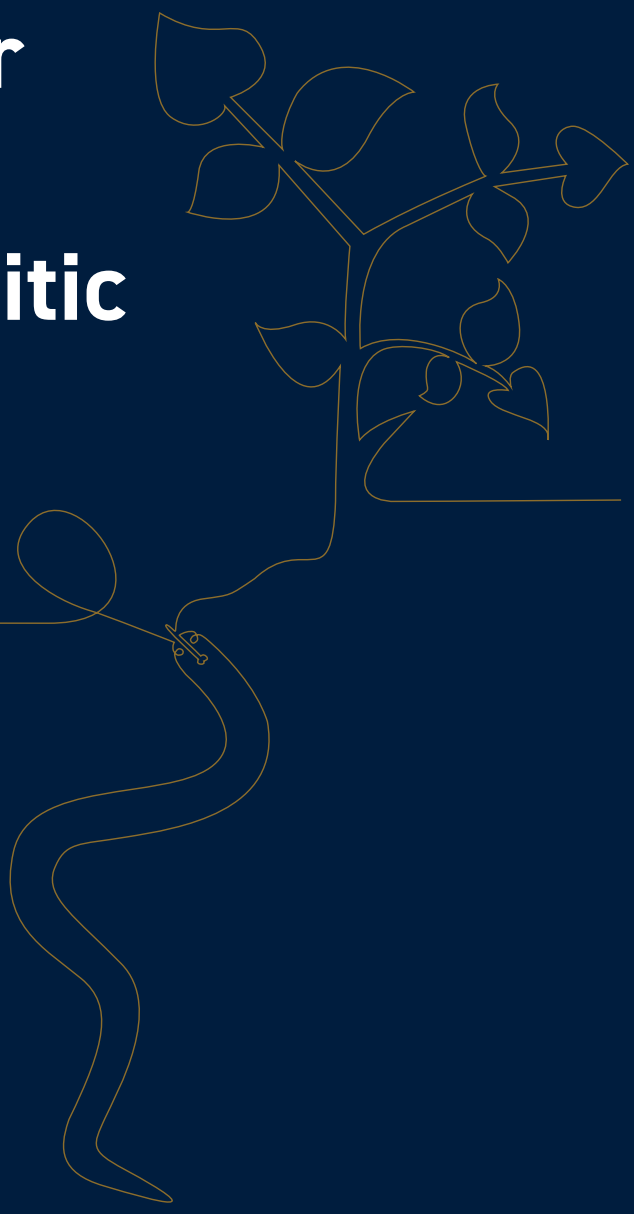
Genomic Insights into effector diversity in Plant-parasitic Nematodes



Joris J.M. van Steenbrugge

Propositions

1. Protein structure analyses will supersede protein sequence analyses in the next five years.
   (this thesis)
2. Nematode promoter regions that orchestrate effector expression might exist but previously proposed DOG boxes fulfil other roles.
   (this thesis)
3. It is unethical to implement quantum computing without adequate data encryption standards.
4. Preparation for the impact of non-human intelligence requires the development of exo-consciousness as a scientific field.
5. Altering the human gut microbiome as a medical treatment will increase inequality.
6. Switching from a numerical to a binary grading system benefits the Dutch school system.

Propositions belonging to the thesis, entitled

Genomic Insights into effector diversity in Plant-parasitic Nematodes

Joris J.M. van Steenbrugge

Wageningen, 13 March 2024

Genomic Insights into effector diversity in Plant-parasitic Nematodes

Joris J.M. van Steenbrugge

Thesis committee

Promotors
Prof. Dr Geert Smant
Professor of Nematology
Wageningen University & Research

Dr Johannes Helder
Associate Professor, Laboratory of Nematology
Wageningen University & Research

Co-promotor
Dr Mark G. Sterken
Assistant Professor, Laboratory of Nematology
Wageningen University & Research

Other members
Prof. Dr Dick de Ridder, Wageningen University & Research
Dr Vivianne Vleeshouwers, Wageningen University & Research
Dr Basten Snoek, Utrecht University
Dr Josselin Montarry, INRAE, Rennes, France

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences

Genomic Insights into effector diversity in Plant-parasitic Nematodes

Joris J.M. van Steenbrugge

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 13 March 2024
at 4 p.m. in the Omnia Auditorium.

Joris J.M. van Steenbrugge
Genomic Insights into effector diversity in Plant-parasitic Nematodes,
198 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English

ISBN 978-94-6447-986-7
DOI https://doi.org/10.18174/642305

Contents

CHAPTER 1

General Introduction

Joris J.M. van Steenbrugge[1]

[1] Laboratory of Nematology, Wageningen University and Research, The Netherlands

1.1 Plant parasitic nematodes and Control strategies

As the human population continues to expand, there are increasingly significant concerns regarding the safeguarding of food security. Various factors, including climate change, pathogens, and pests [Cole et al., 2018], pose additional challenges to food production. Among the most destructive pests are plant-parasitic nematodes, small roundworms that primarily target plant roots [Jones et al., 2013]. In agro-ecosystems, these nematodes are responsible for significant annual yield losses of approximately 12.3% [Nicol et al., 2011, Singh et al., 2015]. While control measures have previously managed the damages, the phasing out of nematicides [Commission, 2022] combined with the rising global population intensifies the demand for food security [Cole et al., 2018]. Therefore, addressing the threat posed by nematodes is crucial for establishing a stable and secure food production system. To achieve this, it is essential to acquire knowledge about the nematodes' parasitic abilities and effective strategies to combat them.

Most of the crop yield losses caused by plant-parasitic nematodes are caused by root-knot nematodes (*Meloidogyne* spp.) and cyst nematodes (*Globodera* spp. and *Heterodera* spp.). They are sedentary endoparasites, necessitating their migration into the host to feed and complete their life cycle. Root-knot nematodes tend to have a relatively large host range. For example, *M. chitwoodi* can parasitise multiple economically important plant species, including the food crops potato, wheat and tomato [Santo, 1980]. Cyst nematodes generally have a smaller host range. For example, the potato cyst nematode *Globodera pallida* can only infect several solanaceous species, of which its impact on potato is economically the most relevant [Price et al., 2021]. Cyst nematodes are most commonly present in temperate climate zones, while root-knot nematodes are typically present in (sub)tropical climates. Once settled in a field, they are nearly impossible to exterminate. Several strategies to control a plant-parasitic nematode population in the field exist, and among the most widely used are:

- **Nematicides** - Probably the most widely used method to control plant-parasitic nematodes is the application of chemical nematicides. The application of nematicides to the soil has many adverse side effects on human health, and the environment. [World Meteorological Organization, 2003], and are therefore unfavourable as long-term solutions [Commission, 2022]. In addition, no nematicide is effective enough to eradicate a population from a field completely, which requires regular reapplication of the pesticide to maintain high yields [Desaeger et al., 2020].
- **Crop rotation** - Additionally, crop rotation is used to limit the growth of a nematode population in the field [Bullock, 1992]. Crop rotation works by rotating between

suitable and non-suitable host crops every few years. As a result, the nematode is regularly deprived of a suitable host for feeding and reproduction. However, this method does have a few drawbacks as well. Firstly, the availability of unsuitable hosts may be limited, depending on the specific nematode species present in the soil. The broader the range of hosts that the nematode can infect, the more challenging it becomes to implement effective crop rotation strategies. Secondly, this method does not directly exterminate individuals, but limits the population size through natural decline. This means that when an unsuitable host is present in the field, a proportion of the population remains in the soil. This is especially an issue for cyst nematodes, as they are known to stay dormant in the soil for up to 20 years [Evans and Stone, 1977]. Consequently, relying solely on crop rotation as a control method will only serve to postpone the issue. Crop rotation is therefore commonly combined with other control strategies.

- **Resistant crops** - An effective method that is relatively straightforward to apply for growers is the usage of resistant crop varieties. Resistance is usually conferred by a specific allele of a gene or multiple genes that target a subset of plant-parasitic nematode species. In commercial crop varieties, this resistance is often derived from wild varieties. These wild varieties obtained this resistance through mutations and recombination events, driven by natural selection [Williamson and Kumar, 2006]. Resistance proteins typically serve as immune receptors that detect the presence of nematodes and initiate a hypersensitive response, thereby inhibiting the formation of a feeding site and the reproduction of the nematode [Castagnone-Sereno, 2002]. A notable concern regarding the reliance on resistant varieties as a control strategy is the uncertainty of their durability [Davies and Elling, 2015]. For example, a resistance gene such as *Mi-1* in tomato, which confers resistance against *M. incognita*, *M. arenaria* and *M. javanica* [Williamson et al., 1994], has been an effective control strategy for multiple decades [Milligan et al., 1998]. However, since 2005, resistance-breaking populations have been reported [Tzortzakakis et al., 2005]. In contrast, the *H1* resistance gene in potato confers resistance against specific pathotypes of *G. rostochiensis* [Mullin and Brodie, 1988]. This resistance gene has proven durable for decades as well, and there have been no reports of host-plant resistance populations. It is possible that, under the evolutionary pressure to overcome *H1* resistance, there may emergence *G. rostochiensis* populations that can break the *H1* resistance in the future, similar to *Mi-1*. The primary challenge is that the identification of resistance genes remains limited in number [McDonald and Linde, 2002]. To enhance our understanding of how resistance is circumvented and make informed decisions regarding the deployment of resistant varieties, extensive knowledge on nematode virulence genes is necessary.

