the contents to the bottom of the tube and then the following components (SuperScript™ III Reverse Transcriptase; Invitrogen) were mixed to the reaction:

- 4µl 5x First-strand buffer,
- 1µl 0.1M DTT,
- 1µl RNAseOUT™ recombinant RNAse Inhibitor,
- 1µl SuperScript™ III RT (200u/µl)

The reaction was then incubated at 50°C for 1hr followed by a heating at 70°C for 15min to stop the reaction. The cDNA sample was stored at -20°C until further use.

2.2.5. Cloning candidate effector genes into pCR™8/GW/TOPO™ TA

All the candidate effector genes were cloned into the vector pCR™8/GW/TOPO™ TA Invitrogen) using the following protocol.

2.2.5.1. PCR using KOD hot-start DNA polymerase

The DNA sequence was amplified with PCR by using the proof-reading KOD hot-start polymerase (Merck Millipore) from either gDNA or cDNA template. PCR components and cycling conditions are shown in Table 2-4 and Table 2-5.

Table 2-4 PCR concentrations and volumes for using KOD hot-start polymerase.

Reagent	Initial concentration	Volume per reaction	Final concentration
KOD Buffer	10x	5µl	1x
MgSO4	25mM	3µl	1.5mM
dNTPs	2mM each	5µl	0.2mM
FOR-(5')-Primer	10µM	1.5µl	300nM
REV-(3')-Primer	10µM	1.5µl	300nM
SDW	-	adjusted to 50µl	-
KOD polymerase	5u/µl	1µl	0.2u/µl
Template (DNA/cDNA)	-	Xµl	-
Total Volume		50µl	

Table 2-5 PCR programme using KOD hot-start polymerase (*calculated based on each gene-specific primer set. It is equal to the lowest annealing temperature plus 3°C. It is typically between 59°C and 62°C).

Step	Temperature	Time	
Initial denaturation	95°C	3min	
Denaturation	95°C	30sec	**x35**
Annealing	Depending on the primer*	30sec	
Extension	72°C	1min/kb	
Final extension	72°C	3min	
Hold	12°C	∞	

2.2.5.2. PCR product visualisation, purification and addition of 3' A-overhangs

PCR products were visualised by gel electrophoresis on a 1% agarose gel (1xTBE buffer) and visualised with SYBR™Safe (Invitrogen) staining under UV light. The purification of the PCR product was performed directly from the gel using the QIAquick Gel Extraction kit (Qiagen) following the manufacturer's protocol with a final elution in 30µl Elution buffer (Tris-HCl 10mM, pH 8.5).

Since the KOD hot-start polymerase creates blunt ends, A-overhangs on the 3' end of each DNA strand had to be added before the cloning by using GoTaq® polymerase (Promega). For this, the following reaction was prepared (Table 2-6) and incubated for 10 minutes at 72°C in a thermocycler.

Table 2-6 PCR concentration and volumes for the addition of 3' A-overhangs for the TA-cloning reaction.

Reagent	Initial concentration	Volume per reaction	Final concentration
GoTaq buffer	5x	2µl	1x
MgCl2	25mM	0.8µl	2mM
dATPs	10mM	0.4µl	0.4mM
SDW	-	adjusted to 10µl	-
GoTaq polymerase	5u/µl	0.1µl	0.05u/µl
PCR purified product	-	up to 6.7µl	10-30ng/µl
Total Volume		10µl	

2.2.5.3. Transformation of plasmid into E. coli *DH5α competent cells*

The candidate genes were cloned into the vector pCR™8/GW/TOPO™ (Invitrogen) (Figure 2-3).

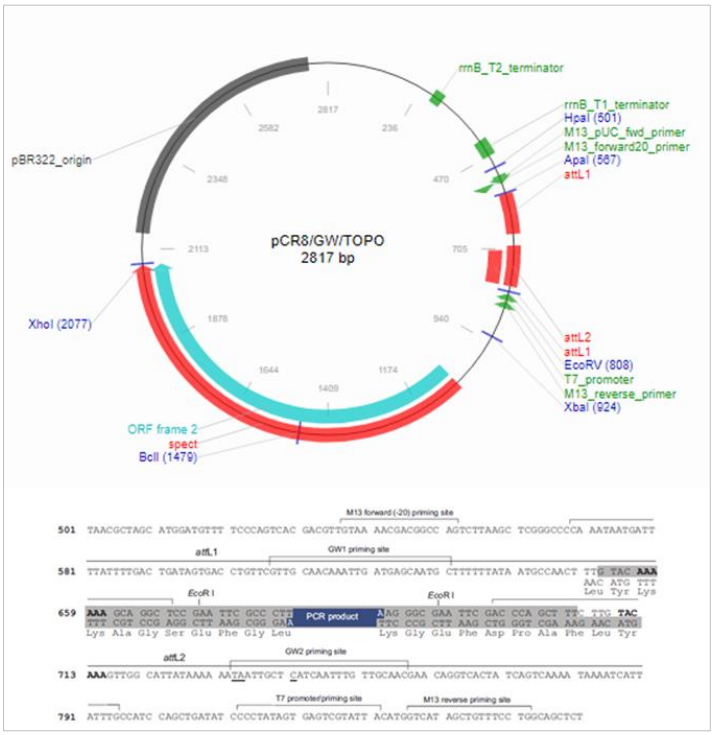


Figure 2-3 The circular and the linear map of the vector pCRTM8/GW/TOPO™ (Invitrogen) (reprinted by the Addgene online vector database).

The TOPO reaction was firstly prepared by mixing the reagents shown in Table 2-7 in a 0.5ml Eppendorf tube and was incubated overnight at room temperature.

Table 2-7 Volumes of the reagents for the preparation of the TOPO TA cloning reaction.

Reagent	Volume per reaction
PCR product (A-overhang reaction)	4µl
4x diluted Salt Solution	1µl
SDW	0.5µl
TOPO vector	0.5µl
Total Volume	6µl

Cloning was carried out by transforming competent *E. coli* cells strain DH5*α* with an aliquot of the TOPO reaction. Two types of *E. coli* DH5*α* cells were used; either electro-competent or chemically competent cells.

Bacteria transformation using electroporation

Two microlitres of the TOPO reaction was added to the *E. coli* DH5*α* cells followed by gentle mixing. The reaction (containing the TOPO reaction with the competent cells) was transferred to a MicroPulser™ electroporation cuvette (Bio-Rad) and a pulse at 1,800V (for 5.8ms) was provided in the MicroPulser™ Electroporator (Bio-Rad). Immediately after, 0.5ml of SOC medium (0.5% yeast extract, 2% tryptone, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose) was added to the electroporated cells, and the solution was transferred into a sterile 2ml Eppendorf tube. The cells were incubated at 37°C for 1hr and 30min, with the tube lying down for increased surface area on a shaking surface (approximately 250rpm).

Bacteria transformation using chemically competent cells:

Three to 4µl of the TOPO reaction were added into the chemically competent *E. coli* DH5*α* cells followed by gentle tapping. The reaction (containing the TOPO reaction with the competent cells) was incubated on ice for 30min. After this incubation time, it was transferred quickly to a 42°C water bath for 30sec and subsequently back to the ice for another 2min. Two hundred microlitres of SOC medium were added to the reaction and then incubated at 37°C for 1hr and 30min, with the tube lying down for increased surface area on a shaking surface (approximately 250rpm).

Transformed cells were subsequently plated at 37°C overnight on LB agar petri dishes containing the appropriate selection antibiotic (i.e. here, Spectinomycin at 100μg/ml). Colonies growing on the selective plates were screened for plasmid insertion by a standard colony PCR. Colonies were picked up with a sterile tip and re-suspended into 30μl of sterile water, using this as a DNA template for the following PCR reaction and programme (Table 2-8 and Table 2-9). Two reactions were performed per sample; one using the primers M13-FOR and M13-REV to ensure the presence of the full insert into the vector and a second one using the M13-FOR and the gene-specific-REV (for the insert) to ensure the correct orientation of it.

Table 2-8 Standard PCR concentrations and volumes for using Taq polymerase.

Reagent	Initial concentration	Volume per reaction	Final concentration
GoTaq green buffer	5x	4μl	1x
MgCl2	25mM	1.6μl	2mM
dNTPs	10mM each	0.8μl	0.4mM
FOR-(5')-Primer	10μM	0.6μl	300nM
REV-(3')-Primer	10μM	0.6μl	300nM
SDW	-	adjusted to 20μl	-
GoTaq polymerase	5u/μl	1μl	0.2u/μl
Template	-	2μl	-
Total Volume		20μl	

Table 2-9 Standard PCR programme for using Taq polymerase.

Step	Temperature	Time	
Initial denaturation	95°C	5min	
Denaturation	95°C	45sec	
Annealing	Depending on the primer*	30sec	**x25**
Extension	72°C	1min/kb	
Final extension	72°C	5min	
Hold	12°C	∞	

All the successful recombinants were inoculated into 5ml of liquid LB media containing the appropriate antibiotic (i.e. here, Spectinomycin at 100μg/ml) in a shaking incubator at 37°C overnight (approximately 350rpm). All the successful recombinants were also stored in 25% final glycerol stocks at -80°C.

2.2.5.4. Plasmid purification

For plasmid DNA isolation and purification, the GeneJET Plasmid Miniprep kit (Thermo Scientific) was used following the manufacturer's instructions with final elution in 50µl Elution buffer EB (10Mm Tris-HCl, pH 8.5).

2.2.6. Recombining TOPO-based ENTRY clones into the PVX vector pGR106GW

The plasmid pGR106GW is a result of a Gateway® cassette insertion into the vector pGR106 that had been designed and synthesised by Dr. Sean Chapman (The James Hutton Institute). All candidate effectors had been previously cloned into a TOPO-based vector that contains *attL1* and *attL2* sites and can be recombined with the *attR1* and *attR2* sites of the destination vector (i.e. pGR106GW) through an LR reaction and using Gateway® technology. The antibiotic selection of the backbone is Kanamycin (Figure 2-4).

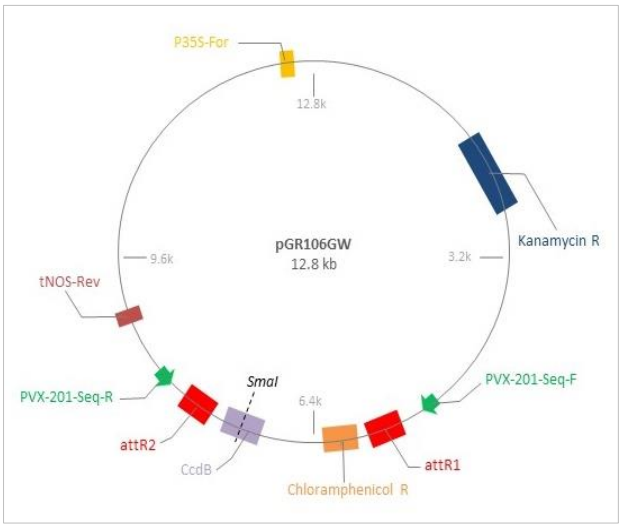


Figure 2-4 The map of the Gateway®-compatible PVX vector pGR106GW. Between the *attR* sites, there is the toxin gene *ccdB* and the gene for resistance to chloramphenicol. *SmaI*-digestion cuts the plasmid in the *ccdB* gene leaving the attachment sites intact. Just outside the *attR* sites, the specific primers PVX-201-Seq are located.

The following protocol was used for cloning the candidate effector genes from the TOPO-based vector into the PVX vector pGR106GW:

2.2.6.1. Digestion of the destination vector pGR106GW

The destination vector pGR106GW was linearized via *SmaI*-digestion for higher efficiency. The enzyme recognises the sequence CCC-GGG in the *ccdB* gene site and deactivates it. Since the vector is linearised following this process, the only circular plasmid present will be that where the recombination reaction has successfully occurred. In a 0.5ml Eppendorf tube, the following reagents were mixed and the reaction was then incubated for 3hrs at 25°C:

- Xμl pDNA pGR106GW (about 0.5μg)
- 2μl of 10x buffer J (Promega)
- 1μl *SmaI* restriction enzyme (Promega; 10u/μl)
- Xμl (up to 20μl) SDW

To dephosphorylate the vector and avoid recircularization, 1μl of phosphatase TSAP (thermosensitive alkaline phosphatase) (Promega; 0.5u/μl) was added to the reaction followed by an incubation at 37°C for 15min. To deactivate TSAP and *SmaI*, the reaction was incubated at 74°C for 15 min. The *SmaI*-digested vector was purified using the QIAquick PCR purification kit (Qiagen), following the manufacturer's instructions with a final elution in 30μl of Elution buffer EB (10mM Tris-HCl, pH 8.5). The vector was then run on a 1% agarose gel (1xTBE) to check that it was successfully digested.

2.2.6.2. LR reaction in pGR106GW and transformation into bacteria cells

LR reaction and Gateway® technology enables the recombination of the *attL* sites and the insert of the ENTRY clone with the *attR* sites of the destination vector. In a 0.5ml Eppendorf tube, the LR reaction was set with a ratio 1:1 ENTRY clone:destination vector, by mixing the following reagents and incubate the reaction overnight at room temperature:

- Xμl purified ENTRY clone (i.e. TOPO-based candidate effector pDNA) (~33ng or 18fmol)
- 6μl *SmaI*-digested pGR106GW (25ng/μl; 150ng or 18fmol)
- 1μl LR clonase II (Invitrogen)
- Xμl (filled up to 10μl) Elution buffer EB (10mM Tris-HCl, pH 8.5) (Qiagen)

The next day, 1µl Proteinase K (2 µg/µL; Invitrogen) was added and incubate at 37°C for 10min to deactivate any nucleases.

Each LR reaction was transformed into the chemically competent *E. coli* cells (strain DH5α) OneShot™ TOP10 (Invitrogen) following the same protocol described above. Transformed cells were plated at 37°C overnight on LB agar petri dishes containing Kanamycin at 50µg/ml. Colonies growing on the selective plates were screened for plasmid insertion by a standard colony PCR (as shown in Table 2-8) using the primers PVX-201-Seq-F and PVX-201-Seq-R. The PCR programme was used is shown in Table 2-9 (the initial denaturation time was adjusted to 10min). For the preparation of liquid cultures, the procedure described in the section 2.2.5.3 was followed using Kanamycin at 50µg/ml as a selective antibiotic.

2.2.6.3. Plasmid purification

The plasmid DNA isolation and purification of the PVX-cloned candidate effector genes was performed using the GeneJET Plasmid Miniprep kit (Thermo Scientific) following the manufacturer's instructions with final elution in 50µl Elution buffer EB (10Mm Tris-HCl, pH 8.5).

2.2.7. Sequencing of the isolated plasmid DNA, curation and analysis of the sequences

The TOPO-based cloned DNA samples were sequenced on capillary-based Sanger sequencing platform (Applied Biosystems) in the sequencing facilities of The James Hutton Institute. The sequences were curated for ambiguous bases in the Sequencher® DNA sequence analysis software (version 5; Gene Codes Corporation, Ann Arbor, MI USA) and a consensus sequence was created for each sample. Each consensus was then BLAST searched to verify and identify sequences resemblance in the WormBase ParaSite database (Howe et al., 2017). Once the alignment of the query sequence was verified, the consensus sequence of each pathotype for each gene was aligned each other in pairwise in BioEdit sequence alignment editor (version 7) (Hall, 1999) using the similarity matrix BLOSUM62 to identify differences in the amino acids.

2.2.8. PVX-mediated transient expression of the candidates *in planta*

Agrobacterium tumefaciens strain GV3101 was transformed with the pGR106GW constructs. Selection for *Agrobacteria* was Rifampicin (25μg/ml). The strain used contains also two helper vectors; pSOUP (selected on Tetracyclin 5μg/ml) and pMP90 (selected on Gentamycin 25μg/ml). The first helper vector is needed because pGR106GW is a pGreen-derived vector that cannot replicate on its own, while the use of the latter aims to provide more efficient expression of the genes within the T-DNA in the *Agrobacterium*-mediated plant transformation (Hellens et al., 2000; Koncz and Schell, 1986). The protocol was used is the following:

2.2.8.1. Transformation of the PVX-based constructs into Agrobacteria

To transform Agrobacteria cells with the PVX-based constructs, each purified pGR106GW-based plasmid construct was diluted down to approximately 5ng/μl. Then, 1μl of the diluted pGR106GW-based plasmid construct was added to electro-competent *Agrobacteria* cells (strain GV3101 containing the helper vectors pSOUP and pMP90) and the reaction was transferred to a MicroPulser™ electroporation cuvette (Bio-Rad) to provide a pulse at 1,800V (for 5.8ms) in the MicroPulser™ Electroporator (Bio-Rad). Immediately after, 0.5ml of SOC medium (0.5% yeast extract, 2% tryptone, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose) was added into the electroporated cells, and the solution was transferred into a sterile 2ml Eppendorf tube. The cells were incubated at 28°C for 2hr, with the tube lying down for increased surface area on a shaking surface (approximately 250rpm). The transformed cells were subsequently plated on low salt LB agar medium (10g/L tryptone, 5g/L yeast extract, 2.5g/L NaCl, 10g/L mannitol) and the appropriate selection antibiotics (i.e. Kanamycin, Rifampicin, Tetracycline, Gentamycin) for 3 days at 28°C.

After this period, a single colony from each construct was picked and inoculate 5ml of LB liquid culture (10g/L tryptone, 5g/L yeast extract, 2.5g/L NaCl, 10g/L mannitol) containing (Rifampicin (25μg/ml). The liquid cultures were then incubated overnight at 28°C on a shaking surface (approximately 250rpm). The next day, 500ml of the overnight liquid culture were transferred into cryotubes and mixed with 500ml of glycerol 50%. The tubes with the constructs were stored at -80°C for future use.

2.2.8.2. Toothpick inoculations of potato plants with Agrobacteria*-mediated PVX*

The inoculations of the plants took place in the greenhouse facilities at The James Hutton Institute 3 days after the incubation of the *Agrobacteria* cultures. *Agrobacteria* colonies were transferred from each plate with a sterile spreader. Then, a sterile toothpick was dipped in the bacteria and subsequently were transferred into the lower side of a leaflet of a 3-week old potato plant by picking gently 5-6 times (i.e. technical replicates). The same procedure was followed for 2 leaflets per leaf (i.e. technical replicates) and at least 3 leaves per plant (i.e. biological replicates). In each experiment, at least 5 plants per cultivar were inoculated in this way. The first symptoms appeared at 7dpi and observations were taken regularly. Scoring is based on visual necrosis (- for no symptoms, + if 1-2 of the inoculation sites showed necrosis, 2+ if half of the sites showed necrosis, 3+ if the majority of the sites showed necrosis).

2.2.8.3. Vacuum infiltrations of potato leaves with Agrobacteria*-mediated PVX*

For vacuum agroinfiltrations of potato leaves, 10ml of YEB liquid cultures (containing 25μg/ml Rifampicin) were inoculated with each *Agrobacterium* construct and incubated for 2 days at 28°C on a shaking surface (approximately 250rpm). On the day of infiltrations, the infiltration buffer (100ml sterile distilled water, 1ml 1M MES, 1ml 1M $MgCl_2$, 250μl 0.1M acetosyringone; per construct) was prepared for all the constructs and kept in the dark at room temperature.

Meanwhile, the liquid *Agrobacterium* cultures were centrifuged at 3,500rpm for 5min to discard the YEB medium. The *Agrobacterium* cells (pellet) were washed twice with 5ml of infiltration buffer (3 centrifuge cycles in total). The cells were then resuspended in infiltration buffer. The first optical density (OD) measurement of 1ml from each construct was taken at A = 600nm in a spectrometer. The desired OD_{600nm} should be approximately at 0.3 to 0.4. If needed, infiltration buffer was added to the construct (inoculum) to dilute down to OD_{600nm} 0.3 to 0.4. Tubes were rocked at room temperature for 2-3hrs. During this time the tubes were kept in the dark by covering them with aluminium foil. After 2-3hr, the OD_{600nm} was measured again. Each construct was transferred into a clean labelled beaker.

Potato leaves were dipped into the construct solution until completely submerged (the top two leaflets were left outside the solution as a negative control) (Figure 2-5). The beaker with the leaf was placed into a vacuum manifold and it was turned on for 30sec (bubbles appeared on the leaf surface at this stage). After 30sec, vacuum was slowly released and was turned on again for an additional 30sec (Figure 2-5). The vacuum was then released; the leaf was removed from the solution and was carefully dried on a paper to remove any excess solution (inoculum). At least 4 leaves were inoculated per construct (i.e. biological replicates). The leaves were then transferred to transparent, labelled plastic boxes with their bottom covered with damped folded tissue paper to keep high humidity inside the box. Lastly, the boxes were shut tightly and wrapped in cling film to avoid any water evaporation. The boxes were left in sunlight at room temperature. The first symptoms started appearing at 4dpi and regular observations were taken for another week.

Figure 2-5 Vacuum infiltrations set up. *Left*: A potato leaf is dipped in the inoculum. All the leaflets are placed inside the inoculum apart the two bottom side leaflets that are kept outside as a 'non-treated' control. *Right*: The beaker with the inoculum and the leaf is placed inside the vacuum manifold.

The ratio of the 'infected leaf area/total leaf area' was calculated on ImageJ (Schneider et al., 2012). The mean number of the ratio from each replicate was calculated for each biological replicate. Statistical analysis was done on SigmaPlot Version 13.0 (Systat Software Inc., San Jose California USA, www.systatsoftware.com). The leaves were also checked for HR or non-HR-specific symptoms.

2.2.8.4. Constructs used in Agrobacterium*-mediated transient expression* in planta

For either the toothpick and vacuum infiltrations of potato leaves, the same pGR106GW-based bacterial constructs were used. The variety Désirée has no resistance against PCN (and specifically does not carry the *H1* gene) and therefore was used as a susceptible potato control, whereas Maris Piper is a variety widely used because of its high resistance to *G. rostochiensis* (since it holds the *H1* resistance *locus*). Different control constructs were used along with the candidate *Avr* genes Ro1_*g13394* and Ro5_*g13394*. In the susceptible potato variety, the widely used in screening tests *P. infestans*-origin effector CRN2 (crinkling and necrosis-inducing protein 2) was used as a (non-recognised) negative control and the enhanced GFP (eGFP) protein was used as a negative control. In the case of the resistant Maris Piper, various controls were used during each experimental tests and series. In addition to the above CRN2 and eGFP, in some technical repeats, the *P. infestans*-origin RXLR effector Avr1 (that causes HR-specific necrosis in Maris Piper) and Avr3a^{EM} were used as positive and negative controls respectively.

2.2.9. Library preparation, target enrichment and sequencing

2.2.9.1. Probe design

Biotinylated RNA-based oligos, 120-nucleotides in length, were designed to fully cover the 700 target genes representing all known potential effectors in *G. pallida*. This list was composed of previously characterised effector genes from *G. pallida* (Thorpe et al., 2014), effectors that are downstream of the gland promoter DOG box motif (Eves-van den Akker et al., 2016) and all the characterised effectors from other closely related nematode species. As a next step, any duplicates were removed as well as sequences with high numbers of ambiguous bases. All the remaining sequences were then cut into 120-mers with a 60-base step size (i.e. 2x tilling density). By choosing 120 base-long baits, they were long enough to be uniquely mapped to the reference as well as short enough to achieve optimum hybridisation. Designed baits (Supplementary Data 2) were subjected to two filtering steps; the first one to remove any duplicates and the second one to remove baits with high numbers of hits to the genome after a BLAST search. Eventually, 24,744 baits (probes) were designed and synthesised by MYcroarray MYbaits (Figure 2-6).

Figure 2-6 An overview of the steps followed for designing the baits to capture the target-genes.

2.2.9.2. DNA fragmentation and purification of the sheared DNA

Prior to the fragmentation of the DNA samples, at least 5ng of the extracted DNA was adjusted to a volume of 50μl by adding nuclease-free water to the eluate in AFA microTUBE tubes (Covaris). The following conditions were used for all 10 samples to shear DNA into approximately 500bp-long fragments, using the M220 Ultrasonicator (Covaris):

Peak incident power	50W
Duty factor	20%
Cycles/burst	200
Treatment time	60sec
Temperature	20°C

Purification of the sheared DNA samples was performed with the bead-based method using 50μl AMPure® XP beads (Beckman Coulter) in order to remove DNA molecules longer than the desired size. The manufacturer's protocol was followed with a final elution in 58μl of Elution buffer AE (Qiagen) (10mM Tris-Cl, 0.5mM EDTA).

2.2.9.3. Library preparation and ligation of adaptors

DNA libraries for each nematode population were generated before proceeding with target enrichment. End repair was performed and adapters were added to the DNA fragments by using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New England BioLabs) following the next steps:

For each DNA library, the following components were mixed in a sterile 0.5ml Eppendorf tube:

Component	Volume (µl)
End Prep Enzyme Mix	3
End Repair Reaction Buffer (10x)	6.5
Sheared gDNA	55.5
Total volume	65

The components were mixed well and incubated in a thermocycler at 20°C for 30min followed by 65°C for 30min (with the heated lid to minimise evaporation). Immediately afterwards, the following components were added to each End Repair reaction and mixed well in order to ligate the Illumina® adaptors:

Component	Volume (µl)
End-repaired reaction	65
Blunt/TA Ligase Master Mix	15
NEBNext Adaptor for Illumina	2.5
Ligation Enhancer	1
Total volume	83.5

All the components were mixed well, followed by a brief centrifugation at low speed to draw any liquid from the sides of the tube to the bottom. The reaction was incubated at 20°C for 15min in a thermocycler. Then, 3µl USER™ enzyme (New England BioLabs) was added to the ligation mixture followed by incubation at 37°C for 15min in a thermocycler. The recovered adaptor-ligated DNA was then purified with 1x (86.5µl) AMPure® XP beads (Beckman Coulter) following the manufacturer's instructions with a final elution in 23µl Elution buffer AE (Qiagen) (10mM Tris-Cl, 0.5mM EDTA).

The NEBNext® Multiplex Oligos for Illumina® kit (Index Primers Set 1) (New England BioLabs) was used to amplify the adaptor-ligated DNA and barcode the libraries (i.e. indexing). For each adaptor-ligated DNA sample, the following reagents were mixed in a sterile PCR tube and incubated in a thermocycler using the programme shown below:

Component	Volume (µl)
Adaptor-ligated DNA fragments	20
NEBNext Ultra Q5 Master Mix	25
10µM Index Primer**	2.5
10µM Universal Primer	2.5
Total volume	50

Step	Temperature	Time	
Initial denaturation	98ºC	30sec	
Denaturation	98ºC	10sec	**x8**
Annealing/extension	65ºC	60sec	
Final extension	65ºC	5min	
Hold	4ºC	∞	

To index the libraries, specific primers [NEBNext® Multiplex Oligos for Illumina® kit - Index Primers Set 1 (New England BioLabs)] were used before pooling them together. The universal Illumina® primer (5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC*T-3′) was used along with a specific and unique index primer for each individual library. The difference between the 10 different indexes used was the unique 6-nucleotide combination in the middle of the primer sequence (Table 2-10).

Table 2-10 The 10 Index primers used for producing barcoded libraries and the populations used [modified from the Instruction Manual of NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) (New England Biolabs)].

Library code	Population	Illumina Index Primer Sequence	Expected Index primer sequence read
Lib1	Newton x Désirée	5′-CAAGCAGAAGACGGCATACGAGAT-CGTGATGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	ATCACG
Lib2	Newton x Désirée	5′-CAAGCAGAAGACGGCATACGAGAT-ACATCGGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	CGATGT
Lib3	n11305 x Sv_8906	5′-CAAGCAGAAGACGGCATACGAGA-TGCCTAAGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	TTAGGC
Lib4	n11305 x Sv_8906	5′-CAAGCAGAAGACGGCATACGAGAT-TGGTCAGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	TGACCA
Lib5	n11305 x Sv_11305	5′-CAAGCAGAAGACGGCATACGAGAT-CACTGTGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	ACAGTG
Lib6	n11305 x Sv_11305	5′-CAAGCAGAAGACGGCATACGAGA-TATTGGCGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	GCCAAT
Lib7	n11415 x Sa_11415	5′-CAAGCAGAAGACGGCATACGAGA-TGATCTGGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	CAGATC
Lib8	n11415 x Sa_11415	5′-CAAGCAGAAGACGGCATACGAGA-TTCAAGTGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	ACTTGA
Lib9	n11415 x Sa_12674	5′-CAAGCAGAAGACGGCATACGAGA-TCTGATCGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	GATCAG
Lib10	n11415 x Sa_12674	5′-CAAGCAGAAGACGGCATACGAGAT-AAGCTAGTGACTGGAGTTCAGACGT-GTGCTCTTCCGATC-s-T-3′	TAGCTT

After the amplification of the indexed libraries, the reaction was purified with 1x (50µl) AMPure® XP beads (Beckman Coulter) following the manufacturer's instructions with a final elution in 35µl Elution buffer AE (Qiagen) (10mM Tris-Cl, 0.5mM EDTA).

2.2.9.4. Verification of the fragmentation and quantification of the purified barcoded DNA libraries

After the fragmentation of each DNA sample, fragments longer than 800bp were removed with the purification step of the sheared DNA samples. Small fragments (<200bp) were removed during the amplification of the adapter-ligated DNA libraries. To verify the success of the fragmentation, 1ng of each library was run on 2100

Bioanalyzer using the Agilent DNA 1000kit (Agilent Technologies). Each DNA library was quantified on a Qubit™ 3.0 Fluorometer (Invitrogen).

2.2.9.5. Target enrichment using the MYcroarray® MYbaits® kit

Before the enrichment, equimolar amounts of all the 10 DNA libraries were pooled together in a 1.5ml Eppendorf tube. Enrichment required a starting hybridisation reaction of 500ng of the pooled DNA library diluted in a maximum volume of 7μl. Therefore, the pooled library was lyophilised in a vacuum concentrator for 25min (low heat; <45°C) and to reconstitute it, 7μl of RNAse-free water (Qiagen) was added.

For the hybridisation of the baits to the pooled library, the instructions provided in the MYbaits kit (user manual v3) were followed, with a hybridisation temperature set at 65°C for 22hr. To bind the bait-targeted hybrids, Dynabeads™ MyOne™ Streptavidin C1 beads (Invitrogen) were used following the instructions provided in the MYbaits kit (user manual v3) with a final elution of the captured enriched on-bead library in 30μl Elution buffer AE (Qiagen) (10mM Tris-Cl, 0.5mM EDTA). To amplify the captured library, the following PCR was performed using Herculase II Fusion DNA polymerase (Agilent Technologies):

Reagent	Volume (μl)
(10x) Herculase II buffer	10
dNTPs (25mM each)	0.5
Forward library Primer #1 (10μM)	1.25
Reverse library Primer #2 (10μM)	1.25
Herculase II polymerase (5u/μl)	1
On-bead enriched library	14
RNAse-free water	22
Total volume	50

Step	Temperature	Time	
Activation	98°C	2min	
Denaturation	98°C	30sec	
Annealing	65°C	30sec	**x10**
Extension	72°C	30sec	
Final extension	72°C	10min	
Hold	8°C	∞	

Lastly, the amplified captured library was purified with 1x (50µl) AMPure® XP beads (Beckman Coulter) following the manufacturer's instructions with a final elution in 35µl Elution buffer AE (Qiagen) (10mM Tris-Cl, 0.5mM EDTA). A final quantification was done on Qubit™ 3.0 Fluorometer (Invitrogen).

2.2.10. Real-time PCR (qPCR) of the enriched library

The enriched captured libraries pool was tested using qPCR. Primers were used that amplify known effectors (*SPRY-414-2*, *SPRY-1719-1*, *G16H02*) and the non-effector *GAPDH* as an endogenous control (Table 2-3). As a control, the pre-enrichment libraries were used. All the primers used are shown in Table 2-3.

Two DNA templates were used; a pre-enrichment and a post-enrichment pooled equimolar library. In a 1.5ml Eppendorf tube, the master mix for the qPCR was prepared as follows in volumes enough for all the reactions (3 replicates for each primer set):

Reagent	Volume (µl) per reaction
SYBR green	12.5
FOR-primer (5µM)	2
REV-primer (5µM)	2
DNA template	1
SDW	7.5
Total volume	25

The reactions were loaded into a 96-well plate and the following programme was run on a StepOne real-time PCR system (ThermoFisher Scientific) (set in quantitation-comparative $C_T\Delta\Delta C_T$ programme):

Step	Temperature	Time	
Hold 1	95ºC	10min	
Cycling 1	95ºC	15sec	**x40**
Cycling 2	60ºC	1min	
Melt curve 1	95ºC	15sec	
Melt curve 2	60ºC	1min	
Melt curve 3	95ºC	15sec	

All the results of the qPCR were analysed on the StepOne Software (version 2.3) (ThermoFisher). The relative quantification (RQ) of each gene was considered. The RQ is a fold change of each gene compared and normalised to the reference, i.e. endogenous control.

2.2.11. High-molecular weight (HMW) DNA preparation from *G. pallida* Newton J2s

To create HMW DNA for long-read sequencing, a 1.5ml Eppendorf tube containing a pellet (approximately 0.1ml-0.2ml) of freshly-hatched J2s (tube#1) was frozen in liquid nitrogen. When the pellet started thawing, 260µl Extraction buffer (0.1M Tris, pH 8, 0.5M NaCl, 50mM EDTA, 1% SDS) and 40µl Proteinase K (20mg/ml) (Qiagen) were added. The pellet was crushed by twisting an autoclaved micro-pestle (for about 30sec). The solution (tube#1) was incubated at 55°C for 24hrs in a heating block and occasionally mixed by gentle inversion.

The next day, 10µl RNAse A (100mg/ml) (Qiagen) were added to the tube#1, followed by a 5-minute incubation at room temperature. An equal volume of Phenol:Chloroform:Isoamyl alcohol (25:24:1) solution (Sigma-Aldrich) was added and mixed by 1-minute vortexing at mid speed. The tube (tube#1) was centrifuged at 13,000rpm for 5min at 4°C. The top aqueous solution was then transferred into a fresh 1.5ml Eppendorf tube (tube#2) (avoiding any of the phenol solution). An equal volume of Elution buffer AE (5mM Tris-HCl, pH 8.5) was added to tube#1 in order to re-extract more DNA, followed by 1-minute vortexing. The tube (tube#1) was centrifuged again at 13,000rpm for 5min at 4°C. The top aqueous solution was then transferred into tube#2 (avoiding any of the phenol solution). Subsequently, an equal volume of Chloroform:Isoamyl alcohol (24:1) solution (Sigma-Aldrich) was added to tube#2, where the phenol-free aqueous phases were pooled together, and mixed by vortexing for 1min. The tube (tube#2) was centrifuged at 13,000rpm for 5min at 4°C. As much of the top aqueous phase as possible was transferred into a fresh 1.5ml Eppendorf tube (tube#3) (without picking up any of the chloroform phase).

Then, 1µl Glycogen (20mg/ml) (ThermoFisher) and NH_4OAc (final concentration of 0.75M) were added to the tube with the chloroform-free phase (tube#3). The solution

was mixed very well by pipetting. Absolute Ethanol was added to tube#3 equal to 2.5 times of the volume that the tube contained, followed by mixing by inversion. The tube was then centrifuged at 13,000rpm for 20min at 4°C. The supernatant was poured and decanted without disturbing the pellet (which contained the DNA). The pellet was washed by adding 300µl 80% Ethanol, followed by a brief vortexing for 3 times. The tube was centrifuged again at 13,000rpm for 15min at 4°C. The supernatant was decanted without disturbing the pellet. The washing step with 80% Ethanol was repeated once more (for in total 2 washes). To draw residual ethanol to the bottom, the tube (tube#3) was briefly centrifuged and subsequently any excess was removed by pipetting without disturbing the pellet of DNA, which was then air-dried for 2min without over-drying it. Lastly, the DNA was re-suspended in 55µl of Elution buffer EB (Tris-HCl pH 8) and incubated overnight at room temperature to allow complete dissolution of the precipitated DNA.

Two different HMW DNA samples were prepared with this method in parallel. At the end, both samples were pooled together to obtain the required amount of DNA. The concentration of each sample was measured on a Qubit™ 3.0 Fluorometer (Invitrogen) using the QubitTM dsDNA HS Assay kit (Invitrogen) and following the manufacturer's instructions. The 260/280 (values between 1.8 and 2.0) and 260/230 (values between 2.0 and 2.2) ratios were checked on a NanoDrop Spectrophotometer (NanoDrop Technologies). To check the integrity of the extracted molecules, 100ng of each sample were run in parallel lanes on 0.7% agarose gel at low voltage (25V) for 5hrs.

2.2.12. Whole-genome re-sequencing of the highly virulent, selected *G. pallida* populations

All the 4 DNA samples (n11305 x Sv_8906, n11305 x Sv_11305, n11415 x Sa_11415 and n11415 x Sa_12674) were extracted as described in section 2.2.2. The libraries were prepared and sequenced on Illumina HiSeq2500 at the McGill University and Genome Quebec Innovation Centre (McGill University, Montreal, Quebec, Canada) using one lane of 125-base paired reads. A summary of the populations used for this analysis is given next:

Library	Population	Selection source	Selected population/name used in Chapter 5
1	n11305 x Sv_8906	*S. vernei*	vernei_8906
2	n11305 x Sv_11305	*S. vernei*	vernei_11305
3	n11415 x Sa_11415	*H3*	H3_11415
4	n11415 x Sa_12674	*H3*	H13_12674

2.3. Generation of the new *G. pallida* genome assembly

Assembly of the PacBio raw reads (generated from all four cells combined) was done using Canu version 1.6 (Koren et al., 2017). Correction of the raw reads was kept at 15% to improve the accuracy of the read bases and low-quality sequences and SMRT adapters were removed before the assembly. The generated consensus sequences (in contig-level) were re-assembled again using wtdbg2 (https://github.com/ruanjue/wtdbg2) to correct errors and create a new consensus sequence. The contigs created during the assembly phase were 'finished' to merge them together and the 'haplotigs' (i.e. contigs with the same haplotype) were purged together to create a curated de-duplicated representation of the genome. The purging was accomplished using 'Purge Haplotigs' (Roach et al., 2018) using a more stringent threshold of nucleotide diversity (i.e. 80%).

The 'finished' and purged contigs were then re-assembled together into scaffolds using SSPACE (Boetzer et al., 2010); minimum numbers of contigs that were re-assembled into a scaffold was set at 10 with minimum overlap of 500bp. Subsequently, all the gapped regions created during scaffolding were removed using the tool gapFinisher (Kammonen et al., 2017). Three rounds of the above step (i.e. re-assembling into scaffolds and removal of gapped regions) were performed. The output assembly was polished (i.e. correcting errors, removal and filling of gaps) using Arrow (https://github.com/PacificBiosciences/GenomicConsensus) in two steps.

The correction of the polished assembly was done with Pilon (Walker et al., 2014) in 5 rounds, using both PacBio and available Illumina reads, and the final assembly was generated. Completeness of the assembly was checked three times with BUSCO (Simão et al., 2015); after the first assembly step, after the purging of the 'haplotigs' and in the generation of the final assembly. At the end, contaminant sequences (i.e. sequences of

non-nematode origin) were removed from all the scaffolds using BlobTools (Laetsch and Blaxter, 2017).

Genes were called in the final version of the new *G. pallida* assembly using Braker (Hoff et al., 2016). The RNA-seq data (trimming quality at Q30) available from the previous old assembly (Lindley population) (Cotton et al., 2014) were mapped against the new assembly using STAR (Dobin et al., 2013) to create a guide for the gene boundaries. Braker uses the GeneMark-ET algorithm for initial gene training along with the created boundaries. When training was completed, Augustus (Stanke and Morgenstern, 2005; Stanke et al., 2006) was run along with the boundaries for the final calling. Pairwise differences between each lifecycle stage were analysed with edgeR (Robinson et al., 2010).

2.4. Variant calling analysis of PenSeq data

2.4.1. Curation and quality control of the raw NGS data

As a first step, quality control (QC) was done on all the raw reads using the tool FastQC v0.11.7 (Andrews, 2010) to calculate their quality (script used fastqc.sh). Then, the reads were processed with the tool Trimmomatic v0.38 (Bolger et al., 2014) to trim the Illumina adapters and remove low-quality reads (with Phred quality score under 30) and to trim to a minimum length of 36bp (script used Gpal_trimming.sh). FastQC was again run on the newly-generated trimmed reads in order to check the success of the trimming parameters (script used fastqc_trimmed.sh).

2.4.2. Alignment of the reads against the reference

The tool bowtie2 v2.3.4.2 (Langmead and Salzberg, 2012) was used (base quality Phred score at 33, minimum acceptable alignment score of "L, -0.12, -0.12" and maximum fragment size of 1000) to align all the trimmed reads against the *G. pallida* reference genomes (below) (script used mapping.sh). Three different *G. pallida* assemblies were used as references, and thus, 3 different alignment jobs were run in parallel. Along with the old reference assembly (Gpal.v1.0) (Cotton et al., 2014), two draft versions (unpublished data, by Peter Thorpe, November 2018) of the newly-generated assembly

from the Newton population were used. Secondary alignments were excluded in the bowtie2 code. In addition, the tool SAMtools v1.9 (Li et al., 2009) was also used to remove secondary and non-mapped reads.

2.4.3. Variant calling and filtering

To run the variant caller, the samples were merged into a single (binary) BAM format file. To be able to identify each sample in the merged file, all the reads had been previously tagged with unique read groups (RG). The merging was performed using the function '-merge' in the SAMtools toolkit (script used merge_bam.sh). Variant calling was done using the tool FreeBayes v1.2.0 (Garrison and Marth, 2012) (script used freebayes.sh) with a minimum mapping quality set to 20, minimum alternate count to 5 and minimum alternate fraction to 0.2. Variant calling was run for the mapped reads in all the 3 reference assemblies in parallel.

Further filtering of the FreeBayes-generated variants for minimum read coverage of 10, no missing data and variant type (e.g. differences between the three biological groups) was done using the PyVCF script (script used filter_with_pycvf.py) (https://github.com/jamescasbon/PyVCF). To annotate the variants and predict their effect on genes, the tool SnpEff v4.3 (Cingolani et al., 2012) was used (Table 2-11). On the SnpEff-annotated variants, the tool bedtools v2.27 (Quinlan and Hall, 2010) (function getfasta) was run for all the 3 annotated references, as well as the custom script SearchAugustus.jar (by Etiene Lord; Agriculture & Agri-Food Canada), to generate and extract all the protein sequences containing called variants. These sequences were then functionally annotated against genomic datasets (i.e. global BLAST) using Blast2GO (Gotz et al., 2008). For predicting the presence of a signal peptide in the extracted amino acid sequence, the SignalP v4.1 server (Gunther and Coop, 2013) was used with default settings.

All the used scripts can be found in the Supplementary Data 3.

Table 2-11 Suggested putative impacts that were used by SnpEff to annotate the identified variants. Each suggested putative impact results from specific Sequence Ontology (SO) terms (re-printed and modified by Cingolani et al. (2012)

Putative impact	SO term
HIGH	Frameshift variant
HIGH	exon loss variant
HIGH	Splice acceptor/donor variant
HIGH	Start/stop codon loss
HIGH	Stop codon gained
MODERATE	Missense variant
MODERATE	Splice region variant

2.5. Variant calling analysis of the re-sequenced populations

2.5.1. Curation and quality control of the raw NGS data

All the raw reads were checked for QC and then trimmed to remove Illumina adapters and any low-quality bases using the same procedure as described in the section 2.4.1; minimum read length for this analysis was set to 50bp.

2.5.2. Alignment of the reads against the reference

The trimmed reads were mapped against the two drafts of the new reference assembly (i.e. new/bigger and new/smaller). The tool BWA v0.7.12 (Li and Durbin, 2009) was used with default settings and all non-mapped reads were discarded. The samples were tagged with unique read groups as described in section 2.4.2.

2.5.3. Variant calling and filtering

Two populations selected on *S. vernei* and two populations selected on *H3* were chosen. Variant calling was performed on those four populations in parallel, using FreeBayes as described in the section 2.4.3; minimum mapping quality was set at 30, minimum base quality at 20, minimum alternate count at 2, minimum alternate fraction at 0.05 and minimum coverage at 5. The variants generated by FreeBayes were filtered for minimum read depth of 10 and no missing data between the samples. As a last step for the basic filtering, they were functionally annotated with SnpEff and then for variant type (i.e. all-type of variants, SNPs only). After this step, two different filtering procedures were followed:

2.5.3.1. Filtering of variants depending on the selection source

Here, the four populations were categorised depending on their selection source, i.e. into *S. vernei*-selected and *H3*-selected. The basic-filtered SnpEff-annotated variants were further being filtered using R, based on the change in the allele frequency (here abbreviated as CAF) between two biological groups (script used 'compare_diff_entre_pop', by Etienne Lord, Agriculture & Agri-Food Canada). Two CAF thresholds were used to discriminate these groups; the 'maximum CAF1' for comparisons within the same biological group and the 'minimum CAF2' for comparisons between different biological groups. Only variants exhibiting a specific CAF value were kept. At this step, maxCAF1 was set to 0.1 (i.e. the allele frequencies of the populations within the same biological group do not differ significantly to each other) and minCAF2 was set to either 0.9 (more stringent) or 0.7 (less stringent) (i.e. the allele frequencies of the populations from different biological group differ significantly).

At the same time, two types of variants were searched; SNPs only and all-type of variants (including SNPs, indels and mixed variants). All the protein sequences containing called variants were extracted using a custom java-based script (script used 'SearchAugustus') (by Etienne Lord, Agriculture & Agri-Food Canada). The extracted sequences were then BLASTp searched using Blast2GO (Staiger and Brown, 2013) for functional annotation.

2.5.3.2. Search for variants selected independently within the same resistance source

To identify variants (all-type variants) that were selected differently among the populations selected on the same resistance source, the same process was followed as before (section 2.5.3.1). This time, the basic-filtered SnpEff-annotated variants, were filtered depending on CAF values with one biological group consisting of one population only and the second biological group consisting of the remaining three populations. Four different combinations were performed in this analysis series. Retrieval of the variant amino acid sequences was done as before (section 2.5.3.1).

All the used scripts can be found in the Supplementary Data 4.

3. Screening of *G. pallida* selected populations

3.1. Background

3.1.1. Resistance to *Globodera pallida*

A great deal of effort has gone into developing potato varieties resistant to potato cyst nematodes (PCN) since natural resistance is the most durable, cost effective and environmentally friendly way of controlling plant parasitic nematodes. Initially, *Globodera pallida* was considered as a pathotype of *G. rostochiensis*, and only in the 1970's it was characterised as a new species (Stone, 1972). It is believed that *Globodera rostochiensis* was the predominant PCN species present in the United Kingdom and other European countries till the 1960's when the first commercial variety resistant to any of the PCN species became available. The variety 'Maris Piper' was bred, which incorporates the major resistance gene *H1* (named after the previous name of the species - *Heterodera rostochiensis*) that gives full resistance to the *G. rostochiensis* Ro1 and Ro4 pathotypes. Since then, *H1*-containing potato varieties have been used effectively in the UK to suppress and control the spread of *G. rostochiensis* (Castelli et al., 2003; Finkers-Tomczak et al., 2011).

However, the extensive use of resistant varieties containing *H1* has resulted in selection in favour of *G. pallida,* which is not controlled by this resistance source. The situation in the field has consequently reversed and *G. pallida* is now the more abundant PCN species (Minnis et al., 2002). In the last decades, more genes conferring resistance to PCN species have been identified and mapped, such as the resistance gene *Hero*, which gives full resistance to most *G. rostochiensis* pathotypes and partial resistance to *G. pallida*, the *locus Gpa2* which is specific for a *G. pallida* population found in the Netherlands and the *locus Gro1* against the *G. rostochiensis* Ro1 pathotype (Ernst et al., 2002; Paal et al., 2004; van der Voort et al., 1997).

In contrast to *G. rostochiensis*, there is no major gene conferring resistance to the predominant types of *G. pallida* found in Europe. The identification of potential resistance sources to *G. pallida* is complicated by the fact that the *G. pallida* presence in Europe is derived from a more heterogeneous genetic pool when compared to *G.*

rostochiensis, resulting in the presence of multiple different pathotypes. Resistance genes have been found to be clustered within the genome of potato (Caromel et al., 2005). Data confirm the presence of several linked major quantitative trait *loci* (QTL) on chromosomes V, IX and XI forming resistance ''hotspots'' in the offspring generated between the wild potato *S. vernei* and other species (Bryan et al., 2002; Bryan et al., 2004). All these linked QTL may function synergistically. Caromel et al. (2005) described a substantial reduction of female development in potato clones containing two QTL derived from *S. sparsipilum* $GpaV^{S}_{spl}$ (from chromosome V) and $GpaXI^{S}_{sp}$ (from chromosome XI); the effect was indeed additive in the presence of both QTL (R^2 = 89%), contrary to the individual effects (R^2 = 76.6% and R^2 = 12.7% respectively) with a stronger necrotic reaction close to the infection site. In addition, pyramiding of two QTL linked with partial resistance to *G. pallida* derived from *S. tuberosum* ssp. *andigena* ($GpaIV^{S}_{adg}$) and one from *S. vernei* ($GpaV_{vrn}$) resulted in higher relative resistance levels, compared to the effects of the single QTL, with multiple Pa2/3 pathotype populations (Rigney et al., 2017).

A range of *S. vernei* accessions have been used to introgress resistance into potato varieties or breeding clones, and this will affect their resistance level against different *G. pallida* populations. *S. vernei* has three main sources; from the breeding clone LGU8 (originated from German breeding material), the European clones V24/20 (from Scotland) and VRN1-3 (from the Netherlands) (Rigney et al., 2017; van Eck et al., 2017). Consequently, different populations of the same *G. pallida* pathotype may display different virulence ranges on resistance from *S. vernei* depending on the background of each source.

3.1.2. Selection for increased virulence of *G. pallida* populations

It has been shown that continuous rearing of *G. pallida* populations on partially resistant host genotypes can increase their reproductive ability (Beniers et al., 1995; Fournet et al., 2016; Phillips and Blok, 2008; Turner and Fleming, 2002; Turner et al., 1983). Many of the virulent *G. pallida* populations are able to bypass or evade recognition by products produced by host resistance genes (Turner and Fleming, 2002). This is another challenge the potato breeding industry needs to address towards the development of fully

resistant and durable varieties. This is reflected by the observation done by Phillips and Blok (2008) that sub-populations selected on a specific resistance source showed a higher reproduction rate on other potato genotypes containing the same resistance source. For example, several *G. pallida* field populations (Farcet, Halton, Bedale, Newton) were grown for successive generations on partially resistant lines containing resistance from *S. vernei* or *S. tuberosum* ssp. *andigena* CPC2802 (*H3*) which are the most widely used in the potato breeding industry in the United Kingdom. The *S. vernei*-containing genotypes tested were Morag, Santé and 62-33-3 while the breeding lines 11415 and 12674 were used for the *S. tuberosum* ssp. *andigena* CPC2802 source. After 12 generations of successive rearing, sub-populations that had been selected on potato clones containing resistance from *S. vernei* showed significantly higher numbers of developed cysts on *S. vernei*-containing clones only, but generally not on the *H3* clones (Phillips and Blok, 2008). This was the case for every selected population tested in this study, except for two Farcet and Bedale sub-populations selected on *H3*, which both showed increased reproduction ability on some *S. vernei*-containing potato clones.

3.2. Chapter objective

In this chapter, two *G. pallida* populations (i.e. Newton and Farcet) and their sub-populations selected for enhanced virulence on either *Gpa5* from *S. vernei* or *S. tuberosum* ssp. *andigena* CPC2802 (*H3)* were challenged on potato genotypes containing those resistance sources. We aimed to confirm differences in reproductive ability and virulence of those selected populations depending on their selection background.

3.3. Results

3.3.1. Hatching assays for cysts response to root diffusates

A series of hatching tests were performed to test how cysts responded to TRD addition prior their use to virulence tests. Different Newton and Farcet field populations' batches from the James Hutton Institute PCN collection were tested along with the selected populations. In the first series, Newton and Farcet field (standard) populations (batch harvested in 2010) were tested along with the derived sub-populations selected on the two resistance sources (batch harvested in 2013).

All the 4 Newton selected populations hatched 3.2 to 4.5 times more after addition of TRD in the medium compared to the control treatment of water. However, no hatching was observed with the Newton field population with any treatment. In addition, contamination by fungi was found in most of the cysts of this Newton field population. On the other hand, the hatching rate of the selected Farcet populations varied from 1.7 to 6 times higher in the TRD-treated cysts compared to the untreated cysts. The Farcet field population cysts showed a 1.6-fold increase in hatching response to TRD compared to the water-treated cysts. In the first virulence test (November till December 2015) only Farcet populations were challenged (results can be found in Supplementary Data 5).

Once a new batch of the Newton field population became available (harvested in 2015), a new hatching test was performed to ensure cyst vitality. The TRD-treated cysts hatched 2 times more than water-treated cysts. The Newton populations were included in the next virulence tests (May till September of 2016).

3.3.2. The reproductive ability of nematodes is dependent on the genetic background of the potato line tested on

Newton and Farcet field and selected populations were tested against 9 potato genotypes in order to assess their reproductive ability in the presence of two different resistance sources (*S. vernei* and *H3*) and to collect nematode material for further analyses (see Chapter 4 and 5). During the screening of Newton sub-populations (Figure 3-1), all sub-populations showed a very high reproductive ability on the susceptible Désirée varying from 84 to 150 developed females 7wpi (weeks post infections). Royal showed

average levels of susceptibility to most of the selected populations whereas Arsenal and Innovator displayed very low susceptibility (i.e. less than 8 females at 7wpi). The sub-populations n-8906 and n-11305, both selected on *S. vernei*, showed substantially increased reproduction rates (up to 4.6 times) on potato lines containing the same resistance source (i.e. Sv_8906 and Sv_11305) when compared to the unselected population on those lines. In contrast, the sub-populations selected on *H3*, i.e. n-11415 and n-12674, displayed a very low multiplication rate on the lines Sv_8906 and Sv_11305.

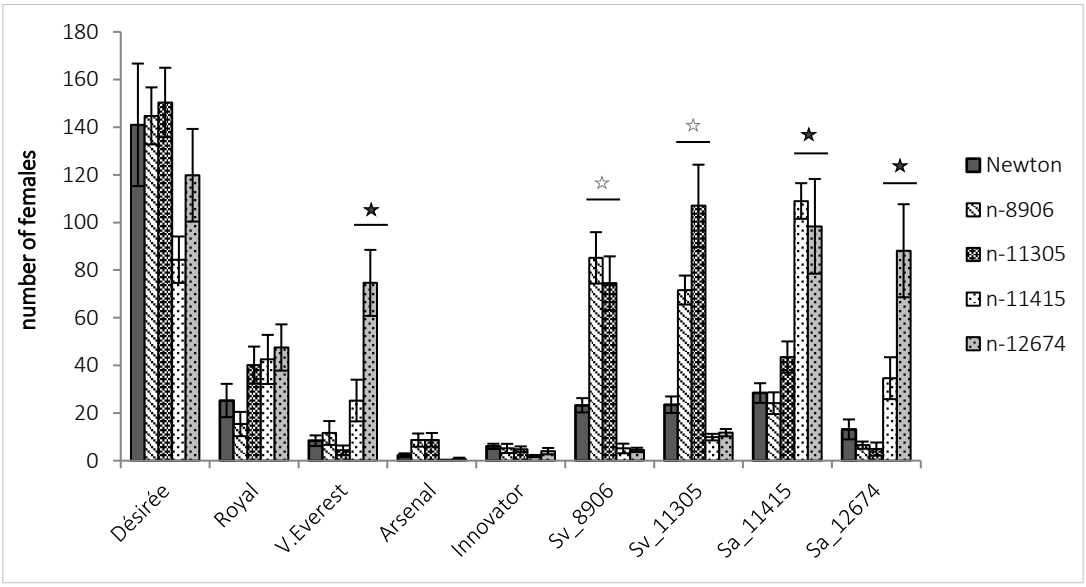


Figure 3-1 Screening test of Newton populations. Number of *G. pallida* females of the Newton field population and its 4 sub-populations selected on *S. vernei* (n-8906, n-11305) and *S. tuberosum* ssp. *andigena* CPC2802 (*H3*) (n-11415, n-12674) tested on nine different potato lines. Each bar represents the mean number of females from 4 biological replicates, from two experimental replications. Désirée is the most susceptible potato line, Arsenal and Innovator the least susceptible. Open stars indicate significantly higher virulence of the *S. vernei*-selected populations when tested on *S. vernei*-containing genotypes. Filled stars indicate significantly higher virulence of the *H3*-selected populations when tested on *H3*-containing genotypes. Error bars stand for standard error of the mean (*p*-value = 0.05).

Regarding the populations selected on *H3*, n-11415 demonstrated approximately 4 times higher reproduction rate on the corresponding potato line Sa_11415 when compared to Désirée and an approximate 3-fold increase on Sa_12674. In addition, the sub-population n-12674 had a 3.5-fold and 6.8-fold rise in reproduction rate on the same *H3*-containing potato clones. The same sub-population showed an increased reproductive ability (approximately 6 times) on Vales Everest, which has the same resistance source background, when compared to the virulence level of the unselected population on the same potato variety.

As a first overall conclusion, the populations selected on *S. vernei* (i.e. n-8906 and n-11305) showed clearly enhanced virulence on the *S. vernei*-containing potato clones Sv_8906 and Sv_11305 compared to the unselected Newton population when tested on those clones. The *H3*-containing selected population n-11415 and n-12674 were virulent only on the *H3*-containing potato genotypes Sa_11415 and Sa_12674 and Vales Everest (in the case of n-12674). Vales Everest was also more susceptible to n-11415 compared to the unselected populations but in a significantly lower level when compared to virulence level by n-12674. The increase in virulence is therefore specific to the resistance source that the population was selected on, rather than representing a general increase in aggressiveness or reproductive ability.

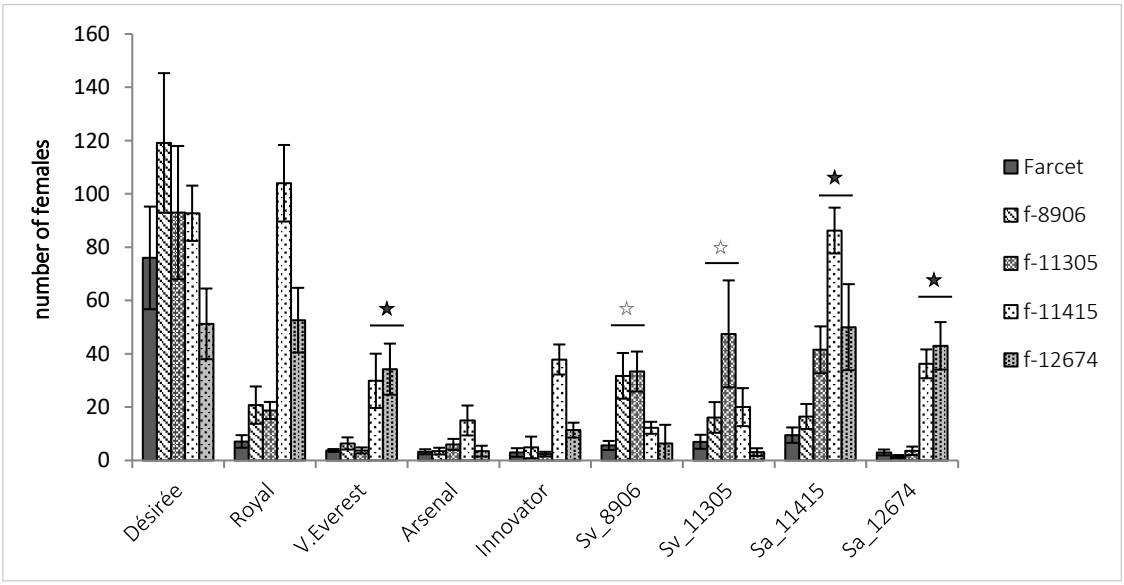


Figure 3-2 Screening test of Farcet populations. Number of *G. pallida* females of the Farcet field population and its 4 sub-populations selected on *S. vernei* (f-8906, f-11305) and *S. tuberosum* ssp. *andigena* CPC2802 (*H3*) (f-11415, f-12674) tested on nine different potato lines. Each bar represents the mean number of females from 4 biological replicates, from two experimental replications. Désirée is the most susceptible potato line, Arsenal and Innovator the least susceptible. Open stars indicate significantly higher virulence of the *S. vernei*-selected populations when tested on *S. vernei*-containing genotypes. Filled stars indicate significantly higher virulence of the *H3*-selected populations when tested on *H3*-containing genotypes. Error bars stand for standard error of the mean (*p*-value = 0.05).

Along with the Newton populations, the Farcet field population and sub-populations, which were selected on the same potato lines as Newton, were also tested on the same potato lines except for Arsenal due to lack of tuber availability. Arsenal was tested only in 4 replicates in total in one technical repeat. The results (Figure 3-2) confirmed differences in the reproductive ability among the nematode populations. All the

populations showed high reproductive ability on the susceptible Désirée as expected. Royal was moderately susceptible to the populations selected on *H3* (i.e. f-11415 and f-12674). Overall, the mean number of females was lower for the Farcet population than the Newton populations.

Arsenal was again the least susceptible to all the Farcet sub-populations. Interestingly, Innovator showed very low levels of susceptibility as was expected, except with f-11415, which is selected on *H3*. Even though Innovator contains resistance derived from *S. vernei*, the specific selected population was able to develop well on it (up to 38 females at 7wpi). As with the Newton populations, the populations selected on *S. vernei* (i.e. f-8906, f-11305) displayed an approximate 6-fold and 7-fold increase respectively when compared to the unselected population on the same potato lines, while the populations f-11415 and f-12674 (both selected on *H3*) did not show any significant increase compared to the unselected. On the other hand, the latter populations gave higher number of females when tested on the *H3*-containing clones (Sa_11415 and Sa_12674) compared to the unselected population (9 times and 14 times respectively). All the screening tests' raw results can be found in Supplementary Data 6.

Table 3-1 The mean female number of the Newton and Farcet populations on Désirée and the four CPC potato clones (Sv_8906, Sv_11305, Sa_11415 and Sa_12674). The underlined numbers show the mean number of females in the cases when selection source and clone's resistance source was the same. The superscripted letters stand for the least significant difference (LSD) (*p*-value = 0.05) within each population. Standard errors are also indicated (±SE).

	Désirée	**Sv_8906**	**Sv_11305**	**Sa_11415**	**Sa_12674**
Newton	141.0±30.3^{a}	23.3±0.7^{cb}	23.5±3.8^{dc}	28.4±6.2^{bd}	13.1±3.3^{bd}
n-8906	144.8±12.3^{a}	85.1±15.2^{c}	71.6±7^{c}	24.1±1.6^{b}	6.5±1.5^{b}
n-11305	150.4±18.9^{a}	74.5±5^{dc}	107.0±10.5^{e}	43.5±2.3^{cd}	4.9±0.3^{b}
n-11415	84.4±13.2^{a}	5.1±3.4^{bd}	9.9±2.5^{c}	109.0±6.2^{a}	34.6±10^{db}
n-12674	119.8±39^{a}	4.5±1.5^{b}	11.8±2.7^{b}	98.4±27.3^{ca}	88.1±27.8^{a}
Farcet	76±19.3^{a}	5.6±1.7^{b}	7±2.6^{b}	9.5±2.9^{b}	3±1.1^{b}
f-8906	119.1±26.2^{a}	31.8±8.6^{b}	16.1±5.8^{b}	16.5±4.7^{b}	1.4±0.7^{b}
f-11305	93±25^{a}	33.4±7.5^{bc}	47.5±20^{c}	41.5±8.8^{c}	3.6±1.6^{b}
f-11415	92.8±10.4^{a}	12.3±2.3^{b}	20±7.1^{b}	86.3±8.6^{a}	36.3±5.4^{b}
f-12674	51.3±13.3^{a}	6.4±7^{bc}	3.1±1.5^{b}	50±16.1^{ac}	43±8.9^{abc}

In summary, Désirée was very susceptible to every Newton and Farcet sub-population as was expected, while the unselected populations had in general lower reproduction rates on the tested clones. Nevertheless, the most important finding was that all the selected *G. pallida* populations (derived from either Newton or Farcet) showed increased virulence on the potato clones containing the same resistance source (Table 3-1).

3.4. Discussion

Globodera pallida is the biggest pest threat of potato in the UK. Multiple resistance sources have been identified and introgressed into potato varieties that confer partial resistance to *G. pallida*. However, nematode populations are able to overcome those resistance sources due to virulence (*vir*) alleles present in their gene pool (Turner and Fleming, 2002; Turner et al., 1983). Two main resistance sources are currently used by the breeding industry against PCN in the UK; *S. vernei* (*Gpa5*) and *S. tuberosum* ssp. *andigena* CPC2802 (also known as *H3*). Both sources are described as polygenic with more than two QTL thought to be involved in the resistance against *G. pallida* (Bryan et al., 2002; Dale and Phillips, 2009; van der Voort et al., 2000). Both resistances are highly efficient against the most common *G. pallida* pathotypes found in the UK i.e. Pa2/3 (Bryan et al., 2002; Bryan et al., 2004). However, the result of these interactions between *G. pallida* virulence alleles with the different QTL in potato can differ depending on the selection pressure that each resistance source creates (Beniers et al., 1995; Phillips and Blok, 2008; Turner and Fleming, 2002)

It has been demonstrated that *G. pallida* populations showed increased reproductive ability after 8 years of successive rearing on the partially resistant variety 'Darwina', which contains a *S. vernei*-derived resistance source (Beniers et al., 1995). The same trend was also confirmed when populations selected for several generations were tested on *S. vernei*, *H3*-containing potato clones and other *Solanum sp.* species (e.g. *S. multidissectum* and *S. sanctae-rosae*) (Phillips and Blok, 2008; Turner and Fleming, 2002; Turner et al., 1983). This shows that successive selection on partially resistant clones creates a strong selection in favour of virulence alleles to overcome resistance sources.

Additionally, the increase in virulence levels was not the same along the different selection sources. Populations selected on *H3* had a significantly higher virulence increase compared to those selected on *S. vernei* clones. For example, populations selected on *H3* showed a 7-fold virulence increase when compared to the unselected populations tested on the same resistance source, whereas those selected on *S. vernei* increased their virulence level by 3 times; this implies that initially, *H3* is more readily overcome than *S. vernei* (Phillips and Blok, 2008). Furthermore, it is highly likely that

the resistance from *S. vernei* is triggered by the additive effect of more genes compared to that from the *H3* and therefore, overcoming this resistance is more difficult (Vivian Blok, personal communication).

Selection pressure creates genetic differences in nematode populations after their selection process, as seen in a random amplification of polymorphic DNA (RAPD)-based study (Phillips and Blok, 2008) (Figure 3-3). This analysis showed that each population, including the derived selected sub-populations, were clustered together. In each group, sub-populations selected on *S. vernei* were phylogenetically closer to the corresponding unselected population, whereas those selected on *H3* were always more distant. This suggests that *H3* may select for a genetic subset of the starting population more readily than *S. vernei*.

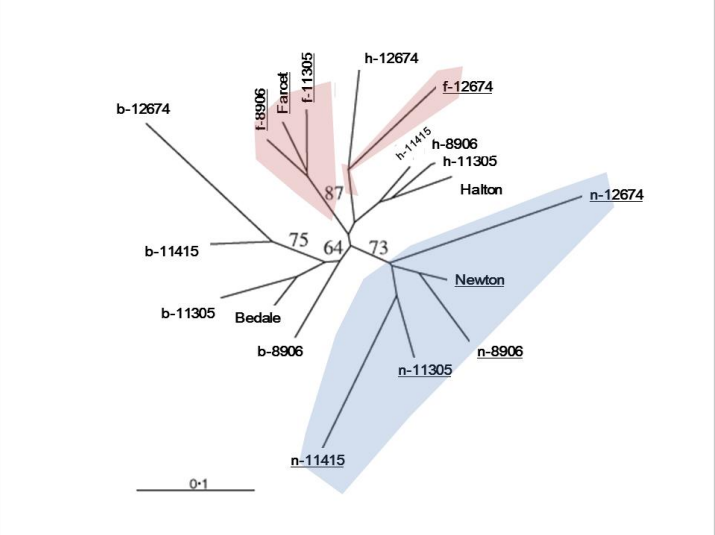


Figure 3-3 Unrooted dendrogramme of the genetic similarity in percentage of the four populations (Newton-n, Farcet-f, Halton-h and Bedale-b) tested based on molecular characterisation of RAPD markers. Numbers show bootstrap values. 8906 and 11305 contain *S. vernei* resistance source, 11415 and 12674 contain *H3* resistance. The underlined codes stand for the populations used in the current study, while the coloured areas point out where Newton (blue) and Farcet (pink) are located compared to each other and to the other two studied populations (Halton and Bedale) (modified and adjusted from Phillips and Blok (2008)).

The screening tests in the current study showed significant differences in reproduction rates among the different genotypes. Désirée was the most susceptible line as expected, whereas Arsenal and Innovator the least. The *S. vernei*-containing potato clones (i.e. Sv_8906 and Sv_11305) showed very high susceptibility to the Newton and Farcet sub-populations selected on the same resistance source (i.e. n-8906, n-11305, f-8906 and f-

11305), contrary to the susceptibility level to the unselected populations. These results are comparable to the findings of a recent study (Fournet et al., 2016). In that study, a *G. pallida* population that had been grown on Désirée for multiple successive generations (i.e. 6, 8 and 10 times) when tested on the *S. vernei*-containing variety 'Iledher' showed very low virulence levels. In contrast, when the *G. pallida* lineage had been selected on Iledher for the same number of generations was tested on Iledher, the virulence level rose vertically with the increase being greater in the cases where the number of generations of selection was higher (Fournet et al., 2016). Similarly, the *H3*-containing potato clones (i.e. Sa_11415 and Sa_12674) were highly susceptible to the Newton and Farcet sub-populations with the same selection background (i.e. n-11415, n-12674, f-11415 and f-12674). These results were consistent and very specific for each resistance source. We also showed and verified that *H3* selection led to approximately higher levels of virulence compared to that from *S. vernei* in both tested populations (i.e. Newton and Farcet).

Conversely, no increased numbers of developed females from populations selected on a specific resistance source were observed when they were grown on potato clones containing a different resistance source. The unselected populations had also low reproductive ability on most of the potato lines containing resistance. In the screening of Farcet sub-populations, we also observed a greater variability in their reproductive ability. However, it should be pointed out that the first repeat of the screening test took place under different environmental conditions that prevailed in the greenhouse during the experiments. The first screening of Farcet sub-populations was performed during winter when the light conditions are not at their optimum. This may have impeded the growth rate of the potato tubers and their emerging roots, which in turn may affect negatively the egg hatching and the performance of nematodes. Nevertheless, the significance of the outcome was not affected.

Interestingly, the sub-population selected on *H3*, f-11415, showed significantly higher reproduction rate on the *S. vernei*-containing Innovator when compared to the other selected sub-populations and the unselected population. This was consistent in both repeats, with up to 8 times higher reproduction rate compared to the rate of a sub-

population with the same resistance background as Innovator, and up to 13 times higher when compared to the unselected population. The same selected population was also able to overcome an individual resistance source (that of *S. vernei*) in other studies, proving that our finding is confirmatory (Phillips and Blok, 2008). The second *H3*-selected population, i.e. f-12674, showed similar behaviour when tested on Innovator but to a lesser extent compared to that of f-11415. Innovator was registered as a commercial variety in 2004 and is listed as highly resistant to both *G. rostochiensis* and to *G. pallida* populations Pa2 and Pa3 (Trudgill et al., 2003) (AHDB potato variety database). There is already evidence that resistance in Innovator has been broken by some *G. pallida* populations (Vivian Blok, personal communication). Three *S. vernei* sources have been described that have been introgressed into potato varieties (van Eck et al., 2017). Innovator was originally bred in the Netherlands and it is therefore likely that contains *S. vernei* source type originating from the Dutch clone type VRN1-3. Moreover, both the selected and unselected populations used in these experiments are a result of recurrent multiplication of the cysts on potato plants (either on the correspondent potato clone or on the susceptible Dèsirèe) for many years. Nevertheless, the initial cysts originate from fields in England where a mix of populations was possible; unfortunately, a further in-depth investigation regarding the purity of these initial populations was not feasible during the current project.

To summarise, *G. pallida* populations selected on a specific resistance source showed high reproductive ability on potato genotypes that contain that resistance source. The increase in virulence was greater in the populations selected on *H3* compared to those selected on *S. vernei*, which may reflect a stronger initial selection pressure on *H3*. We also suggest a possible resistance breakdown on the commercially important variety 'Innovator', which holds resistance from *S. vernei*, by selected populations when specific selection levels have been applied. During these populations screening tests, the developed females were collected in single tubes and stored at -20^{O}C. This unique generated nematode material will be used for downstream analyses and more specifically for constructing DNA libraries intended for target gene enrichment sequencing in order to analyse their effector sequences and identify possible polymorphisms among the sequences of different selected populations (see Chapter 4).

4. Identification of *Avr* genes in *G. pallida* using PenSeq, in conjunction with a new genome assembly

4.1. Background

4.1.1. Effector diversity

In chapter 1, it was described how the multiple introductions of *G. pallida* into Europe have led to the emergence of many different populations with varying levels of virulence against resistance sources (Picard et al., 2007; Plantard et al., 2008). The interactions between these nematodes and their host are mediated by a wide range of effectors, which in the case of the plant-parasitic nematodes (and thus the PCN) are produced mainly in the oesophageal gland cells. The continuous interaction between the parasite and host can lead to the adaptation of certain populations towards specific hosts with resistance *loci* though the acquisition of virulence (Phillips and Blok, 2008; Turner and Fleming, 2002). These interactions can be often very specific and unique in specific populations (Rosso and Grenier, 2011).

The way these adaptations are mediated or have taken place is intriguing. Activation of resistance responses is usually achieved through the recognition, either directly or indirectly, of an effector. Single amino acid changes in the effector-encoding genes can change their recognition status turning them into *avr* genes if recognised, such as that of the SPRYSEC effector *Gp-RBP1* that has a single amino acid polymorphism in the protein N-terminus that determines recognition by the *R* gene *Gpa2* (Diaz-Granados et al., 2016; Sacco et al., 2009). Gene duplications within gene families can create diverse functions of proteins as well as domain changes such as the multiple motifs that *G. rostochiensis CLE* genes can have (Lu et al., 2009). It is suggested that cell wall degrading enzymes (CWDEs) (e.g. cellulases, pectate lyases etc.) that play an important role in parasitism of the plant-parasitic nematodes have been acquired from bacteria through multiple lateral (or horizontal) gene transfer events (Danchin and Rosso, 2012; Jones et al., 2005; Smant et al., 1998). Lastly, alternative splicing of precursor mRNA has also been reported in the

plant-parasitic nematodes and in *G. rostochiensis* specifically (e.g. in *G. rostochiensis* chorismate mutase) creating different gene variants (Lu et al., 2009).

A great diversity of effectors is present in the genomes of pathogens and several mechanisms exist that promote effector diversity. Thines and Kamoun (2010) reported that 74% of the genome of the oomycete *P. infestans* consists of repetitive DNA with low gene numbers. Effector sequences are preferentially located in these regions. It was suggested that this region is highly dynamic, and functions in a way to accelerate effector evolution. Additionally, Eoche-Bosy et al. (2017b) suggested that this may also be the case in *G. pallida*, where a large part of the genome (Cotton et al., 2014) could be involved in host adaptation.

4.1.2. An overview of the published PCN genome assemblies

The first genome of a PCN species was published recently, specifically that of *G. pallida* (Cotton et al., 2014). This was followed with an assembly from *G. rostochiensis* opening new insights into PCN parasitism. The first assembly of *G. pallida* (124.6Mb) is larger than that of *G. rostochiensis* (95.9Mb) (Eves-van den Akker et al., 2016) and similar to that of the newly described PCN species from the North Americas, *G. ellingtonae* (119Mb) (Phillips et al., 2017). The *G. pallida* assembly shows lower CEGMA-based completeness (74%, compared to 94% for the other two species). It is composed of 6,873 scaffolds with the longest scaffold being about 600kb-long. On the other hand, the *G. rostochiensis* assembly contains 4,377 scaffolds (with the longest scaffold being 688kb long), while that of *G. ellingtonae* contains only 2,248 scaffolds (with the longest scaffold 2.8Mb long).

Regarding the number of genes, 16,466 and 14,378 genes were identified in *G. pallida* and *G. rostochiensis* respectively (Table 4-1). The low completeness of the assembly and the differences in its organisation in *G. pallida*, might reflect the different sequencing technologies and bioinformatics tools used at that time, as well as the greater genetic heterogeneity of the *G. pallida* populations used for sequencing (Eves-van den Akker et al., 2015). Specifically, sequencing of *G. pallida* and *G. rostochiensis* was performed on Illumina® HiSeq™ that produces significantly shorter reads with higher read depth compared to the long-read sequencing PacBio SMRT™ technology used in *G. ellingtonae* (Cotton et al., 2014; Eves-van den Akker et al., 2016; Phillips et al., 2017).

Table 4-1 Comparison of the published genome assemblies of the three PCN species.

	G. pallida	*G. rostochiensis*	*G. ellingtonae*
Assembly size (Mb)	124.6	95.9	119.1
Scaffolds (n)	6,873	4,377	2,248
Scaffold N50 (kbp)	121.7	88.5	360
Longest scaffold (bp)	600,076	688,384	2,800,000
GC content (%)	37	38	37
CEGMA (complete/partial %)	74/81	94/96	92/96
Genes (n)	16,466	14,378	14,104
Proteins (n)	16,417	14,309	13,946
Gene density (per Mb)	132.2	149.9	n/a

The analyses of the PCN genome assemblies shows the complex evolutionary background and the level of the genomic re-arrangement that cyst nematodes have passed through. The gene synteny of *G. pallida* with the root-knot nematodes and that of *C. elegans* is restricted and only a small proportion of the scaffolds seem to have orthologues in other species (e.g. in *C. elegans* or even *M. hapla*) (Cotton et al., 2014); however, its low completeness and the fragmented nature of the *G. pallida* genome does not allow a safe and unbiased conclusion. Previous studies (Bird et al., 2009; Keeling, 2004) showed that small and compact genomes are characteristic of obligate parasites, possibly because of gene elimination. The root-knot nematode species have a similar number of genes in a genome smaller than 100Mbp. For instance, *M. hapla* contains 14,200 genes in a 54Mbp genome, while *M. incognita* has 19,200 genes along its 86Mbp genome (Bird et al., 2009).