1.2 Pathotyping and Effectors

Importantly, certain resistance genes are effective only against specific subsets of populations within the same nematode species. It is therefore important to understand the differences in virulence between populations, to use resistant varieties effectively. To characterise these differences, nematode field populations are *pathotyped*. In this thesis, I use the definition of a pathotype as having a common gene or multiple genes for virulence, as defined by [Andersen and Andersen, 1982], originally for cyst nematodes. Currently, there are multiple pathotyping schemes in place that assess the reproductive capabilities of nematode populations on a predetermined set of host-plant races exhibiting varying levels of resistance. For example, *G. pallida* and *G. rostochiensis* populations are tested on the same standardised set of eight potato hybrids [Kort et al., 1977]. Five pathotypes are defined for *G. rostochiensis* (Ro1 - Ro5), while originally three pathotypes were described for *G. pallida* (Pa1 - Pa3) [Kort et al., 1977]. However, in a later study, a distinction between Pa2 and Pa3 could not be reliably made [Trudgill, 1985], hence nowadays Pa2/3 is used as denominator. This highlights a flaw of the predefined differentials. The choice of these differentials typically has historical origins. For example, when Kort et al. [1977] proposed the pathotyping scheme for *G. rostochiensis* and *G. pallida*, differentials carrying the *H1, /H2*, *K1* and *K2*, and two polygenic resistances were selected based on previous pathotyping schemes used in Europe [Kort et al., 1977, Andersen and Andersen, 1982]. Such a pathotyping method therefore serves as a proxy for virulence, as the actual virulence gene(s) are not necessarily known. Using a phenotypical pathotyping scheme, it is impossible to determine whether different populations of the same pathotype share the same virulence gene/genes. Consequently, when new resistant material is introduced, it is possible for populations of the same pathotype to exhibit varying levels of virulence. The development of a molecular-based pathotype scheme, in which virulence genes are mapped against corresponding resistance genes, would be highly desirable. Such a scheme would enable the informed utilisation of the limited host resistances available. However, to achieve this, extensive knowledge of the nematode's virulence genes and the genetic variation among populations is necessary.

Root-knot and cyst nematodes deploy specialised virulence proteins, known as effector proteins, inside the host. These effectors serve a broad range of functions and are aimed towards the manipulation of the host, including the modification of the plant cell wall (e.g., pectate lyases [Kikuchi et al., 2006] and cellulases [Smant et al., 1998]), the suppression of plant immunity (e.g., SPRYSEC [Sacco et al., 2009] and Gr-1106 [Barnes et al., 2018, Finkers-Tomczak, 2011]), and the formation of a permanent feeding site (e.g., CLE [Mitchum et al., 2012] and glutathione synthetase [Lilley et al., 2018]). The genes that code for effector proteins are often members of (expanded) gene families. Effector genes have two likely evolutionary origins: horizontal gene transfer, and gain-of-function after gene duplication. Multiple cell-wall modifying enzymes have been previously hypothesised to be of bacterial

origin [Keen and Roberts, 1998]. In contrast, other effector families are the product of diversified non-effector genes that gained a novel function through selection [Lilley et al., 2018]. For example, a single nucleotide polymorphism in the *Gp-rbp1* gene can make the difference between either recognition or avoidance of the plant immune system in *G. pallida* [Sacco et al., 2009]. To advance the molecular pathotyping of root-knot and cyst nematode populations, it is essential to have a comprehensive understanding of their genetic diversity and its connection to virulence. Reference genome assemblies serve as valuable tools for conducting this research.

1.3 Nematode Genome assembly

The first assembled genome sequences of parasitic plant nematodes were of *Meloidogyne incognita* [Abad et al., 2008] and *M. hapla* in 2008 [Opperman et al., 2008]. The genome of *M. incognita* was assembled into 2,817 scaffolds, with a total size of 86 Mb, whereas the assembly of *M. hapla* represents a genome of 53 Mb across 1,523 scaffolds. Both studies provided diverse genomic insights, including on gene content and information on virulence genes. Notably, they highlighted the probable occurrence of horizontal gene transfer, specifically the transfer of cell wall-degrading enzymes from bacteria to these plant-parasitic nematode species. [Bellafiore et al., 2008, Opperman et al., 2008]. At the time both studies were conducted, no other genome sequences of plant-parasitic nematodes were available. The authors therefore compared the genomes mainly with other nematode species, including the model organism *Caenorhabditis elegans* and the animal parasite *Brugia malayi*. This made it challenging to gain insights into the evolution of effector genes.

In later years, other genome sequences became available, including those of the potato cyst nematodes *G. pallida* [Cotton et al., 2014] and *G. rostochiensis* [Eves-van den Akker et al., 2016]. In contrast to the genome studies of *M. hapla* and *M. incognita*, the availability of the *G. pallida* and *G. rostochiensis* genomes allowed for a comparison with other plant-parasitic nematode genomes. One of these comparisons revealed high similarities between the *G. rostochiensis* genome and that of *G. pallida*, while demonstrating significantly less similarity to the two root-knot nematode species [Eves-van den Akker et al., 2016]. Furthermore, Eves-van den Akker et al. [2016] identified putative effectors based on similarities with previously published effectors. It was found that a large fraction of putative effectors is physically clustered together in gene-rich regions on the genome. In a later study a correlation was found between copy number variation and adaptations to changes in the environment and host resistance [Castagnone-Sereno et al., 2019]. Due to the high degree of fragmentation of these assemblies, 6,873 scaffolds in *G. pallida* and 4,281 scaffolds in *G. rostochiensis*, it remained difficult to examine the full scale of the genomic organisation of effectors genes. Moreover, the utilisation of short-read sequencing data posed an additional challenge in

accurately discerning between sequence variants and copy number variants. To thoroughly investigate the genomic organisation of effector genes, a highly contiguous reference genome is essential to ensure accuracy and comprehensiveness of the analysis.

The process of generating a reference genome assembly is complicated. Through DNA sequencing, sequence reads are obtained which are then assembled in the largest contiguous sequences possible. One could consider the assembly of a genome to be similar to resolving a jigsaw puzzle with millions of pieces, of which many pieces are missing, duplicated, or contain errors. In the past, many genome assemblies were produced using high-throughput short-read sequencing technologies such as Illumina HiSeq (e.g., [Cotton et al., 2014, Eves-van den Akker et al., 2016]). The main advantage of short-read sequencing is that high genome coverage can be reached, and therefore can produce high confidence base calls. However, the main disadvantage is that low-complexity regions in the genome, such as repeats, are difficult to resolve using short reads. As a consequence, *de novo* assemblies generated from short-read data often exhibit significant fragmentation. In recent years, long-read sequencing technologies developed by Pacific Biosciences and Oxford Nanopore have reached a point of maturity and are commonly used. While short-read sequencing technologies yield reads up to 300bp in size, long-read sequencing technologies sequence reads of, on average, 10,000 - 25,000 in size, while the longest read length ever sequenced is measured at 2.3Mb [Payne et al., 2019]. Long read sequencing technologies historically have a lower base calling accuracy compared to short-read sequencing. Therefore, many genome assembly pipelines use a hybrid approach, combining the strengths of both technologies [Deshpande et al., 2013, Antipov et al., 2016, Walker et al., 2014].

Genome assembly for plant-parasitic nematodes presents additional challenges. With the current technologies, it is not possible to extract enough DNA from an individual nematode (due to their microscopic size) to perform high-coverage sequencing. As a result, DNA is extracted and sequenced from multiple genetically different individuals from a population. This introduces many haplotypes to the pool of sequences that make it difficult to distinguish between sequence variants and copy number variation during the assembly process. So, although plant-parasitic nematode genome sizes are often small (ranging from roughly 50 Mb to 120 Mb on average) [Opperman et al., 2008, Eves-van den Akker et al., 2016, Siddique et al., 2022], producing a contiguous genome assembly requires additional measures. To resolve these issues, a number of steps are taken including a more stringent error correction of sequence reads, and the pruning of the assembly artefacts that results from the high levels of heterozygosity in the starting material [Roach et al., 2018].

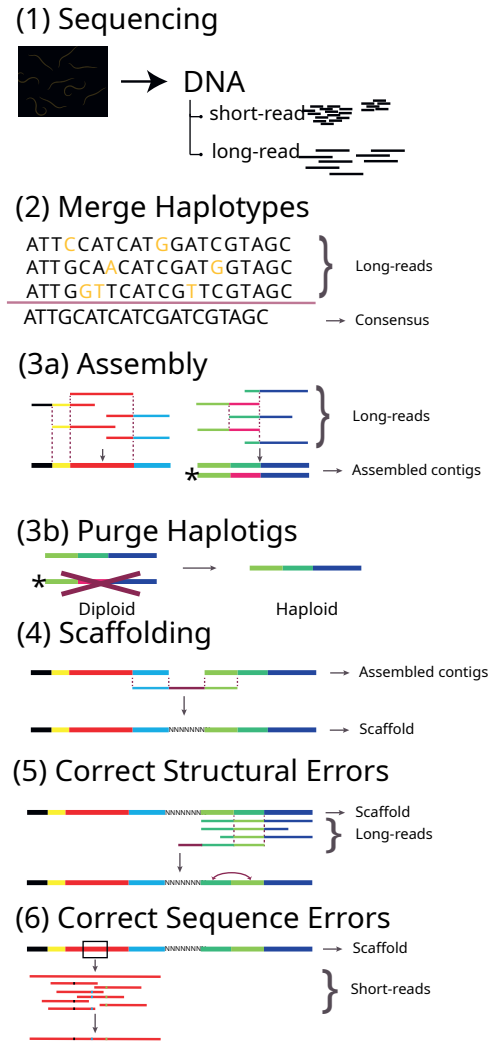


Figure 1.1: Flowchart of the assembly pipeline. 1) Both short- and long-read sequence data is obtained from a nematode population. 2) The overlap in sequence identity between long-reads is determined. Similar long-reads that deviate at most 15% in their nucleotide composition are merged to reduce noise. 3) Long-reads are assembled into contiguous sequences (contigs) based on their overlap. A haploid representation is then extracted by pruning diploid contigs. 4) Contigs are scaffolded in the correct order, introducing gaps in the assembly. 5) Structural errors such as large insertions, deletions, or gaps are detected and possibly resolved using the unmerged long-reads. 6) Smaller insertions, deletions, or gaps are detected and resolved using short-reads.