However, this is not the case for the cyst nematodes of the *Globodera* genus which have larger genome sizes and longer introns. The genes in *G. pallida* are also widely organised into operons, but the majority of them have no adjacent homologues in any other species including *C. elegans,* suggesting loss or re-arrangement of the operon structures (Cotton et al., 2014). Lastly, in comparison with *G. rostochiensis*, *G. pallida* shares only 2,000 out of the total 16,800 orthologue clusters, 60% of which are putatively involved in plant parasitism (Eves-van den Akker et al., 2016).

4.1.3. Transcript profiles in different lifecycle stages of *G. pallida*

The analysis of the PCN genomes indicated different expression levels of gene clusters at different lifecycle stages (Cotton et al., 2014). Genes expressed in subventral gland cells had higher expression levels during the pre-parasitic stages (e.g. J2), whereas those in dorsal gland cells were highly activated later in the post-infection and sedentary stages (Eves-van den Akker et al., 2016; Palomares-Rius et al., 2012). Transcriptional profiling of different lifecycle stages showed higher numbers of differentially regulated genes during the pre-parasitic (i.e. J2) stages of the juveniles. In the egg and J2 stages, the genes related to survival against environmental conditions (e.g. heat shock proteins, proteins relating to regulate redox balance) are more active, since these are stages that occur outside of the host. As the J2 becomes infective and enters the host, proteins and other molecules facilitate these interactions as it enters its parasitic life-stages.

While the juvenile feeds on the syncytium, it grows, moults, and sex determination takes place (Sobczak and Golinowski, 2011). The number of differentially regulated genes is gradually decreased from that stage on. In the females, a small increase in gene regulation is observed in the late stages due to the embryo development. At the same time, proteins that might be correlated with the transition from the parasitic into a free-living lifestyle are expressed in the males (Cotton et al., 2014; Jones et al., 2009; Palomares-Rius et al., 2012).

4.1.4. Target enrichment sequencing

Knowledge of the specific *G. pallida* effectors that distinguish virulent and avirulent populations can be a useful tool for predicting durability of cognate resistances in potato. However, sequencing of entire genomes from different populations is very laborious and not cost-effective. Additionally, genomes may contain large repetitive or non-coding regions making their study even more difficult (Jupe et al., 2013; Mamanova et al., 2010).

Target enrichment sequencing (Mamanova et al., 2010) can be a solution to this, as it offers a cost-effective and high-throughput approach to identify and study sequence polymorphisms in a subset of the genome as it focuses on selectively capturing, enriching and sequencing only the targeted genomic regions. In plants, it has been

developed and applied for capturing *R* genes (i.e. *R* gene enrichment sequencing or RenSeq) (Chen et al., 2018; Jupe et al., 2014; Jupe et al., 2013) and subsequently as a diagnostic tool to show presence or absence of functional NB-LRR *R* genes in potato (i.e. dRenSeq) (Van Weymers et al., 2016). It has also been used for re-annotating NB-LRR genes in previously non-annotated genomic regions, such as the 331 previously uncharacterised NB-LRR *R* genes identified in the potato *S. tuberosum* DM clone (Jupe et al., 2013). Similarly, target enrichment sequencing can also facilitate the study of *Avr* genes between populations (i.e. pathogen enrichment sequencing – PenSeq).

The principle is that unique biotinylated RNA-based oligonucleotide probes (or 'baits') of a specific length are designed to bind to complementary, previously characterised or potential target genes. The genomic DNA (gDNA) samples are firstly fragmented into small fragments of a specific size prior the production of DNA libraries. The DNA libraries consist of purified fragmented DNA that has been end repaired (to ensure that each fragment is free of overhangs) and have an adaptor ligated specific for the sequencing platform. The adapters are then used to bind to specific indexes (barcodes) in the case of multiplexed libraries (Figure 4-1; step i) (Jupe et al., 2014). The enrichment is done through the hybridisation of the PCR-amplified barcoded DNA library with the biotinylated baits (Figure 4-1; step ii). The baits are subsequently captured by the added streptavidin-coated magnetic beads (Figure 4-1; step iii). Eventually, any non-targeted DNA fragments are washed away and discarded (Figure 4-1; step iv), the magnetic beads are washed away from the captured library, which in turn is amplified and sequenced (Figure 4-1; step v) (Jupe et al., 2014).

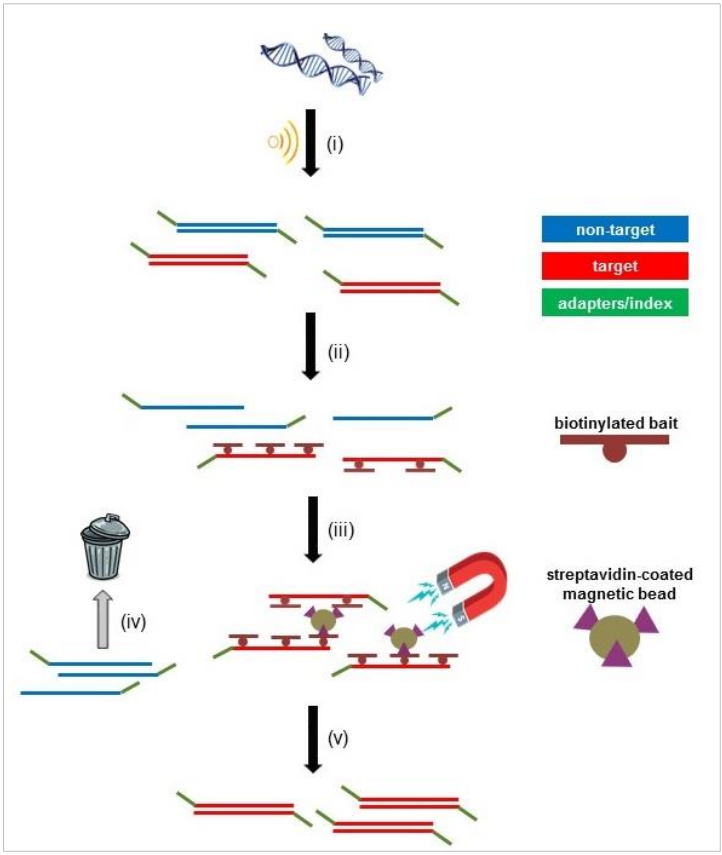


Figure 4-1 The workflow of the target enrichment sequencing.

4.2. Chapter objective

In this chapter, target enrichment sequencing was applied to highly virulent *G. pallida* populations that had been collected during the screening tests of chapter 3. Using this method, effector-encoding genomic regions were targeted, captured and sequenced. Sequencing was performed on the Illumina MiSeq platform in order to achieve high read depth. At the same time, HMW DNA was extracted from the unselected Newton population and was sequenced with the long-read PacBio sequencing platform to generate a new, improved and more complete *G. pallida* genome assembly. Two draft versions of this new assembly were used for variant calling analysis of the captured sequenced reads from the selected populations. This analysis was aimed at the identification of variants between the different virulent populations, and ultimately in candidate *avr* genes.

4.3. Results

4.3.1. HMW DNA extraction for the new *G. pallida* reference assembly

High-molecular weight (HMW) DNA was extracted from *G. pallida* J2s (unselected Newton population) following the protocol described in chapter 2 (section 2.2.11). Two extractions and elutions were performed in parallel, and the final samples contained 6.03µg and 5.15µg of DNA respectively. Both samples were checked for protein and salt contamination using a Nanodrop spectrophotometer. The absorbance 260/280 ratios for the samples were 2.01 and 1.95, while the 260/230 ratios were 2.06 and 2.07, which indicated that salt and protein contaminants were below the detection threshold of the instrument in both eluates. To verify the integrity of the extracted samples, 100ng from each eluate were run on a 0.7% agarose gel. Both samples showed similar estimated sizes of 20kb to 22kb with low amounts of degradation and high integrity (Figure 4-2).

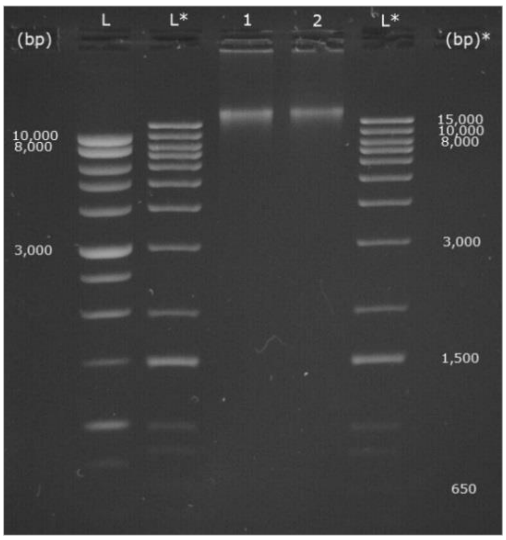


Figure 4-2 An 0.7% agarose gel electrophoresis of the two HMW DNA samples extracted (lanes 1, 2). Two DNA ladders were used; 1kb DNA ladder (L) and 1kb+ (L*) (ThermoFisher). Both samples showed no fragmentation, high integrity and sizes estimated between 20kbp and 22kbp.

4.3.2. PacBio sequencing and generating of a new *G. pallida* reference assembly

The sequencing of the HMW DNA samples was done in the Norwegian Sequencing Centre in Oslo. Prior to sequencing, both samples were pooled together and were subsequently sequenced on 4 PacBio SMRT® cells aiming at 300x total coverage. In total,

2,789,551 reads were generated with subread N50 equal to 10,250bp to 10,750bp depending on the SMRT cell and, on average, yielded about 9GB from each SMRT cell. In parallel with this work an estimation of the physical size of the stained nucleus DNA was performed with flow cytometry by Dr. Sebastian Eves-van den Akker (Cambridge University) using samples provided by me. This analysis showed an estimated physical size between 106Mb and 113Mb (Figure 4-3).

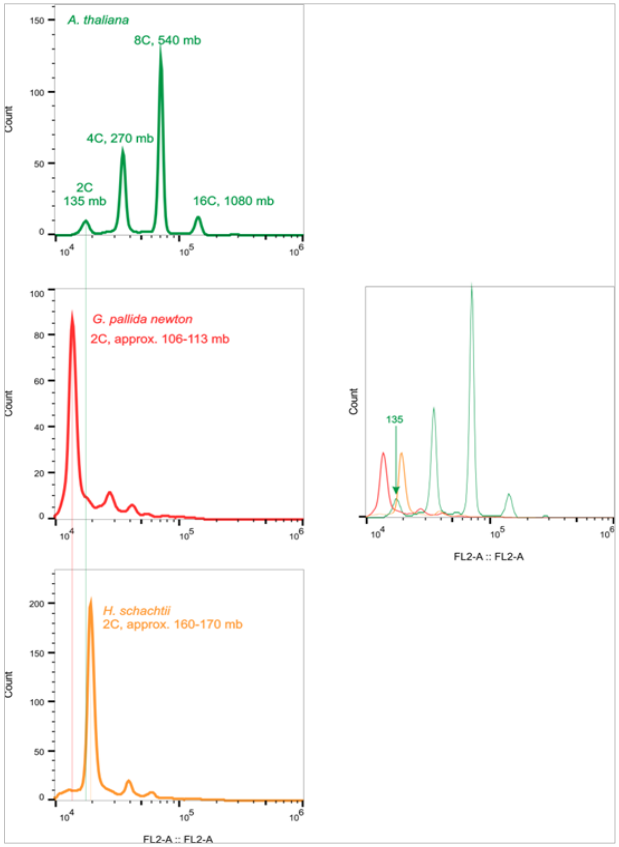


Figure 4-3 Estimation of the physical size of stained nuclear DNA in *G. pallida* using flow cytometry. Estimates were between 106mb and 113mb haploid. The 2C peak is written as haploid length of the diploid nucleus. DNA from *Arabidopsis thaliana* and *Heterodera schachtii* were used as controls.

The *de novo* assembly and gene calling were done following the steps described in chapter 2 by Dr. Peter Thorpe and was completed in December 2018 (unpublished data). The final phased assembly had a size of 119,544,868bp organised in 163 scaffolds, with an N50 of approximately 2.3million bp. BUSCO analysis showed that 89% of the analysed sequences were complete with a further 5% present as partial sequences. Only 37,000nt from 4 scaffolds did not show similarity to nematode-origin sequences when BlobTools was run on the final polished assembly showing that the final assembly was largely free of any contaminant sequences (Figure 4-4).

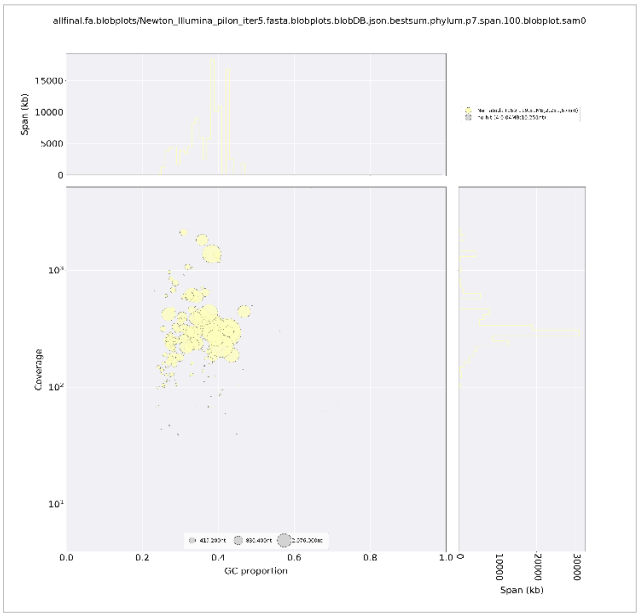


Figure 4-4 Taxonomic affiliation of the sequences generated by BlobTools. The figure was used for removing contaminant sequences on the final phased new *G. pallida* assembly. In this two-dimensional plot, the tested sequences (here in a scaffold-level) are represented by dots and coloured by taxonomic affiliation based on BLAST searches. For each sequence, the position on the x-axis is determined by the GC content, while the position on the y-axis is determined by the base coverage. The yellow coloured dots represent sequences of nematode origin, while the dot diameter is relative to the sequence size.

Over 92% of the available RNA-seq reads available for *G. pallida* were mapped back to the assembly. Eventually, 19,088 genes were predicted, which is about 3,000 more than the predicted genes in the published genome. Differential expression analysis of the genes was done using RNA-seq data from 8 different lifecycle stages (egg, J2, 7dpi, 14dpi, 21dpi, 28dpi, 35dpi and male) as part of the original *G. pallida* genome project. For each stage, data from 2 replicates were analysed, except that of J2 (3 replicates). Principal component analysis (PCA) showed that the stages 21dpi, 28dpi and 35dpi could not be separated depending on their expression profiles, contrary to the rest of the samples (Figure 4-5 and Figure 4-6). Similarly, pairwise analysis showed that the replicates for each life stage clustered together, with the exception of the 21dpi, 28dpi and 35 dpi samples which were clustered together. These data suggest that gene expression at these three life stages does not change dramatically compared to the changes occurring at other stages. This reflects the biology of the nematode as these life stages are all settled at the syncytium, feeding and moulting.

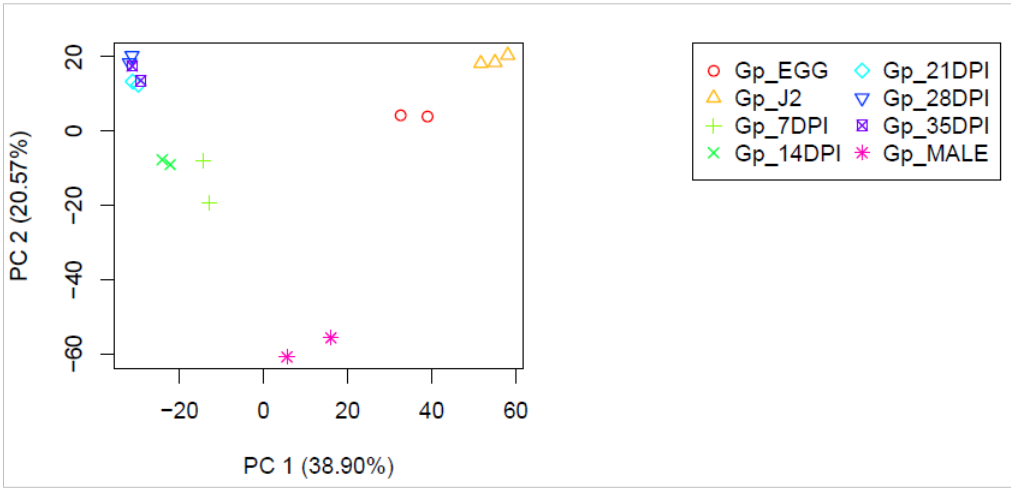


Figure 4-5 PCA matrix of the RNA-seq data from the eight different *G. pallida* lifecycle stages. It shows the possibly correlation of the observations (variables) after pairwise comparisons of the data sets. The first principal component (PC1) accounts for as much of the variability in the data set as possible, and the second PC (PC2) for as much of the remaining variability as possible. Each dot represents an RNA-seq data set from a specific lifecycle stage replicate. Here, in the PC1 (most possibly correlation), the samples from the stages 21dpi, 28dpi and 35dpi are closely clustered together. In other words, gene expression in these stages do not differ significantly, but varied significantly from the gene expression level during the stages egg, J2, 7dpi, 14dpi and male.

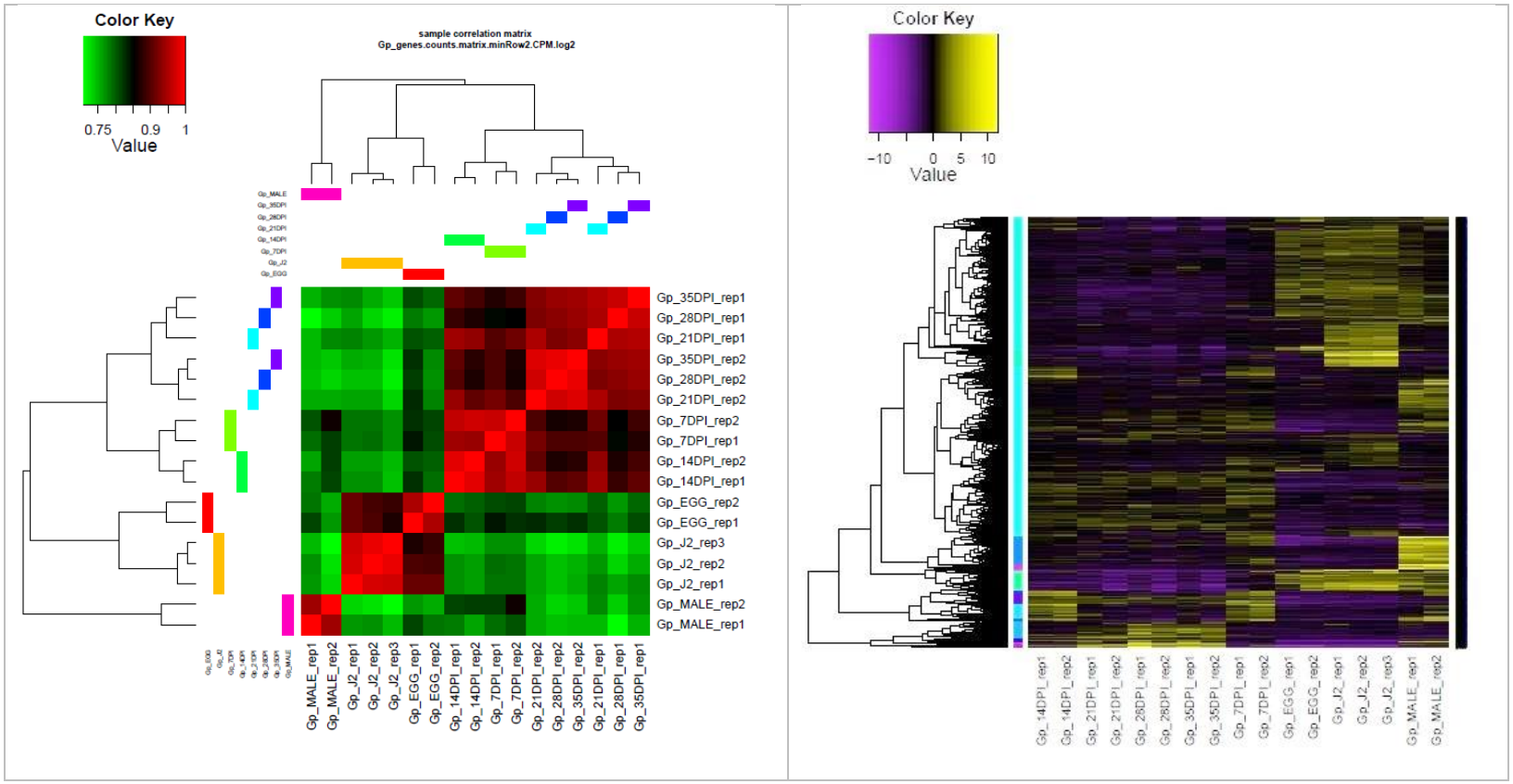


Figure 4-6 Plotting RNA-seq data on the new *G. pallida* genome assembly. *Left*: Heatmap demonstrating the pairwise correlation of the RNA-seq data based on the expression levels in each lifecycle stage. Red indicates higher levels of expression and green indicates lower levels of expression (log fold-change > 2, $p < 0.001$). *Right*: Differential expression (DE) analysis of the clustered called genes per lifecycle stage (p = 40). In both analyses, the stages egg, J2, 7dpi, 21dpi, 28dpi, 35dpi and male were used.

Table 4-2 An overview of the two drafts of the new *G. pallida* assembly used as well as the final one. The published genome (Cotton et al., 2014) is also presented as comparison. In the final column, the final version of the new assembly is presented (* in December 2018).

	Published *G. pallida*	**New *G. pallida***		
		Bigger/more complete	**Smaller/less complete**	**Final new assembly***
Size (Mb)	124.6	192.5	120.5	119.6
Scaffolds (n)	6,873	1,923 [contigs]	267	163
Scaffold N50 (bp)	121,687	207,432	1,194,397	2,251,599
Longest scaffold (bp)	600,076	1,613,714	6,176,216	8,303,766
GC content (%)	37	36	36	37
Span of Ns (bp)	21,024,229	N/A	662,155	1,245,593
BUSCO (% complete)	74	N/A	89	94
Predicted genes (n)	16,000	N/A	19,000	19,088

4.3.3. Applying PenSeq to *G. pallida*

In total, 24,744 biotinylated RNA-based oligonucleotides for selecting *G. pallida* potential effectors were designed *in silico* by Joanne Lim (The James Hutton Institute) and used for target enrichment sequencing of genomic DNA (PenSeq as described in chapter 2). Seven hundred potential effectors were targeted; 611 were fully covered by the designed baits (Table 4-3).

Table 4-3 Predicted number of target-genes with a corresponding bait coverage (%) of the designed baits. From the total 700 targets, 611 were fully covered by the designed baits.

# of total	**Predicted number of targets with x% bait coverage (Total target genes = 700)**									
baits	**100%**	**90%**	**80%**	**70%**	**60%**	**50%**	**40%**	**30%**	**20%**	**10%**
24,744	611	647	667	683	689	694	695	700	700	700

DNA was extracted from the highly virulent females of 2 selected Newton populations following the protocol described in chapter 2 (section 2.2.2). The population n11305 selected on *S. vernei*-containing line Sv_11305 and tested on both Sv_8906 and Sv_11305, the population n11415 selected on *H3*-containing line Sa_11415 and tested on both Sa_11415 and Sa_12674, as well as the starting unselected Newton population (grown on Désirée). All populations were represented in two replicates (Table 4-4). Since the kit used (NEBNext® Multiplex oligos for Illumina®) allowed the indexing of up to 12 samples per pooled library in the given cost, only one of the population sets were

included in the PenSeq workflow (i.e. Newton sub-populations). In addition to this, they were chosen over Farcet populations because of the greater number of the initial nematode material that was available, as well as the fact that these populations had shown greater and more clear increase of virulence level during the screening tests. All 10 DNA libraries were uniquely barcoded and pooled together before their hybridisation with the baits, while the captured genomic regions were sequenced on a single lane of Illumina MiSeq.

Table 4-4 The 5 selected *G. pallida* Newton populations used for the DNA extractions to prepare DNA libraries. Each sample is duplexed. In the third and fourth columns the selection source is shown and the total extracted DNA.

Lib	Population	Selection source	Total input DNA
1	Newton x Désirée	-	190ng
2	Newton x Désirée	-	191ng
3	n11305 x Sv_8906	*S. vernei*	226ng
4	n11305 x Sv_8906	*S. vernei*	276ng
5	n11305 x Sv_11305	*S. vernei*	177ng
6	n11305 x Sv_11305	*S. vernei*	250ng
7	n11415 x Sv_11415	*H3*	338ng
8	n11415 x Sv_11415	*H3*	424ng
9	n11415 x Sv_12674	*H3*	278ng
10	n11415 x Sv_12674	*H3*	364ng

In total 23,590,596 reads were generated by the Illumina MiSeq platform. The majority of them (19,895,213, or 97.48%) were assigned to one of the 10 indexed libraries (PF reads). All the libraries were represented approximately equally (1/10 each) in the final enriched and sequenced captured library pool (Figure 4-7).

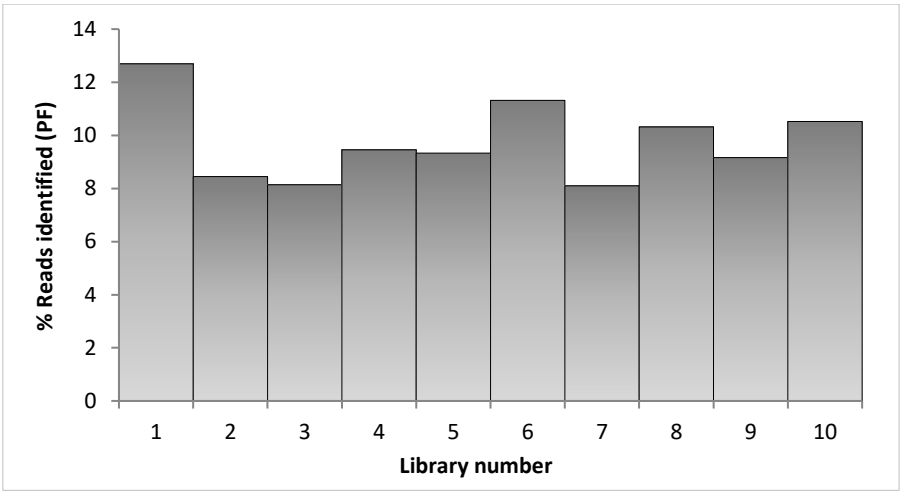


Figure 4-7 Representation of each barcoded (indexed) library in the final sequenced enriched pooled library. PF stands for the total fraction of the reads assigned to a specific index.

4.3.4. Post-capture libraries were enriched for effectors

A real-time PCR (qPCR) was performed on the pre- and post-enrichment libraries using primer sets that amplify known effectors (i.e. SPRY-414-2, SPRY-1719-1 and G16H02) in *G. pallida*. All three effector genes showed an approximate 100-fold to 1000-fold increase in amplification in the post-enrichment library compared to the pre-enriched (Figure 4-8). All the effector genes were represented at very low levels in the pre-enrichment library. This showed that effector sequences were highly enriched in the post-capture libraries.

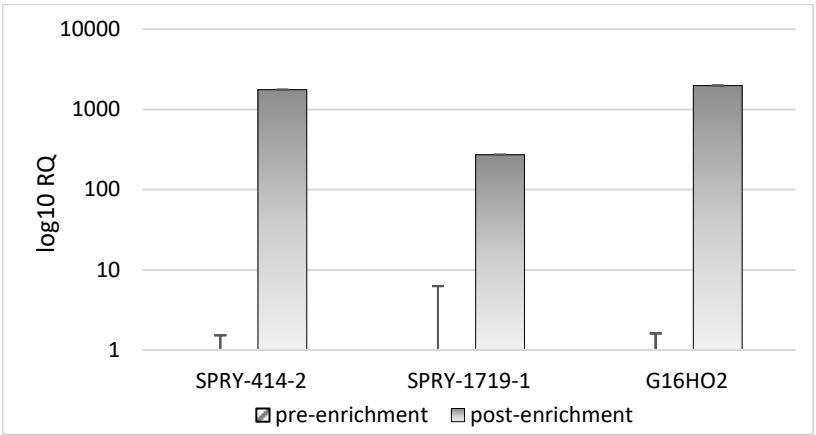


Figure 4-8 Comparative relative quantification (RQ) of the three known effector genes *(SPRY-414-2*, *SPRY-1719-1* and *G16H02)* in the pre- and post-enrichment libraries. The non-effector gene *GAPDH* was used as endogenous control. Error bars stand for 95% confidence interval (CI). RQ values are presented in a logarithmic scale log10.

4.3.5. Curation of the MiSeq raw reads and mapping to the reference

Firstly, all the raw reads were subjected to quality control (QC). The sequencing quality score was kept stringent (i.e. Q30) in order to avoid any noise increase in the variant calling analysis due to sequencing errors. In other words, during the trimming of the raw reads, only the bases with a quality score more than Q30 were kept and hence, the probability of calling a base incorrectly was 1 in 1,000.

According to the basic statistics generated by the tool FastQC, about 76% of the bases of the reads met the set quality score. By using Trimmomatic, all the raw reads that did not meet the set parameter as well as the Illumina adapters were removed (Figure 4-9). Trimming was done for both forward and reverse 250bp-long reads generated in sequencing.

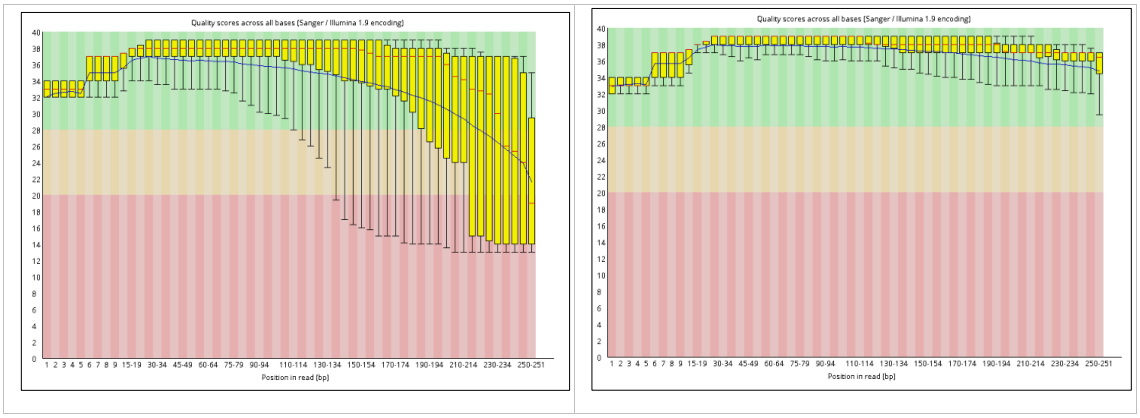


Figure 4-9 An example of the FastQC control of the same sample before (left) and after (right) trimming the Illumina adapters and low-quality bases (Q30). The horizontal axis shows the position of each base (bp) in the read, while the vertical axis represents the quality score. Similar patterns were shown in the other 9 samples for both paired reads.

At the time the mapping of the sequence reads was performed, the final version of the new assembly was not available. At this time, two new assemblies were available – a smaller but less complete (as assessed by BUSCO) version and a larger but more complete version (Table 4-2). During trimming of the raw reads, only the successfully paired reads were kept. These reads were mapped against the three reference genomes; the old published assembly (Cotton et al., 2014) and the two drafts of the new one. In the mapping, secondary alignments were excluded to avoid any reads mapped back to more than one site. Table 4-5 shows the number of the mapped paired reads of each assigned library/sample for all three reference genomes used. In every case, a larger number of reads were mapped to the larger of the new genome assemblies than did to the original *G. pallida* assembly, suggesting that this may be a more complete representation of the *G. pallida* genome. It can also be said that the significantly larger size of the new/bigger assembly is due to repetitive regions, which may cause some reads to be multiply mapped against it.

The success of the target enrichment sequencing can also be seen in Figure 4-10 where the reads generated in sequencing were specifically matched to targeted genomic regions. These regions were highly enriched in the final, enriched library what was eventually sequenced. Conversely, genomic regions that were not targeted by the designed biotinylated baits were not present in the library that was sequenced.

Table 4-5 Number of the mapped paired reads to each library/sample. The percentage and number of the sequenced (pass-filter) PF (indexed) raw reads assigned to each sequenced library/sample is shown. The total number of the PF reads shows the number of the sequenced raw reads generated during sequencing and assigned to one of the 10 used Illumina indexes. The total number of reads is the number of all the sequenced raw reads containing also raw reads without any Illumina index ligated (e.g. low-quality reads). See also Figure 4-7. Library 2 had lower PF number compared to the library 1 possibly because of an error during the final elution of the library before sequencing.

Library	%PF reads	Read (PF) number	Total mapped reads		
			Big	Small	Old
1	12.69	2,525,239	785,612	570,380	558,061
2	8.45	1,680,648	316,203	232,100	223,166
3	8.14	1,619,430	734,400	518,516	532,718
4	9.45	1,880,973	585,346	413,104	420,133
5	9.33	1,856,541	786,115	553,747	531,197
6	11.32	2,251,700	781,037	548,281	559,849
7	8.1	1,611,174	651,925	474,599	487,504
8	10.32	2,052,390	597,584	435,068	443,878
9	9.16	1,822,640	650,849	467,519	489,690
10	10.52	2,093,334	621,578	450,032	468,283
	TOTAL PF reads	19,895,213			
	TOTAL reads	23,590,596			

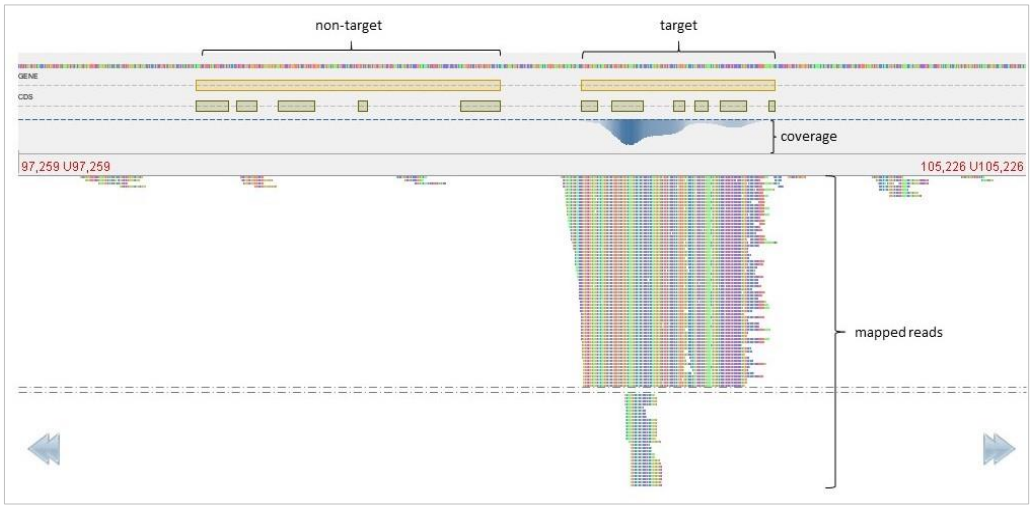


Figure 4-10 Visualisation on Tablet (Milne et al., 2012) of the reads to a successfully captured enriched target gene. In the figure the difference between the mapped reads on a target and non-target gene is visible. The coverage of the mapped reads is also shown.

4.3.6. Variant calling

For the variant calling, the 10 samples were divided into three biological groups; two groups containing the populations selected on *S. vernei* or *H3*, and one group containing

the unselected Newton population on the susceptible genotype. Two further filtering steps were done for a minimum read depth of 10 and differences in the allele frequencies between the biological groups. At this point, only variants with consistent alleles present in all the 3 biological samples were kept in the analyses.

In total, 445 variants (e.g. SNPs, indels etc.) were found in the mapping against the old reference, while 309 and 374 were identified against the new bigger and smaller draft references respectively. From those, approximately 1/3 were synonymous changes. The annotation created by SnpEff (see also Table 2-11) indicated that there were few markers with high-impact on the predicted gene product (e.g. variants implicated in splicing, gaining or loss of a start/stop codon, frameshift variants); 3 in the mapping to the new/smaller assembly reference, 2 to the new/bigger assembly and only one to the old reference. Regarding the identified variants with moderate effect (e.g. missense variants), a significantly larger number were found in the old assembly mapping compared to the two new draft versions. This could be a result of the different population background used in the old and the new reference (Table 4-6). The rest of the non-synonymous all-type variants had moderate impact (i.e. missense variants) (full list of the identified variants for both assemblies can be found in Supplementary Data 7).

Table 4-6 An overview of the identified all-type variants in all three reference assemblies used for mapping.

	Big assembly	Small assembly	Old assembly
Total variants	309	374	445
Synonymous variants	104	119	119
High-impact variants	2	3	1
Moderate-impact variants	55	80	122

4.3.7. Identified variant genes

As a next step, all the amino acid sequences containing at least one high- or moderate-impact (all-type) variant were extracted and BLASTp searched. From this, 37 different BLASTp hits were identified in the mapping to the new/bigger assembly and 54 different hits to the new/smaller assembly; the BLASTp search failed for the old assembly mapping. In the new/bigger assembly mapping, 11 amino acids were aligned to SPRY-containing proteins (PF00622), while this number was larger (16) in the new/smaller

reference. Five in total SPRY-containing proteins (two identified in both drafts and 3 only in the new/smaller draft) were truncated (i.e. the N- or C-terminus was eliminated). Several were annotated as RBP-1, a member of the SPRYSEC multigene family. Many amino acid sequences gave no BLAST hit or matched a sequence of unknown function; specifically, 13 out of 37 in the new/bigger reference and 21 out of 54 in the new/smaller one.

When comparing the BLASTp search hits from both references, 2 high-impact genes were identified in common; a SPRYSEC protein (similar to the gene GPLIN_000725400) and a C2H2-type zinc finger-containing protein (similar to GPLIN_000245500), which is also similar to a SPRY-containing protein. An additional third high-impact gene, similar to the SPRYSEC GPLIN_000908700, was also flagged in the mapping of the new/smaller reference.

The analysis also gave 16 variants with moderate impact identified in both references. Four of them had an unknown function or no BLAST hit, and 6 were similar to a SPRY-containing protein again. Two hits showed high similarity to previously known esophageal gland-localised proteins from cyst nematodes of the *Heterodera* genus, while one of them showed similarity to a chorismate mutase (PF01817). The list also included representatives of another two proteins similarly previously identified in cyst nematodes, such as the *G. rostochiensis* effector 1106 and a β-1,4-endoglucanase (PF00150). Table 4-7 and Table 4-8 list all the identified variant genes containing a variant with high or moderate impact according to SnpEff annotation (full list of extracted amino acid sequences in Supplementary Data 8).