1.4 In this thesis

In this thesis, I produce five high-quality reference genomes of economically relevant plant-parasitic nematode species (*G. rostochiensis*, *G. pallida*, *H. schachtii*, *M. chitwoodi*) by overcoming the challenges of generating highly contiguous plant-parasitic nematode genome assemblies. Long-read sequencing technologies from Oxford Nanopore and Pacific Biosciences were used to generate assemblies that are highly contiguous and provide a solid basis to study copy number variation and genetic diversity. In my experimental chapters, my objective is to identify and analyse genetic diversity within different pathotypes and species, with a particular focus on effector gene families. The acquired knowledge on effector gene diversification will contribute to the identification of pathotype and species-specific molecular markers. This knowledge contributes to the development of molecular pathotyping in the future, as it can improve the delineation between pathotypes, and enable a more effective usage of the limited resistance genes.

In **Chapter 2**, I use a comparative genomics approach to study the differences between two inbred lines of the potato cyst nematode *Globodera rostochiensis* - Gr-Line19 and Gr-Line22. These two inbred lines were originally derived from different field populations, and are mainly characterised by their difference in virulence. Gr-Line19 is completely avirulent against the *H1* resistance gene, whereas Gr-Line22 is fully virulent against the *H1* resistance gene. The genetic aspect of their difference in virulence is however poorly understood. I therefore present reference genome assemblies of the two inbred lines and identify numerous genetic variations in effector gene families.

In **Chapter 3**, I study the history of the introduction of *G. rostochiensis* in Indonesia through a population genetics approach. While potato was already being cultivated in Indonesia in 1711, the first report of *G. rostochiensis* appeared in 2003. However, it remained unclear when *G. rostochiensis* was introduced. I conducted a comparison of genetic diversity at both the genome-wide level and within specific effector gene families across fourteen Indonesian populations of *Globodera rostochiensis*. These populations were compared against a population from the United Kingdom, using the Gr-Line19 genome assembly from Chapter 2 as a reference. Based on the diversity that was present between the Indonesian and United Kingdom populations, we hypothesise that *G. rostochienis* was likely introduced to Indonesia concurrently with the introduction of potato cultivation in the country.

In **Chapter 4**, the focus of the study is to characterise evolutionary contrasts within effector gene families with the aim of gaining a better understanding of their origins. By comparing the assembled genomes of the other potato cyst nematode *G. pallida*, the beet cyst nematode *Heterodera schachtii*, and the genome of *G. rostochiensis* we find that effector gene families have different diversification patterns.

In **Chapter 5**, I present a reference genome of a Dutch population of the root-knot nematode *Meloidogyne chitwoodi*, and study the genetic diversity by comparing the genome to three other populations from the United States. We discovered that the over-all genetic variation is remarkably low, and most of this diversity is concentrated in five polymorphic regions, which harbour putative effectors.

In **Chapter 6**, I study the overlap in secreted proteins between the cyst nematode *G. pallida* and the root-knot nematode *M. chitwoodi*. Since the parasitic ability of these species likely has evolved independently, it is challenging to identify effector homology through sequence similarity searches. I therefore conducted a novel approach where I predicted the protein structures of all secreted proteins in both species. Alignments were then made based on the structurally similarities. This resulted in several structurally similar proteins between the two species that would not have been identified based on sequence similarity

CHAPTER 2

Comparative genomics of two inbred lines of the potato cyst nematode *Globodera rostochiensis* reveals disparate effector family-specific diversification patterns

Joris J.M. van Steenbrugge[1], Sven van den Elsen[1], Martijn Holterman[1,2], Mark G. Sterken[1], Peter Thorpe[3], Aska Goverse[1], Geert Smant[1], and Johannes Helder[1]

[1] Laboratory of Nematology, Wageningen University and Research, The Netherlands
[2] Solynta, Wageningen, The Netherlands
[3] School of Medicine, Medical & Biological Sciences, University of St. Andrews, United Kingdom

This chapter is published in BMC Genomics [van Steenbrugge et al., 2021].

Abstract

Background

Potato cyst nematodes belong to the most harmful pathogens in potato, and durable management of these parasites largely depends on host-plant resistances. These resistances are pathotype specific. The current *Globodera rostochiensis* pathotype scheme that defines five pathotypes (Ro1 - Ro5) is both fundamentally and practically of limited value. Hence, resistant potato varieties are used worldwide in a poorly informed manner. We generated two novel reference genomes of *G. rostochiensis* inbred lines derived from a Ro1 and a Ro5 population. These genome sequences comprise 173 and 189 scaffolds respectively, marking a 24-fold reduction in fragmentation as compared to the current reference genome. We provide copy number variations for 19 effector families. Four dorsal gland effector families were investigated in more detail. SPRYSECs, known to be implicated in plant defence suppression, constitute by far the most diversified family studied herein with 60 and 99 variants in Ro1 and Ro5 distributed over 18 and 26 scaffolds. In contrast, CLEs, effectors involved in feeding site induction, show strong physical clustering. The 10 and 16 variants cluster on respectively 2 and 1 scaffolds. Given that pathotypes are defined by their effectoromes, we pinpoint the disparate nature of the contributing effector families in terms of sequence diversification and loss and gain of variants. Two novel reference genomes allow for nearly complete inventories of effector diversification and physical organisation within and between pathotypes. Combined with insights we provide on effector family-specific diversification patterns, this constitutes a basis for an effectorome-based virulence scheme for this notorious pathogen.

2.1 Introduction

Plant-parasitic nematodes have a significant impact on food and feed production worldwide. Every cultivated crop can be parasitized by at least one nematode species, resulting in a net loss of over 70 billion US dollar annually [Nicol et al., 2011]. From an economic point of view root-knot and cyst nematodes have the highest impact [Jones et al., 2013]. Whereas root-knot nematodes have a higher impact in warmer climate zones, cyst nematode problems mostly occur in the temperate regions. Unlike root-knot nematodes, most cyst nematodes have a defined center of origin. For example, soybean cyst nematodes originate from north-east Asia and have spread as a successful and highly harmful parasite to all major soybean-growing areas. Potato cyst nematodes diversified in the Andes in South America, and have now proliferated to all major potato production areas in the world (e.g., [Plantard et al., 2008]). Outside of their centers of origin, cyst nematodes belong to the most harmful pathogens of the crops mentioned above.

One of the most widely applied control measures is the use of resistant host plants. Resistances against potato cyst nematodes tend to have a long agronomic life span due to cyst nematodes' unique biology. Potato cyst nematodes usually have only one generation per year through obligate sexual reproduction, go into diapause for months, and - once hatched - their motility is in the range of a few cm per day. Apart from this, remarkably low effective population sizes have been reported for multiple cyst nematode species [Montarry et al., 2019, Jan et al., 2016]. Together these characteristics drastically slow down the process of selection and proliferation of virulent individuals, a process that happens underground and therefore often goes unnoticed for years. Potato breeders have introgressed the resistance gene H1 from *Solanum tuberosum ssp. andigena* CPC 1674 into numerous potato cultivars from the 1960's onwards [Fuller and Howard, 1974]. The H1 gene confers resistance against *G. rostochiensis* pathotypes Ro1 and Ro4 [Toxopeus and Huijsman, 1952], and this resistance gene is still effective in virtually all major potato producing countries.

Based on a number of Solanum differentials, pathotypes have been defined within the two potato cyst nematode species *G. rostochiensis* and *G. pallida*. Five pathotypes named Ro1 - Ro5 have been proposed for *G. rostochiensis*, whereas three pathotypes (Pa1 - Pa3) were discriminated within *G. pallida* [Kort et al., 1977]. Apart from being laborious and time-consuming, the current pathotype scheme has limited value as it lacks a solid genetic basis. The distinction between for instance the *G. pallida* pathotypes Pa2 and Pa3 is elusive [Phillips and Trudgill, 1983]. For *G. rostochiensis*, genome-wide allele frequencies correlate with the geographical distribution of populations, regardless of pathotype [Mimee et al., 2015, Thevenoux et al., 2020]. This indicates that the genetic basis of the predefined pathotypes is small. A robust pathotyping scheme for potato cyst nematodes is highly desirable because it would lead to far more efficient and durable use of the limited number of host plant

resistances currently available. The availability of high-quality reference genome sequences from individual pathotypes would be an ideal starting point for pathotypes' molecular characterization.

Resistant plant species deploy R proteins as surveillance molecules that recognize either directly or indirectly specific effector molecules - or their activities - secreted by nematodes. Nematodes use a protrusible stylet to inject effector proteins into plant cells. Effectors are diverse and fulfil functions ranging from plant cell wall degradation to the induction of a feeding site and suppressing the plant's innate immune system [Ali et al., 2015a]. The nematode produces effectors mainly in the subventral and the dorsal esophageal glands. Effectors are usually members of diversified gene families, and potato cyst nematode typically produces multiple variants per effector. An example is the SPRYSEC gene family that codes for a highly expanded set of proteins that act as activators and suppressors of plant defence [Diaz-Granados et al., 2016]. One variant of this family, *RBP-1*, was shown to trigger the activation of the potato resistance gene *Gpa2* [Sacco et al., 2009] resulting in local hypersensitive response. Effector proteins secreted by the cyst nematode parasite are most likely responsible for the activation of plant resistance proteins. However, this was demonstrated for only a small number of resistance genes (*Gpa2*; [Sacco et al., 2009], Cf-2; [Lozano-Torres et al., 2012]).

Sequencing the genome of plant-parasitic nematodes is more challenging than for other, larger organisms. With the currently available methods, it is practically impossible to isolate and sequence DNA from an individual nematode to gain enough coverage to generate a high-quality reference genome sequence especially when isolating high molecular weight DNA required for long-read sequencing technologies. Reference genomes of plant-parasitic nematodes are therefore often based on the genetic material from a population. Consequently, the reference genome includes a substantial heterozygosity level, as the starting material includes a high degree of allelic variation. The current reference genome sequences of potato cyst nematodes *Globodera rostochiensis* [Eves-van den Akker et al., 2016] and *G. pallida* [Cotton et al., 2014] were each generated using heterozygous starting material (selected field populations), and are relatively fragmented (respectively 4,377 and 6,873 scaffolds). In *G. rostochiensis*, Eves-van den Akker et al. [2016] predicted 138 high confidence effector genes based on sequence similarity with previously described effector gene families. Furthermore, a third of these genes were identified to cluster on effector gene islands. Among these expanded gene families, sequence divergence between different pathotypes was estimated as well. While many single nucleotide polymorphisms and insertions/deletions were observed [Eves-van den Akker et al., 2016], the highly fragmented reference genome sequence made it challenging to distinguish between sequence and copy number variation. Less fragmentation in the genome sequence would similarly make it possible to display the degree of clustering of effector genes more accurately.

We generated a new set of reference genome sequences to allow for the accurate organization of effector genes and to compare copy number variation and sequence variation between the Ro1 and Ro5 pathotypes. A precise representation of these two sources of genetic variation is essential for developing molecular pathotyping methods in the future. The current reference genome sequence of *G. rostochiensis* shows a haploid genome size of 95.9 Mb [Eves-van den Akker et al., 2016] and is expected to spread over eighteen diploid chromosomes [van der Voort et al., 1996]. For this, we used two *G. rostochiensis* lines, one fully avirulent and one fully virulent with regard to the H1 gene. The starting materials for these lines were two distinct field populations sampled from The Netherlands (Ro1-Mierenbos) and Germany (Ro5-Hamerz) [Janssen et al., 1990]. The selection process started with a single cross between an individual male and a female. After multiple generations, fully avirulent Ro1 (Gr-line19) and fully virulent Ro5 (Gr-line22) lines were generated regarding the H1 resistance in potato [Janssen et al., 1990]. As a result, both Gr-Line19 and Gr-Line22 harbour limited genetic variation, with a theoretical maximum of 4 alleles per locus. For diploid sexually reproducing species, this is the minimum level of heterozygosity that can be present in a population.

New genome assemblies were generated for each of the inbred lines based on PacBio long read-sequencing technology. Using these newly generated *G. rostochiensis* reference genome sequences with a substantially reduced number of scaffolds, we investigated the genomic organisation and the diversification of 19 effector families. A large number of differences in the number of paralogs and variation in sequence content were identified between the effector arsenals of the avirulent Gr-line19 and the virulent Gr-line22. These pathotype-specific effector variants form the basis for the generation of a virulence scheme for potato cyst nematodes.

2.2 Results

2.2.1 Genome Assemblies

Two inbred lines of the potato cyst nematode *G. rostochiensis* were initially derived from crossings between individuals from two populations, Ro1-Mierenbos and Ro5-Harmerz [Janssen et al., 1990]. DNA from these lines, Gr-Line19 and Gr-line22, were sequenced using PacBio sequencing technology with respectively 119X and 132X coverage and assembled into two reference genome sequences (Table 2.1). Benefiting from this long read technology and the significantly smaller genetic background, the two newly generated *G. rostochiensis* genome assemblies are less fragmented than the first genome sequence that was published (nGr.v1.0) [Eves-van den Akker et al., 2016] while maintaining a comparable assembly size. The number of scaffolds in the new assemblies is about 24-fold lower than in the original *G. rostochiensis* reference genome sequence (Table 2.1). At the same time, the scaffold N50 increased about 20-fold from 0.085 to around 1.7 Mb. Regarding the assembly size and BUSCO score, the novel assemblies are comparable to the current reference. The assemblies of Gr-Line19 and Gr-Line22 harbor 2,733 and 6,572 gaps, respectively, covering in total 130 Kb and 150 Kb. As compared to the current *G. rostochiensis* reference [Eves-van den Akker et al., 2016], the number and lengths of gaps showed a 29-fold reduction.

Table 2.1: BUSCO (Benchmarking Universal Single-Copy Orthologs) - eukaryota_odb10

***G. rostochiensis* population**	**Size (Mb)**	**No. scaffolds**	**N50**	**BUSCO results**			
				Single	**duplicated**	**fragmented**	**missing**
JHI-Ro1	96	4377	0.085	82.8	1.0	8.3	7.9
Gr-Line19	92	173	1.70	82.2	1.7	7.9	8.2
Gr-Line22	101	189	1.80	81.5	1.3	8.3	8.9

The repeat content in both reference genome sequences is relatively low, 2.6% for Gr-Line19 and 1.6% for Gr-Line22. The GC content in repeat regions for Gr-Line19 (40.3%) was comparable to this genotype's overall GC content (39.1%). In Gr-Line22, the GC content in repeat regions (32.5%) was lower than the overall GC content (38.3%). In predicted protein-coding regions, the GC content is comparable between both reference genome sequences (Gr-Line19: 50.8%, Gr-Line22: 50.9%). Using BRAKER2 as a gene-prediction tool, 17,928 and 18,258 genes were predicted in the Gr-Line19 and Gr-Line22 genome assemblies, coded for 21,037 21,514 transcripts, respectively. The protein-coding regions take up approximately 33% (Gr-Line19) and 30% (Gr-Line22) of the genomes at an average density of 89.3 (Gr-Line19) and 86.6 (Gr-Line22) genes per Mb.

Synteny between the newly generated genomes and the current reference genome [Eves-van den Akker et al., 2016] was evaluated using a progressive genome alignment. Homologous

regions larger than 3 kb and their genomic organization are presented in Figure 2.1. A broad span of regions in the nGr.v1.0 reference assembly shows homology to both new assemblies (respectively 67%, 72%, and 61% of the total assembly sizes for JHI-Ro1, Gr-Line19 and Gr-Line22). While the total numbers of base pairs that are covered in a homologous region are roughly within a 10% range of each other, both new assemblies show substantially larger contiguous, and so far, uncovered regions.

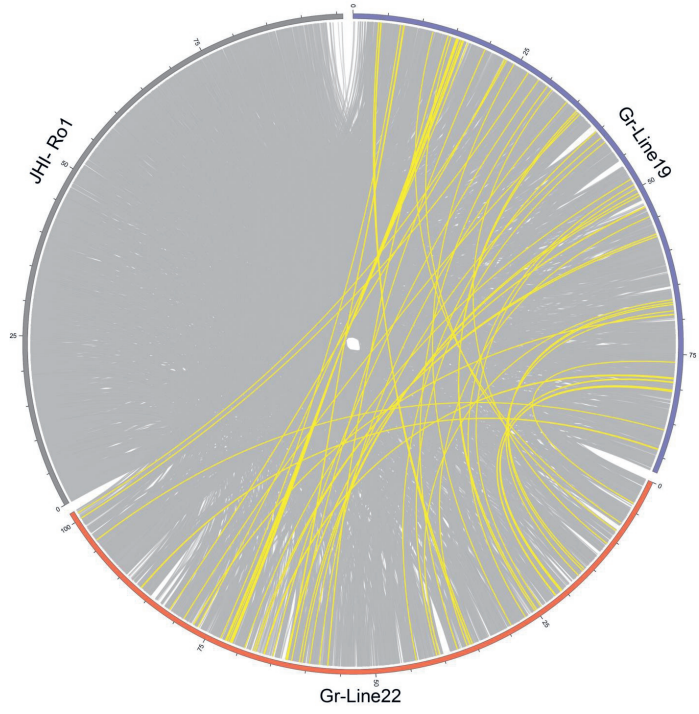


Figure 2.1: Synteny between Gr-Line19, Gr-Line22, and JHI-Ro1 based on a progressive genome alignment in Mauve. Only syntenic regions larger than 3 kb are shown. Yellow lines represent regions that are exclusively syntenic between Gr-Line19 and Gr-Line22

2.2.2 Heterozygosity and structural variation between the two *Globodera rostochiensis* genomes

The inbred lines Gr-Line19 and Gr-Line22 originate from single crossings of individuals, and, as a result, the genetic variation is expected to be smaller than for field populations. To pinpoint the effect of this genetic bottleneck caused by a single crossing, a comparison was made between the proportion of heterozygous and homozygous single nucleotide variants

between Gr-Line19 (Ro1) and JHI-Ro1, the selected field population used to generate the current *G. rostochiensis* reference genome [Eves-van den Akker et al., 2016] while using the Gr-Line22 (Ro5) genome as a reference. Among the called variants that passed the quality filter (JHI-Ro1 n = 584,145; Gr-Line19 n = 716,491), 37% of the JHI-Ro1 loci were homozygous, as compared to 47% of the variants in Gr-Line19 (Fig. 2.2 A). The increased level of homozygosity in Gr-Line19 reflects the relatively narrow genetic basis of this inbred line. Secondly, structural variation (e.g. insertions, deletions, inversions) of approximately 1 kb or larger within the individual lines and between Gr-Line19 and Gr-Line22 was determined. The proportions of heterozygous and homozygous structural variants with fragment sizes $>$ 1 kb were compared (Fig. 2B). The structural variation within Gr-Line19 and Gr-Line22 was minimal (Fig. 2.2 B). This observation confirms the low level of structural intra-population heterozygosity. The proportion of homozygous variants was nearly identical while comparing Gr-Line19 with Gr-Line22 and vice versa (Gr-Line22 versus Gr-Line19: 85.09% & Gr-Line19 on Gr-Line22 85.06%).