As can also be seen in the tables below, about 87% and 78% (in the new big and small reference respectively) of the variants with high or moderate impact were included in the bait design. This difference in the percentages between the two draft references may be a result of the smaller completeness (BUSCO%) of the latter. Lastly, it can be seen that the majority of the identified non-synonymous variants (with either high or moderate impact) were found with the populations selected on the *H3* resistance source (Table 4-9 and Table 4-10).

Table 4-7 List of all the variant genes identified when the new/bigger assembly was used as reference. The first two columns show the position of the identified flagged variant. The description column shows the function of each extracted amino acid after the BLASTp search generated by Blast2GO. The best hit is shown in the BLAST column along with the corresponding e-value and alignment percentage; n/a stands for an unknown functional annotation (as a result of the BLASTp search) of the identified candidate gene. The target column indicates whether the variant gene was intended to be targeted during the bait designing.

CHROM	POS	Description	BLAST	Target	e-value	%ID
tig00000288	394527	n/a	GPLIN_001185100	YES	0	91.9
tig00000396	48722	n/a	GPLIN_000148800	YES	2.10E-58	98.1
tig00001809	706403	beta-1,4-endoglucanase [Aphelenchus avenae]	GPLIN_001185800	YES	0	93.3
tig00001810	59622	pectate lyase partial	GPLIN_000673000	YES	2.60E-130	96.4
tig00003402	44026	SPRYSEC effector SPRY- partial	GPLIN_000909700	YES	0	96.4
tig00003402	118479	SPRYSEC effector SPRY- partial	GPLIN_001349800	YES	2.80E-117	94
tig00006097	252075	n/a	GPLIN_001292400	YES	1.80E-175	90.7
tig00006102	118264	Ubiquitin-conjugating enzyme E2 partial [Trichinella pseudospiralis]	GPLIN_001268500	YES	1.10E-127	92.1
tig00007011	35665	truncated secreted SPRY domain-containing partial	GPLIN_001436900	YES	0	94.5
tig00007981	25725	n/a	GPLIN_000560800	YES	0	94.4
tig00008058	46211	Vitellogenin- partial	GPLIN_000945900	YES	0	95.7
tig00010936	45559	esophageal gland-localized secretory 3	GPLIN_000996800	YES	2.70E-145	93.2
tig00044370	159411	n/a	GPLIN_000283500	YES	2.40E-177	91.7
tig00044372	101951	SPRYSEC effector SPRY- partial	GPLIN_001300800	YES	4.50E-102	94.8
tig00044372	169072	SPRYSEC effector SPRY- partial	GPLIN_001048200	YES	8.80E-85	90.7
tig00044393	72334	esophageal gland-localized secretory 16	GPLIN_000666500	YES	2.10E-148	95.1
tig00044435	3490	SPRYSEC effector SPRY- partial	GPLIN_000700300	YES	1.20E-73	95
tig00044517	645244	n/a	GPLIN_000036500	NO	1.40E-11	88.9
tig00044517	309620	RBP-1 partial	GPLIN_000632600	YES	9.40E-81	93.7

tig00044517	363374	SPRYSEC effector SPRY- partial	GPLIN_001506100	NO	1.50E-127	92.9
tig00044578	439096	failed axon connections isoform X2	GPLIN_000604400	YES	0	80.6
tig00044647	1027985	n/a	GPLIN_000376600	YES	1.50E-90	96
tig00044742	132663	glutathione synthase	GPLIN_000241200	YES	0	94.7
tig00044865	193016	n/a	GPLIN_000049600	NO	2.80E-116	95.2
tig00044865	167702	truncated secreted SPRY domain-containing partial	GPLIN_000312300	YES	0	96.4
tig00044965	86031	Zinc C2H2 domain and Zinc finger C2H2-type integrase DNA-binding domain and Zinc finger, C2H2-like domain-containing	GPLIN_000245500	YES	3.30E-76	71
tig00044969	228042	n/a	GPLIN_001244700	NO	3.70E-81	72.3
tig00045060	63349	dopey-1 [Stronglyloides ratti]	GPLIN_000328200	YES	0	96.2
tig00045073	87751	n/a	GPLIN_000909200	YES		
tig00045124	45954	RBP-1 [Globodera pallida]	GPLIN_001032500	YES	1.30E-68	91.3
tig00045208	46655	n/a	GPLIN_001171800	YES	1.00E-30	83.6
tig00045365	93575	SPRYSEC effector SPRY- partial	GPLIN_000626800	YES	1.90E-75	85.3
tig00045412	189986	n/a	GPLIN_001153200	YES	1.70E-155	93.1
tig00529440	11676	Tyrosine- kinase Fps85D [Stronglyloides ratti]	GPLIN_000490600	NO	0	91.7
tig00529478	109177	RBP-1 partial	GPLIN_000725400	YES	0	97.3
tig00531362	361787	n/a	GPLIN_000673400	YES	1.70E-124	89.7
tig00531427	55571	1106 effector family variant 22 5	GPLIN_000768400	YES	1.40E-54	89.7

Table 4-8 List of all the variant genes identified when the new/smaller assembly was used as reference. The first two columns show the position of the identified flagged variant. The description column shows the function of each extracted amino acid after the BLASTp search generated by Blast2GO. The best hit is shown in the BLAST column along with the corresponding e-value and alignment percentage; n/a stands for an unknown functional annotation (as a result of the BLASTp search) of the identified candidate gene. The target column indicates whether the variant gene was intended to be targeted during the bait designing.

CHROM	POS	Description	BLAST	Target	e-value	%ID
scaffold1	2729897	von Willebrand factor type D domain [Necator americanus]	GPLIN_000945900	YES	1.00E-61	99.1
scaffold1	1044903	n/a	GPLIN_001059500	YES	0.0073	95.2
scaffold114	109167	RBP-1 partial	GPLIN_000725400	YES	0	97.3
scaffold14	96464	n/a	GPLIN_000501800	NO	2.90E-24	94.0
scaffold14	132668	glutathione synthase	GPLIN_000706900	NO	5.40E-69	95.7
scaffold15	1539941	Electron transfer flavo -ubiquinone mitochondrial [C. elegans]	GPLIN_000242100	YES	1.60E-23	95.8
scaffold15	311100	n/a	GPLIN_001444100	YES	0	95.4
scaffold15	313605	n/a	GPLIN_001444200	YES	3.60E-64	93.4
scaffold159	87748	Hypothetical protein	GPLIN_000909200	YES	1.60E-04	100.0
scaffold19	921905	SPRYSEC effector SPRY- partial	GPLIN_000252200	YES	1.60E-06	95.8
scaffold2	838707	truncated secreted SPRY domain-containing partial	GPLIN_000177900	YES	7.10E-103	95.6
scaffold2	845840	n/a	GPLIN_000178200	NO	2.40E-104	93.9
scaffold2	1167978	SPRYSEC effector SPRY- partial	GPLIN_000413600	YES	5.90E-171	94
scaffold2	1158865	truncated secreted SPRY domain-containing partial	GPLIN_000413700	YES	8.20E-85	92.9
scaffold2	6019574	RBP-1	GPLIN_000433800	YES		
scaffold2	1544927	n/a	GPLIN_000466900	YES	4.40E-14	98
scaffold2	1081901	SPRYSEC effector SPRY- partial	GPLIN_001048200	YES	5.40E-126	89.2
scaffold2	1014613	SPRYSEC effector SPRY- partial	GPLIN_001300800	YES	5.10E-101	94.2

scaffold20	55072	1106 effector family variant 22 5	GPLIN_000768400	YES	2.1E-53	89.7
scaffold24	833469	n/a	GPLIN_001153300	YES		
scaffold24	388528	n/a	GPLIN_001296800	YES	2.80E-48	98.8
scaffold25	322890	n/a				
scaffold3	1241499	n/a	GPLIN_000017600	NO	0.0022	86.4
scaffold3	2838416	n/a	GPLIN_000036500	NO	1.00E-11	88.9
scaffold3	3172248	n/a	GPLIN_000426600	YES	0.03	94.7
scaffold3	838813	SPRYSEC effector SPRY- partial	GPLIN_000626800	YES	4.40E-74	83.1
scaffold3	3118475	n/a	GPLIN_001506100	NO	1.50E-127	92.9
scaffold35	802370	beta-1,4-endoglucanase	GPLIN_001185800	YES	2.60E-35	95.5
scaffold36	998835	truncated secreted SPRY domain-containing partial	GPLIN_000737800	NO	1.50E-37	88.7
scaffold37	172364	n/a	GPLIN_000413900	NO	3.60E-99	96.8
scaffold37	162757	SPRYSEC effector SPRY- partial	GPLIN_000414100	YES	0.0036	100.0
scaffold37	167928	secreted SPRY domain-containing 9	GPLIN_000971300	YES	0	94.5
scaffold4	928518	n/a	GPLIN_000641200	YES	0	91.6
scaffold4	711005	cbp-1				
scaffold49	126099	Zinc C2H2 domain and Zinc finger C2H2-type integrase DNA-binding domain and Zinc finger, C2H2-like domain-containing	GPLIN_000245500	YES	1.90E-93	93.3
scaffold49	369924	beta-1,4-endoglucanase	GPLIN_000313600	YES	1.00E-147	96.3
scaffold5	312526	n/a	GPLIN_000049700	NO		
scaffold5	278893	n/a	GPLIN_000312300	YES	0.00E+00	96.4
scaffold5	510904	failed axon connections isoform X2	GPLIN_000555600	YES	5.80E-54	91.4
scaffold5	1936112	n/a	GPLIN_000560800	YES	0	94.4

scaffold5	1190437	SPRYSEC effector SPRY- partial	GPLIN_001186200	YES	1.30E-109	96.8
scaffold53	628919	phosphoglycerate mutase family	GPLIN_000107900	YES	0	92.1
scaffold53	63322	dopey-1 [Strongyloides ratti]	GPLIN_000328200	YES	5.30E-89	96.6
scaffold6	280053	esophageal gland-localized secretory 3	GPLIN_000996800	YES	2.50E-67	94.6
scaffold61	228354	esophageal gland-localized secretory 16	GPLIN_000666500	YES	5.40E-147	95.8
scaffold7	1369164	n/a	GPLIN_000376600	YES	1.50E-90	96.0
scaffold73	461390	truncated secreted SPRY domain-containing partial	GPLIN_001436900	YES	0	94.3
scaffold8	1624474	SPRY domain-containing SOCS box SSB-4 [Loa loa]	GPLIN_001150700	YES	8.90E-165	94.3
scaffold8	80106	no_blast	GPLIN_001292400	YES	4.60E-68	91.4
scaffold80	321149	SPRYSEC effector SPRY- partial	GPLIN_000908700	YES	4.30E-137	92.7
scaffold89	211144	n/a	GPLIN_000312100	YES	8.60E-28	93.4
scaffold89	83609	n/a	GPLIN_001258100	YES	4.90E-120	93.5
scaffold94	99973	unc-80 [C. elegans]	GPLIN_000425400	NO	9.90E-65	97.2
scaffold97	291103	gag-pol poly	GPLIN_000779000	YES	6.70E-13	94.4

Table 4-9 Presence of the reference (REF) and the alternative (ALT) allele in the three biological groups (Désirée, *S. vernei* and *H3* resistances) after the variant calling analysis of the new/bigger assembly. In the first two columns, the position of the identified variant is shown, and the putative impact on the annotated product (according to the SnpEff annotation). In the "*S. vernei*" and "*H3*" columns, the presence of REF/ALT allele is shown in the mapping against the reference (which coincides with the "Désirée" column). The underlined letters indicate the resistance source where the selection took place (i.e. in the *S. vernei*- or *H3*-selected populations).

CHROM	POS	TYPE	Désirée	*S. vernei*	*H3*
tig00000288	394527	MOD	G/A	G/G	G/A
tig00000396	48722	MOD	A/A	A/G	A/A
tig00001809	706403	MOD	G/A	G/A	G/G
tig00001810	59622	MOD	T/C	T/C	T/T
tig00003402	44026	MOD	G/G	G/G	G/C
tig00003402	118479	MOD	G/G	G/G	G/A
tig00003402	118492	MOD	A/A	A/A	A/G
tig00006097	252075	MOD	G/A	G/G	G/A
tig00006097	252215	MOD	T/T	T/A	T/T
tig00006097	252423	MOD	G/A	G/G	G/A
tig00006102	118264	MOD	G/G	G/T	G/G
tig00007011	35665	MOD	T/C	T/C	T/T
tig00007011	35680	MOD	C/T	C/T	C/C
tig00007981	25159	MOD	G/C	G/C	G/G
tig00007981	25725	MOD	A/G	A/G	A/A
tig00008058	46211	MOD	A/A	A/G	A/A
tig00010936	45559	MOD	G/A	A/A	G/G
tig00010936	46294	MOD	G/T	G/T	G/G
tig00010936	46441	MOD	C/T	C/C	C/T
tig00010936	46565	MOD	C/T	C/C	C/T
tig00010936	46843	MOD	G/G	G/A	G/G
tig00010936	46894	MOD	A/G	A/G	A/A
tig00044370	159411	MOD	A/A	A/A	A/T
tig00044372	101951	MOD	C/T	C/T	T/T
tig00044372	169072	MOD	C/G	C/G	G/G
tig00044372	169532	MOD	A/G	A/G	A/A
tig00044393	72334	MOD	C/C	C/T	C/C
tig00044435	3490	MOD	A/A	A/A	A/G
tig00044435	3585	MOD	T/T	T/T	T/G
tig00044517	309620	MOD	C/G	C/C	C/G
tig00044517	363374	MOD	A/C	A/C	A/A
tig00044517	363393	MOD	T/C	T/C	T/T
tig00044517	645244	MOD	T/G	T/G	T/T
tig00044578	439096	MOD	G/A	G/G	G/A
tig00044647	1027985	MOD	G/G	G/T	G/G
tig00044742	132663	MOD	G/A	G/G	G/A
tig00044865	167702	MOD	G/G	G/G	G/A
tig00044865	193016	MOD	C/G	C/G	G/G

tig00044965	85183	MOD	G/G	G/G	G/A
tig00044965	85818	MOD	G/C	G/C	G/G
tig00044965	86031	HIGH	GCCTTCTCA/GCTTCTCA	GCCTTCTCA/GCTTCTCA	GCTTCTCA/ GCCTCTCA
tig00044969	228042	MOD	A/T	A/A	A/T
tig00044969	228223	MOD	G/A	G/G	G/A
tig00045060	63349	MOD	A/A	A/G	A/A
tig00045073	87751	MOD	C/C	C/C	C/T
tig00045124	45954	MOD	C/C	C/T	C/C
tig00045208	46655	MOD	G/A	G/A	G/G
tig00045208	47516	MOD	T/T	T/T	T/A
tig00045208	47669	MOD	T/T	T/T	T/C
tig00045365	93575	MOD	C/T	C/C	C/T
tig00045412	189986	MOD	A/A	A/C	A/A
tig00529440	11676	MOD	A/A	A/A	A/T
tig00529478	109177	HIGH	TCG/TCG	TCG/TG	TCG/TCG
tig00531362	361787	MOD	C/C	C/T	C/C
tig00531427	55571	MOD	A/G	G/G	A/A

Table 4-10 Presence of the reference (REF) and the alternative (ALT) allele in the three biological groups (Désirée, *S. vernei* and *H3* resistances) after the variant calling analysis of the new/smaller assembly. In the first two columns, the position of the identified variant is shown, and the putative impact on the annotated product (according to the SnpEff annotation). In the "*S. vernei*" and "*H3*" columns, the presence of REF/ALT allele is shown in the mapping against the reference (which coincides with the "Désirée" column). The underlined letters indicate the resistance source where the selection took place (i.e. in the *S. vernei*- or *H3*-selected populations).

CHROM	**POS**	**TYPE**	**Désirée**	***S. vernei***	***H3***
scaffold1	1044903	MOD	A/A	A/G	A/A
scaffold1	1045090	MOD	C/C	C/T	C/C
scaffold1	2729897	MOD	C/G	C/G	C/C
scaffold1	2729931	MOD	G/C	G/C	G/G
scaffold114	109167	HIGH	TCG/TCG	TCG/TG	TCG/TCG
scaffold14	96464	MOD	C/C	C/C	C/T
scaffold14	132668	MOD	G/A	G/G	G/A
scaffold15	311100	MOD	G/T	G/G	G/T
scaffold15	313605	MOD	A/C	A/A	A/C
scaffold15	1539941	MOD	A/A	A/G	A/A
scaffold159	87748	MOD	C/C	C/C	C/T
scaffold19	921905	MOD	A/A	A/A	A/G
scaffold2	838707	MOD	T/T	T/T	T/A
scaffold2	845840	MOD	A/G	A/A	A/G
scaffold2	1014613	MOD	C/T	C/T	T/T

scaffold2	1081901	MOD	G/C	G/C	G/G
scaffold2	1158865	MOD	C/C	C/C	C/T
scaffold2	1167978	MOD	A/G	A/G	A/A
scaffold2	1544927	MOD	C/G	C/C	C/G
scaffold2	6019574	MOD	C/C	C/A	C/C
scaffold20	55072	MOD	A/C	A/C	A/A
scaffold20	55084	MOD	T/C	T/C	T/T
scaffold20	55574	MOD	A/G	A/G	A/A
scaffold24	388528	MOD	A/T	A/A	A/T
scaffold24	388709	MOD	G/A	G/G	G/A
scaffold24	833469	MOD	A/A	A/C	A/A
scaffold24	833510	MOD	T/T	T/C	T/T
scaffold25	322890	MOD	C/T	C/T	T/T
scaffold3	838813	MOD	G/G	G/G	G/C
scaffold3	838909	MOD	A/A	A/A	A/G
scaffold3	839350	MOD	G/T	G/G	G/T
scaffold3	840716	MOD	TGC/CGT	TGC/TGC	TGC/CGT
scaffold3	840784	MOD	A/G	A/G	A/A
scaffold3	1241499	MOD	A/C	A/C	A/A
scaffold3	2838416	MOD	A/C	A/C	A/A
scaffold3	3118475	MOD	A/G	A/G	A/A
scaffold3	3118494	MOD	T/G	T/G	T/T
scaffold3	3172248	MOD	G/C	G/G	G/C
scaffold35	802370	MOD	G/A	G/A	G/G
scaffold36	998835	MOD	G/C	G/C	C/C
scaffold37	162757	MOD	AG/AG	AG/AG	AG/GA
scaffold37	162849	MOD	A/G	A/A	A/G
scaffold37	163441	MOD	A/A	A/A	A/T
scaffold37	167928	MOD	C/T	C/T	C/C
scaffold37	172364	MOD	A/G	G/G	A/G
scaffold4	711005	MOD	G/A	G/A	G/G
scaffold4	711058	MOD	A/G	A/G	A/A
scaffold4	928518	MOD	GTG/TTA	GTG/TTA	GTG/GTG
scaffold49	125251	MOD	G/G	G/G	G/A
scaffold49	125886	MOD	G/C	G/C	G/G
scaffold49	369924	MOD	A/G	G/G	A/A
scaffold49	126099	HIGH	GCCTTCTCA/ GCTTCTCA	GCCTTCTCA/ GCTTCTCA	GCCTTCTCA/ GCCTCTC

scaffold5	278893	MOD	G/G	G/G	G/A
scaffold5	312526	MOD	G/G	G/G	G/A
scaffold5	312670	MOD	T/T	T/T	T/C
scaffold5	510904	MOD	C/T	C/C	C/T
scaffold5	1190437	MOD	C/C	C/C	C/T
scaffold5	1936112	MOD	G/C	G/C	G/G
scaffold5	1936678	MOD	A/G	A/G	A/A
scaffold53	63322	MOD	A/A	A/G	A/A
scaffold53	628919	MOD	A/A	A/A	A/C
scaffold6	280053	MOD	T/C	T/C	T/T
scaffold6	280104	MOD	C/C	C/T	C/C
scaffold6	280382	MOD	G/A	G/G	G/A
scaffold6	280506	MOD	G/A	G/G	G/A
scaffold6	280653	MOD	C/A	C/A	C/C
scaffold6	281388	MOD	C/T	T/T	C/C
scaffold61	228354	MOD	G/G	G/A	G/G
scaffold7	1369164	MOD	G/G	G/T	G/G
scaffold73	461390	MOD	G/A	G/A	G/G
scaffold73	461405	MOD	A/G	A/G	A/A
scaffold8	80106	MOD	C/T	C/C	C/T
scaffold8	80314	MOD	A/A	A/T	A/A
scaffold8	80454	MOD	C/T	C/C	C/T
scaffold8	1624474	MOD	T/A	T/T	T/A
scaffold80	321149	HIGH	G/A	G/G	G/A
scaffold89	83609	MOD	A/C	A/C	A/A
scaffold89	211144	MOD	GC/GC	GC/GC	GC/AG
scaffold89	211162	MOD	ACAGT/ACAGC	ACAGT/GCAGC	ACAGT/ACAGC
scaffold89	212535	MOD	C/A	C/A	C/C
scaffold89	212560	MOD	G/A	G/A	G/G
scaffold94	99973	MOD	T/T	T/C	T/T
scaffold97	291103	MOD	C/C	C/T	C/C

4.4. Discussion

In the last few years, new sequencing technologies have been developed, which in combination with up-to-date available bioinformatic tools can help in delivering better and more complete genomes with higher prediction accuracy. This can be very important especially for complicated genomes that come from diverse and non-uniform populations, including microscopic, outbreeding organisms like PCN. Although the genome assembly of *G. pallida* (Lindley population) was published relatively recently, a large part of it remains still unknown mainly due to low completeness (74% CEGMA).

Therefore, for the current project, a new improved genome assembly of *G. pallida* (Newton population) was generated using long-read sequencing technology. This allows sequencing of long, unfragmented DNA molecules. The relatively higher error rate observed compared to other technologies, can be significantly decreased by the sequencing of multiple SMRT cells (achieving very high coverage) combined with the correction of the sequenced reads with the shorter Illumina-generated reads. It was also important to generate a reference assembly for the starting population being used for selection analysis as we were keen to ensure that differences due to selection were not masked by polymorphisms between the Lindley and Newton populations.

As shown in Table 4-2, the new *G. pallida* assembly (unpublished data) is about 5Mb smaller than the published one but with a significantly smaller number of ambiguous bases (~1.2million bases only, compared to ~21million ambiguous bases in the published). The assembly was organised in 163 scaffolds with the longest one being over 8 million bp long and 50% of them were longer than 2.2 million bp. By contrast, the published genome assembly was organised in 6,873 scaffolds with the longest one being about 600Kbp long. Lastly, 3,000 more genes were predicted in the new genome, and this is reflected in its higher completeness score (94% BUSCO, compared to 74% CEGMA). In the current project, draft versions of the new assembly were used since the optimisation of the final one was still ongoing.

Effectors play a crucial role in the plant-nematode interactions in promoting infection (Dangl et al., 2013; Haegeman et al., 2012; Thorpe et al., 2014). In chapter 3, it was shown that *G. pallida* populations selected on two resistance sources showed increased virulence

on all the potato genotypes containing the same resistance source used in the initial selection. This selection process created a strong bottleneck effect that may also lead to genetic drift (Eoche-Bosy et al., 2017a; Phillips and Blok, 2008). In this chapter, these sub-populations were used in order to identify candidate effectors with a significant role in virulence on a specific resistance source.

To analyse polymorphisms of the effector-encoding genomic regions with high confidence and bypassing the complexities of the genome, target enrichment sequencing (Jupe et al., 2013; Mamanova et al., 2010) was applied. In the last few years, this method has been used for successfully sequencing specific genes or gene families from either plants (RenSeq) (Chen et al., 2018; Jupe et al., 2013) or plant pathogens (Strachan, 2018), or as a diagnostic tool for the presence or absence of specific genes (dRenSeq) (Armstrong et al., 2018; Van Weymers et al., 2016).

PenSeq was adapted and applied to capture and sequence candidate effectors from the above selected PCN populations. Baits were designed to capture 700 targets as described in chapter 2. Subsequently, DNA libraries were constructed from two Newton sub-populations selected on *S. vernei* (*Gpa5*), two on *H3* and the non-selected standard Newton population (grown on Désirée). The sequencing was performed on the short-read generating sequencing platform Illumina MiSeq to achieve high gene coverage. For the mapping of the generated reads, two draft versions of the new *G. pallida* (Newton population) genome assembly were used; a larger in size but more complete and a smaller in size (close to the predicted physical size) but less complete. In addition to that, the reads were also mapped against the published genome assembly.

Initial analyses showed the presence of 309 all-type variant calls (mostly SNPs) in the new/bigger assembly spread in 96 different scaffolds and 374 all-type variant calls on 67 scaffolds of the new/smaller. When the published assembly was used as a reference, 445 all-type variant calls were found. However, the significantly greater number of calls with the latter may include incorrect calls because of the different population; for this reason, we focused more on the analysis using the new genome assemblies as reference.

A closer look at the identified variant calls, revealed that 104 and 119 (in the new/bigger and new/smaller assembly respectively) did not cause any change in the encoded amino acids, i.e. synonymous variants. The rest of the (non-synonymous) calls, were SnpEff-annotated based on the putative impact on their product to those with high and moderate impact, narrowing the numbers to 57 and 83 variant calls (in the mapping of the new/bigger and new/smaller assembly respectively) including mainly SNPs. BLASTp search showed that these variant calls were located on 37 (in the mapping of the bigger reference) and 54 (in the mapping of the smaller reference) annotated genes with known or unknown function. From these, 32 (or 42 respectively) were highly similar to targeted genes against which the baits were designed.

SPRY-containing proteins (PF00622) were the most abundant protein family found. The significance and role of the SPRY-containing proteins and the secreted SPRY-containing proteins (i.e. SPRYSECs) in suppression and activation of plant immune responses is well known and has been covered in the introduction of the current chapter as well as in chapter 1. Here, in total 23 candidates highly similar to 23 different secreted SPRY-containing proteins previously annotated in *G. pallida* were identified in BLASTp search. Four of them were similar to the first SPRYSEC that was identified in *G. pallida*, the RBP-1, which constitutes a known effector produced in the dorsal glands of the J2s (Diaz-Granados et al., 2016; Mei et al., 2018). From the identified SPRY-containing proteins, the proteins similar to GPLIN_001436900, GPLIN_000312300, GPLIN_000177900, GPLIN_000413700 and GPLIN_000737800 were truncated. Only 2 of them (GPLIN_001032500 and GPLIN_000433800) appeared to be complete (Table 4-7 and Table 4-8), while the rest were partial. According to previous studies (Carpentier et al., 2012; Sacco et al., 2009), the hypervariable surface of the SPRY domain contains several sites with polymorphisms as a result of positive selection processes, which might be the case in these selected populations too.

Furthermore, two candidate cellulases were found which have putative functions in catalysing cell wall components (e.g. fungi glucans, plants pectate etc.); a cellulase similar to β-1,4-endoglucanase (PF00150) and to a pectate lyase (PF03211), both found as effectors in many plant parasitic nematode species including that of the genera

Heterodera, *Globodera*, *Meloidogyne* and *Bursaphelenchus* (Haegeman et al., 2012; Hamamouch et al., 2012; Jones et al., 2009; Kikuchi et al., 2005; Yan et al., 1998). They are both expressed in the subventral gland cells of pre- and parasitic J2 stages, and they facilitate their penetration into the host tissue as well as in intracellular migration (Geric Stare et al., 2011; Goellner et al., 2000; Kikuchi et al., 2005; Smant et al., 1998; Vanholme et al., 2007). Most of them have been shown to have high homology with bacterial (Jones et al., 2009; Yan et al., 1998), or fungal sequences (Haegeman et al., 2012) and have possibly been acquired by horizontal gene transfer.

About 70 esophageal gland-localised secretory protein effectors (GLAND proteins) have been identified in *Heterodera glycines*, expressed in the dorsal glands of the infective and parasitic juvenile stages (Gautier, 2015; Noon et al., 2015). Our analysis showed the presence of two sequences very similar to GLAND3 and GLAND16. The first one, also previously known as G12H04, is a putative gland protein implicated in the invasion of the J2 into the host (Kumar et al., 2014; Noon et al., 2015). The second may be a chorismate mutase catalysing several pathways which lead to the production of secondary metabolites, salicylic acid and vitamins in plants (Jones et al., 2003). Glutathione synthase (PF03917) was listed here as candidate too. With similarity with the GLAND proteins, *G. pallida*-secreted glutathione synthase may also manipulate host developmental pathways by stimulating its host to supply the nematode with adequate nutrients (Cotton et al., 2014; Lilley et al., 2018).

Zinc finger-like proteins have been previously characterised as novel effectors in *G. pallida* (Thorpe et al., 2014) though their function in this species is not well known. The C2H2 zinc finger (PF00096) transcription factors have been largely found in *C. elegans*, possibly implicated in neurotransmission (Doitsidou et al., 2018). The versatile zinc motifs of those small proteins are able to bind to nucleic acids and lipid substrates mediating several biological processes such as gene transcription, mRNA transferring, cytoskeleton organisation and protein folding (EMBL-EBI). In the same category, dopey-1 (similar to a mouse protein), also found in fungal species of the genus *Aspergillus*, might be implicated in protein trafficking and signalling. A last candidate that might be implicated in various developmental processes and signalling in the nervous system was

similar to a tyrosine-kinase receptor (Tanaka et al., 2014; Yoda et al., 2003). However, no signal peptide was predicted in the extracted amino acid sequences of the above three potential candidates (i.e. zinc finger-like proteins, dopey-1 and the tyrosine-kinase receptor protein), suggesting that they can be important for the survival of the invading nematode inside the host plant.

Lastly, the outcome of our analysis also reported the presence of two other proteins with unknown or limited knowledge of their role in plant parasitic nematodes; the egg yolk protein vitellogenin (PF01347) and a member of the von Willebrand factor domain (PF00092). The former has also been found in the females of *M. incognita* (Wang et al., 2012) and in *C. elegans* is known for protecting nucleic acids and other proteins from harmful metal compounds derived from the environment (Nakamura et al., 1999). The latter has also been reported as a novel effector in *G. pallida* by Thorpe et al. (2014), while in *C. elegans* it maintains the body integrity (e.g. cuticle attachment to the hypodermis) (Bercher et al., 2001).

In this chapter of the thesis, target enrichment sequencing was successfully adjusted and applied for the identification of candidate *avr* genes in the selected *G. pallida* populations. BLASTp searches showed that many of the variant sequences were very similar to known effectors being reported as crucial factors for the survival of the nematode and the infection of the host plant. Others, of which less is known about their function in plant parasitic nematodes, a possible role in the interactions with the host were proposed. Although all these genes were found that may play significant roles in virulence of the selected populations on either *S. vernei* or *H3* resistance sources, only their functional validation can confirm their role as *Avr* genes. To enlist candidate genes as priority for functional analysis, we took into consideration only the candidate genes identified in common in both mapped references (new/bigger and new/smaller assembly). Hence, the following 7 candidate *Avr* genes can be suggested; the vitellogenin (GPLIN_000945900), the effectors GLAND3 (GPLIN_000996800) and 1106 (GPLIN_000768400) as well as the SPRY domain protein similar to GPLIN_000312300. In addition to them, another two genes with unknown function (similar to GPLIN_000560800 and GPLIN_000909200) can also be potential candidate genes.

5. Whole genome re-sequencing of the selected, highly virulent *G. pallida* populations

5.1. Background

5.1.1. Whole genome re-sequencing of plant-parasitic nematodes

In chapter 4, target enrichment sequencing (Jupe et al., 2013) was described as a successful high-throughput approach for the identification and study of polymorphisms in the sequences of the highly virulent *G. pallida* selected populations. By targeting effector-encoding regions, analysis led to identification of candidate *Avr* genes that could play role in the virulence activity against specific resistance sources. Although sequencing of specific targeted sequences is one of the advantages of that method, it also constitutes a limitation, since interesting outlier genomic regions are excluded from the analysis. These genomic regions may be important for the understanding of the selection, evolution or adaptation on host resistances (Eoche-Bosy et al., 2017a; Gautier, 2015; Gunther and Coop, 2013).

It has also been found that effectors in PCNs can be located inside or in the proximity of non-effector *loci* organised in genomic "islands" (Eves-van den Akker et al., 2016). Examination of these "islands" can reveal aspects of the evolution of adaptation in PCN species. Recent RNA-seq analyses of the transcriptomic activity of different populations of PCN species revealed that a single gene (i.e. *nep-1*) was able to initiate effector expression (such as that of RBP-1, pectate lyases and expansin) at a specific lifecycle stage. This might indicate that expression of certain genes possibly promotes the up-regulation of other gene or group of genes, including potential effectors (Duceppe et al., 2017). These effects can only be observed by whole-genome re-sequencing of populations.

Populations able to overcome resistance barriers can parasitize plants, which were previously non-hosts. This increased potential in host adaptation can also be depicted in their genomic backgrounds (Eoche-Bosy et al., 2017a; Eoche-Bosy et al., 2017b; Luikart et al., 2003). Depending on the selection pressure exerted on nematode populations, a directional change in a specific *locus* may occur in favour of virulence (Luikart et al.,

2003). Therefore, genomic analysis can provide additional information regarding the intra-species variance in virulence. Whole-genome comparisons of different populations is now a common approach in population genomics, including that of the cyst nematodes (Clement et al., 2013; Eoche-Bosy et al., 2017b; Eves-van den Akker et al., 2016; Futschik and Schlotterer, 2010; Gendron St-Marseille et al., 2018; Mimee et al., 2015). This approach reduces the likelihood of noise due to redundant DNA reads and increases the reliability of allele frequency analysis (Futschik and Schlotterer, 2010; Gunther and Coop, 2013).

5.2. Chapter objective

In this chapter, whole-genome re-sequencing was applied on the same four *G. pallida* selected populations used in chapter 4 for variant calling and genome-wide allele frequency analysis. This approach was applied as an alternative approach for identifying *G. pallida* effector-encoding genes under selection towards the *S. vernei* and *H3* resistance sources, as well as providing us with important information about the impact of selection on PCN genomes. As described for the downstream analysis in chapter 4, here the two newly generated draft versions of the *G. pallida* assembly were used for mapping the Illumina HiSeq-generated reads.

5.3. Results

5.3.1. Sequencing, curation of the HiSeq-generated reads and variant calling

DNA from the four selected populations collected in the screening tests (see chapter 3) was extracted and re-sequenced on the Illumina HiSeq.

Table 5-1 The 4 selected *G. pallida* Newton populations used for the DNA extractions for re-sequencing. Two were selected on *S. vernei* and two on *H3*. In the last column, the total number of reads generated in sequencing is shown

	Population	Selection source	Named as	Total reads
1	n11305 x Sv_8906	*S. vernei*	vernei_8906	96,847,261
2	n11305 x Sv_11305	*S. vernei*	vernei_11305	93,499,267
3	n11415 x Sa_11415	*H3*	H3_11415	85,030,469
4	n11415 x Sa_12674	*H3*	H3_12674	89,827,138

Sequencing was performed on all 4 samples (Table 5-1). During the trimming of the raw reads, approximately the 79% - 83% (depending on the sample) were successfully trimmed and survived. About 12% of the raw reads represented only the forward read and about 3% only the reverse. Mapping was done against the two drafts of the new assembly reference (i.e. bigger and smaller genome assembly; see also chapter 4).

Freebayes identified about 3.3 million variants in both mapped assemblies and more specifically, about 1 variant per 50 bases (the bigger assembly) and 1 variant per 28 (in the smaller assembly). The majority (about two thirds) were SNPs, whereas the rest consisted of indels, multiple-nucleotide polymorphisms (MNPs) or a combination of both (mixed). About 10% of the variants had a silent effect (synonymous mutations). In both draft assemblies, about 64% of the left, non-synonymous variants were modifiers for either the upstream or downstream transcripts of an annotated gene and 10% (in the bigger assembly) or 5% (in the smaller assembly) were located in exons. The rest ~25% (in the bigger assembly) or 30% (in the smaller assembly) were variants located on non-coding genomic regions (e.g. intron or intergenic regions) (Figure 5-1).

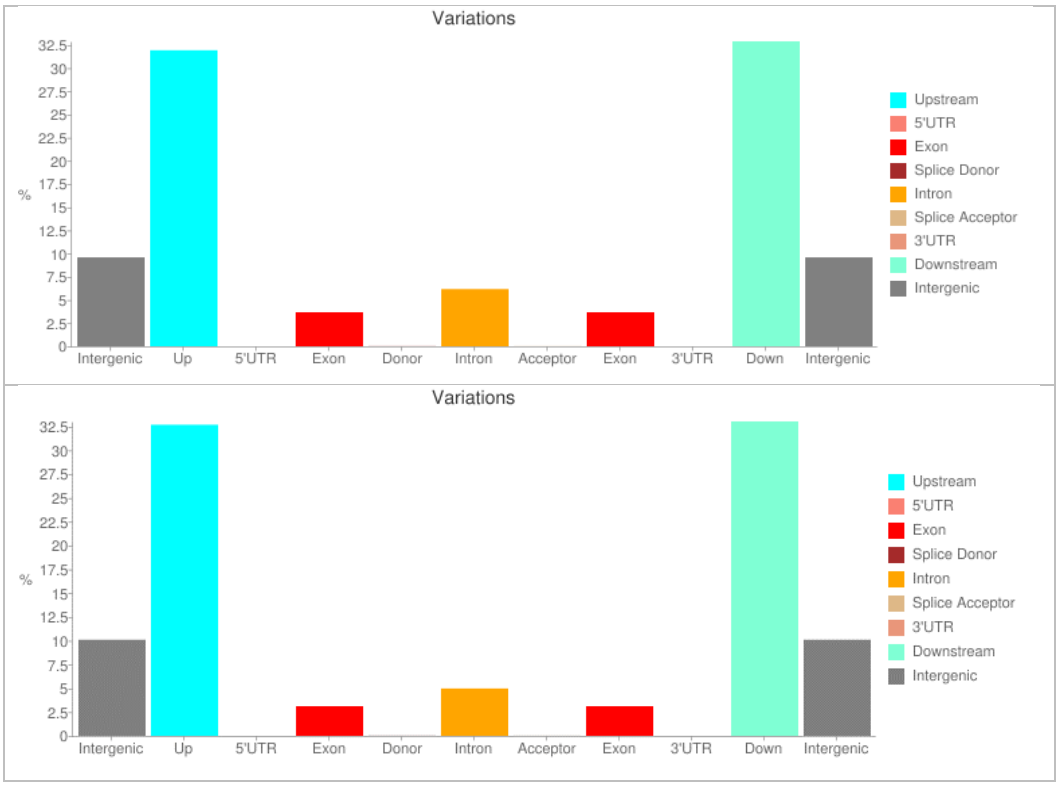


Figure 5-1 Proportion of the effects of the SnpEff-annotated variants by region after mapping against the draft version of the new bigger (top) and the new smaller (bottom) *G. pallida* reference assembly.