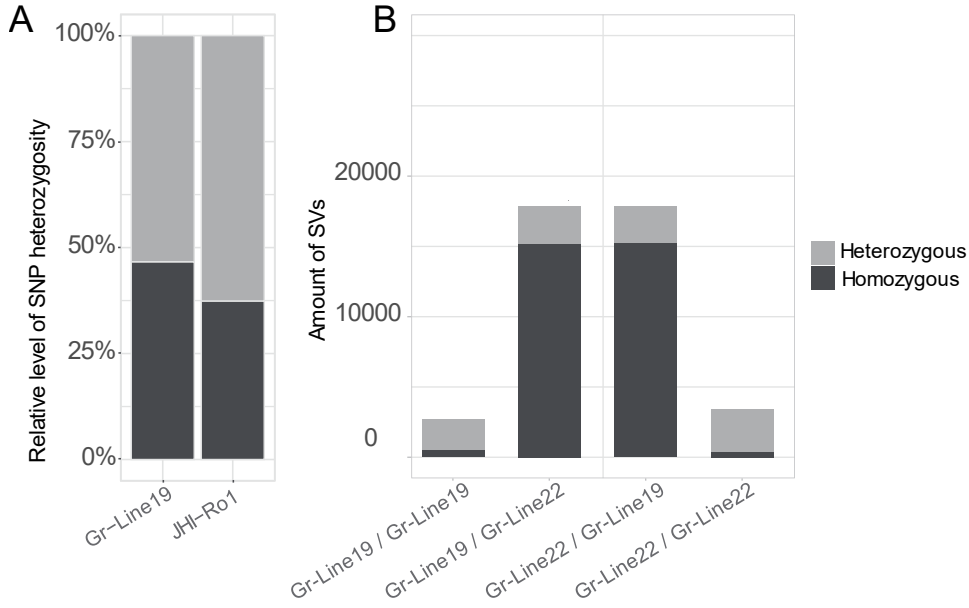


Figure 2.2: (A) Relative level of SNP heterozygosity of Gr-Line19 and JHI-Ro1 (JHI – James Hutton Institute, Scotland, UK). Both lines were compared to Gr-Line22. JHI Ro1 was used as the current *G. rostochiensis* reference genome. (B) Comparison overall genomic constitution of *G. rostochiensis* inbred lines 19 and 22. In this overview structural variants are shown. Structural variants are DNA region of approximately 1 kb and larger in size, and can include inversions, insertions, deletions

2.2.3 Expansion of Effector gene Families

We identified homologs of 19 known effector gene families from which at least one member was shown to be expressed in the subventral (6) or the dorsal (11) oesophageal gland cells, or in the amphids (1) (Table 2.2; Fig. 2.3 A). For each of these gene families, the copy number differences between Gr-Line19 and Gr-Line22 were determined. The number of paralogs per effector families varied from 99 SPRYSEC variants in Gr-Line22 to a single Hg-GLAND14 gene with signal peptide in the same line. Among the 19 effector families, six have a lower number of paralogs in Gr-Line22, seven have an equal number of paralogs, whereas six have a higher number of variants in Gr-Line22 (Fig. 2.3 A).

Four effector families show a relatively large difference in the number of paralogs between Gr-Line19 and Gr-Line22. SPRYSEC is by far the most speciose effector family in both lines, but Gr-Line22 harbor 36 more paralogs with signal peptide than pathotype Ro1. Similarly, 11 Hg-GLAND5 homologs were present in Gr-Line19, while Gr-Line22 comprised 18 paralogs with a signal peptide. The reverse was also observed for the subventral gland effector family GH30. Whereas six variants with signal peptide were identified in Gr-Line19, only two were found in Gr-Line22. It is noted that the GH30 family harbors various glycoside hydrolases that were previously categorized as GH5.

Table 2.2

Expression	**Effector family**	**Functionality / similarity**		**Reference**
	GH5a	Beta 1,4 endoglucanase		[Smant et al., 1998]
	GH30	xylanase, glucosylceramidase, etc.		[Mitreva-Dautova et al., 2006]
Subventral esophageal glands	GH43	candidate arabinanase	CWDE	[Eves-van den Akker et al., 2016]
	GH53	candidate arabinogalactanase		[Eves-van den Akker et al., 2016]
	PL3b	Pectate lyase		[Popeijus et al., 2000]
	Hg-GLAND 10	cellulose binding protein		[Hewezi et al., 2008, Noon et al., 2015]
	VAL	Venom allergen-like protein	Immune	[Wilbers et al., 2018]
	SPRYSEC	Suppression and activation of plant innate immunity	Immune	[Rehman et al., 2009]
	GSS	glutathione synthetase-like effectors involved in redox regulation	Feeding site	[Lilley et al., 2018]
	CLE	CLAVATA3/ESR-related peptides, mimic plant CLEs	Feeding site	[Lu et al., 2009]
	1106	PTI and ETI suppressor		Finkers-Tomczak [2011]
	Hg16B09	Suppression plant innate immunity		[Hu et al., 2019]
Dorsal esophageal gland	Hg-GLAND1	ETI suppressor	Immune	[Noon et al., 2015, Pogorelko et al., 2020]
	Hg-GLAND5	PTI suppressor		[Noon et al., 2015, Pogorelko et al., 2020]
	Hg-GLAND6 (4D06)	PTI suppressor		[Noon et al., 2015, Pogorelko et al., 2020]
	Hg-GLAND 12	Pioneer (function unknown)		[Noon et al., 2015]
	Hg-GLAND 13	Invertase (Rhizobium)	Feeding site	[Noon et al., 2015]
	Hg-GLAND 14	Endopeptidase (Ascaris suum)		[Noon et al., 2015]
Amphids	HYP	hyper-variable extracellular effector	Unknown	[Eves-van den Akker et al., 2014]

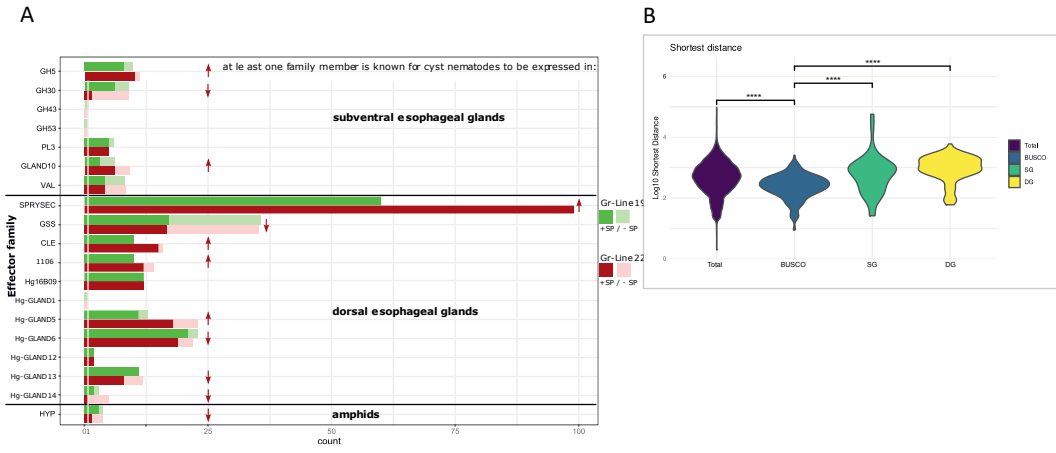


Figure 2.3: (A) Copy numbers of effector gene families expressed in the secretory gland cells. A comparison was made between Gr-Line19 (Green) and Gr-Line22 (Red). Bright shades of green/red indicate copies that contain a signal peptide for secretion. Faded shades of green/red indicate the absence of a signal peptide for secretion. Upward and downward pointing arrows are used to indicate increase or decrease of the number of variants with a predicted signal peptide for secretion in Gr-Lin22 as compared to Gr-line19. (B) Gene density comparison of all genes (total), eukaryotic universal single copy genes (BUSCO), effector genes secreted from the subventral esophageal gland cells (SG), and effector genes secreted from the dorsal esophageal gland cells (DG). The shortest distance of each gene was based on the closest adjacent gene (3' or 5') and is measured as the log10 number of base pairs. Statistical significance for each group was determined by comparison with the BUSCO gene set using a Wilcoxon test ($P < 0.0001$)

2.2.4 Genomic organisation of effector genes

To characterise the genomic organisation of effector genes, the shortest distance between each gene and the closest adjacent gene was calculated (either at the 3' or 5'-end of the

full genomic sequence). This was done for effector genes, as well as for known non-effector genes (i.e., BUSCO gene set). The distances based on the full set of predicted genes ranged from extremely gene sparse to extremely gene dense regions (Fig. 2.3-B). BUSCO genes are generally located in regions that are more gene dense than expected at random (Wilcoxon Rank Sum test P $<$ 0.0001). Effector genes expressed in either the dorsal or the subventral esophageal gland cells are often located in more gene sparse regions both as compared to non-effector genes and to any random gene (Wilcoxon Rank Sum test, $P < 0.0001$).

Furthermore, the spatial organization and diversification between two pathotypes of *G. rostochiensis* lines is presented for four selected effector families that are expressed in the dorsal esophageal glands during parasitic life stages. Hg-GLAND5 effectors are known as plant triggered immunity suppressors [Pogorelko et al., 2020]. Members of the effector family 1106 were demonstrated to suppress both plant triggered immunity and effector triggered immunity [Finkers-Tomczak, 2011]. The highly speciose SPRYSEC family was shown to be involved in both the suppression and the activation of the plant immune system [Ali et al., 2015a]. CLE-like effectors were demonstrated to be involved in feeding site induction by mimicking the functionality of endogenous host-plant CLE peptides [Mitchum et al., 2012]. Concentrating on the distribution of individual family members over the relevant scaffolds, large differences in the level of clustering per family are observed (Fig. 2.4). While the 60 and 99 SPRYSEC variants are distributed over respectively 18 and 26 scaffolds, the moderately diversified CLE family is concentrated on two scaffolds in case of Gr-Line19, and on a single scaffold for Gr-Line22.

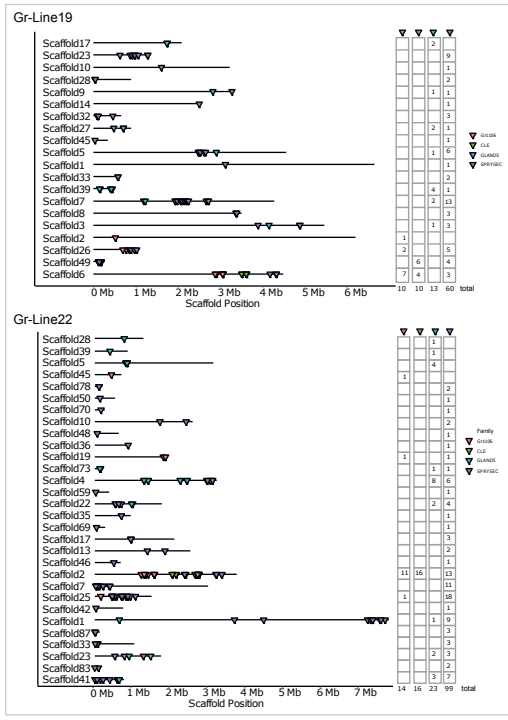


Figure 2.4: Spatial distribution of genes belonging to the effector families Gr1106 (Red), CLE (Green), GLAND5 (blue), and SPRYSEC (purple). Each triangle indicates the genomic position of a single gene. At the right, the number of variants per effector family are given for each scaffold.

2.2.5 A. Hg-GLAND5

Hg-GLAND5, also referred to as 'putative gland protein G11A06' [Gao et al., 2003], has first been discovered in soybean cyst nematode *Heterodera glycines*. This effector is expressed in the dorsal gland during a range of parasitic life stages, and it functions as a PTI suppressor [Pogorelko et al., 2020]. In a transcriptional analysis of two *H. glycines* races, the expression level of Hg4J4-CT26, a GLAND5 family member, was shown to be highly race-dependent [Wang et al., 2014]. Searches in public genome database revealed that both PCN species harbour homologs of HG-GLAND5 [Yang et al., 2019b].

In *G. rostochiensis*, the GLAND5 effector family comprises 13 and 23 members in Gr-Line19 and Gr-Line22, respectively. Among these variants, three and five are unlikely to be involved in parasitism as the corresponding protein sequences are not preceded by a signal peptide

(SP) for secretion (Fig. 2.3 A). A phylogenetic analysis of the GLAND5 family based on the coding sequences using RaxML revealed an initial split between four effectors without an SP in one clade, and all functional GLAND5 effectors in the other (Fig. 2.5). Four clusters with mainly secreted GLAND5 variants could be discerned. Differences in numbers of effector paralogs were observed, and in three clusters more Gr-Line22 paralogs are present. Most notable is the diversification in Box I, where eight related GLAND5 representatives from Gr-Line22 surround a single Gr-Line19 variant. Box III and Box IV illustrate expansion in Gr-Line22 (or gene loss in Gr-Line19) as well, although less extreme. On the other hand, in Box II two paralogs of Gr-Line19 are present and a single variant of Gr-Line22.

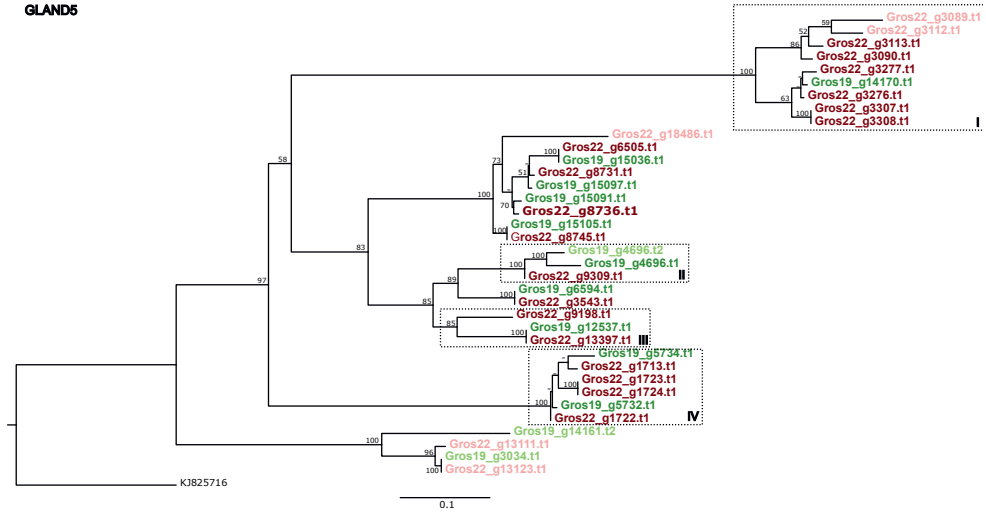


Figure 2.5: Phylogeny of GLAND5 effector genes of both Gr-Line19 (Green) and Gr-Line22 (dark red). A multiple sequence alignment was made using MUSCLE on the coding sequence. A phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a "-". Lighter shades of green or red indicate effector variants that lack a signal peptide for secretion. Boxed clusters (Roman I – IV) highlight asymmetric representations of variants among one of the two *G. rostochiensis* lines

2.2.6 B. 1106

The 1106 gene family encodes mainly secreted proteins, and members were demonstrated to suppress both PTI and ETI responses [Finkers-Tomczak, 2011]. In this previous study by Finkers-Tomczak [2011], a conserved region of 1106 variants was shown to hybridize in the dorsal gland of infective juveniles of *G. rostochiensis*. Gr-Line19 contains ten paralogs, whereas 14 1106 paralogs were found in Gr-Line22. In terms of organization, the genes in Gr-Line22 and Gr-Line19 show a comparable degree of physical clustering (Fig. 2.4). To investigate the diversification of the effector family 1106, the phylogenetic relationship between the variants identified in Gr-Line 19 and Line 22 was examined (Fig. 2.6). In many cases, an 1106 variant in Gr-line 19 had a single, orthologous equivalent in Gr-line22 (see e.g. Gros19_g2102 and Gros22_g4744, and Gros19_g2104 and Gros22_g4746). Notably, the relationship between the small clusters of 1106 variants was largely unresolved. Within cluster I in Fig. 2.6, four Gr-Line22 variants were present, and only one representative from Gr-Line19. Two variants, Gros22_g4703 and Gros22_g4696, deviate substantially from the other 1106 family members. It is noted that these variants are not preceded by a signal peptide for secretion and thus are unlikely to act as effectors. Clusters II is highlighted as it represents a local expansion of this effector family in Gr-Line22.

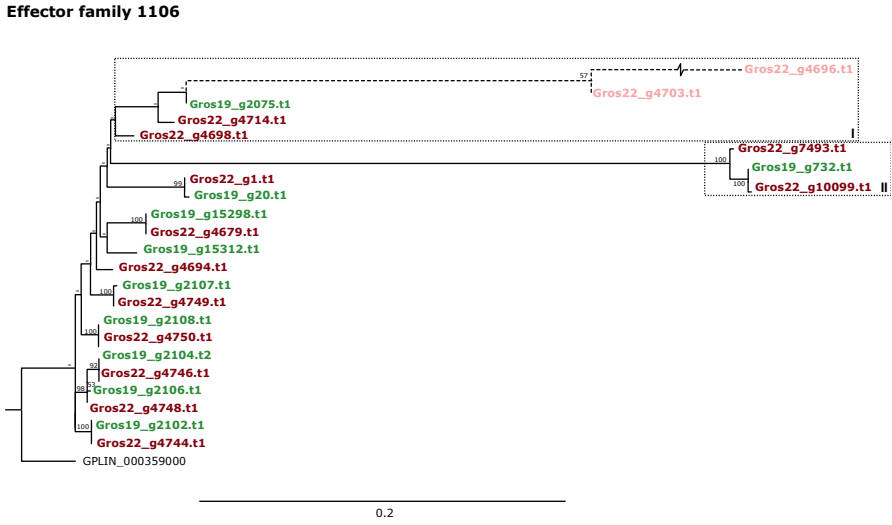


Figure 2.6: Phylogeny of Gr1106 effector genes of both Gr-Line19 (green) and Gr-Line22 (dark red). A multiple sequence alignment was made using MUSCLE on the coding sequence. A phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a "-". Lighter shades of green or red indicate effector variants that lack a signal peptide for secretion. Boxed clusters (Roman I - II) highlight asymmetric representations of variants among the two *G. rostochiensis* lines. Dashed lines are used in case there was uncertainty about the location of the 5'end of a given effector family variant.