For the next analysis step, only variants with a read depth above 10 (i.e. min number of unique reads that include a specific identified variant) and no missing data (i.e. sequenced in all populations) were kept (files abbreviated as *_10_1.vcf). After this first basic filtering step, about 55% of the initial, unfiltered variants were discarded and about 1.8 million remained.

5.3.2. Variant calls identified under different allelic frequencies

In the next step of filtering, we focused on the change of the allele frequency (CAF) between the populations. For this, we set two biological groups; the *S. vernei*-selected populations and the *H3*-selected populations. A CAF threshold of 0.1 (maxCAF1 = 0.1) was applied on the variants of the file '*_10_1.vcf' between the populations evolved on the same resistance source; in other words, a maximum difference of 10% between the allele frequencies (AF) of those populations was set as a toleration threshold on the populations of the same biological group. At this stage, we ensured that populations selected on the same resistance source (or biological group) exhibit the same (or very similar) AF. Afterwards, the CAF parameter between the two different biological groups

was set to the more stringent CAF value (minCAF2 = 0.9) (i.e. the difference in AF between the two groups is at least 90%; alternatively, the populations selected on different resistance sources differ significantly in their AF). Under these parameters (files named as '*_10_1_0.1_0.9.vcf'), 8,803 variant calls were found in the mapping to the bigger assembly from which 5,541 were SNPs. In the mapping to the smaller assembly, 5,579 all-type variants were found with 3,323 identified as SNPs. As expected, when the minCAF2 between the different biological groups was set to the less stringent value of 0.7 (files named as '*_10_1_0.1_0.7.vcf'), the number of variants showed an approximate 4-fold increase. In the mapping to the bigger assembly, 31,729 all-type variants were found from which 18,915 were SNPs, and 26,735 characterised variants in the mapping to the smaller assembly with 14,893 being SNPs (Table 5-2).

Using SnpEff, all the identified variants were annotated depending on their putative impact on the predicted product, as described in chapter 2, and those with high and moderate impact (i.e. missense variants) were chosen. When the variants were filtered using the more stringent minCAF2 = 0.9, 65 all-type variants with high impact were identified in the bigger assembly and 18 in the smaller assembly reference; from those, 15 and 4 were respectively SNPs, 33 and 10 were indels and the rest were complex or a mix of different type changes. Regarding the identified variants with moderate effect (missense variants), 645 and 208 all-type of variants were flagged in the new and small drafts respectively. From those, the majority were SNPs. A more detailed overview is presented in Table 5-2.

Table 5-2 An overview of the total number of variant calls resulting from each variant calling step. The numbers show the identified all-type variants (e.g. SNPs, MNPs, indels, mixed etc) from which number of SNPs only is shown in parentheses. The variants belonging to the file '*_10_1' are filtered variants for min read depth of 10 and no missing data; variants belonging to the file '*10_1_0.1_0.9 (or 0.7)' are further filtered variants for maxCAF1 = 0.1 and minCAF2 = 0.9 (or 0.7). The variants were annotated for high and moderate (i.e. missense variants) putative impact using SnpEff.

	Big assembly	**Small assembly**
Total variants (unfiltered)	3,266,394	3,369,686
***_10_1**	1,636,372	1,814,193
***_10_1_0.1_0.9**	8,803 (5,541)	5,579 (3,323)
high impact	65 (15)	18 (4)
moderate impact/missense	645 (543)	208 (145)
***_10_1_0.1_0.7**	31,729 (18,915)	26,735 (14,893)
high impact	233 (35)	155 (19)
moderate impact/missense	2,469 (1,799)	1,726 (1,177)

5.3.3. Identification of variant genes in the re-sequenced *G. pallida* populations

All the amino acid sequences containing a variant call with high or moderate impact on the predicted gene product (according to the SnpEff annotation) were retained and BLASTp searched. When the stringent minCAF2 value was used (file as '*_10_1_0.1_0.9.vcf'), 65 (all-type) variant calls with high (Table 5-3) and 645 with moderate predicted impact were identified in the mapping of the bigger assembly and 18 (Table 5-4) and 208 (with high and moderate impact respectively) in that of the smaller assembly.

For practical reasons, only the amino acid sequences containing high-impact calls were extracted and BLASTp searched for the current study. Therefore, 65 and 18 candidate (high-impact) variant genes were identified in the bigger and smaller assembly respectively, of which 35 and 11 respectively gave a hit against a gene with known function. For each variant call, the AF (the ratio $'count\ of\ reference\ allele/\ total\ read\ depth'$) value generated by SnpEff was used to estimate the selection preference to each source (see also Table 5-3 and Table 5-4). In the presence of the reference allele only, the AF takes values closer to 1. Likewise, when a variant is called, the AF takes lower values, with the minimum being 0. According to this, most of the variants identified were preferably selected to *H3* resistance source (full lists for variant

calls with both high and moderate impact in Supplementary Data 9 and Supplementary Data 10).

From those, 5 high-impact variant genes were found in common in both assemblies; an ankyrin repeat-containing protein (PF00023), an invected (inv) transcription factor, a putative effector and two SPRY-containing proteins of which one was highly similar to an RBP-1. The protein family with the most hits (6 calls) in the mapping of the bigger assembly was the SPRY family (PF00622). In particular, 4 SPRY-containing proteins were flagged as high-impact variants, of which 3 were similar to RBP-1 and one similar to SPRYSEC-19. About half of the variant calls in both mapping series caused disruption of the translated reading frame (i.e. frameshift variants). Intriguingly, 14 different variant genes in the bigger (and 2 in the smaller) with a BLAST search hit, contained a call that was implicated in gene splicing, either as splice acceptors or donors. Most of them were located in an intron area (either at the start of the 5'-end or 3'-end of it) affecting splicing of the gene. Four calls in the bigger assembly (and 2 in the smaller one) caused loss of a stop codon leading in turn to different gene splicing, such as in that of the gene highly similar to the *SPRYSEC-19*.

When the minCAF2 value was reduced to 0.7 (file as '*_10_1_0.1_0.7.vcf'), another 107 high impact calls in the bigger assembly and 82 in the smaller assembly were identified on top of the previous candidates. In addition to the variants similar to the *RBP-1* that were previously reported, this analysis now showed the presence of another 6 different variant genes highly similar to this member of the SPRY domain family in the mapping of the bigger assembly, and 4 in that of the smaller assembly (Table 5-5) (full lists for variant calls with both high and moderate impact in Supplementary Data 11 and Supplementary Data 12). Interestingly, by decreasing the minCAF2 value from 0.9 to 0.7 in the mapping of the smaller assembly (in the scaffold 5), we were able to identify a gene encoding for an RBP-1-like protein, which had previously been found in the mapping of the bigger assembly (tig00001014) under more stringent minCAF2.

Table 5-3 List of the variants with high impact (according to SnpEff annotation) identified in the mapping of the bigger assembly after filtering them using maxCAF1 = 0.1 and minCAF2 = 0.9 (i.e. file '*_10_1_0.1_0.9.vcf'). The position in scaffold is shown in the first column. The reference (REF) allele and that of the alternate (ALT) allele is indicated in the third column. In the fourth, fifth and sixth column, the type of the sequence ontology (SO) term used for the functional characterisation of the variant gene as well as the functional annotation of the extracted protein sequence resulted from the BLASTp search are shown. In the last two columns, it is indicated where the selection took place (i.e. where the ALT allele was identified). In the cases where multiple alternate alleles were found within the populations of the same biological group, an asterisk (*) has been added. Each biological group (*S. vernei*, *H3*) consists of two populations.

Scaffold	Pos	REF/ALT	Type (SO) term	Description	Transcript	*S. vernei*	*H3*
tig00002413	369606	TA/TCA	upstream_gene_variant	ankyrin repeat protein	g4245.t1	REF	ALT
tig00045762	82148	G/T	stop_gained	collagen alpha-5(IV) chain	g28274.t1	REF	ALT
tig00004958	38992	CTT/CTTT	frameshift_variant	collagen-like protein	g7095.t1	ALT	REF
tig00003567	31916	G/T	splice_acceptor_variant_intron	CRE-STO-5 protein	g5893.t1	REF	ALT
tig00531494	20393	CT/CAAT	frameshift_variant	dorsal gland cell-specific expression protein	g32720.t1	ALT	REF
tig00044628	318780	GCA/GA	frameshift_variant	DUF4982 domain-containing protein	g17309.t1	ALT	REF
tig00045384	171629	CCAG/CCACAG	splice_acceptor_variant_intron	EPHeXin (Eph-interacting GEF) homolog	g25998.t1	ALT	REF
tig00013778	5158	TAACGCGCA/TA, TAACGCGTA	frameshift_variant	eukaryotic aspartyl protease	g11465.t1	ALT*	REF
tig00044960	28333	T/C	stop_lost_splice_region	hypothetical protein LOTGIDRAFT_194748, partial	g21670.t1	REF	ALT
tig00045307	175410	ATAGAT/CCAGAA, CTAGAT	splice_acceptor_variant & missense_variant_splice_region	Hypothetical protein SRAE_2000413500	g25031.t1	REF	ALT*
tig00044865	104086	CTTTTTTTTC/CTTTTTTTC, CTTTTTTC	frameshift_variant	inv protein	g20931.t1	REF	ALT*
tig00045324	28843	GTT/GTC	splice_donor_variant_splice_region_variant_intron	Laminin subunit alpha-2	g25245.t1	ALT	REF
tig00529538	1661	CGATTTTTA/CTATTTA, CGATTTA	splice_acceptor_variant_splice_region_variant_intron	Lipid phosphate phosphohydrolase 1	g30362.t1	ALT*	REF

tig00001014	129531	AGA/AA, AAA	frameshift_variant	---NA---	g2264.t1	REF	ALT*
tig00001014	129686	GAT/GAAA	frameshift_variant_missense	---NA---	g2264.t1	REF	ALT
tig00002059	23988	ATTT/GTTC, GTTT	start_lost	---NA---	g3928.t1	ALT*	REF
tig00002413	25510	T/C	stop_lost_splice_region	---NA---	g4153.t1	REF	ALT
tig00002413	268405	GAA/GA	frameshift_variant	---NA---	g4226.t1	REF	ALT
tig00002660	69281	GTTTTTA/GTTTTTTA	splice_acceptor_variant_intron	---NA---	g4661.t1	ALT	REF
tig00003856	23513	CAATAATA/CAATA	splice_donor_variant_splice_region	---NA---	g6243.t1	ALT	REF
tig00005079	119827	GCCAGAAGTCTCTTCT/GTCAAACGTCTCTGCC, GCCAAAAGTCTCTTCT	frameshift_variant	---NA---	g7274.t1	REF	ALT*
tig00006550	3076	GTTTTTTC/GTTTTTC	frameshift_variant	---NA---	g8560.t1	REF	ALT
tig00007011	138034	TCA/TA	frameshift_variant	---NA---	g8948.t1	REF	ALT
tig00007500	149991	TGA/TA	frameshift_variant	---NA---	g9305.t1	REF	ALT
tig00009681	5106	CTGG/TTGT, TTGG	stop_gained	---NA---	g10603.t1	ALT*	REF
tig00015915	23169	AAA/AAGAA, AAGA, AAG	frameshift_variant	---NA---	g11802.t1	REF	ALT*
tig00016675	8606	G/A	splice_donor_variant_intron	---NA---	g11854.t1	REF	ALT
tig00044435	62723	TTAGT/CTAGGT, TTGGT	frameshift_variant_stop_lost_stop_retained_variant_splice	---NA---	g13754.t1	REF	ALT*
tig00044435	63329	CCTC/CT	frameshift_variant	---NA---	g13755.t1	REF	ALT

tig00044517	720816	ATGTAACATGTA/AATGA	splice_donor_variant_intron	---NA---	g15580.t1	ALT	REF
tig00044586	35190	TA/TCA	splice_acceptor_variant_intron	---NA---	g16470.t1	ALT	REF
tig00044828	142918	GC/GTC	frameshift_variant	---NA---	g20566.t1	ALT	REF
tig00044828	201234	ATT/AT	frameshift_variant	---NA---	g20586.t1	ALT	REF
tig00045156	28129	A/T	stop_gained	---NA---	g23778.t1	REF	ALT
tig00045206	239956	CTT/CT	frameshift_variant	---NA---	g23898.t1	REF	ALT
tig00045211	7853	AC/ATC	frameshift_variant	---NA---	g23934.t1	REF	ALT
tig00045412	8015	CTG/CG	frameshift_variant	---NA---	g26121.t1	ALT	REF
tig00045415	103534	CAAATA/CA	frameshift_variant	---NA---	g26236.t1	ALT	REF
tig00045605	18506	GCA/GCCC	frameshift_variant_missense	---NA---	g27384.t1	REF	ALT
tig00045934	36782	AG/ACT	frameshift_variant_missense	---NA---	g28828.t1	ALT	REF
tig00529214	339368	GAT/GC	frameshift_variant_missense	---NA---	g29694.t1	ALT	REF
tig00044435	48758	C/T	stop_gained	nuclear pore complex protein NUP98A-like isoform X1	g13748.t1	REF	ALT
tig00045298	80183	TAA/TA	splice_acceptor_variant_intron	PAN domain protein	g24786.t1	REF	ALT
tig00045320	205703	C/T	splice_acceptor_variant_intron	Poly [ADP-ribose] polymerase 2	g25227.t2	REF	ALT
tig00529969	148449	ATA/AA, TAA	frameshift_variant	poly(A) polymerase beta	g31225.t1	REF	ALT*
tig00045307	294079	CCTCGCTC/ACTCCCTT, CCTCGTTC	stop_gained	Protein CBG14625	g25059.t1	REF	ALT*
tig00004728	23232	GTAGT/AAAGG	splice_acceptor_variant	protein DENND6A-like isoform X4	g6958.t1	ALT	REF
tig00007498	22543	TCACCCACCA/TCACCCACCCACCA	frameshift_variant_splice_region	protein FAM49A isoform X3	g9263.t2	REF	ALT
tig00044435	37041	AGG/AG	frameshift_variant	protein fem-1 homolog CG6966 isoform X2	g13746.t1	REF	ALT

tig00045298	55143	GTTTTTTG/GTTTTTG	frameshift_variant	protein starmaker-like isoform X1	g24782.t1	REF	ALT
tig00044587	352155	TAA/TG	frameshift_variant_synonymous	protein-arginine deiminase	g16529.t1	ALT	REF
tig00045232	135716	T/G	stop_lost_splice_region	putative effector protein	g24147.t2	REF	ALT
tig00003648	42265	CT/CATT	frameshift_variant	RanBPM-like protein	g5955.t1	REF	ALT
tig00001014	128579	T/C	splice_donor_variant_intron	RBP-1 protein	g2263.t1	REF	ALT
tig00001014	129077	A/C	splice_acceptor_variant_intron	RBP-1 protein	g2263.t1	REF	ALT
tig00001014	129167	ATT/ATTT	frameshift_variant_stop_gained	RBP-1 protein	g2263.t1	REF	ALT
tig00044964		AACA/AA	frameshift_variant	RBP-1 protein	g21717.t1	REF	ALT
tig00530659	30136	G/A	splice_donor_variant_intron	RBP-1 protein	g31725.t1	REF	ALT
tig00001992	26191	TGTA/GGTG	splice_donor	RNA binding repeat protein, Pumilio-family	g3822.t1	ALT	REF
tig00001992	75277	GCG/GG	frameshift_variant	RNA binding repeat protein, Pumilio-family	g3828.t1	ALT	REF
tig00529478	99720	C/A	stop_lost_splice_region	secreted SPRY domain-containing protein 19	g30331.t1	REF	ALT
tig00045350	47654	AATG/AG	frameshift_variant	transmembrane protein	g25469.t1	REF	ALT
tig00045414	138124	GTTTTTTA/GTTTTTA, GTTTTTTTA	splice_acceptor_variant_intron	von Willebrand factor A domain-containing protein 7-like isoform X1	g26212.t1	REF	ALT*
tig00013045	12087	GTT/GT	frameshift_variant	zinc finger BED domain-containing protein 1-like	g11378.t1	REF	ALT
tig00045305	58184	ACCCCCCA/ACCCCCCA	splice_donor_variant_intron	Zonadhesin	g24914.t1	ALT	REF

Table 5-4 List of the variants with high impact (according to SnpEff annotation) identified in the mapping of the smaller assembly after filtering them using maxCAF1 = 0.1 and minCAF2 = 0.9 (i.e. file '*_10_1_0.1_0.9.vcf'). The position in scaffold is shown in the first column. The reference (REF) allele and that of the alternate (ALT) allele is indicated in the third column. In the fourth, fifth and sixth column, the type of the sequence ontology (SO) term used for the functional characterisation of the variant gene as well as the functional annotation of the extracted protein sequence resulted from the BLASTp search are shown. In the last two columns, it is indicated where the selection took place (i.e. where the ALT allele was identified). In the cases where multiple alternate alleles were found within the populations of the same biological group, an asterisk (*) has been added. Each biological group (*S. vernei*, *H3*) consists of two populations.

Scaffold	**Position**	**REF/ALT**	**Type (SO term)**	**Description**	**Transcript**	***S. vernei***	***H3***
scaffold5	215261	CTTTTTTTTTC/CTTTTTTTT, CTTTTTTTC	frameshift_variant	inv protein	g37.t1	REF	ALT*
scaffold1	1306744	C/T	splice_donor_variant_intron_variant	---NA---	g2648.t1	REF	ALT
scaffold1	5027715	GAA/GA	frameshift_variant	---NA---	g3215.t1	ALT	REF
scaffold1	5085172	GC/GAC	frameshift_variant	---NA---	g3236.t1	ALT	REF
scaffold2	903710	TAA/CAC	upstream_gene_variant	---NA---	g3423.t1	REF	ALT
scaffold2	5996628	TGTGCC/CGTGT, TGGGC	frameshift_variant_missense_variant	predicted protein	g4207.t1	ALT*	REF
scaffold3	1186724	TA/TCA	frameshift_variant	ankyrin repeat protein	g4476.t1	REF	ALT
scaffold6	46136	AGG/AGGG, AGGGGAGGG	frameshift_variant	shTK domain protein	g5825.t1	REF	ALT*
scaffold27	282211	TCA/TCCA	frameshift_variant	PR domain zinc finger protein 14-like	g9989.t1	REF	ALT
scaffold36	1144393	T/C	splice_donor_variant_intron_variant	RBP-1 protein	g11607.t1	REF	ALT
scaffold69	508750	T/G	stop_lost_splice_region_variant	putative effector protein	g15196.t1	REF	ALT
scaffold73	90756	TACACA/TA, TATTTA	frameshift_variant	AP-1 complex subunit beta-1 isoform X2	g15596.t1	REF	ALT*

scaffold73	293695	CTC/CC	frameshift_variant	---NA---	g15649.t1	ALT	REF
scaffold80	307995	CC/CGT	frameshift_variant_synonymous_variant	autotransporter domain-containing protein	g16162.t1	REF	ALT
scaffold86	174638	GCA/GCCC	frameshift_variant_missense_variant	---NA---	g16555.t1	REF	ALT
scaffold89	308438	CTT/CT	frameshift_variant	---NA---	g16722.t1	REF	ALT
scaffold105	9726	TCCCCCCCCCA/TCCCCCCCCCCCA	splice_acceptor_variant_intron_variant	Protein CBG05069	g17442.t1	REF	ALT
scaffold114	99710	C/A	stop_lost_splice_region_variant	secreted SPRY domain-containing protein 19	g17963.t1	REF	ALT

Table 5-5 List of the additional variant genes similar to *RBP-1* found in both assembly references using the less stringent minCAF2 = 0.7. In the last two columns, it is indicated where the selection took place (i.e. where the alternate-ALT allele was identified).

Assembly	Scaffold	Position	Transcript	*S. vernei*	*H3*
Bigger/Smaller	tig00044372/scaffold2	64976/977634	g12758.t1/ g3437.t1	REF	ALT
Bigger/Smaller	tig00044372/scaffold2	71575/984233	g12759.t1	REF	ALT
Bigger	tig00044431	157425	g13719.t1	REF	ALT
Bigger	tig00044964	104094	g21717.t1	REF	ALT
Bigger	tig00044964	104797	g21717.t1	REF	ALT
Bigger	tig00045365	10353	g25577.t1	ALT	REF
Bigger	tig00529214	262458	g29678.t1	ALT	REF
Smaller	scaffold5	151520	g29.t1	REF	ALT
Smaller	scaffold3	1014365	g4426.t1	REF	ALT
Smaller	scaffold114	170320	g17977.t1	REF	ALT
Smaller	scaffold114	170337	g17977.t1	REF	ALT

5.3.4. The identified SNPs are possibly organised in "islands"

The physical position of the SNPs identified using minCAF2=0.9 (variants '*_10_1_0.1_0.9') was visualised on both draft assembly references. The SNPs were allocated in 534 scaffolds of the bigger assembly (out of 1923) and 114 scaffolds of the smaller (out of 267). Many identified SNPs were located in a small region of a scaffold, next to other SNPs, forming SNP "islands". Characteristic examples are the scaffolds tig00044435 and tig00531427 of the bigger draft assembly (Figure 5-3) and the scaffolds 14 and 15 of the smaller draft assembly (Figure 5-4). On the other hand, several SNPs were scattered along the genome, such as those found in the scaffolds tig00001809 and 12 of the big and smaller draft assembly respectively, as well as scaffolds with a proportionally large number of SNPs relative to their total size (e.g. scaffolds tig00044372 and tig00004674 of the bigger assembly).

Most of the identified SNPs in the mapping of the bigger assembly were found in the scaffolds tig00531427 and tig00044423 with 145 and 139 SNPs respectively. As can be seen from Figure 5-2, even though the number of the calls was similar in both scaffolds, the ratio '$number\ of\ called\ SNPs/scaffold\ length$' was significantly different, with the first case almost doubled. Moreover, the called SNPs in the scaffold tig00044423 are more uniformly located along its whole length, contrary to the scaffold tig00531427 where a main SNP "island" seems to be formed at the beginning of the scaffold. Within the

335,567bp of the total length, 121 SNPs were identified in the first 92kbp-long region, from which 93 caused non-synonymous amino acid changes. From those in turn, 14 SNPs had moderate functional impact affecting the function of the start or stop codon of the predicted product. Notably, all of these moderate-impact SNPs, along with another 90 (out of the 121) in that scaffold region were mostly with *S. vernei*-selected populations (i.e. null frequency of the reference allele). In that region 6 different non-synonymous SNPs were identified on the intragenic region of a gene that encodes for dorsal gland cell-specific proteins of *Heterodera* origin.

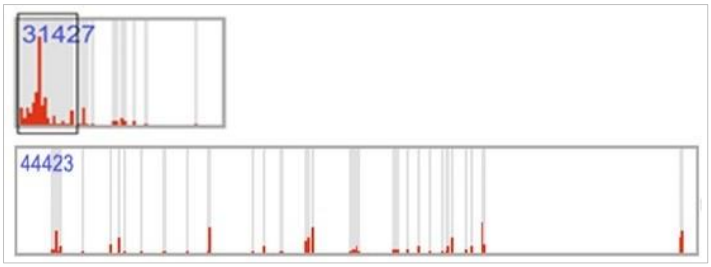


Figure 5-2 Comparison of the SNP distribution. The comparison is between the scaffolds tig00531427 and tig00044423 in the mapping of the bigger assembly when maxCAF1 = 0.1 and minCAF2 = 0.9. In the first case, the 'number of SNPs/scaffold length' is about $4.3e^{-4}$ whereas $8.6e^{-5}$ in the second case meaning that even the number of called SNPs are approximately similar to both scaffolds, the SNPs density is significantly higher in the first scaffold. The black area shows the presence of the 92kb-long SNP island (taken from Figure 5-3).

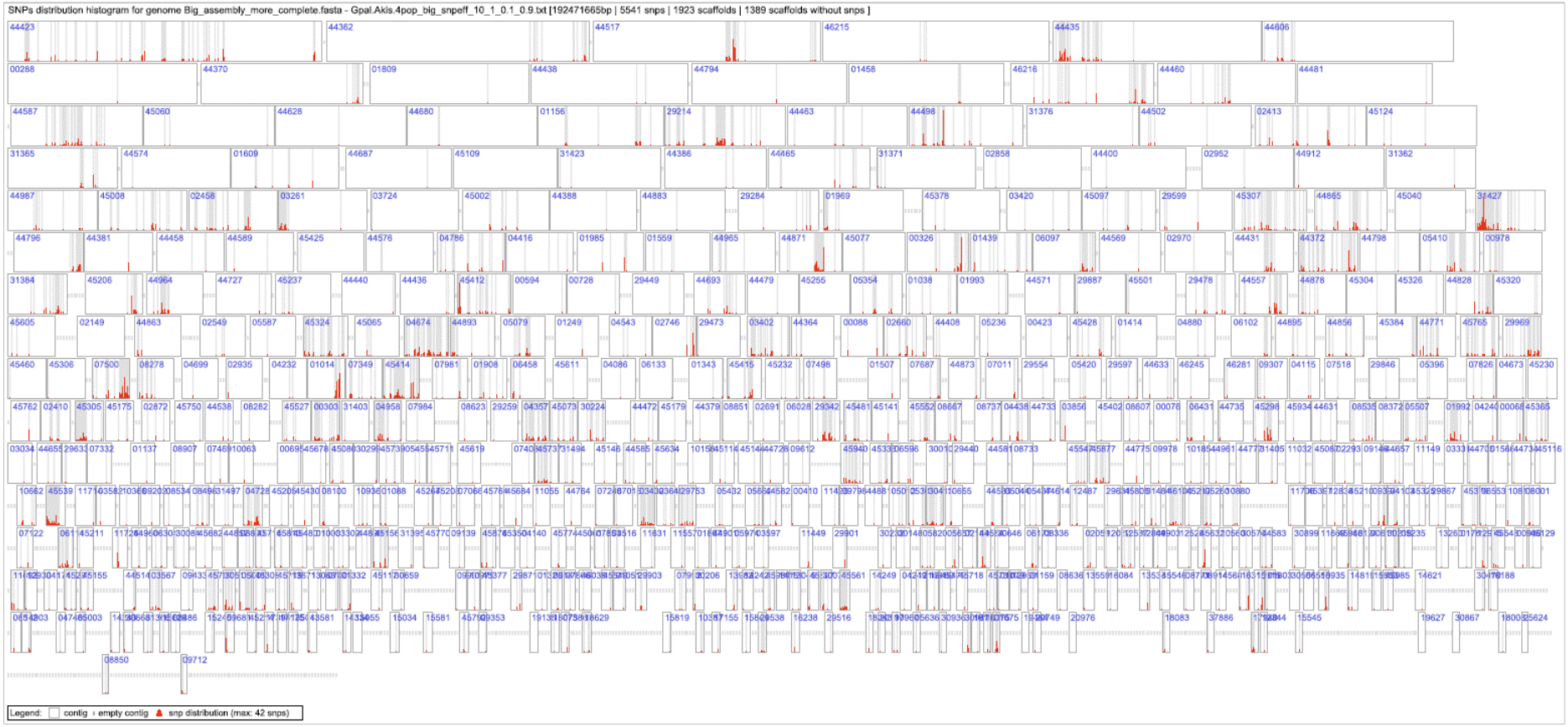


Figure 5-3 Distribution histogram of the identified SNPs in the mapping of the new/bigger assembly. The SNPs were identified using maxCAF1 = 0.1 and minCAF2 = 0.9. In total, 5,541 SNPs (red areas) were distributed along 1923 in total scaffolds of the draft assembly. The 1389 scaffolds without any SNP are not shown in the figure.

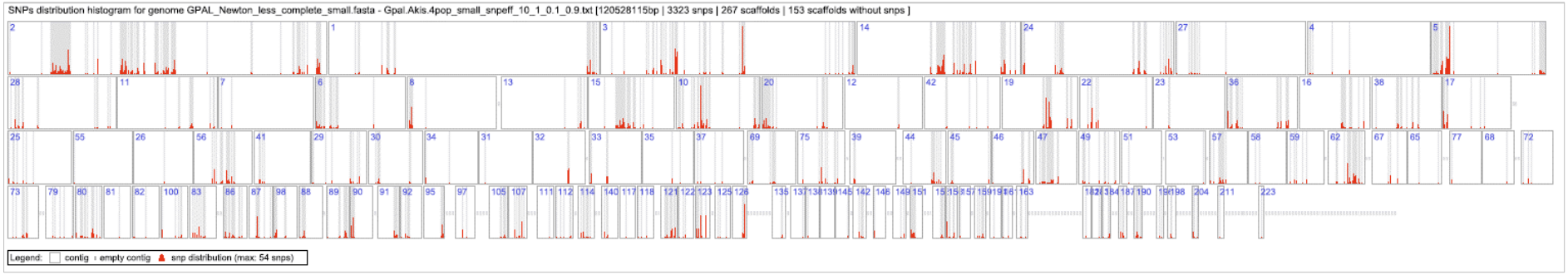


Figure 5-4 Distribution histogram of the identified SNPs in the mapping of the new/smaller assembly. The SNPs were identified using maxCAF1 = 0.1 and minCAF2 = 0.9. In total, 3,323 SNPs (red areas) were distributed along 267 in total scaffolds of the draft assembly. The 153 scaffolds without any SNP are not shown in the figure.

Similarly, another "island" with 70 SNPs with the *S. vernei*-selected populations on was identified in the scaffold tig00044517 (approx. positions 685kbp–750kbp). SNPs specifically selected on *H3* were observed in the "islands" of the scaffold tig00044435 (58 SNPs located between 18kbp – 92kbp and 25 SNPs in the region 150kbp – 250kbp), tig00044871 (71 SNPs in the region of 173kbp – 220kbp) and tig00044796 (38 SNPs in the region 276kbp – 331kbp). Regarding the mapping of the smaller assembly, typical examples of SNP "islands" specifically to the *S. vernei* selection were located in the scaffolds 14 (106 SNPs approx. between 1,425kbp – 2,054kbp) and 15 (71 SNPs approx. between 500kbp – 1,030kbp), whereas "islands" exclusively selected towards *H3* were discovered in scaffolds 2 (33 SNPs between 838kbp – 970kbp), 5 (144 SNPs between 2,135kbp – 2,160kbp) and 19 (87 SNPs between 760kbp – 930kbp).

5.3.5. Identification of variant genes selected differently to both resistance sources

As a last step of this chapter, it was checked whether variant calls in the same identified annotated gene (or neighbouring genes) were selected towards the different resistance sources at the same time. For this, the R-based custom script 'compare_diff_entre_pop' was run on the variants '*_10_1_allvar' in 4 different combinations for each mapping. In each case, every selected population was used as a single biological group on its own, and the other three (regardless of their selection source) were grouped together into the second biological group. The minCAF2 was set to the value of 0.7.

In the scaffold 3 (total size of approximately 4.9million bp) of the small assembly, two variants (SNPs), with a distance of about 144Kbp, were selected differently within the *H3*-selected populations. The first SNP (position 849,981) was identified on the transcript *g4387*.t1 of a gene encoding an RBP-1 protein, and the second SNP (position 997,297) on a gene encoding a zinc-containing metalloproteinase similar to dpy-31. Some genes that were found within that region encode for known effector proteins in *G. pallida*, such as RBP-1, SPRYSECs and an ankyrin repeat protein that was also found in the previous variant calling analyses (Figure 5-5). However, since the two markers are located quite far apart from each other, the specific finding is possibly a result of a random event rather than specific.

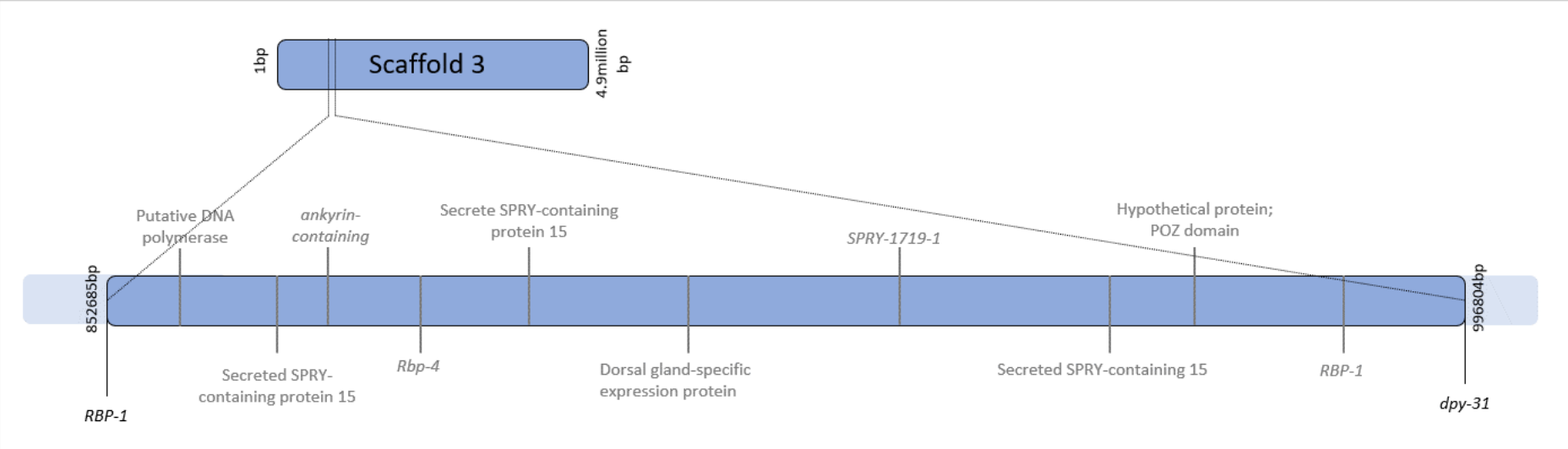


Figure 5-5 Possible selection differentiation of identified genes. The zoomed in area of the scaffold 3 of the smaller assembly where the two identified genes (*RBP-1* and *dpy-31*) that are differently selected within *H3*-selected populations were identified. Within this region, all the neighbouring annotated genes are shown.

5.4. Discussion

Since introduction into Europe, *G. pallida* populations have spread establishing trade-off relations with host plants in terms of virulence. Several artificially-selected populations have increased virulence towards specific resistance sources due to the availability of *vir* alleles in their gene pool (Phillips and Blok, 2008; Turner and Fleming, 2002). This large pool allows them to genetically shift in favour of adaptation to resistances upon strong selection pressure that is continuously applied (Eoche-Bosy et al., 2017a; Phillips and Blok, 2008). Several populations of *G. pallida* that have overcome resistance from *S. vernei* have been reported from fields of starch producers in continental Europe (J. Jones, personal communication). This host adaptation can also be depicted in the genomic background of these populations, and their study can be crucial for the better understanding of nematode virulence, its molecular mechanisms and how this relates to the durability of a resistance.

Here, we used *G. pallida* populations subjected to strong selection pressure that have increased virulence against two different resistance sources. It is likely that these *vir* alleles are already present in the initial genetically diverse "founder" population but at very low frequencies; the continuous selection process then could result in the flow of those alleles within the population and eventually in their over-representation in the final selected populations. It is also possible that novel mutations could also occur during the selection process (Fournet et al., 2016; Fournet et al., 2013; Luikart et al., 2003). In the current chapter, the potential of these selected nematode populations and modern genomic tools to study these regions and gain a better insight into selection of *G. pallida* is investigated. On top of the PenSeq used in chapter 4 for the identification of *Avr* genes, the approach used here constitutes a second, alternative approach to this goal and it has been previously used on PCN population genetics studies (Eves-van den Akker et al., 2016; Mimee et al., 2015).

Contrary to the more targeted approach of PenSeq used in the previous chapter, here we performed genome scanning of selected nematode populations. With this approach, *loci* and genomic regions in general that may play important role in virulence and host adaptation were not excluded from the analysis pipeline. This approach also allowed

identification of genomic regions that have become highly modified following selection. NGS was performed on the two populations selected on *S. vernei* and two selected on *H3*, which had been generated during the screening tests of the current study (see also chapter 3). Then, variant calling analyses was used to identify markers that distinguished between them. Similar to the downstream analyses done in the chapter 4, here we used the two draft versions of the new *G. pallida* genome assembly as references.

In our pipelines, we kept variants that exhibit similar AF when selected on a specific resistance source and show high or moderate impact on the predicted product. We analysed all-type of variants (e.g. SNPs, indels, mixed) as well as SNPs more specifically. In the mapping to the bigger assembly, 710 moderate- and high-impact variants were identified when the more stringent minCAF2 value was used (i.e. minCAF2 = 0.9) and 2,702 variants when minCAF2 was reduced to 0.7. When the smaller assembly was used as reference, 226 and 1,881 variants respectively were identified. The identified markers were distributed in 191 and 48 scaffolds of the bigger and smaller draft assemblies respectively, while the distribution width was doubled in the more relaxed minCAF2 value.

In Table 5-3 and Table 5-4, the type the of marker identified based on Sequence Ontology (SO) is shown (Eilbeck et al., 2005). Most of the high-impact variants disrupted the translation of the reading frame (frameshift variant) or they caused a change of the start/stop codon. Our analyses (when the less stringent minCAF2 parameter was used) also showed that a significant number of genes (70 and 30 in the bigger and smaller assembly respectively) contained variants that occurred in putative splice sites, either close to the exon start (splice acceptor site) or exon end (splice donor site). Another 77 moderate-impact variants in the bigger assembly and 52 in the smaller were found close to splice sites (close to the intron ends).

Alternative splicing is an interesting phenomenon where multiple mRNA transcripts are generated from a single precursor gene. This leads to different proteins or protein isoforms from a single gene. This phenomenon is a common mechanism to generate genomic diversity in many eukaryotic organisms. Another interesting feature of this mechanism is the fact that its products can differ in function or cellular localisation

(Staiger and Brown, 2013). One chorismate mutase gene in *G. rostochiensis* encodes for two different transcripts as a result of alternative splicing (Lu et al., 2008; Yu et al., 2010). There is also evidence for alternative splicing in many identified genes that are similar to the *Heterodera* known effectors *4D06* and *G20E03* (Thorpe et al., 2014). Although our data indicate that a significant number of variants occurred in putative splice sites, further research on mRNA levels is essential along with RNA-seq analysis on the generated number of transcripts.

SPRYSEC proteins were the most abundant protein family found in our analyses and BLASTp searches showed that most of the flagged proteins were highly similar to the SPRY-containing RBP-1 proteins, which as described in previous chapters constitute known effectors in PCNs (Diaz-Granados et al., 2016; Sacco et al., 2009). It is known that the SPRY domain is structurally highly versatile and genetically diverse. It is believed that this large diversity has been exploited by *G. pallida* to expand its pool of effectors during positive selection events and ultimately avoid host recognition (Carpentier et al., 2012; Cotton et al., 2014; Rehman et al., 2009). From the high-impact SPRY proteins identified in total, about half of them contained variants on splice sites (scaffolds tig00001014, tig00044964, tig00529214, tig00530659 and tig0052947 of the bigger assembly and 36 and 114 in the smaller assembly). During the selection process, many gene (and effector) families are expanded in order to avoid host recognition and this is believed to be the case in SPRY domain proteins too (Cotton et al., 2014). This gene family expansion may also be facilitated by alternative splicing phenomena in order to create new gene products that are able to break and overcome resistance barriers, but more research is needed to look at the mRNAs in the selected populations.