2.2.7 C. SPRYSEC

The SPRYSEC gene family encodes for secreted proteins that contain an SPla and RYanodine receptor. SPRYSECs are produced in the dorsal esophageal glands, and this effector family is by far the most expanded one among the plant-parasitic nematodes [Diaz-Granados et al., 2016]. Several SPRYSECs from *G. rostochiensis* were shown to be implicated in the suppression and the activation of defence-related cell death [Postma et al., 2012]. Suppression was demonstrated for the variants SPRYSEC-4, -5, -8, -15, -18, and −19, whereas only

SPRYSEC15 elicited a defence response in tobacco [Ali et al., 2015a]. In the closely related cyst nematode species G. pallida, a single SPRYSEC variant - RBP-1 - was shown to be responsible for the evasion of the potato resistance gene Gpa2, thus preventing a local HR Sacco et al. [2009]. No direct ortholog of *RBP-1* could be found among the *G. rostochiensis* SPRYSECs (identity $< 50\%$), with the used filtering criteria.

The diversification of the SPRYSEC-like variants in Gr-Line19 and Gr-Line22 was investigated by analysing the phylogenetic relationships. Although the number of SPRYSECs in Gr-Line19 (n = 60) was already higher than for any other effector family, Gr-Line22 was shown to harbour even more members of this effector family (n = 99) (Fig. 2.3 A). Maximum-likelihood-based inference revealed several SPRYSEC clusters (Fig. 2.7). Due to the poor backbone resolution, no statements can be made about the relationship between these clusters. It is noted that the support values for the more distal parts of the SPRYSEC tree are substantially higher than the support values for most of the more proximal bifurcations. Three large (A, B, and D) and two smaller (C, E) SPRYSEC clusters could be identified. The majority of Gr-Line19 gene family members have a single orthologous equivalent in Gr-Line22, while Gr-Line22 contains additional paralogs in each of the clusters.

As compared to B and D, cluster A shows the highest level of diversification. Both types of asymmetric SPRYSEC expansion were found in this cluster. Box I in Cluster A comprises a single Gr-Line22 and three Gr-Line-19 SPRYSECs. Box II exemplifies Gr-Line22 expansion, where four closely related Gr-Line22 SPRYSECs surround a single Gr-Line19 variant.
Cluster B harbours three of the SPRYSEC variants described in [Ali et al., 2015a] (SPRYSEC-4, -5 and -8), all of which seem to be represented by a single orthologous pair.
Cluster C is characterized by a set of genes homologous to SPRYSEC-15 that are considerably expanded in Gr-Line22. It is noted that Gros19_g2329.t1 also is the closest match of SPRYSEC-18, however only with 60% identity.
Cluster D unites SPRYSEC variants with a low degree of diversification. Although most Gr-Line-19 variants have a single equivalent in Gr-Line22, there are a few examples of further diversification in Gr-Line22. Box III shows a notable example of a diversification event where a single Gr-Line19 variant has five closely related equivalents in Gr-Line22.

SPRYSEC-19, a variant that was demonstrated to suppresses programmed cell death mediated by several immune receptors [Postma et al., 2012], localized in cluster E. SPRYSEC-19 was first identified in a *G. rostochiensis* Ro1 Mierenbos population [Rehman et al., 2009], which is the population Gr-Line19 was originally derived from. Cluster E shows the Gr-line 22 equivalent of SPRYSEC-19 (g7323).


A
B
C
D
E
II
III
SPRYSEC-5
SPRYSEC-4
SPRYSEC-15
SPRYSEC-19
H_glycines_Hetgly.G000009390
H_glycines_Hetgly.G000028771

Figure 2.7:

Figure 2.7: Phylogeny of SPRYSEC effector genes of both Gr-Line19 (green) and Gr-Line22 (dark red). Only SPRY proteins with a signal peptide for secretion are included A multiple sequence alignment was made using MUSCLE on the coding sequence. A phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a "-". Closest homologs to the functionally described SPRYSEC-4, SPRYSEC-5, SPRYSEC-8, SPRYSEC-15, and SPRYSEC-19 [Mimee et al., 2015] are shown. Clusters of SPRYSEC variants are boxed (A- E). Boxed clusters (Roman I – III) highlight asymmetric representations of variants among the two *G. rostochiensis* lines. Dashed lines are used in case there was uncertainty about the location of the 5'end of a given effector family variant.

2.2.8 D. CLE-like

The CLE-like gene family is an unusual effector family coding for pro-peptides that are delivered via the stylet of the infective J2 to the syncytial cell. For CLEs from the related cyst nematode species *Heterodera glycines* with domain structures similar to *G. rostochiensis*, it was shown that the mature pro-peptide comprised a nematode-specific translocation signal that facilitated the export from the developing syncytium to the apoplast [Wang et al., 2021]. Subsequently, the protein is cleaved outside the plant cell, and bioactive CLEs are released [Gheysen and Mitchum, 2019]. Two classes of CLE-like proteins were found to be expressed in the dorsal gland of *G. rostochiensis*. Members of the Gr-CLE-1 class showed moderate (10 fold) up-regulation in early parasitic life stages (peak in parasitic J-3), whereas Gr-CLE-4 representatives showed an over 1,000 fold in later parasitic stages (at 21 dpi) Lu et al. [2009].

The two *G. rostochiensis* lines 19 and 22 harbour 10 and 16 CLE variants, and both lines comprise members of the Gr-CLE-1 and the Gr-CLE-4 class. As shown in the phylogenetic analysis (Fig. 2.8), members of class Gr-CLE-4 show little variation among each other, while Gr-CLE-1 s show a higher level of diversification. As compared to Gr-Line19, the number of Gr-CLE-4 variants had doubled from four to eight in Gr-Line22. On the contrary, each of the Gr-Line19 representatives of Gr-CLE-1 had a single homolog in Gr-Line22. In Fig. 2.8, clusters A and B include four Gr-Line22 variants with no immediate ortholog in Gr-Line19. In cluster C, a Gr-Line19 variant is present that deviates substantially from the closest Gr-Line22 orthologous sequence. Cluster D contains an example of a homologous gene pair, with a tentative duplication in Gr-Line22. In addition to sequence similarity within the Gr-CLE function classes, there is also a high degree of physical clustering (Fig. 9). The Gr-CLE-4 variants are all located adjacent to each other, and not interspersed by any other gene. Based on this remarkable physical organization, we hypothesize that one or more duplication events in this region underlies the copy number difference of Gr-CLE-4 effectors (n = 4 in Gr-Line19; n = 8 in Gr-Line22).

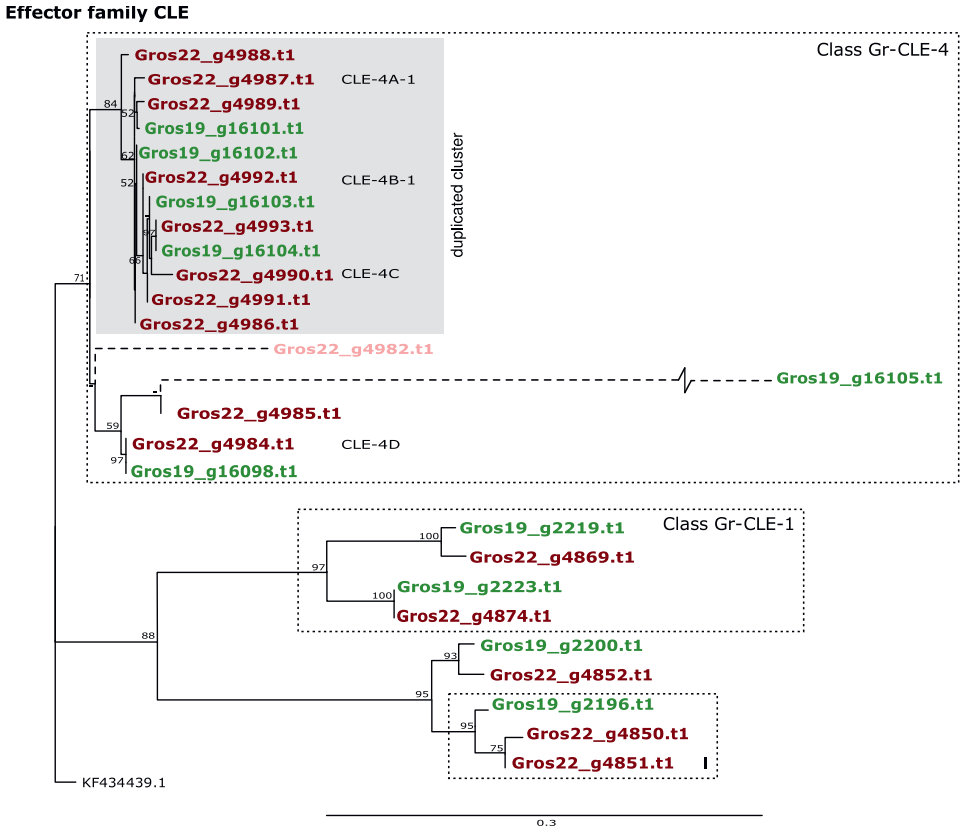


Figure 2.8: Phylogeny of CLE effector genes of both GrLine19 (Green) and GrLine22 (dark red). A multiple sequence alignment was made using MUSCLE on the coding sequence. A phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a "-". Lighter shades of green or red indicate effector variants that lack a signal peptide for secretion. Genes belonging to the functional classes Gr-CLE-1 and Gr-CLE-4 [Lu et al., 2009] are labeled with dashed boxes. A boxed cluster (Roman I) highlights an asymmetric representation of variants among of the two *G. rostochiensis* lines. Dashed lines are used in case there was uncertainty about the location of the 5'end of a given effector family variant.