The ubituitin ligase, a protein that also works as an effector in PCNs and has been reported as a suppressor of plant immune responses in *G. rostochiensis* was also listed as a candidate *Avr* gene (Haegeman et al., 2012). Protein ubiquitination has been reported as a mechanism exploited by plant pathogens to suppress host defences (Birch et al., 2009). Our analysis identified proteins with possible participation in protein ubiquitination, such as BTB/MATH domain-containing protein. Previously reported PCN effectors were also present, such as two dorsal gland-expressed proteins similar to

that of the *Heterodera* genus, as well as a member of the von Willebrand factor domain (PF00092) that may act as a body integrator in *C. elegans* and has also been reported in *G. pallida* by Thorpe et al. (2014). Both candidate families were also reported as candidate effectors in the downstream analyses in chapter 4.

Ankyrin repeat (PF00023) appears in many bacterial and eukaryotic organisms and plays a role in protein-protein interactions and protein recognition (Mosavi et al., 2004). Even though their function is not well understood, it is believed that members of this family are required for the moulting process and cell fate in *C. elegans* (Lažetić and Fay, 2017). Similarly, the RNA-binding proteins (PF00806) constitutes a large and diverse gene family in nematodes as well as other organisms. Zinc finger-like proteins also belong to this family and were also found as candidate effector genes in our analyses in chapter 4. Here, a specific RNA-binding family was identified, the so-called Pumilio family. Like the zinc finger-like proteins, members of the Pumilio family also mediate and regulate mRNA trafficking, protein-protein interactions, protein folding and cytoskeleton organisation (Kaymak et al., 2010; Klug, 2010; Parisi and Lin, 2000).

One of the advantages of the genome-wide re-sequencing method is the study of allele frequencies from multiple virulent and avirulent populations with reliability (Mimee et al., 2015). Before filtering the initial variants based on CAF parameters, we removed all of these with low-frequency and read depth (Ferretti et al., 2013). In this way, we increased the reliability of our results by eliminating any variants that arose due to sequencing errors. Additionally, the use of two independent populations per selection increased the confidence level of the outcome.

The majority of the variant calls were selected towards *H3* in both mappings and filtering parameters. In the more stringent filtering parameters (i.e. minCAF2 = 0.9), about 51.1% of the variants in the bigger assembly and 63% of the smaller were found in *H3*-selected populations. When the filtering parameters were relaxed, both fractions came into similar levels (~52%). Dorsal gland-expression genes were selected towards both resistances (13 identified genes selected towards *S. vernei* and 11 towards *H3*), RBP-1-encoding genes were mainly selected towards *H3* (57 in comparison to 22 that were selected to *S. vernei*), while the two pectate lyases genes were selected to *S. vernei* only.

The selection processes on a specific resistance causes the selection of nematode populations containing variants that facilitate overcoming the continuous applied resistance. This is the case for both *S. vernei* and *H3* with the majority of the observed changes found in the populations selected in the second resistance source. This can be linked with either the fact that *H3* might show higher levels of resistance and hence it creates stronger selection pressure on the evolving populations (Phillips and Blok, 2008), or the speculation that *H3* is genetically less complicated (e.g. only a single QTL) and hence easier to be overcome. Another alternative explanation could be the fact that some essential effectors or part of them, are more difficult to subject to structural changes; and this might be the case on *S. vernei*. Fournet et al. (2016) has also shown that adaptation of experimentally evolved populations on resistant plants affects the phenotype (and fitness) of the virulent populations after several successive generations compared to the avirulent parent population.

In several plant pathogens, the effectors are located in specific genomic regions organised in "islands", for example *P. infestans*. This genomic organisation accelerates genomic plasticity and host adaptation through the development of expanded gene families in these regions (Raffaele et al., 2010). Eves-van den Akker et al. (2016) reported that effectors in both *G. rostochiensis* and *G. pallida* are clustered together into effector "islands" in specific genomic regions. Analysis of these regions can allow the identification of *loci* associating with the adaptation of *G. pallida* populations on resistances (Eoche-Bosy et al., 2017b; Gendron St-Marseille et al., 2018). In the present study, we created a physical map of SNP distribution along the two references and we identified the presence of SNP "islands" where hundreds of high-impact SNPs were allocated in small regions that contain multiple genes encoding candidate effectors. It needs to be pointed out that only SNP variants were used for the identification of the "islands". Research should also be expanded towards "islands" containing other than SNP variants too. More research and analysis are needed, since only the physical "islands" located in scaffolds with many SNPs were also examined.

To conclude, this chapter aimed at the identification of genetic variants linked to adaptation to the resistance sources *S. vernei* and *H3*. Using different filtering

stringencies, we identified potential candidates. Many of them are very similar to the candidate effector list generated in target enrichment sequencing (see chapter 7). Apart from this, we suggest that the majority of the variant calls and genes are preferably selected in *H3*-selected populations and might have resulted from the stronger selection pressure and/or different genetic basis of the *H3* exerted selection. Since these analysis pipelines were applied on draft versions of the new and improved *G. pallida* genome assembly, it will be very interesting for these to be repeated in the final version.

6. Functional validation of *G. rostochiensis* candidate *avirulence* genes

6.1. Background

6.1.1. Functional validation of candidate *avirulence* genes

According to the 'gene-for-gene' model, the recognition of a plant pathogen by a plant is based on the interaction between the products of genes derived from both sides (Dangl et al., 2013; Jones and Dangl, 2006). This interaction results in the recognition of a pathogen-derived *avirulence* (*Avr*) gene by a cognate host-derived *R* gene. This theory was then expanded with the 'guard hypothesis' that involves indirect recognition of *Avr* genes. Small changes in the amino acid sequence of putative effectors can cause loss of their function and may turn them into *Avr* genes. For example, the *Phytophthora infestans* RXLR effector AVR3a has two forms (alleles) differing by two amino acids. The $AVR3a^{KI}$ is recognised by the potato resistance protein R3a leading to resistance responses activation, whereas $AVR3a^{EM}$ does not. At the same time, $AVR3a^{KI}$ also suppresses hypersensitive response (HR) triggered by the *P. infestans* elicitin INF1 more strongly than $AVR3a^{EM}$ (Bos et al., 2010; Bos et al., 2009). This suggests a delicate balance of the effectors to preserve their function and avoid recognition by host resistance proteins.

In plants, a common system used for studying and validating a candidate *Avr* gene is by transferring it into the soil bacterium *Agrobacterium tumefaciens* (or *Rhizobium radiobacter*, as it has been recently renamed) through cloning and expressing it transiently *in planta*. The delivery and expression of the gene-of-interest into the host plants is mediated by the bacterial Ti (tumour-inducing) plasmid that contains a T-DNA that in turn is able to be translocated and incorporated into the host transcriptional and translational machinery during infection (Du et al., 2014; Kapila et al., 1997; van der Hoorn et al., 2000).

6.1.2. *G. rostochiensis* candidate effectors

In Chapter 1, it was described how *G. rostochiensis* has been successfully controlled with potato cultivars containing the major *R locus H1*. Of the 5 different pathotypes that have been identified for *G. rostochiensis* (Ro1 to Ro5), Ro1 is the most abundant in the UK and

continental Europe. *H1*-containing varieties can successfully fully control the *G. rostochiensis* pathotypes Ro1 and Ro4, while the pathotypes Ro2, Ro3 and Ro5 are able to overcome *H1*-mediated resistance (Eves-van den Akker et al., 2016; Finkers-Tomczak et al., 2011; Kort et al., 1977). Generation of a high-quality draft genome assembly of *G. rostochiensis* (pathotype Ro1) (Eves-van den Akker et al., 2016) allowed identification and study of putative effectors of this PCN species.

Eves-van den Akker et al. (2016) analysed data from different population of all five *G. rostochiensis* pathotypes that had been re-sequenced and mapped against the reference genome assembly (pathotype Ro1). In total, 1,081,802 variants were found, which from those the majority (approximately 80%) were SNPs and the rest consisted of indels (inserts or deletions). In total 190 genes had potential for modified or loss of function between the avirulent and virulent populations; in these the variant alleles were homozygous absent in all the avirulent populations and homozygous or heterozygous present in all the virulent populations. When these 190 genes were then cross-referenced with the effector list, it was shown that only two of them contained a signal peptide (SP) and encoded for putative proteins similar to those on the effector list. These two genes (GROS_*g13394* and GROS_*g12477*) were therefore characterised as candidate *Avr* genes (Eves-van den Akker et al., 2016). The gene GROS_*g13394* encodes a putative cellulose binding protein produced in the subventral gland cells and its sequence is similar to *GLAND10*, which is expressed in the esophageal glands of the pre-parasitic *H. glycines* J2s; GROS_*g12477* encodes an ubiquitin-like protein produced in the dorsal gland cells and is similar to a protein found in the parasitic *G. rostochiensis* J2s involved in suppression of immune responses (Chronis et al., 2013; Eves-van den Akker et al., 2016; Noon et al., 2015).

6.2. Chapter objective

The objective of this chapter was to amplify the candidate *G. rostochiensis Avr* genes from two pathotypes; one that is recognised (avirulent) by the resistance source *H1* (Ro1) and the virulent to *H1*, Ro5. Then, the sequences of the genes amplified from these two pathotypes were analysed for possible differences that may cause alteration in virulence

activity. An *A. tumefaciens* transient expression system was used in order to transfer and express the genes into potato plants for functional validation.

6.3. Results

6.3.1. DNA extractions from the *G. rostochiensis* populations Ro1 and Ro5

DNA was extracted from 10 individual cysts from *G. rostochiensis* pathotypes of either Ro1 or Ro5 separately. Since the discrimination of the cysts of *G. pallida* with those of *G. rostochiensis* is impossible using a microscope (other than for a trained specialist) and due to the co-existence of both species in many fields around the UK, a diagnostic PCR was carried out on all the extracted DNA samples using species-specific primers. Initially, DNA extraction was performed for each single cyst separately. Diagnostic PCR was then performed to confirm no cross-contamination between the two species.

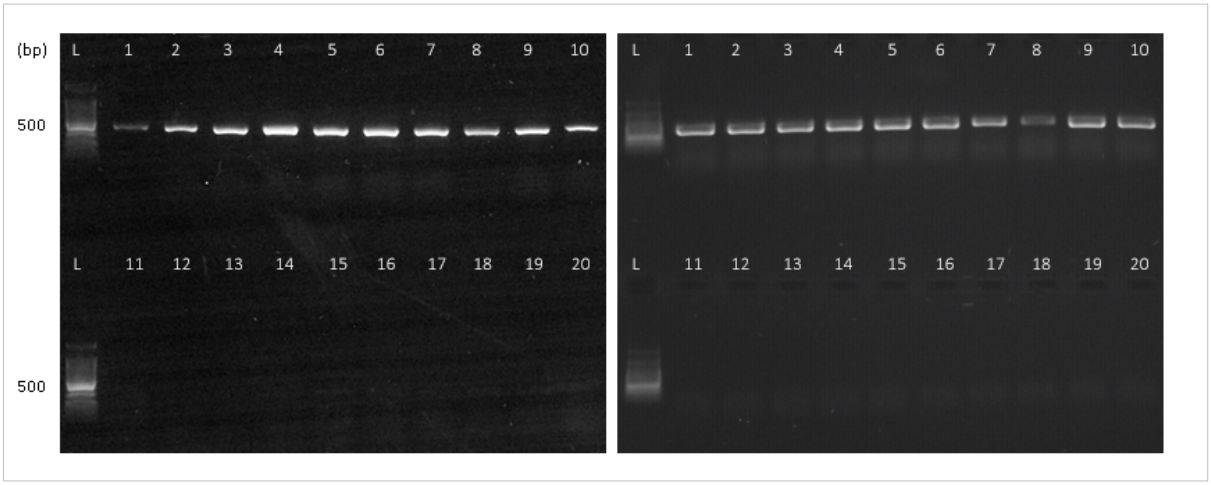


Figure 6-1 Diagnostic PCR of the cysts used for DNA extractions to confirm *G. pallida*-free DNA. In the lanes 1-10 the primers 18S-UNI and PITSr3 were used to amplify DNA from *G. rostochiensis*, whereas the lanes 11-20 the primers 18S-UNI and PITSp4 for amplifying DNA from *G. pallida* were used. In the left agarose gel, template from Ro1 were used while Ro5 on the right one. Amplification of DNA was only observed in the *G. rostochiensis*-specific primers.

As the Figure 6-1 shows, in both pathotypes only the primers set 18S-UNI and PITSr3 that are *G. rostochiensis*-specific amplified a fragment with the expected size of 500bp. This showed that indeed, all the cysts used for the DNA extractions were *G. rostochiensis*. Once this was achieved, the extracted DNA samples from each pathotype were pooled together in equal volumes for further use.

6.3.2. Cloning the candidates *Avr* genes into pCR™8/GW/TOPO® vector

6.3.2.1. Amplification of the candidate sequences from genomic DNA

The primer sets G13394WLF_sp and G13394WLR, and G12477WLF_sp and G12477WLR were used to amplify the sequences GROS_g13394 and GROS_g12477 respectively from genomic DNA from both pathotypes. Both sets of amplified products included a SP. A similar fragment size of about 650bp for GROS_*g13394*SP and one of about 400bp for GROS_*g12477*SP from both pathotypes was amplified (Figure 6-2).

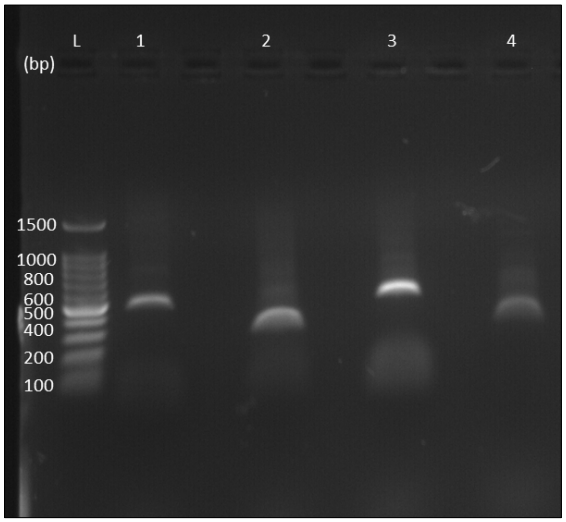


Figure 6-2 Purified PCR products of the genes GROS_*g13394*SP and GROS_*g12477*SP amplified from genomic DNA from either Ro1 or Ro5 *G. rostochiensis* genomic DNA. Gene-specific primers were used for each PCR reaction. PCR products were purified and separated by electrophoresis on 1% agarose gel. Amplification of Ro1_*g13394*SP is shown on lane 1, Ro1_*g12477*SP on lane 2, Ro5_*g13394*SP on lane 3 and Ro5_*g12477*SP on lane 4.

6.3.2.2. Amplification of the candidate fragments from cDNA

cDNA from J2s of both pathotypes Ro1 and Ro5 was used as template for amplifying the candidate genes *g13394* and *g12477* without the introns but including the SP. For this, the same primers as before were used. The purified PCR products had an approximate size of 450bp and 350bp for *g13394* and *g12477* respectively (Figure 6-3).

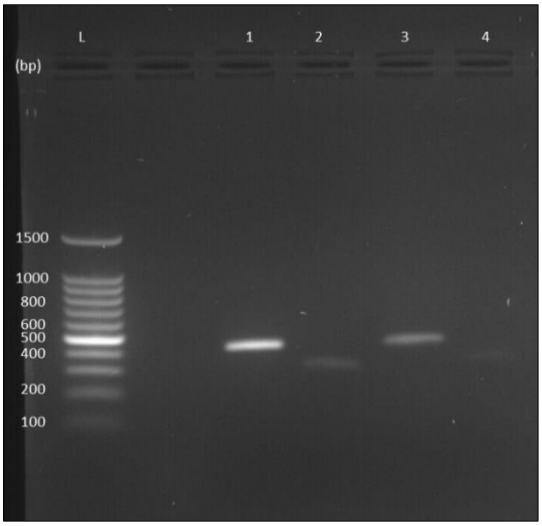


Figure 6-3 Purified PCR products of the genes GROS_*g13394*SP and GROS_*g12477*SP amplified from cDNA from either Ro1 or Ro5 *G. rostochiensis* J2s. Gene-specific primers were used for each PCR reaction. PCR products were purified and separated by electrophoresis on 1% agarose gel. Amplification of Ro1_*g13394*SP is shown on lane 1, Ro1_*g12477*SP on lane 2, Ro5_*g13394*SP on lane 3 and Ro5_*g12477*SP on lane 4.

Amplified PCR products were then cloned into pCR™8/GW/TOPO™ (Invitrogen) for sequence analysis.

6.3.3. Differences in amino acids sequences between the avirulent and the virulent pathotype were shown only in the gene *g13394*

Alignment of the amplified sequence *g13394*SP between the Ro1 and Ro5 pathotypes showed differences in 5 amino acids (Figure 6-4). Specifically, in the position 58 and 59 glutamine (Q) and valine (V) in Ro1 were replaced with histidine (H) and leucine (L) in Ro5 respectively. Proline (P) and aspartic acid (D) in the positions 66 and 67 respectively in Ro1 were replaced with glutamine (Q) and asparagine (N) in Ro5. The last change was identified in the position 124 where an ambiguous amino acid in Ro1 was replaced with lysine (K) in Ro5.

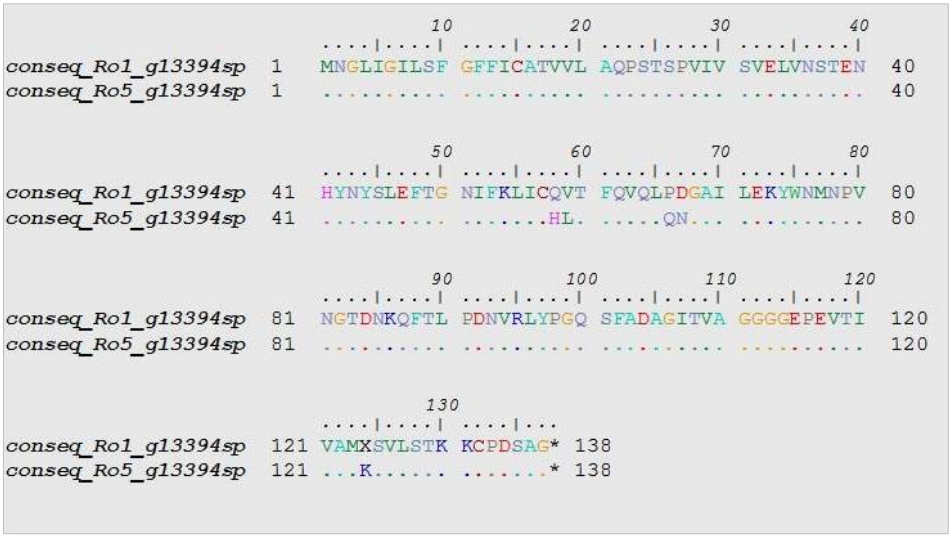


Figure 6-4 Pairwise global alignment between Ro1_*g13394*SP and Ro5_*g13394*SP. The shown sequences constitute a consensus (conseq) from several individual sequences of plasmid DNA samples. Dots (.) represent amino acid similarity and the asterisk (*) a stop codon.

The sequenced variants were then cross-checked with the available GBS data published by Eves-van den Akker et al. (2016). According to this study, in the coding sequence of GROS_*g13394* there is only one missense (non-synonymous) mutation at the nucleotide level (specifically, Guanine in the reference pathotype Ro1 changed into Adenine in pathotype Ro5) at the position 391. In the alignment done here, no nucleotide change was found at this position (Figure 6-5); moreover, the nucleotide at the position 391 was Adenine in both alleles.

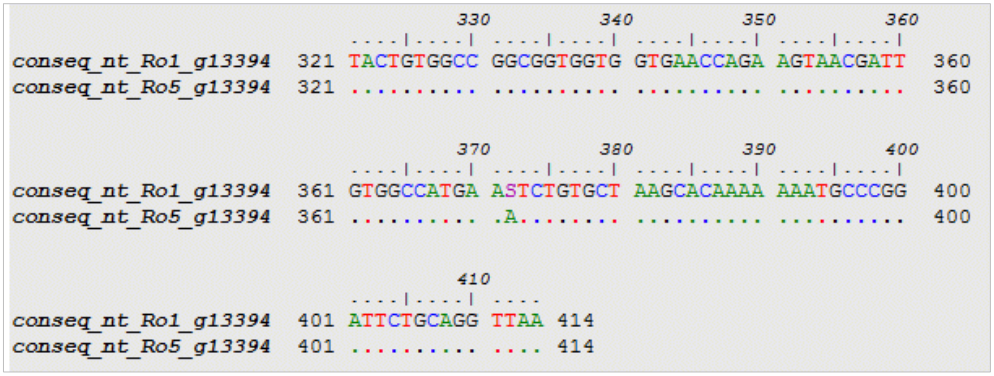


Figure 6-5 Pairwise global alignment between Ro1_*g13394*SP and Ro5_*g13394*SP (cloned from cDNA) at nucleotide (nt) level (here only the section between 321 and 414 is shown). At the position 391, no variant nucleotide was found in the sequenced *g13394*.

By contrast to *g13394*SP, the cloning and sequencing of *g12477*SP was not successful. The sequenced fragments were translated into amino acids, but the analysis of these sequences showed the presence of several different ORFs in samples derived from both

pathotypes. Cloning of the gene *g12477*SP into the pCR™8/GW/TOPO® vector was repeated using different cDNA templates batches with similar results (i.e. no consistent amino acid sequences similar to those expected from the reference genome sequence).

When the *G. rostochiensis* genome assembly was published (Eves-van den Akker et al., 2016), RNA-seq data were also available. The following table (Table 6-1) shows the differential expression values of the two genes in query in 4 different conditions (from cysts to 14dpi). It can be seen that the gene *g13394* showed the highest expression during the cyst stage and less during the infective stages (J2 onwards). On the other hand, the gene *g12477* seemed to be expressed at very low levels during the infective stages (J2) with slightly higher expression in parasitic stages.

Table 6-1 Differential expression levels of the genes GROS_g13394 and GROS_g12477. The levels are shown in during 8 lifecycle stages according to the RNA-seq data analysis (Eves-van den Akker et al., 2016).

Gene name	cyst_1	cyst_2	egg_1	egg_2	J2_1	J2_2	14dpi_1	14dpi_2
GROS_g13394	51.74	68.19	16.97	24.59	5.17	5.26	3.33	1.58
GROS_g12477	0	0	0.17	0	0.86	0.44	0	0

Visualising the RNA-seq data for the gene *g12477* showed that the annotation of this predicted gene overlaps with the annotation of another gene, *g12517* (Figure 6-6).

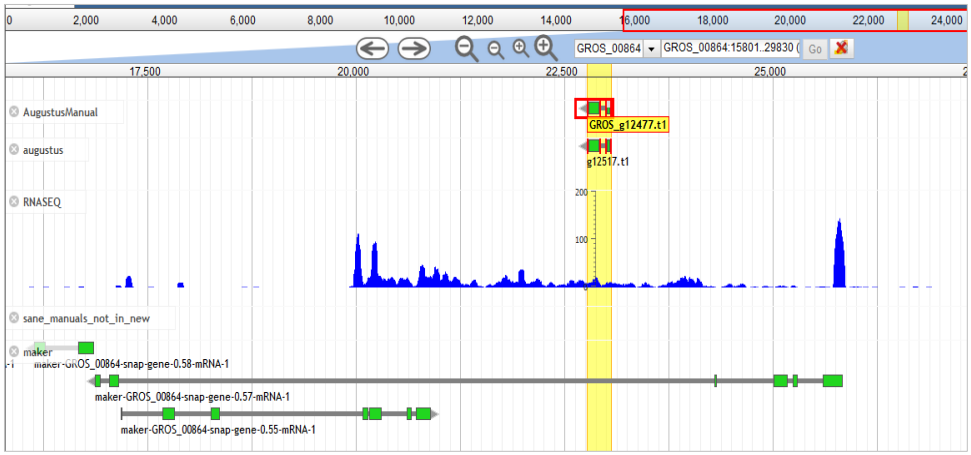


Figure 6-6 Visualisation of the alignment of the RNA-seq data of the candidate annotated gene GROS_*g12477*. Visualisation was done in the Apollo genomic annotation editor.

In a BLASTn search, the nucleotide sequence of the predicted gene *g12477* was aligned to the predicted gene *g12517* and gave a 100% match (Figure 6-7). Taking into

consideration all the above, it can be deduced that the candidate GROS_*g12477* is not a well-predicted gene and thus it was not further used for functional validation.

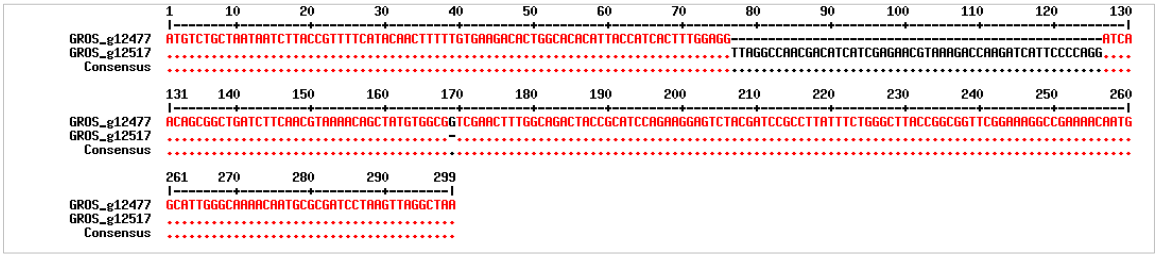


Figure 6-7 Pairwise alignment of GROS_*g12477* with GROS_*g12517*. The non-aligned portion (black letters) belongs to an intron.

Following this, it was necessary to clone the GROS_*g13394* sequence without the SP in order to prevent the protein from being secreted out of the cells flowing expression. To do this, a new primer that amplifies the gene *g13394* without the SP (G13394WLF) was used along with the primer G13394WLR.

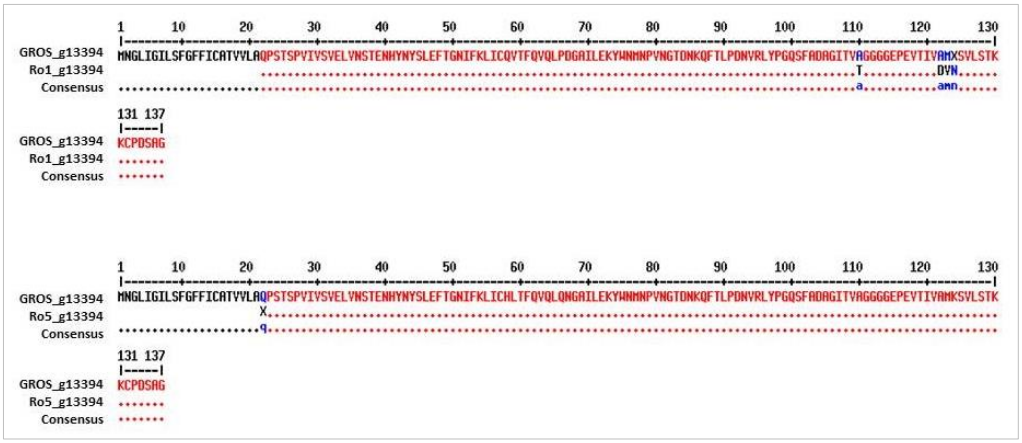


Figure 6-8 Pairwise alignment of the cloned Ro1_*g13394* (top) and Ro5_*g13394* (bottom) with the reference GROS_*g13394* to verify the successful removal of the signal peptide (SP). The first 21 amino acids from both consensus pathotype-specific sequences were successfully removed with the new sets of primers.

Following the same procedure outlined above but using the cloned full-length gene as a template rather than cDNA, the gene without the SP was successfully cloned from both Ro1_*g13394* and Ro5_*g13394* (Figure 6-8).

6.3.4. Functional validation of the candidate *g13394* on potato leaves

To perform *Agrobacterium*-mediated transient expression *in planta* of the candidate *Avr* gene *g13394*, the cloned genes (without their signal peptide) were transferred from the

ENTRY clone into the viral Gateway®-compatible PVX vector pGR106GW according to the procedure described in the section 2.2.6. The viral vector, which is able to infect potato plants and is widely used for effector screening on potato, was transformed into *A. tumefaciens* (strain GV3101) bacterial cells in order to allow expression of the gene in potato plants.

Two potato varieties were used; Désirée (as a susceptible control variety) and Maris Piper (which contains *H1*). More details about the constructs used for the inoculations are described in Chapter 2 (section 2.2.8.4).

6.3.4.1. Toothpick PVX Agroinfection in potato plants

Firstly, potato leaves were inoculated with a toothpick following the procedure described in Chapter 2 (section 2.2.8.2). The observations started at 7dpi and were taken regularly for the next 10 days. In total 3 series of inoculations were performed. In the case of recognition, local programmed cell death was expected around the inoculation sites indicated by a thin ring of blackened cells; these symptoms were HR-specific. On the other hand, in the case of no recognition, yellowing might appear around the inoculation sites as a result of viral infection and response to bacteria. For example, in the case of the *H1*-containing Maris Piper, when a negative, non-recognised control (*Avr1*:pGR106GW) was applied, yellowing would be seen spreading quickly around the inoculation site. On the other hand, in the case of the avirulent allele (Ro1_*g13394*:pGR106GW) brown areas (HR-specific symptoms) would be seen caused by its recognition by the *locus H1*, contrary to the virulent allele (Ro5_*g13394*:pGR196GW) where no response would be seen. In the susceptible Désirée leaves where no *R* gene exists, no recognition was expected to take place at all.

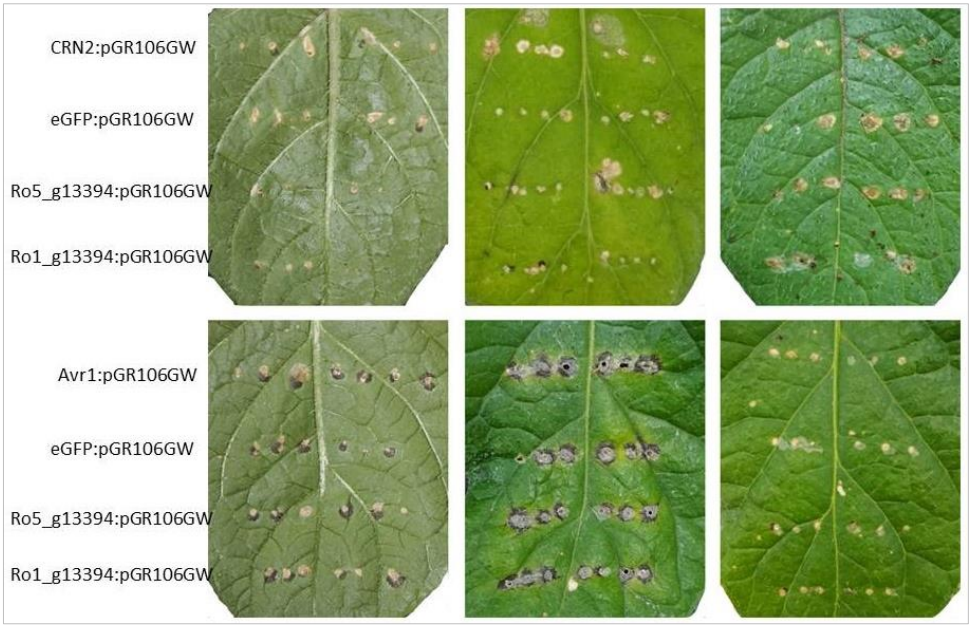


Figure 6-9 Toothpick inoculations. The susceptible potato variety Désirée (top) and the *H1*-holder Maris Piper (bottom) were toothpick-inoculated with the candidate *Avr* gene Ro1_*g13394*:pGR106GW and Ro5_*g13394*:pGR106GW. As positive control *CRN2*:pGR106GW (for Désirée) and *Avr1*:pGR106GW (for Maris Piper) were used, while eGFP:pGR106GW as negative controls. Each column is an example representation of the three PVX agroinfection experimental replicates performed in total. In each experiment, at least 3 biological replicates were performed.

As can be seen in Figure 6-9, in Désirée no HR was observed in any of the constructs as expected. Instead, yellowing was shown adjacent to the inoculation sites regardless of the applied construct including that of the negative control (*eGFP*:pGR106GW); this might be a sign of wound-induced symptoms caused by the toothpicks during the inoculations. In Maris Piper, black necrotic areas similar to HR-specific symptoms were shown in all the inoculation sites in the majority of the biological repeats. However, no difference was observed between the controls (*Avr1*:pGR106GW and *eGFP*:pGR106GW) and the two *g13394* alleles. This demonstrates the technical difficulties of this technique (e.g. wounds caused by toothpicks) and therefore no conclusion could be safely reached.

6.3.4.2. Vacuum infiltration of detached leaves

As an alternative approach for expression of the candidate *Avr* genes in potato, PVX Agroinfiltration was performed on detached potato leaves using vacuum manifold and following the protocol described in Chapter 2 (section 2.2.8.3). In total the experiment was repeated twice using the same potato varieties and constructs as the PVX Agroinfections. Briefly, the leaves were dipped in the inoculum and vacuum was applied to allow entrance of *Agrobacteria* through stomata. The first observations were

taken at 4dpi and then regularly thereafter. The top leaflets were not infiltrated and were used as an internal control. Therefore, in the cases where no recognition takes place (i.e. the negative controls and the virulent allele in Maris Piper, as well as all the constructs in Désirée), no HR-specific programmed cell death would be expected throughout the infiltrated area. Non-HR specific symptoms (e.g. weak chlorosis) can be induced by the viral vector upon infection. In the leaves where activation of HR responses take place (e.g. Ro1_*g13393*:pGR106GW in Maris Piper), strong programmed cell death (i.e. dark brown areas) would be expected in the areas between the leaf veins.

In the first experimental series (Figure 6-10), all Désirée leaves showed high level of response at 4dpi regardless of which inoculum was used, which is also translated into a large infected leaf surface. No significant difference was observed between the response of Désirée to either of the controls or to either of the *g13394* alleles. On Maris Piper leaves, a difference in the area showing response between the avirulent and the virulent allele was observed. The calculated mean necrotic leaf area caused by the virulent allele (Ro5_*g13394*:pGR106GW) was approximately 4 times larger compared to the one caused by the avirulent one (i.e. Ro1_*g13394*:pGR106GW). However, no statistically significant difference was found in the infected leaf area caused by Ro1_*g13394*:pGR106GW compared to the negative, non-recognised control *Avr1*:pGR106GW, which in turn caused the same symptoms in quantitative terms with Ro5_*g13394*:pGR106GW. The symptoms in the leaves of Maris Piper also developed more slowly compared to the Désirée.

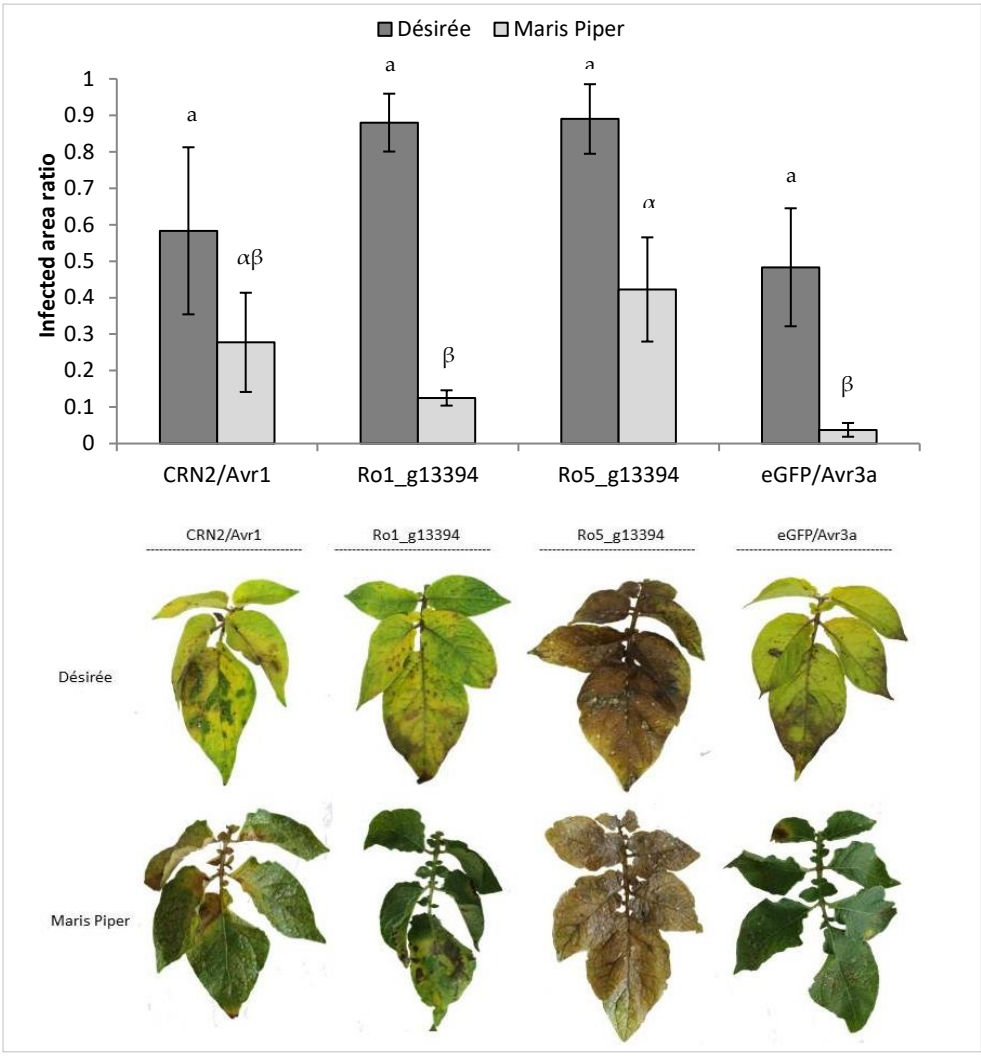


Figure 6-10 First experimental series of the vacuum PVX infiltrations. In the chart, each bar represents the ratio of the infected leaf area/total leaf area. Dark grey bars show the infected area by each construct on Désirée leaves and the light grey bars on Maris Piper. CRN2 was used as a non-recognised, negative control in Désirée and Avr1 in Maris Piper, whereas eGFP was used as a negative control in Désirée and Avr3a in Maris Piper. Error bars stand for standard error of the means. Latin characters indicate statistical differences within Désirée and Greek characters within Maris Piper determined with a two-way ANOVA test (Duncan's method, *p*-value < 0.05, n = 4). The photos show indicative examples of the inoculated leaves for each construct.

The second repeat of the experiment (Figure 6-11) showed no significant difference again between the leaves of the susceptible Désirée as before. Most of the Désirée detached leaves showed extensive non-HR specific necrotic areas as a result of the infection. In Maris Piper, the two alleles (Ro1_*g13394*:pGR106GW and Ro5_*g13394*:pGR106GW) did not show significantly different infected areas when compared to each other. The control *CRN2*:pGR106GW, despite its increased variance between the biological repeats, caused

significantly higher infected area compared to the second control and the avirulent *g13394* allele, whereas *eGFP*:pGR106GW did not.

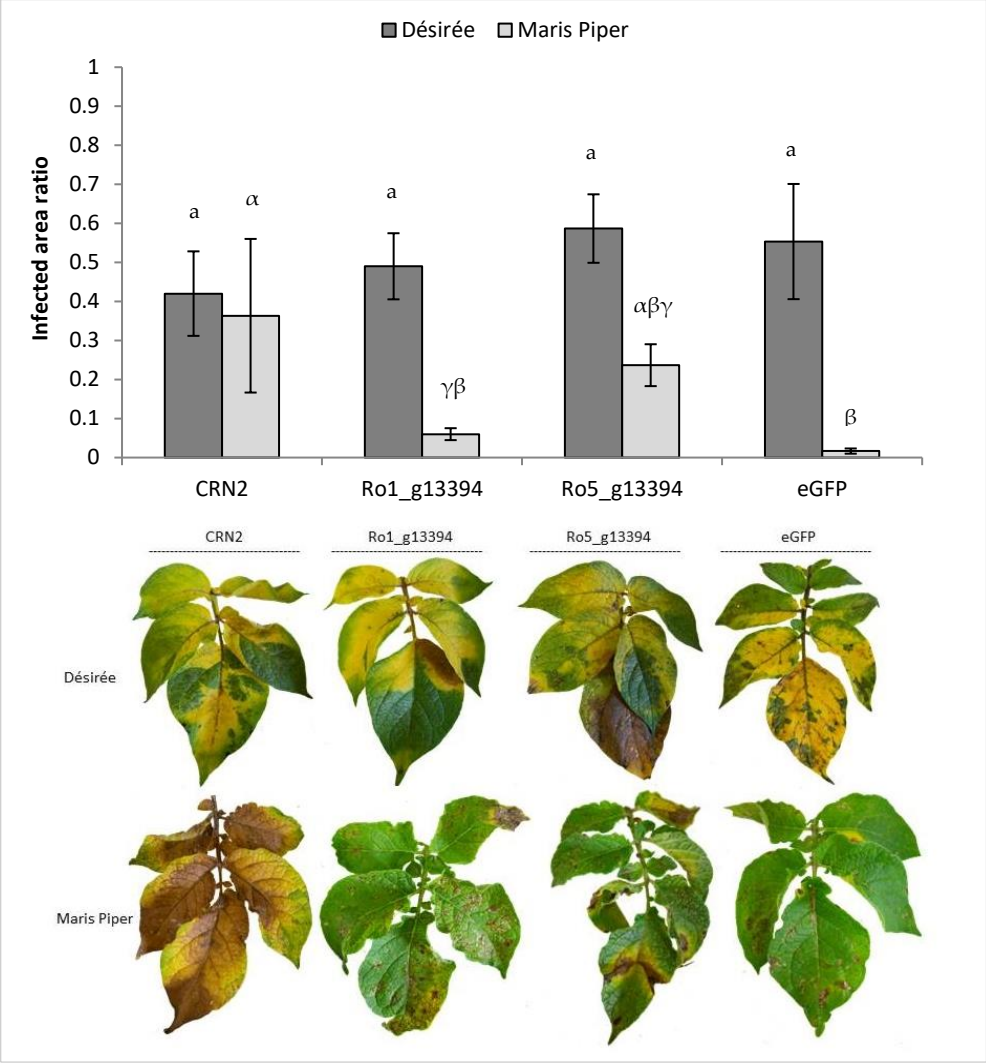


Figure 6-11 Second repeat of the experiment of the vacuum PVX infiltrations. In the chart, each bar represents the ratio of the necrotic area to the total leaf area. Dark grey bars show the infected area by each construct on Désirée leaves and the light grey bars on Maris Piper. CRN2 and eGFP were used as a non-recognised, negative control and negative control respectively. Error bars stand for standard error of the means. Latin characters indicate statistical differences within Désirée and Greek characters within Maris Piper determined with a two-way ANOVA test (Duncan's method, p-value < 0.05, n = 4). The photos show indicative examples of the inoculated leaves for each construct.

6.4. Discussion

The last few decades, the breeding industry has successfully exploited the resistance *locus H1* which originated from *S. tuberosum* ssp. *andigena* against *G. rostochiensis* (pathotypes Ro1 and Ro4). A few years ago, the genome of *G. rostochiensis* was published (Eves-van den Akker et al., 2016) allowing us to better comprehend and study the molecular mechanisms underlying the virulence behaviour of PCN. Effectors are key players for the infection, parasitism and virulence of PCN. They mediate a plethora of multiple and complex interactions between the host and the parasite executing several biological functions. Candidate *Avr* genes can be identified by analysing effector sequences predicted from genome sequences and by identifying possible differences between virulent and avirulent populations or isolates (Domazakis et al., 2017; Du and Vleeshouwers, 2014). Using this approach, candidate *Avr* genes were identified by resequencing virulent and avirulent *G. rostochiensis* populations (Eves-van den Akker et al., 2016).

Initial experiments showed that one of the candidate genes identified (GROS_*g12477*) was unlikely to be a real gene. Although the DNA fragment could be amplified from genomic DNA, it was not possible to clone this sequence from cDNA. Further analysis of the RNA-seq data generated during the *G. rostochiensis* genome project suggested that the prediction of *g12477* was not correct. All further efforts were therefore focused on the other candidate, GROS_*g13394*. This gene was cloned from genomic DNA and cDNA from an avirulent to *H1* pathotype (Ro1) and a virulent pathotype (Ro5) into an entry vector to be sequenced. By analysing the sequences, five amino acid changes (positions 58, 59, 65, 66 and 124) were found in the sequenced *g13394* (Figure 6-4). However, these changes did not match with the missense mutation identified by the GBS analysis between the virulent and the avirulent populations (Eves-van den Akker et al., 2016). This may be due to the different sequencing technologies used to generate the original DNA sequences or possibly may reflect differences within the populations used for amplifying the sequences cloned here. Numerous clones were sequenced from the cDNA cloning experiments and the sequence presented here was ubiquitous. It is possible that had additional sequencing been undertaken of further clones the original sequence reported in Eves-van den Akker *et al.* (2016) may have been identified.

In plant pathology, a widely used method for functional validation of candidate *Avr* genes is by expressing them in plants in the presence of the corresponding *R* gene in order to determine whether a cell death response is invoked. The only remaining candidate *Avr* gene (i.e. *g13394*) was therefore transferred into *A. tumefaciens* bacterial cells and then potato plants were infected. To express the gene, the PVX-based viral vector pGR106GW was used which had previously been modified as a destination plasmid vector in a Gateway® cloning system. The potato virus X (PVX) is an advantageous system for gene expression since it is able to infect potato carrying large foreign sequence inserts and *Agrobacterium*-origin sequences assisting their transfer and integration into host genome machinery (Chapman et al., 1992; van der Hoorn et al., 2000; Wagner et al., 2004).

In potato, *Agrobacterium*-mediated transient expression can be challenging possibly because of the leaf anatomy (e.g. thick epidermis, waxed surface, presence of dense trichomes and low stomatal density, dense main veins). Therefore, creation of an entry site on the leaf surface is necessary. First, two different potato varieties (the susceptible Désirée and the resistant *H1*-containing Maris Piper) were inoculated with *Agrobacteria* using toothpicks in a system that has been used in other studies (Du et al., 2014; Vleeshouwers et al., 2006). However, none of the replicates were successful. The positive control did not successfully infect the plant, causing no necrotic symptoms. Damaged areas adjacent to the infection sites were due to the wounds caused by the toothpick and the same response was observed by all the constructs, including the negative control.

As an alternative method, vacuum infiltration was used to allow bacteria to enter the plant through the stomata of detached leaves (Kapila et al., 1997). Once inside the plant leaf, *Agrobacteria* were able to infect and allow expression of the PVX encoding construct with the additional candidate genes or controls. The experiment was repeated twice, with 4-5 biological replicates in each. The results were interpreted qualitatively (by observing leaf phenotypes) and quantitatively (by expressing their phenotype as 'infected leaf area/total inoculated leaf area'). In all the Désirée leaves, extensive non-HR specific necrotic areas were developed at 5dpi regardless of the construct that was used. In Maris Piper, differences in the phenotypes were caused by the different constructs. In

summary, on the leaves inoculated with Ro1_*g13394*:pGR106GW, HR-induced cell death was observed in some leaves only, whereas the virulent allele Ro5_*g13394*:pGR106GW gave no HR-specific response at all. The resulting phenotypes were also often variable among the different biological repeats. The ratio 'infected area/total area' was calculated to express the phenotype in quantitative terms. Indeed, the symptomatic area between the two candidate *Avr* variants was significantly different only in the one of the experimental repeats in Maris Piper leaves. Also, no statistically significant difference was shown when compared to the negative controls. Additionally, in the second experimental repeat, the two controls differed significantly to each other but not compared to the *g13394* variants as was seen in (Figure 6-11).

In both series, the negative control *eGFP*:pGR106GW caused extensive non-HR specific symptoms on the susceptible Désirée; this is contradicting what was expected, since the protein eGFP cannot cause any symptoms. This can be explained with susceptibility of the variety to pathogens and pests, including viruses such as PVX, which according to available databases (ECPD; SASA, 2018) can be very high. By combining all the above, the *H1 locus* recognised the protein *g13394* allele from the avirulent Ro1 population only in some biological repeats, causing HR symptoms. However, no statistically significant evidence that *g13394* is indeed an *Avr* gene was obtained. During its cloning and analysis of the sequences from both pathotypes, 5 amino acid changes were identified that might be responsible for the loss of the virulence activity. These changes though are not matched with the findings of the GBS analysis.

In this chapter, different transient expression assays on potato were done proving how challenging this can be compared to infiltrations on *Nicotiana benthamiana* plants. Vacuum infiltrations can be a potential solution for the *Agrobacterium*-transient expression of genes in potato, although further optimisation is necessary. During the experiments, a great phenotypic variability of the control treatments was observed especially on Maris Piper. Additionally, when the infected leaf surface was calculated, in some cases, the threshold between the chlorotic and non-chlorotic areas was not always clear leading to increased results subjectivity and thus of both the type I and type II statistical errors. *Agrobacteria* infections need more than 4-5 days to be phenotypically

visible and can also differ among the different potato varieties. Vacuum infiltrations are done on detached leaves and therefore physical stresses on leaves, such as water loss, may affect the infection rate. Other studies have also shown that vacuum infiltrations of detached leaves may give variable results (Wroblewski et al., 2005). Weak expression of *Agrobacteria* after vacuum infiltrations can be caused by the change of the inoculum temperature due to the vacuum application or the uneven penetration of it into the leaf. Also, the presence of many veins in the leaves can limit the spread of the inoculum in the parenchymatic cells. Lastly, non-specific defence responses to PVX and *Agrobacteria* should be always taken into consideration as has been also shown on other studies (Du et al., 2014).

Therefore, further optimisation of the experimental procedure while more biological replicates could help. It can also be suggested that more *H1*-containing varieties can also be used in parallel with Maris Piper in order to eliminate unexpected effects caused by different molecular backgrounds that potato varieties have. Although the technical difficulties of this technique, here it was proved that *Agrobacteria*-mediated transient expression of candidate genes on detached potato leaves under vacuum conditions can be a potential method for future functional validation assays.

7. General Discussion & Future Work

Potato is the fourth most widely cultivated crop in Europe and has a significant input on the annual GDP for many countries, including the UK. Although the areas in Europe and the Americas cultivated with potato have steadily decreased in the last decades, in developing countries the area cultivated with potato has increased (FAOSTAT, 2016). This increase in developing nations reflects increased demand for potato due to it being a high-yielding and nutritious crop. Concomitantly, this increase in demand creates the need for healthy, high-yielding potato crops, including varieties resistant to pathogens and pests. One of the most economically significant potato pathogens is PCN, i.e. *G. pallida* and *G. rostochiensis*. It is estimated that every year, about 9% of global potato production is lost because of these pests, while in the UK the majority of the fields used for potato cultivation are infested by one or both species (Minnis et al., 2002; Turner and Subbotin, 2013). New EU legislation (European Union, 2009) on fumigant and chemical use for controlling PCN, makes the development of durable resistant varieties a priority (Whitehead and Turner, 1998).

In the last few decades, this extensive use of *H1*-containing varieties has reduced the problem of *G. rostochiensis* in the UK; however, it has allowed a strong selection towards *G. pallida*, which is now the major PCN problem (Minnis et al., 2002). The presence of many different populations of *G. pallida* even in the same field, combined with the absence of a major gene that confers resistance to them, makes the control of this species a challenge (Eves-van den Akker et al., 2015). Until now, only partial resistance has been identified against some *G. pallida* pathotypes and populations. Moreover, virulent alleles, which are present in the initial gene pool, are able to accumulate under a high selection pressure that may lead to some populations overcoming resistance (Fournet et al., 2016; Turner and Fleming, 2002).

Plants have an innate immune system, which includes R proteins that are able to recognise specific nematode-derived proteins and molecules, the so-called effectors. Recognised effectors are consequently referred to as *avirulence* (*avr*) genes (Dangl et al., 2013; Jones and Dangl, 2006). The current research aimed at the identification and validation of candidate *avr* genes in PCNs. This included an identification of candidate

vir genes from *G. pallida* populations previously selected for high virulence, as well as functional validation of a candidate *avr* gene from *G. rostochiensis*.

7.1. Analysing virulence levels of the selected populations

To determine the phenotypic differences in virulence levels of the populations selected on different resistances, these sub-populations were screened against a collection of potato genotypes containing a resistance source either from *S. vernei* or *H3* (described in Chapter 3). Both resistance sources have been used in breeding programmes against the pathotype Pa2/3, which is the most common one found in the UK (Blok et al., 1997; Bryan et al., 2004). Two founder populations (Newton and Farcet) were used along with their 4 sub-populations selected on the above resistances, which a previous project had generated (Phillips and Blok, 2008).

The tests showed that the virulence level of the sub-populations was specifically dependent on the genetic background of the potato genotype tested. In other words, nematode populations selected on a specific resistance source had higher virulence on another potato genotype containing the same resistance source, when compared to the populations selected on a different source of resistance. This increase in virulence was specific; populations selected on one source showed no change in virulence on the other source. The populations selected on *H3* showed a higher increase in virulence level when compared to the unselected founder population. There was an approximate 7-fold increase in the case of the Newton sub-populations and up to 14-fold in the case of Farcet sub-populations. The corresponding increase in the *S. vernei*-selected populations was approximately 4.6- and 7-fold in Newton and Farcet sub-populations respectively (Figure 3-1 and Figure 3-2). On the top of that, our results did not only showed adaptation on the resistance source, but also on the specific potato genotype where the selection took place. For instance, the *S. vernei*-selected populations n-8906 and n-11305 were highly virulent on both *S. vernei*-containing Sv_8906 and Sv_11305, but also n-11305 was even more virulent on Sv_11305 specifically.

Fournet et al. (2016) has shown that 6 generations of continuous adaptation of *G. pallida* populations on resistance sources are sufficient to create highly virulent lineages. In other cases, this number is slightly higher (i.e. 8 generations) (Beniers et al., 1995), while after 10-12 generations the virulence level seems to be stabilised at its maximum (Fournet et al., 2016; Phillips and Blok, 2008). It is obvious that this process creates a strong selection environment that allows to *vir* alleles to overcome resistant barriers. Since this increase was higher in the *H3*-containing varieties, it can also be suggested that this specific resistance is more readily overcome. Moreover, the observation of a Farcet population overcoming the *S. vernei* resistance source of the highly resistant variety "Innovator", also raises the possibility of a future breakdown of this specific resistance. The speed at which virulence could be established against both sources suggests that virulent individuals exist, albeit at low levels, in both founder populations against both resistance sources. The changes in virulence levels against each source may therefore reflect the relative abundance of virulent individuals in the founder populations.

7.2. The new *G. pallida* genome assembly

The publication of the genome assemblies of PCNs has allowed the use of genome-based approaches towards the identification of *avr* genes in PCNs. These studies (Cotton et al., 2014; Eves-van den Akker et al., 2016) have also provided important information about effector complements and aspects of parasitism. The published *G. pallida* genome assembly (described in Chapter 4) has a size of approximately 125Mb, organised in 8,873 scaffolds with the scaffold N50 at almost 122kbp, which shows its high level of fragmentation. Furthermore, it shows a relatively low level of gene completeness at around 74%, compared to the 94% estimated for *G. rostochiensis*. It has been suggested that the large unmapped genomic area could be involved in host adaptation (Eoche-Bosy et al., 2017b). Besides, these figures can constitute limitations for genomic studies including the present one.

For the next step of this study, the *G. pallida* genome assembly was needed as a reference for NGS downstream analyses (in the Chapters 4 and 5). In addition to the limitations described above, the published genome sequence originates from a different British

population (Lindley) than the one used here (Newton). This could constitute an additional limitation in the analyses. New modern sequencing technologies (PacBio SMRT sequencing) were exploited for the development of a new, improved and more complete *G. pallida* genome assembly from the unselected founder population "Newton" (described in the Chapter 4). This long-read sequencing technology allows the sequencing of significantly longer DNA molecules, with allegedly high accuracy and relatively low cost (Roberts et al., 2013).

The final version of the new assembly had a size of 119.6Mb, very close to the published version and close to the predicted physical size of the genome as estimated using nucleus flow cytometry. This can be explained by the large number of ambiguous bases in the published assembly (approx. 21 million bases), which was considerably decreased by about 20 million bases in the final version of the newly developed assembly (Table 4-2). The advantages of the PacBio sequencing technology are reflected in the improvements in the number of scaffolds (163 scaffolds) and the N50 of about 2.3 million bp. Additionally, the gene completeness of the new assembly is drastically improved to 94% (BUSCO). RNA-seq mapping predicted about 3,000 more genes compared to the old assembly – a higher proportion of the RNA-seq data mapped to the new assembly compared to the old one. Transcripts were mainly expressed in the early stages (i.e. pre-parasitic stages). Along with the final version, two drafts of it prepared earlier were used for the analyses of the present PhD project.

7.3. Variant calling analysis from the application of PenSeq and re-sequencing

Populations selected for virulence have been subjected to strong selection pressure. This pressure allows the emergence of *vir* alleles that might be already present in the initial population gene pool (Milne et al., 2012) (also described in Chapter 5). To determine the effector genes that may determine virulence activity against the resistance sources *Gpa5* from *S. vernei* and *S. tuberosum* ssp. *andigena* CPC2802 (*H3*), target enrichment sequencing (PenSeq) was used (described in Chapter 4). In the past, nematode effectors

were identified through EST sequencing analysis, using mRNA molecules from specific tissues or lifecycle stages, combined with bioinformatic tools (Gautier, 2015; Jones et al., 2009; Qin et al., 2000). Information about the bases of genomic pathogenicity can be acquired by analysing diversity from whole-genome sequences from virulent and avirulent populations (Eves-van den Akker et al., 2016), which was also used here as an additional method to PenSeq (described in Chapter 5).

The application of PenSeq significantly decreases genome complexities and focuses only on specific area of interest, which in our case is that of the effectors. Additionally, the use of short-read technology for sequencing the targeted areas, ensures the elimination of bias by achieving high read coverage. On top of that, by whole-genome re-sequencing of these populations, large genomic regions that may be closely related to adaptation of the selected populations against the specific resistances were included in the analyses.

Contrary to the effectors of the oomycete *P. infestans* that contain a characteristic RXLR motif (Birch et al., 2006), effectors in nematodes do not contain a specific motif. However, recent studies have identified promoter regions (e.g. DOG box, STATAWAARS motif), which can be used as a guide for effector identification (Espada et al., 2018; Eves-van den Akker et al., 2016; Eves-van den Akker et al., 2014). Genes containing the above predicted promoter regions upstream of the coding region were included in the initial bait design of the PenSeq target list. Additionally, potential effectors from closely-related species, as well as previously identified *G. pallida* effector genes (Thorpe et al., 2014) were also added. Sequencing of the (PenSeq) captured regions was performed on Illumina MiSeq. The sequencing produced about 23.5 million short reads with high coverage from two duplexed Newton sub-populations selected on S. *vernei*, two on *H3* and the founder population. The mapping of the reads was done on the two drafts of the new *G. pallida* genome assembly to eliminate the sequence polymorphisms due to the population background.

More than 300 markers were identified, with the majority of them being SNPs. The non-synonymous variants were filtered depending on the putative impact on the protein into high- and moderate-impact. Approximately 80% of them were enlisted in the initial target list of the designed PenSeq baits. As can be seen from the Table 4-7 and Table 4-8,

the majority of the identified variant genes were SPRY-containing proteins, gland cell-localised proteins and CWDEs (e.g. β-1,4-endoglucanase, pectate lyases). Proteins of these categories have previously been reported as effector genes in cyst nematodes (Eves-van den Akker et al., 2016; Haegeman et al., 2012; Noon et al., 2015; Thorpe et al., 2014).

Similar to PenSeq, downstream NGS analyses of the re-sequenced populations followed a related approach (Chapter 5). Different filtering parameters were used, mainly based on changes of the AF threshold (CAF) between the different biological groups (populations selected on *S. vernei* and those selected on *H3*). The closer to 0 the allele frequency values were, the most likely the variant was to contain the alternate allele in comparison with the reference. Since the population re-sequencing can identify variants throughout the whole length of their genome (in contrast to PenSeq), we had to narrow down the number of the candidate variants; therefore, only those with high or moderate putative impact were kept. Again, SPRYSECs were the most abundant protein family in the re-sequenced populations (Table 5-3 and Table 5-4).

As reviewed in the discussion sections of the Chapters 4 and 5, the highly versatile SPRY domain is an expanded gene family. Selected populations are able to exploit this characteristic in order to bypass selection pressure. The mechanism that nematodes use for this is not yet fully understood, but both of the analyses described in this thesis showed that alternative splicing might be implicated. Further research using mRNA of those variants and analysing their transcription profiles in different lifecycle stages, could validate this hypothesis.

In both pipelines (i.e. PenSeq and re-sequencing), the drafts of the new *G. pallida* genome assembly were used as references. Hence, it would be useful for the same steps to be repeated using the final version of the genome as a reference for mapping the raw reads. Moreover, the identified genes constitute candidates, which in turn means that functional validation assays are required, by cloning and then transiently expressing these candidates *in planta* (using the method described in Chapter 6). During the filtering steps on the re-sequenced populations maxCAF1 was set at 0.1, i.e. max 10% change of the AF of the populations selected on the same resistance. However, as can be seen in

the screening tests of Chapter 3 (Figure 3-1), some populations with the same selection background can still show differences in virulence, which might indicate that these differences represent wider differences in their AF values. Therefore, more filtering steps should be conducted using larger maxCAF1 values.

By comparing the lists generated through both methods, only three candidate effector genes were common to both (Table 7-1). Specifically, in the mapping of the new/bigger assembly, 3 candidates were found in both analyses. One of these was also identified from the mapping of the new/smaller assembly. Two of them constitute known effectors (a SPRY domain protein and the 1106 effector protein), while the third constitutes a novel protein (J. Jones, personal communication). Interestingly, all three were included in the initial PenSeq target list, while the latter two candidates contained a predicted signal peptide in their sequences. These three candidates could be listed as high-confidence candidates and can be prioritised for functional validation. The role of the SPRYSECs in virulence has been reviewed extensively in the previous chapters of this thesis. However, little data existed regarding the 1106 effector protein (Chapter 4 and 5). Finkers-Tomczak (2011) suggested that the members of this gene family, originally found in *G. rostochiensis*, produce multidomain proteins. Therefore, similar to SPRY domain proteins, the specific family could be diversified during positive selection events in order to increase the range of its virulence targets. Although their main function is not well known, transient expression assays showed strong suppression of plant immune responses by inhibiting transcription factors, and therefore this variant family is also reported as NSI-1 (i.e. nematode suppressors of immunity-1). For the above reasons, it will be very interesting to further study whether this identified variant functions similarly in *G. pallida*.

In the Table 7-1, the genes identified in both draft assemblies in common during PenSeq can also be added. As can be seen, many of them are already known potential effectors (e.g. SPRY domain proteins, β-1,4-endoglucanase and esophageal gland proteins), as well as some without a known function.

Table 7-1 List of the suggestive candidate *G. pallida* effector genes that can be prioritised for validation. The third column shows the analysis or analyses they identified from, and the last column the presence or not of a predicted signal peptide.

Gene	Description	Identified in	SP
GPLIN_000909200	N/A (hypothetical protein)	PenSeq + ReSeq	yes
GPLIN_000768400	1106 effector	PenSeq + ReSeq	yes
GPLIN_000909700	SPRYSEC	PenSeq + ReSeq	no
GPLIN_000725400	RBP-1	PenSeq; both assemblies	no
GPLIN_001048200	SPRYSEC	PenSeq; both assemblies	no
GPLIN_000036500	N/A	PenSeq; both assemblies	no
GPLIN_000626800	SPRYSEC	PenSeq; both assemblies	no
GPLIN_001185800	β-1,4-endoglucanase	PenSeq; both assemblies	no
GPLIN_000328200	Dopey-1	PenSeq; both assemblies	no
GPLIN_000996800	GLAND3	PenSeq; both assemblies	yes
GPLIN_000666500	GLAND16	PenSeq; both assemblies	no
GPLIN_000376600	N/A	PenSeq; both assemblies	no
GPLIN_001436900	SPRY domain protein	PenSeq; both assemblies	no
GPLIN_000245500	Zinc finger-like protein	PenSeq; both assemblies	no
GPLIN_000560800	N/A	PenSeq; both assemblies	yes
GPLIN_000945900	Vitellogenin	PenSeq; both assemblies	yes
GPLIN_000312300	SPRY domain protein	PenSeq; both assemblies	yes

7.4. Study of the selection preferences

The study of AF can provide important information on selection and host adaptation. Although analysis of the PenSeq data provided information about AF values, interesting outlier genomic regions are excluded. However, these outlier genomic regions can be crucial for the evolution of selection and host adaptation (Gautier, 2015). Whole genome scanning of populations also allows the reliable study of the AF values from all the markers identified in the whole genome.

As described in Chapter 5, the largest number of the identified candidate effector genes were found in the nematode populations selected on *H3* (Table 4-9 and Table 4-10). Specific effector gene families were selected towards specific resistances, such as the *RBP-1* mainly selected in the *H3*-selected populations, whereas some *CWDEs* were found on the *S. vernei*-selected populations. It seems that specific gene families may be

preferentially selected in the populations depending on the resistance source in order to adapt accordingly and it could also be postulated that this can be related to the timing of the response. For instance, CWDEs are mostly expressed during the penetration and invasion of the juvenile inside the host cells (i.e. pre-parasitic lifecycle stages), whereas SPRYSECs are expressed later to suppress or manipulate the activated host responses (reviewed by Haegeman et al. (2012)). Therefore, in the case of an early activation of the *H3* resistance (e.g. in the nematode pre-parasitic stages), it could be hypothesised that virulent populations to *H3* would have been selected for specific types of effectors that are mainly expressed at that time (i.e. during the invasion). This is not known yet in the case of these resistances, but it is known that in the case of the *Mi-1* resistance to *Meloidogyne* sp. the immune responses are activated upon the formation of the giant cells (Dropkin, 1969). This can also be supported by the formation of genomic "islands" containing potential variant genes, including effectors. These variants found within those SNP "islands" were entirely selected towards a specific resistance source and it is thought that they are selected together in order to facilitate genomic plasticity and host adaptation during selection evolution; this suggestion aligns with other studies performed on other plant pathogens and PPNs (Eoche-Bosy et al., 2017b; Eves-van den Akker et al., 2016; Raffaele et al., 2010). Moving forward, more analysis should be done to identify whether these potential "islands" found in the present study are also formed in the final new genome assembly. Then, potential effectors located on those regions and that are selected towards a specific resistance source can be found by analysing several CAF thresholds (from stringent to less stringent). This can show us whether specific effector families are selected by the nematode populations to overcome specific resistances.

As a last step of the analysis done in the re-sequenced populations, variants that were selected differently against the same resistance source was performed. No specific gene (or closely-neighbouring) was reported that was selected to both resistance sources in the populations with the same selection background. However, in the new/smaller assembly, two SNPs located 144Kbp far from each other were selected differently in the *H3*-selected populations. This might constitute a random event rather than a meaningful genetic change. It would be very interesting for this analysis to be repeated on the final

new assembly using different CAF thresholds. In the screening tests of the chapter 3 (Figure 3-1), some populations selected on the same resistance source showed different virulence levels when tested on the same resistance; the selected populations that were tested on the specific potato clone where the selection process had taken place (e.g. n-12674 from the clone Sa_12674) had even higher virulence levels compared to the virulence shown by the other population selected on the same resistance source. Such cases were observed in the populations n-8906 and n-11305 when tested on Sv_11305, as well as the n-11415 and n-12674 tested on Sa_12674. These especially "selective" preferences within the same resistance source might be also depicted in the genome of those populations.

7.5. Functional validation of candidate *G. rostochiensis avr* genes

In the beginning of this thesis (Chapter 1), the gene-for-gene model was used to explain simply how pathogen effectors could be recognised by specific host *R* genes (Flor, 1971; Jones and Dangl, 2006). In plant pathology, a common technique used to functionally validate a candidate *avr* gene is by co-expressing it with the cognate *R* gene on *A. thaliana* or *N. benthamiana* leaves. Upon recognition of an effector, HR responses are activated in the leaf in order to block the pathogen's spread. Depending on the vector used, the inoculations of the plants can be done by transferring the inoculum that contains *Agrobacteria* that carry and can express the gene of interest inside the leaf. For *N. benthamiana*, this can be achieved simply by using a syringe. However, this method has not been successful in potato, possibly because of the leaf structure (i.e. heterobaric leaf anatomy). Specifically, potato leaves have minor veins that span vertically the mesophyll separating it into compartments (i.e. areoles) in order to reduce gas movement during dry conditions (Alison Roberts, personal communication). In the past, several techniques have been used, including that of cloning the candidate gene into a viral system that is able to infect potato plants (e.g. PVX), which is subsequently expressed from *Agrobacteria*. The bacteria then are transferred and potato leaves infected through a small surface wound caused by a toothpick (Du et al., 2014; Vleeshouwers et al., 2006). The

latter method was used in the current PhD project (described in Chapter 6) without success.

To functionally validate the candidate *G. rostochiensis avr* gene *g13394*, an alternative technique was used, that of the vacuum infiltration on detached leaves. This technique described by Kapila et al. (1997) was tested in the Nematology lab of The James Hutton Institute (Strachan, 2018), gave various and sometimes contradicting results, which did not allow confident conclusion about the virulence status of the gene *g13394*. As described in the discussion of Chapter 6, more leaves need to be tested. In contrast with the other methods, in the vacuum infiltrations the inoculum is being applied to all the leaflets of a single leaf and this may increase the variability of the results if the leaves do not have similar physical characteristics (e.g. size, developmental stage, cuticle structure). Many more biological replicates could help in solving this problem, in combination with the testing of more than two potato varieties, since some non-specific symptoms can be caused by the molecular background of the varieties.

Effectors are crucial for the lifecycle of nematodes and promoting virulence. In the presence of the cognate host *R* gene, recognition takes place and host immune responses are activated; in which case these effectors are then characterised as *avr* genes (Jones and Dangl, 2006). The current study used up-to-date genome-based approaches for the identification of candidate *avr* and effector genes in *G. pallida*. Populations selected for high virulence on the resistance sources *Gpa5* from *S. vernei* and *S. tuberosum* ssp. *andigena* CPC2802 (*H3*) were used. It was proved that virulence of those populations was very specific and dependent on the initial selection source. By using target enrichment sequencing technology for effectors, high-confidence candidate effector genes that determine virulence on these resistances were identified and suggested for further validation. Furthermore, whole genome scanning of these selected populations provided important information on selection and adaptation on the specific resistances. A higher accuracy in the analyses was ensured by the use of the new, more complete *G. pallida* genome assembly that was developed during this project. Lastly, a new method for functional validation of candidate *avr* genes on potato plants was tested.

Knowledge of the effector genes that are recognised by specific *R* genes is crucial for the development of diagnostic markers and their subsequent integration in the future breeding programmes. Nowadays, modern genomic tools can facilitate this goal. By exploiting all the available tools, we can acquire significant knowledge on developing new durable resistances in potato to meet future global demand.

References

Agrios, G. (2005). Plant diseases caused by nematodes. *In* "Plant Pathology" (G. Agrios, ed.). Elsevier.

AHDB (2015). Potato cultivation in the UK. Agriculture and Horticulture Development Board.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data.

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

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., van Eck, H., Coppoolse, E., van der Vossen, E., Bakker, J., and Goverse, A. (2004). A high-resolution map of the H1 locus harbouring resistance to the potato cyst nematode Globodera rostochiensis. *Theor Appl Genet* **109**, 146-52.

Bakker, E., Borm, T., Prins, P., van der Vossen, E., Uenk, G., Arens, M., de Boer, J., van Eck, H., Muskens, M., Vossen, J., van der Linden, G., van Ham, R., Klein-Lankhorst, R., Visser, R., Smant, G., Bakker, J., and Goverse, A. (2011). A genome-wide genetic map of NB-LRR disease resistance loci in potato. *Theor Appl Genet* **123**, 493-508.

Beniers, A., Mulder, A., and Schouten, H. (1995). Selection for virulence of Globodera pallida by potato cultivars. *Fundam. appl. Nematol.* **18**, 497-500.

Bent, A. F., and Mackey, D. (2007). Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annu Rev Phytopathol* **45**, 399-436.

Bercher, M., Wahl, J., Vogel, B. E., Lu, C., Hedgecock, E. M., Hall, D. H., and Plenefisch, J. D. (2001). mua-3, a gene required for mechanical tissue integrity in Caenorhabditis elegans, encodes a novel transmembrane protein of epithelial attachment complexes. *The Journal of Cell Biology* **154**, 415-426.

Birch, P. R. J., Gilroy, E. M., Pritchard, L., Armstrong, M., Boevink, P., Whisson, S. C., Taylor, R. M., Bos, J., Kamoun, S., Sadanandom, A., Conti, L., Ewan, R., van West, P., and Wawra, S. (2009). Towards understanding the virulence functions of RXLR effectors of the oomycete plant pathogen Phytophthora infestans. *Journal of Experimental Botany* **60**, 1133-1140.

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

Bird, D. M., Jones, J. T., Opperman, C. H., Kikuchi, T., and Danchin, E. G. (2015). Signatures of adaptation to plant parasitism in nematode genomes. *Parasitology* **142 Suppl 1**, S71-84.

Bird, D. M., Williamson, V. M., Abad, P., McCarter, J., Danchin, E. G., Castagnone-Sereno, P., and Opperman, C. H. (2009). The genomes of root-knot nematodes. *Annu Rev Phytopathol* **47**, 333-51.

Blaxter, M., De Ley, P., Garey, J., Liu, L., Scheldeman, P., Vierstraete, A., Vanfleteren, J., Mackey, L., Dorris, M., Frisse, L., Vida, J., and Thomas, W. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature* **392**, 71-75.

Blok, V. C., Phillips, M. S., and Harrower, B. E. (1997). Comparison of British populations of potato cyst nematodes with populations from continental Europe and South America using RAPDs. *Genome* **40**, 286-293.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **30**, 2114-2120.

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

Bos, J. I., Armstrong, M. R., Gilroy, E. M., Boevink, P. C., Hein, I., Taylor, R. M., Zhendong, T., Engelhardt, S., Vetukuri, R. R., Harrower, B., Dixelius, C., Bryan, G., Sadanandom, A., Whisson, S. C., Kamoun, S., and Birch, P. R. (2010). Phytophthora infestans effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proc Natl Acad Sci U S A* **107**, 9909-14.

Bos, J. I., Chapparo-Garcia, A., Quesada-Ocampo, L. M., McSpadden Gardener, B. B., and Kamoun, S. (2009). Distinct amino acids of the Phytophthora infestans effector AVR3a condition activation of R3a hypersensitivity and suppression of cell death. *MPMI* **22**, 269-281.

Bozkurt, T., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., Cano, L., Jones, A., Huitema, E., van der Hoorn, R. A., and Kamoun, S. (2011). Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *PNAS* **108**, 20832-20837.

Bradshaw, J. E., and Ramsay, G. (2005). Utilisation of the Commonwealth Potato Collection in potato breeding. *Euphytica* **146**, 9-19.

Bryan, G. J., McLean, K., Bradshaw, E., De Jong, S., Phillips, M., Castelli, L., and Waugh, R. (2002). Mapping QTLs for resistance to the cyst nematode Globodera pallida derived from the wild potato species Solanum vernei. *Theor Appl Genet* **105**, 68-77.

Bryan, G. J., McLean, K., Pande, B., Purvis, A., Hackett, C. A., Bradshaw, J. E., and Waugh, R. (2004). Genetical dissection of H3-mediated polygenic PCN resistance in a heterozygous autotetraploid potato population. *Mol Breed* **14**, 105-116.

Bulman, S. R., and Marshall, J. W. (1997). Differentiation of Australasian potato cyst nematode (PCN) populations using the polymerase chain reaction (PCR). *New Zealand Journal of Crop and Horticultural Science* **25**, 123-129.

Caromel, B., Mugniery, D., Kerlan, M. C., Andrzejewski, S., Palloix, A., Ellisseche, D., Rousselle-Bourgeois, F., and Lefebvre, V. (2005). Resistance quantitative trait loci originating from Solanum sparsipilum act independently on the sex ration of Globodera pallida and together for developing a necrotic reaction. *MPMI* **18**, 1186-1194.

Carpentier, J., Esquibet, M., Fouville, D., Manzanares-Dauleux, M. J., Kerlan, M. C., and Grenier, E. (2012). The evolution of the Gp-Rbp-1 gene in Globodera pallida includes multiple selective replacements. *Mol Plant Pathol* **13**, 546-55.

Casteel, C. L., Walling, L. L., and Paine, T. D. (2006). Behavior and biology of the tomato psyllid, Bactericerca cockerelli, in response to the Mi-1.2 gene. *Entomologia Experimentalis et Applicata* **121**, 67-72.

Castelli, L., Ramsay, G., Bryan, G., Neilson, J., S., and Phillips, M. (2003). New sources of resistance to the potato cyst nematodes Globodera pallida and G. rostochiensis in the Commonwealth Potato Collection. *Euphytica* **129**, 377-386.

Chapman, S., Kavanagh, T., and Baulcombe, D. C. (1992). Potato virus X as a vector for gene expression in plants. *The Plant Journal* **2**, 549-557.

Chen, S., Chronis, D., and Wang, X. (2013). The novel GrCEP12 peptide from the plant-parasitic nematode Globodera rostochiensis suppresses flg22-mediated PTI. *Plant Signal Behav* **8**.

Chen, X., Lewandowska, D., Armstrong, M. R., Baker, K., Lim, T. Y., Bayer, M., Harrower, B., McLean, K., Jupe, F., Witek, K., Lees, A. K., Jones, J. D., Bryan, G. J., and Hein, I. (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. *Theor Appl Genet* **131**, 1287-1297.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nurnberger, T., Jones, J. D., Felix, G., and Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**, 497-500.

Chronis, D., Chen, S., Lu, S., Hewezi, T., Carpenter, S. C., Loria, R., Baum, T. J., and Wang, X. (2013). A ubiquitin carboxyl extension protein secreted from a plant-parasitic nematode

Globodera rostochiensis is cleaved in planta to promote plant parasitism. *Plant J* **74**, 185-96.

Cingolani, P., Platts, A., Wang Ie, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80-92.

Clement, J. A., Toulza, E., Gautier, M., Parrinello, H., Roquis, D., Boissier, J., Rognon, A., Mone, H., Mouahid, G., Buard, J., Mitta, G., and Grunau, C. (2013). Private selective sweeps identified from next-generation pool-sequencing reveal convergent pathways under selection in two inbred Schistosoma mansoni strains. *PLoS Negl Trop Dis* **7**, e2591.

Coaker, G., and Baker, D. (2013). Signal Transduction Pathways Activated by R Proteins. *In* "Molecular Plant Immunity" (G. Sessa, ed.). Wiley-Blackwell.

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

Dale, M. F. B., and Phillips, M. S. (2009). An investigation of resistance to the white potato cyst-nematode. *The Journal of Agricultural Science* **99**, 325.

Danchin, E. G., and Rosso, M. N. (2012). Lateral gene transfers have polished animal genomes: lessons from nematodes. *Front Cell Infect Microbiol* **2**, 27.

Dangl, J. L., Horvath, D. M., and Staskawicz, B. J. (2013). Pivoting the Plant Immune System from Dissection to Deployment. *Science* **341**, 746-751.

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

Davis, E. L., Hussey, R., Mitchum, M. G., and Baum, T. (2008). Parasitism proteins in nematode-plant interactions. *Current Opinion in Plant Biology* **11**, 360366.

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

Decraemer, W., and Hunt, M. (2013). Structure and Classification. *In* "Plant Nematology" (R. N. Perry and M. Moens, eds.). CAB International.

Diaz-Granados, A., Petrescu, A. J., Goverse, A., and Smant, G. (2016). SPRYSEC Effectors: A Versatile Protein-Binding Platform to Disrupt Plant Innate Immunity. *Front Plant Sci* **7**, 1575.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21.

Doitsidou, M., Minevich, G., Kroll, J. R., Soete, G., Gowtham, S., Korswagen, H. C., Sebastiaan van Zon, J., and Hobert, O. (2018). A Caenorhabditis elegans zinc finger transcription factor, ztf-6, required for the specification of a dopamine neuron-producing lineage. *G3: Genes, Genomes, Genetics* **8**, 17-26.

Domazakis, E., Lin, X., Aguilera-Galvez, C., Wouters, D., Bijsterbosch, G., Wolters, P., and Vleeshouwers, V. G. (2017). Effectoromics-based identification of cell surface receptors in potato. *In* "Plant Pattern Recognition Receptors: Methods and Protocols" (L. Shan and P. He, eds.). Humana Press.

Dropkin, V. H. (1969). Cellular Responses of Plants to Nematode Infections. *Annual Review of Phytopathology* **7**, 101-122.

Du, J., Rietman, H., and Vleeshouwers, V. G. (2014). Agroinfiltration and PVX agroinfection in potato and Nicotiana benthamiana. *J Vis Exp*, e50971.

Du, J., and Vleeshouwers, V. G. (2014). The do's and don'ts of effectoromics. *In* "Plant-Pathogen Interactions: Methods and Protocols" (P. R. Birch, J. T. Jones and J. I. Bos, eds.). Humana Press.

Duceppe, M. O., Lafond-Lapalme, J., Palomares-Rius, J. E., Sabeh, M., Blok, V., Moffett, P., and Mimee, B. (2017). Analysis of survival and hatching transcriptomes from potato cyst nematodes, Globodera rostochiensis and G. pallida. *Sci Rep* **7**, 3882.

ECPD. The European Cultivated Potato

Eilbeck, K., Lewis, S. E., Mungall, C. J., Yandell, M., Stein, L., Durbin, R., and Ashburner, M. (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biology* **6**, R44.

EMBL-EBI InterPro: Protein sequence analysis & classification.

Eoche-Bosy, D., Gauthier, J., Juhel, A. S., Esquibet, M., Fournet, S., Grenier, E., and Montarry, J. (2017a). Experimentally evolved populations of the potato cyst nematode Globodera pallida allow the targeting of genomic footprints of selection due to host adaptation. *Plant Pathology* **66**, 1022-1030.

Eoche-Bosy, D., Gautier, M., Esquibet, M., Legeai, F., Bretaudeau, A., Bouchez, O., Fournet, S., Grenier, E., and Montarry, J. (2017b). Genome scans on experimentally evolved populations reveal candidate regions for adaptation to plant resistance in the potato cyst nematode Globodera pallida. *Mol Ecol* **26**, 4700-4711.

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

Escobar, C., Brown, S., and Mitchum, M. G. (2011). Transciptomic and Proteomic Analysis of the Plant Response to Nematode Infection. *In* "Genomics and Molecular Genetics of Plant-Nematode Interactions" (J. T. Jones, G. Gheysen and C. Fenoll, eds.). Springer.

Espada, M., Eves-van den Akker, S., Maier, T., Vijayapalani, P., Baum, T., Mota, M., and 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.

EU (2009). Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. *Official Journal of the European Union* **309**, 1-50.

Evans, K. (1993). New approaches for potato cyst nematode management. *Nematropica* **23**.

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

Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M. O., Boyes, I., Marvin, J. M., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., Esquibet, M., Sabeh, M., Rott, M., Overmars, H., Finkers-Tomczak, A., Smant, G., Koutsovoulos, G., Blok, V., Mantelin, S., Cock, P. J., Phillips, W., Henrissat, B., Urwin, P. E., Blaxter, M., and Jones, J. T. (2016). The genome of the yellow potato cyst nematode, Globodera rostochiensis, reveals insights into the basis of parasitism and virulence. *Genome Biol* **17**, 124.

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

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

FAOSTAT (2016). Food and Agriculture Organization of the United Nations.

Ferretti, L., Ramos-Onsins, S. E., and Perez-Enciso, M. (2013). Population genomics from pool sequencing. *Mol Ecol* **22**, 5561-76.

Finkers-Tomczak, A., Bakker, E., de Boer, J., van der Vossen, E., Achenbach, U., Golas, T., Suryaningrat, S., Smant, G., Bakker, J., and Goverse, A. (2011). Comparative sequence analysis of the potato cyst nematode resistance locus H1 reveals a major lack of co-linearity between three haplotypes in potato (Solanum tuberosum ssp.). *Theor Appl Genet* **122**, 595-608.

Finkers-Tomczak, A. M. (2011). Co-evolution between Globodera rostochiensis and potato driving sequence diversity of NB-LRR resistance loci and nematode suppressors of plant immunity, Wageningen University, Wageningen.

Flor, H. H. (1971). Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology* **9**, 275-296.

Fournet, S., Eoche-Bosy, D., Renault, L., Hamelin, F. M., and Montarry, J. (2016). Adaptation to resistant hosts increases fitness on susceptible hosts in the plant parasitic nematode Globodera pallida. *Ecol Evol* **6**, 2559-68.

Fournet, S., Kerlan, M. C., Renault, L., Dantec, J. P., Rouaux, C., and Montarry, J. (2013). Selection of nematodes by resistant plants has implications for local adaptation and cross-virulence. *Plant Pathology* **62**, 184-193.

Futschik, A., and Schlotterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* **186**, 207-18.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907 [q-bio.GN]*.

Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics* **201**, 1555-79.

Gebhardt, C., Mugniery, D., Ritter, E., Salamini, F., and Bonnel, E. (1993). Identification of RFLP markers closely linked to the H1 gene conferring resistance to Globodera rostochiensis in potato *Theor Appl Genet* **85**, 541-544.

Gebhardt, C., and Valkonen, J. P. (2001). Organization of genes controlling disease resistance in the potato genome. *Annu. Rev. Phytopathol.* **39**, 79-102.

Gendron St-Marseille, A. F., Lord, E., Veronneau, P. Y., Brodeur, J., and Mimee, B. (2018). Genome Scans Reveal Homogenization and Local Adaptations in Populations of the Soybean Cyst Nematode. *Front Plant Sci* **9**, 987.

Geric Stare, B., Fouville, D., Širca, S., Gallot, A., Urek, G., and Grenier, E. (2011). Molecular Variability and Evolution of the Pectate Lyase (pel-2) Parasitism Gene in Cyst Nematodes Parasitizing Different Solanaceous Plants. *Journal of Molecular Evolution* **72**, 169-181.

Gheysen, G., and Mitchum, M. G. (2011). How nematodes manipulate plant development pathways for infection. *Curr Opin Plant Biol* **14**, 415-421.

Goellner, M., Smant, G., De Boer, J. M., Baum, T., and Davis, E. L. (2000). Isolation of beta-1,4-endoglucanase genes from Globodera tabacum and their expression during parasitism. *Journal of Nematology* **32**, 154-165.

Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talon, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420-3435.

Government, S. (2016). Farmland used for potatoes in Scotland. The Scottish Government.

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

Goverse, A., and Smant, G. (2014). The activation and suppression of plant innate immunity by parasitic nematodes. *Annu Rev Phytopathol* **52**, 243-65.

Griffiths, B. S., Robertson, W. M., and Trudgill, D. L. (1982). Nuclear changes induced by the nematodes Xiphinema diversicaudatum and Longidorus elongatus in root-tips of perennial ryegrass, Lolium perenne. *Histochemical Journal* **14**, 719-730.

Grunewald, W., van Noorden, G., Van Isterdael, G., Beeckman, T., Gheysen, G., and Mathesius, U. (2009). Manipulation of Auxin Transport in Plant Roots during Rhizobium Symbiosis and Nematode Parasitism. *The Plant Cell* **21**, 2553.

Gunther, T., and Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205-20.

Guo, Y., Ni, J., Denver, R., Wang, X., and Clark, S. E. (2011). Mechanisms of molecular mimicry of plant CLE peptide ligands by the parasitic nematode Globodera rostochiensis. *Plant Physiol* **157**, 476-84.

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

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp.* **41**, 95-98.

Hamamouch, N., Li, C., Hewezi, T., Baum, T. J., Mitchum, M. G., Hussey, R. S., Vodkin, L. O., and Davis, E. L. (2012). The interaction of the novel 30C02 cyst nematode effector protein with a plant β-1,3-endoglucanase may suppress host defence to promote parasitism. *Journal of Experimental Botany* **63**, 3683-3696.

Hawkes, J. G., and Francisco-Ortega, J. (1993). The early history of the potato in Europe. *Euphytica* **70**, 1-7.

Hearne, R., Lettice, E. P., and Jones, P. W. (2017). Interspecific and intraspecific competition in the potato cyst nematodes Globodera pallida and G. rostochiensis. *Nematology* **19**, 463-475.

Hellens, R., Edwards, A., Leyland, N., Bean, S., and Mullineaux, P. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Molecular Biology* **42**, 819-832.

Hijmans, R. J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., and Spooner, D. M. (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (Solanum section Petota). *Global Ecology and Biogeography* **16**, 485-495.

Hijmans, R. J., and Spooner, D. M. (2001). Geographic distribution of wild potato species. *American Journal of Botany* **88**, 2101-2112.

Hockland, S., Niere, B., Grenier, E., Blok, V., Phillips, M., den Nijs, L., Anthoine, G., Pickup, J., and Viaene, N. (2012). An evaluation of the implications of virulence in non-European populations of Globodera pallida and G. rostochiensis for potato cultivation in Europe. *Nematology* **14**, 1-13.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics (Oxford, England)* **32**, 767-769.

Howe, K., Bolt, B., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite - a comprehensive resource for helminth genomics.

Hughes, J., Reay, G., Thomas, L., and Walker, A. (2010). Pesticide use in Scottish potato cultivation 1998-2008: Trends and predictions. *Proceedings Crop Protection in Northern Britain*, 245-250.

Huijsman, C. A. (1955). Breeding for resistance to the potato root eelworm. *Euphytica* **4**, 133-140.

Hussey, R., and Grundler, F. (1998). Nematode Parasitism of Plants. *In* "The Physiology and Biochemistry of Free-living and Plant-parasitic Nematodes" (R. N. Perry and K. M. Wright, eds.). CABI Publishing.

Jacobs, J. M., van Eck, H. J., Horsman, K., Arens, P., Verkerk-Bakker, B., Jacobsen, E., Pereira, A., and Stiekema, W. J. (1996). Mapping of resistance to the potato cyst nematode

Globodera rostochiensis from the wild potato species Solanum vernei. *Mol Breed* **2**, 51-60.

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

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

Jones, J. T., Furlanetto, C., and Kikuchi, T. (2005). Horizontal gene transfer from bacteria and fungi as a driving force in the evolution of plant parasitism in nematodes. *Nematology* **7**, 641-646.

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

Jones, J. T., Kumar, A., Pylypenko, L. A., Thirugnanasambandam, A., Castelli, L., Chapman, S., Cock, P. J., Grenier, E., Lilley, C. J., Phillips, M., and Blok, V. (2009). Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode Globodera pallida. *Molecular Plant Pathology* **10**, 815-828.

Jones, M. G., and Northcote, D. H. (1972). Nematode-induced syncytium-A multinucleate transfer cell. *J Cell Sci.* **10**, 789-809.

Jones, P. W., Tylka, G. L., and Perry, R. N. (1998). Hatching. *In* "The Physiology and Biochemistry of Free-living and Plant-parasitic Nematodes" (R. N. Perry and D. J. Wright, eds.). CABI Publishing.

Jupe, F., Chen, X., Verweij, W., Witek, K., Jones, J. D., and Hein, I. (2014). Genomic DNA Library Preparation for Resistance Gene Enrichment and Sequencing (RenSeq) in Plants. *In* "Plant Pathogen Interactions: Methods and Protocols" (P. R. Birch, J. T. Jones and J. I. Bos, eds.). Humana Press.

Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G. J., Maclean, D., Cock, P. J., Leggett, R. M., Bryan, G. J., Cardle, L., Hein, I., and Jones, J. D. (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.

Juvale, P. S., and Baum, T. J. (2018). "Cyst-ained" research into Heterodera parasitism. *PLoS Pathog* **14**, e1006791.

Kaczmarek, A., MacKenzie, K., Kettle, H., and Blok, V. (2014). Influence of soil temperature on Globodera rostochiensis and Globodera pallida. *Phytopathologia Mediterranea* **53**, 396-405.

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

Kammonen, J. I., Smolander, O.-P., Paulin, L., Pereira, P. A. B., Laine, P., Koskinen, P., Jernvall, J., and Auvinen, P. (2017). gapFinisher: a reliable gap filling pipeline for SSPACE-LongRead scaffolder output. *PeerJ Preprints* **5**, e3467v1.

Kapila, J., de Rycke, R., van Montagu, M., and Angenon, G. (1997). An Agrobacterium-mediated transient gene expression system for intact leaves. *Plant Science* **122**, 101-108.

Kaymak, E., Wee, L. M., and Ryder, S. P. (2010). Structure and function of nematode RNA-binding proteins. *Current opinion in structural biology* **20**, 305-312.

Keeling, P. (2004). Reduction and compaction in the genome of the Apicomplexan parasite Cryptosporidium parvum. *Developmental Cell*, 614-615.

Kikuchi, T., Shibuya, H., and Jones, J. T. (2005). Molecular and biochemical characterization of an endo-β-1,3-glucanase from the pinewood nematode Bursaphelenchus xylophilus acquired by horizontal gene transfer from bacteria. *Biochem. J.* **389**, 117-125.

Klug, A. (2010). The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annual Review of Biochemistry* **79**, 213-231.

Knoetze, R., Swart, A., Wentzel, R., and Tiedt, L. R. (2017). Description of Globodera agulhasensis n. sp. (Nematoda: Heteroderidae) from South Africa. *Nematology* **19**, 305-322.

Koncz, C., and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Mol Gen Genet* **204**, 383-396.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* **27**, 722-736.

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

Kumar, M., Gantasala, N. P., Roychowdhury, T., Thakur, P. K., Banakar, P., Shukla, R. N., Jones, M. G., and Rao, U. (2014). De novo transcriptome sequencing and analysis of the cereal cyst nematode, Heterodera avenae. *PLoS One* **9**, e96311.

Laetsch, D., and Blaxter, M. (2017). BlobTools: Interrogation of genome assemblies [version 1; referees: 2 approved with reservations]. *F1000Research* **6**.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie2. *Nature Methods* **9**, 357-359.

Lažetić, V., and Fay, D. S. (2017). Conserved ankyrin repeat proteins and their NIMA kinase partners regulate extracellular matrix remodeling and intracellular trafficking in Caenorhabditis elegans. *Genetics* **205**, 273-293.

Lee, C., Chronis, D., Kenning, C., Peret, B., Hewezi, T., Davis, E. L., Baum, T. J., Hussey, R., Bennett, M., and Mitchum, M. G. (2011). The Novel Cyst Nematode Effector Protein 19C07 Interacts with the Arabidopsis Auxin Influx Transporter LAX3 to Control Feeding Site Development. *Plant Physiology* **155**, 866.

Lee, D. (2002). Life Cycles. *In* "The Biology of Nematodes" (D. Lee, ed.). CRC Press.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754-1760.

Li, H., Handsaker, B., and Wysoker, A. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

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

Lilley, C. J., Maqbool, A., Wu, D., Yusup, H. B., Jones, L. M., Birch, P. R. J., Banfield, M. J., Urwin, P. E., and Eves-van den Akker, S. (2018). Effector gene birth in plant parasitic nematodes: Neofunctionalization of a housekeeping glutathione synthetase gene. *PLoS Genet* **14**, e1007310.

Lozano-Torres, J. L., Wilbers, R. H., Gawronski, P., Boshoven, J. C., Finkers-Tomczak, A., Cordewener, J. H., America, A. H., Overmars, H. A., Van 't Klooster, J. W., Baranowski, L., Sobczak, M., Ilyas, M., van der Hoorn, R. A., Schots, A., de Wit, P. J., Bakker, J., Goverse, A., and Smant, G. (2012). Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proc Natl Acad Sci U S A* **109**, 10119-24.

Lozano-Torres, J. L., Wilbers, R. H., Warmerdam, S., Finkers-Tomczak, A., Diaz-Granados, A., van Schaik, C. C., Helder, J., Bakker, J., Goverse, A., Schots, A., and Smant, G. (2014). Apoplastic venom allergen-like proteins of cyst nematodes modulate the activation of basal plant innate immunity by cell surface receptors. *PLoS Pathog* **10**, e1004569.

Lozano, J., and Smant, G. (2011). Survival of plant-parasitic nematodes inside the host. *In* "Molecular and Physiological Basis of Nematode Survival" (R. N. Perry and D. A. Wharton, eds.). CAB International.

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

Lu, S. W., Tian, D., Borchardt-Wier, H. B., and Wang, X. (2008). Alternative splicing: a novel mechanism of regulation identified in the chorismate mutase gene of the potato cyst nematode Globodera rostochiensis. *Mol Biochem Parasitol* **162**, 1-15.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., and Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**, 981.

Ma, L., van den Burg, H., Cornelissen, B., and Takken, F. L. (2013). Molecular basis of effector recognition by plant NB-LRR proteins. *In* "Molecular Plant Immunity" (G. Sessa, ed.). John Wiley & Sons.

Machida-Hirano, R. (2015). Diversity of potato genetic resources. *Breed Sci* **65**, 26-40.

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., and Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nat Methods* **7**, 111-8.

Manosalva, P., Manohar, M., von Reuss, S. H., Chen, S., Koch, A., Kaplan, F., Choe, A., Micikas, R. J., Wang, X., Kogel, K. H., Sternberg, P. W., Williamson, V. M., Schroeder, F. C., and Klessig, D. F. (2015). Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. *Nat Commun* **6**, 7795.

Mburu, H., Cortada, L., Mwangi, G., Gitau, K., Kiriga, A., Kinyua, Z. M., Ngundo, G., Ronno, W., Coyne, D. L., Holgado, R., and Haukeland, S. (2018). First report of potato cyst nematode Globodera pallida (Stone, 1973) infecting potato (Solanum tuberosum L.) in Kenya. *Plant Disease*.

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

Mei, Y., Wright, K. M., Haegeman, A., Bauters, L., Diaz-Granados, A., Goverse, A., Gheysen, G., Jones, J. T., and Mantelin, S. (2018). The Globodera pallida SPRYSEC Effector GpSPRY-414-2 That Suppresses Plant Defenses Targets a Regulatory Component of the Dynamic Microtubule Network. *Front Plant Sci* **9**, 1019.

Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., Shaw, P. D., and Marshall, D. (2012). Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics* **14**, 193-202.

Mimee, B., Duceppe, M. O., Veronneau, P. Y., Lafond-Lapalme, J., Jean, M., Belzile, F., and Belair, G. (2015). A new method for studying population genetics of cyst nematodes based on Pool-Seq and genomewide allele frequency analysis. *Mol Ecol Resour* **15**, 1356-65.

Minnis, S., Haydock, P., Ibrahim, S., Grove, I., Evans, K., and Russel, M. (2002). Potato cyst nematodes in England and Wales - occurence and distribution. *Ann. appl. Biol.* **140**, 187-195.

Monaghan, J., and Zipfel, C. (2012). Plant pattern recognition receptor complexes at the plasma membrane. *Curr Opin Plant Biol* **15**, 349-57.

Mosavi, L. K., Cammett, T. J., Desrosiers, D. C., and Peng, Z.-Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein science : a publication of the Protein Society* **13**, 1435-1448.

Mwangi, J. M., Kariuki, G. M., Waceke, J. W., and Grundler, F. M. (2015). First report ofGlobodera rostochiensisinfesting potatoes in Kenya. *New Disease Reports* **31**, 18.

Nakamura, A., Yasuda, K., Adachi, H., Sakurai, Y., Ishii, N., and Goto, S. (1999). Vitellogenin-6 is a major carbonylated protein in aged nematode, Caenorhabditis elegans. *BBRC* **264**, 580-583.

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

Noon, J. B., Hewezi, T., Maier, T. R., Simmons, C., Wei, J. Z., Wu, G., Llaca, V., Deschamps, S., Davis, E. L., Mitchum, M. G., Hussey, R. S., and Baum, T. J. (2015). Eighteen New Candidate Effectors of the Phytonematode Heterodera glycines Produced Specifically in the Secretory Esophageal Gland Cells During Parasitism. *Phytopathology* **105**, 1362-72.

Nurnberger, T., and Kemmerling, B. (2009). Pathogen-Associated Molecular Patterns (PAMP) and PAMP-Triggered Immunity. *In* "Annual Plant Reviews (vol. 34): Molecular Aspects of Plant Disease Resistance" (J. Parker, ed.). Blackwell Publishing.

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

Palomares-Rius, J. E., Hedley, P. E., Cock, P. J., Morris, J. A., Jones, J. T., Vovlas, N., and Blok, V. (2012). Comparison of transcript profiles in different life stages of the nematode Globodera pallida under different host potato genotypes. *Mol Plant Pathol* **13**, 1120-34.

Parisi, M., and Lin, H. (2000). Translational repression: A duet of Nanos and Pumilio. *Current Biology* **10**, R81-R83.

Perfetto, L., Gherardini, P. F., Davey, N. E., Diella, F., Helmer-Citterich, M., and Cesareni, G. (2013). Exploring the diversity of SPRY/B30.2-mediated interactions. *Trends Biochem Sci* **38**, 38-46.

Perry, R. N. (2002). Hatching. *In* "The Biology of Nematodes" (D. Lee, ed.). Taylor & Francis.

Perry, R. N., and Moens, M. (2011a). Introduction to plant-parasitic nematodes; Modes of parasitism. *In* "Genomics and Molecular Genetics of Plant-Nematode Interactions" (J. T. Jones, G. Gheysen and C. Fenol, eds.). Springer.

Perry, R. N., and Moens, M. (2011b). Survival of Parasitic Nematodes outside the Host. *In* "Molecular And Physiological Basis of Nematode Survival" (R. N. Perry and D. A. Wharton, eds.). CAB International

Petersen, T., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: Discriminating singal peptides from transmembrane regions. *2011* **8**, 785-786.

Phillips, M., and Blok, V. (2008). Selection for reproductive ability in Globodera pallida populations in relation to quantitative resistance from Solanum vernei and S. tuberosum ssp. andigena CPC2802. *Plant Pathology* **57**, 573-580.

Phillips, M., and Trudgill, D. L. (1998). Variation of virulence, in terms of quantitative reproduction of Globodera pallida populations, from Europe and South America, in relation to resistance from Solanum vernei and S. tuberosum ssp. andigena CPC2802. *Nematologica* **44**, 409-423.

Phillips, W., Howe, D., Brown, A., Eves-van den Akker, S., Dettwyler, L., Peetz, A., Denver, D., and Zasada, I. (2017). The draft genome of Globodera ellingtonae. *Journal of Nematology* **49**, 127-128.

Picard, D., Sempere, T., and 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. *Mol Phylogenet Evol* **42**, 308-316.

Plantard, O., Picard, D., Valette, S., Scurrah, M., Grenier, E., and 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. *Mol Ecol* **17**, 2208-18.

Postma, W., Slootweg, E. J., Rehman, S., Finkers-Tomczak, A., Tytgat, T., van Gelderen, K., Lozano, J., Roosien, J., Pomp, R., van Schaik, C., Bakker, J., Goverse, A., and Smant, G. (2012). The effector SPRYSEC-19 of Globodera rostochiensis suppresses CC-NB-LRR-mediated disease resistance in plants. *Plant Physiol* **160**, 944-954.

Qin, L., Overmars, H. A., Helder, J., Popeijus, H., van der Voort, J. R., Groenink, W., van Koert, P., Schots, A., Bakker, J., and Smant, G. (2000). An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode Globodera rostochiensis. *MPMI* **13**, 830-836.

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

Raffaele, S., Win, J., Cano, L. M., and Kamoun, S. (2010). Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of Phytophthora infestans. *BMC Genomics* **11**, 637.

Rehman, S., Butterbach, P., Popeijus, H., Overmars, H. A., Davis, E. L., Jones, J. T., Goverse, A., Bakker, J., and Smant, G. (2009). Identification and characterization of the most abundant cellulases in stylet secretions from Globodera rostochiensis. *Nematology* **99**, 194-202.

Rigney, B., Blok, V., Griffin, D., Dalton, E., and Milbourne, D. (2017). Consistent action of two partially effective loci conferring resistance to Globodera pallida Pa2/3 across multiple nematode field populations. *Plant Pathology* **66**, 1031-1040.

Roach, M. J., Schmidt, S. A., and Borneman, A. R. (2018). Purge Haplotigs: Synteny Reduction for Third-gen Diploid Genome Assemblies. *bioRxiv*, 286252.

Roberts, R. J., Carneiro, M. O., and Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome biology* **14**, 405-405.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139-140.

Rojo, E., Sharma, S. K., Kovaleva, V., N., R., and Fletcher, J. (2002). CLV3 is localized to the extracellular space, where it activates the Arabidopsis CLAVATA stem cell signaling pathway. *The Plant Cell Online* **14**, 969-977.

Rooney, H. C., Van 't Klooster, J. W., Van der Hoorn, R. A., Joosten, M. H., Jones, J. D., and De Wit, P. J. (2005). Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science* **308**, 1783-1786.

Rosso, M. N., and Grenier, E. (2011). Other nematode effectors and evolutionary constrains. *In* "Genomics and Molecular Genetics of Plant-Nematode Interactions" (J. T. Jones, G. Gheysen and C. Fenoll, eds.). Springer.

Sacco, M. A., Koropacka, K., Grenier, E., Jaubert, M. J., Blanchard, A., Goverse, A., Smant, G., and Moffett, P. (2009). The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death. *PLoS Pathog* **5**.

SASA (2018). Seed & Ware Potatoes. Science and Advice for Scottish Agriculture.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212.

Smant, G., Goverse, A., Stokkermans, J., de Boer, J., Pomp, R., Zilverentant, J., Overmars, H. A., Helder, J., Schots, A., and Bakker, J. (1997). Potato root diffusate-induced secretion of soluble, basic proteins originating from the subventral esophageal glands of potato cyst nematodes. *Nematology* **87**, 839-845.

Smant, G., Stokkermans, J., Yan, Y., de Boer, J., Baum, T., Wang, X., Hussey, R., Gommers, F. J., Henrissat, B., Davis, E. L., Helder, J., Schots, A., and Bakker, J. (1998). Engenous cellulases in animals: Isolation of β-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc Natl Acad Sci U S A* **95**, 4906-4911.

Sobczak, M., and Golinowski, W. (2011). Cyst Nematodes and Syncytia. *In* "Genomics and Molecular Genetics of Plant-Nematode Interactions" (J. T. Jones, G. Gheysen and C. Fenoll, eds.). Springer.

Sobczak, M., Golinowski, W., and Grundler, F. (1997). Changes in the structure of Arabidopsis thaliana roots induced during development of males of the plant parasitic nematode Heterodera schachtii. *European Journal of Plant Pathology* **103**, 113-124.

Staiger, D., and Brown, J. W. S. (2013). Alternative Splicing at the Intersection of Biological Timing, Development, and Stress Responses. *The Plant Cell* **25**, 3640.

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic acids research* **33**, W465-W467.

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC bioinformatics* **7**, 62-62.

Stone, A. R. (1972). Heterodera pallida n.sp. (Nematoda: Heteroderidae), a second species of potato cyst nematode. *Nematologica* **18**, 591-606.

Strachan, S. M. (2018). Characterisation of the potato H2 resistance gene against Globodera pallida. PhD Thesis, University of St Andrews, St Andrews.

Tanaka, M., Izawa, T., Yamate, J., Franklin, R. J. M., Kuramoto, T., Serikawa, T., and Kuwamura, M. (2014). The VF rat with abnormal myelinogenesis has a mutation in Dopey1. *Glia* **62**, 1530-1542.

Thines, M., and Kamoun, S. (2010). Oomycete-plant coevolution: recent advances and future prospects. *Curr Opin Plant Biol* **13**, 427-33.

Thomma, B. P., Nurnberger, T., and Joosten, M. H. (2011). Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *Plant Cell* **23**, 4-15.

Thorpe, P., Mantelin, S., Cock, P. J., Blok, V., Coke, M. C., Eves-van den Akker, S., Guzeeva, E., Lilley, C. J., Smant, G., Reid, A. J., Wright, K. M., Urwin, P. E., and Jones, J. T. (2014). Genomic characterization of the effector complement of the potato cyst nematode Globodera pallida. *BMC Genomics* **15**, 923.

Trudgill, D. L., Elliott, M. J., Evans, K., and Phillips, M. (2003). The white potato cyst nematode (Globodera pallida) - a critical analysis of the threat in Britain. *Ann. appl. Biol.* **143**, 73-80.

Trudgill, D. L., Phillips, M. S., and Elliott, M. J. (2014). Dynamics and management of the white potato cyst nematodeGlobodera pallidain commercial potato crops. *Annals of Applied Biology* **164**, 18-34.

Turner, S. J., and Fleming, C. C. (2002). Multiple selection of potato cyst nematode Globodera pallida virulence on a range of potato species. I.Serial selection on Solanum-hybrids. *European Journal of Plant Pathology* **108**, 461-467.

Turner, S. J., and Stone, A. R. (1984). Development of potaty cyst nematodes in roots of resistant Solanum tuberosum ssp. andigena and S. vernei hybrids. *Nematologica* **30**, 324-332.

Turner, S. J., Stone, A. R., and Perry, J. N. (1983). Selection of potato cyst-nematodes on resistant Solanum vernei hybrids. *Euphytica* **32**, 911-917.

Turner, S. J., and Subbotin, S. A. (2013). Cyst nematodes. *In* "Plant Nematology" (R. N. Perry and M. Moens, eds.). CAB International.

USDA Worldwide Distribution of Nematodes. United States Department of Agriculture.

van der Hoorn, R. A., and Kamoun, S. (2008). From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* **20**, 2009-17.

van der Hoorn, R. A., Laurent, F., Roth, R., and De Wit, P. J. (2000). Agrioinfiltration is a versitile tool that facilitates comparative analyses of Avr9/Cf-9-induced and Avr4/Cf-9-induced necrosis. *MPMI* **13**, 439-446.

van der Voort, J. R., van der Vossen, E., Bakker, E., Overmars, H. A., van Zandvoort, P., R., H., Klein-Lankhorst, R., and Bakker, J. (2000). Two additive QTLs conferring broad-spectrum resistance in potato to Globodera pallida are localized on resistance gene clusters. *Theor Appl Genet* **10**, 1122-1130.

van der Voort, R. J., Lindeman, W., Foklertsma, R., Hutten, R., Overmars, H., van der Vossen, E., Jacobsen, E., and Bakker, J. (1998). A QTL for broad-spectrum resistance to cyst

nematode species (Globodera spp.) maps to a resistance gene cluster in potato. *Theor Appl Genet* **96**, 654-661.

van der Voort, R. J., Wolters, P., Folkertsma, R., Hutten, R., van Zandvoort, P., Vinke, J. H., Kanyuka, K., Bendahmane, A., Jacobsen, E., Janssen, R., and Bakker, J. (1997). Mapping of the cyst nematode resistance locus Gpa2 in potato using a strategy based on comigrating AFLP markers. *Theor Appl Genet* **95**, 874-880.

van Eck, H. J., Vos, P. G., Valkonen, J. P., Uitdewilligen, J. G., Lensing, H., de Vetten, N., and Visser, R. G. (2017). Graphical genotyping as a method to map Ny (o,n)sto and Gpa5 using a reference panel of tetraploid potato cultivars. *Theor Appl Genet* **130**, 515-528.

van Esse, H. P., Van't Klooster, J. W., Bolton, M. D., Yadeta, K. A., van Baarlen, P., Boeren, S., Vervoort, J., de Wit, P. J., and Thomma, B. P. (2008). The Cladosporium fulvum virulence protein Avr2 inhibits host proteases required for basal defense. *Plant Cell* **20**, 1948-63.

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

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., Armstrong, M. R., McKenzie, G., Bryan, G. J., and Hein, I. (2016). Utilizing "Omic" Technologies to Identify and Prioritize Novel Sources of Resistance to the Oomycete Pathogen Phytophthora infestans in Potato Germplasm Collections. *Front Plant Sci* **7**, 672.

Vanholme, B., W, V. A. N. T., Vanhouteghem, K., J, D. E. M., Cannoot, B., and Gheysen, G. (2007). Molecular characterization and functional importance of pectate lyase secreted by the cyst nematode Heterodera schachtii. *Mol Plant Pathol* **8**, 267-78.

Vleeshouwers, V. G., Driesprong, J. D., Kamphuis, L. G., Torto-Alalibo, T., Van't Slot, K. A., Govers, F., Visser, R. G., Jacobsen, E., and Kamoun, S. (2006). Agroinfection-based high-throughput screening reveals specific recognition of INF elicitins in Solanum. *Mol Plant Pathol* **7**, 499-510.

Wagner, B., Fuchs, H., Adhami, F., Ma, Y., Scheiner, O., and Breiteneder, H. (2004). Plant virus expression systems for transient production of recombinant allergens in Nicotiana benthamiana. *Methods* **32**, 227-34.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963.

Wang, J., Replogle, A., Hussey, R., Baum, T., Wang, X., Davis, E. L., and Mitchum, M. G. (2011). Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode Heterodera schachtii. *Mol Plant Pathol* **12**, 177-86.

Wang, X.-R., Moreno, Y., Wu, H.-R., Ma, C., Li, Y., Zhang, J., Yang, C., Sun, S., Ma, W., and Geary, T. (2012). Proteomic profiles of soluble proteins from the esophageal gland in female Meloidogyne incognita. *International Journal for Parasitology* **42**, 1177-1183.

Whitehead, A. G., and Turner, S. J. (1998). Management and regulatory control strategies for potato cyst nematodes (Globodera rostochiensis and Globodera pallida). *In* "Potato Cyst Nematodes" (R. J. Marks, ed.). CAB International.

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

Wroblewski, T., Tomczak, A., and Michelmore, R. (2005). Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. *Plant Biotechnol J* **3**, 259-73.

Wyss, U. (1992). Observations on the feeding behaviour of Heterodera schachtii throughout development, including events during moulting. *Fundam. appl. Nematol.* **15**, 75-89.

Wyss, U., and Zunke, U. (1986). Observations on the behavior of second stage juveniles of Heterodera schachtii inside host roots. *Revue Nematol* **9**, 153-165.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., Wang, J., Orjeda, G., Guzman, F., Torres, M., Lozano, R., Ponce, O., Martinez, D., De la Cruz, G., Chakrabarti, S. K., Patil, V. U., Skryabin, K. G., Kuznetsov, B. B., Ravin, N. V., Kolganova, T. V., Beletsky, A. V., Mardanov, A. V., Di Genova, A., Bolser, D. M., Martin, D. M., Li, G., Yang, Y., Kuang, H., Hu, Q., Xiong, X., Bishop, G. J., Sagredo, B., Mejia, N., Zagorski, W., Gromadka, R., Gawor, J., Szczesny, P., Huang, S., Zhang, Z., Liang, C., He, J., Li, Y., He, Y., Xu, J., Zhang, Y., Xie, B., Du, Y., Qu, D., Bonierbale, M., Ghislain, M., Herrera Mdel, R., Giuliano, G., Pietrella, M., Perrotta, G., Facella, P., O'Brien, K., Feingold, S. E., Barreiro, L. E., Massa, G. A., Diambra, L., Whitty, B. R., Vaillancourt, B., Lin, H., Massa, A. N., Geoffroy, M., Lundback, S., DellaPenna, D., Buell, C. R., Sharma, S. K., Marshall, D. F., Waugh, R., Bryan, G. J., Destefanis, M., Nagy, I., Milbourne, D., Thomson, S. J., Fiers, M., Jacobs, J. M., Nielsen, K. L., Sonderkaer, M., Iovene, M., Torres, G. A., Jiang, J., Veilleux, R. E., Bachem, C. W., de Boer, J., Borm, T., Kloosterman, B., van Eck, H., Datema, E., Hekkert, B., Goverse, A., van Ham, R. C., and Visser, R. G. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189-95.

Yan, Y., Smant, G., Stokkermans, J., Qin, L., Helder, J., Baum, T., Schots, A., and Davis, E. (1998). Genomic organization of four β-1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications. *Gene* **220**, 61-70.

Yang, X., Deng, F., and Ramonell, K. M. (2012). Receptor-like kinases and receptor-like proteins: keys to pathogen recognition and defense signaling in plant innate immunity. *Frontiers in Biology* **7**, 155-166.

Yoda, A., Oishi, I., and Minami, Y. (2003). Expression and function of the Ror-family receptor tyrosine kinases during development: Lessons from genetic analyses of nematodes, mice and humans. *Journal of Receptors and Signal Transduction* **23**.

Yu, H., Chronis, D., Lu, S., and Wang, X. (2010). Chorismate mutase: an alternatively spliced parasitism gene and a diagnostic marker for three important Globodera nematode species. *European Journal of Plant Pathology* **129**, 89-102.

Zipfel, C. (2008). Pattern-recognition receptors in plant innate immunity. *Curr Opin Immunol* **20**, 10-6.

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E., Jones, J. D., Felix, G., and Boller, T. (2004). Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature* **428**, 764-797.