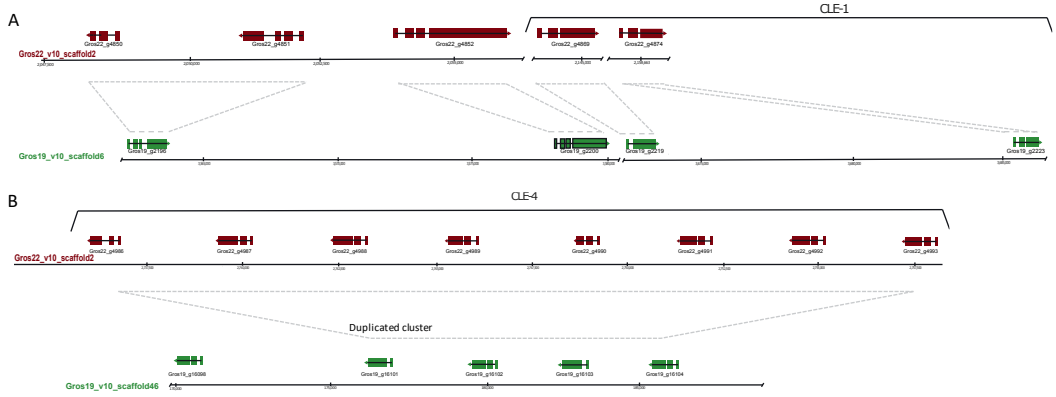


Figure 2.9: Spatial organization of CLE functional classes CLE-1 (A) and CLE-4 (B) [Lu et al., 2009] for both Gr-Line19 (Green) and Gr-Line22 (Dark red). In panel A, the spatial organization of CLEs belonging to Class Gr-CLE-1 are shown for both lines, as well as the other CLEs residing in the same major CLE clade (Fig. 2.8). In panel B, the spatial organization of CLEs belonging to Class Gr-CLE-4 are shown (grey panel in Fig. 2.8), as well as CLE variant Gros19_16098 as this variant was located close to the Gr-Line19 Class-4 CLEs.

2.3 Discussion

Due to specific biological characteristics of plant-parasitic nematodes, host plant resistances tend to be a remarkably durable means to manage this category of soil-borne pathogens. The main challenge is the actual developing and breeding resistant host-plant varieties. As the genetic basis for virulence in plant-parasitic nematodes is unknown, breeding for resistance can only be done on a trial-and-error basis. The whole process is, therefore, inefficient, and thus time consuming and expensive. The availability of molecular-based pathotyping methods of plant-parasitic nematode populations would allow for the deployment of more targeted resistance. Here we concentrated on two *Globodera rostochiensis* inbred lines, Gr-Line19 and Gr-Line22, with distinct pathotypes [Janssen et al., 1990]. Resulting from a single male-female crossing by Janssen et al. [1990], each of these lines' genomic background is small, with a maximum of 4 haplotypes per locus. These small genomic backgrounds significantly simplify the generation of high-quality reference genome sequences, which has been a challenge for sexually reproducing plant-parasitic nematodes in the past. Therefore, we expect that the reference genome sequences of Gr-Line19 and Gr-Line22 are a more accurate representation of the *G. rostochiensis* genome, making the process of molecular pathotyping a step closer. Furthermore, long-read sequencing technology allowed us to generate reference genomes about 24 fold less fragmented than the current reference genome [Eves-van den Akker et al., 2016]. This higher contiguity made it possible to pinpoint the physical distribution and the diversification in a way that was not possible with the highly fragmented JHI-Ro1 reference genome sequence. Four effector families that, together with other effectors, define this potato cyst nematode's pathogenicity were explicitly studied in detail.

One of the main technical challenges we tried to overcome was generating high-quality reference genome sequences of a highly heterozygous nematode species. In terms of assembly sizes, we see a comparable size to the *G. rostochiensis* JHI-Ro1 genome [Eves-van den Akker et al., 2016] as well as to the estimated *G. rostochiensis* genome size [Grisi et al., 1995]. Therefore, it is likely that the high levels of heterozygosity did not negatively impact the assembly by the presence of haplotigs [Roach et al., 2018]. Another possibility is that the presence of many haplotypes negatively influenced the fragmentation of the assembly. Due to more variation in the bases, it might have been more challenging to combine more contigs into scaffolds. To further reduce the number of scaffolds, possibly to a chromosome level, it might be advantageous in the future to supplement long-read sequencing with other techniques such as optical mapping [Deschamps et al., 2018, Field et al., 2020].

We furthermore assessed the effect of generating a genome assembly of a highly inbred line instead of a regular population. A comparison was made between SNPs' zygosities called on short-read data and found that Gr-Line19 had a 10% higher proportion of homozygous

SNPs than the JHI-Ro1 population. This suggests that there is indeed a smaller number of haplotypes present in the inbred-lines than in the JHI-Ro1 population. Since more than 50% of the called SNPs in Gr-Line19 are still heterozygous, it seems reasonable to assume that the measured heterozygosity levels provide a more realistic picture of the heterozygosity that is present in an individual. Which is in line with previous findings in the cyst nematode *Heterodera glycines* by Ste-Croix et al. [2021], who described that individuals can have mixing levels of zygosity.

A more detailed analysis of four effector families for which at least a subset of members are known to be expressed in the dorsal gland of nematodes during feeding site induction or maintenance revealed dozens of novel potential virulence-associated variants.

To some extent, our starting point was comparable to the approach taken by Bekal et al. [2008]. Within the soybean cyst nematode *Heterodera glycines* subsets of populations with comparable pathogenicity have been defined based on their multiplication characteristics on a set of seven soybean indicator lines. Populations that shared multiplication characteristics were coined 'HG types' [Niblack et al., 2002]. Subsequently, Bekal et al. [2008] used two inbred lines that were either avirulent ('TN10'; HG type 0)) or virulent ('TN20'; HG type 1, 2, 3, 4, 5, 6, 7). 454 micro-bead sequencing of these indicator lines resulted in the generation of tens of millions of short reads (110–120 bp), which allowed for whole-genome comparative analysis. These efforts resulted in 239 homozygous SNPs between TN20 and TN10 [Bekal et al., 2008]. Although the relationship between these SNPs and pathogenicity is unclear, these SNPs could be considered one of the first molecular markers for pathogenicity in cyst nematodes. Here we took it one step further, by identifying copy number variation that might serve as potential pathotype specific molecular markers. Copy number variation is relevant, as it has been linked to virulence in various pathogens [Brynildsrud et al., 2016, Zhao and Gibbons, 2018] including plant parasitic nematodes [Castagnone-Sereno et al., 2019].

For potato cyst nematodes, Folkertsma et al. [1996] used AFLP assays [Vos et al., 1995] to characterize pathotypes of the potato cyst nematodes *G. rostochiensis* and *G. pallida*. Almost 1,000 marker loci were employed to genotype populations of both potato cyst nematode species. These analyses revealed genetic markers that can distinguish between the *G. rostochiensis* pathotypes Ro1, Ro3 and Ro4, while such loci appeared to be absent for the *G. pallida* pathotypes Pa2 and Pa3. In a more extensive approach focusing on *G. rostochiensis* only, Mimee et al. [2015] employed a restriction enzyme-based genotyping-by-sequencing approach. The genotypic characterization of 23 populations, covering all five pathotypes, revealed a clear distinction between pathotypes Ro1 and Ro2 on the one hand, and Ro,3, Ro4, and Ro5 on the other. Moreover, their analyses seemed to demonstrate intra-pathotype variation within Ro1. However, it is noted that with 14 populations from 9 different countries, Ro1 was over-represented in this research.

The first reference genome for *G. rostochiensis* was published by Eves-van den Akker et al. [2016], and - in conjunction with this - the intra-species variation regarding members of known effector families was mapped. All five *G. rostochiensis* pathotypes were represented in this study, and whereas homozygous molecular markers to discriminate between Ro4 and Ro5 could be identified, this was not possible for the remaining three pathotypes. Moreover, this research confirmed the large genotypic diversity of populations that are all labeled Ro1, indicating that there are many possible genotypes that yield a similar Ro1-like virulence. Here it might be mentioned that Ro1 and Ro4 share the inability to parasitise potato genotypes that harbor the H1 resistance gene from *Solanum tuberosum ssp. andigena CPC 1673*. Moreover, the H1 resistance genes have been introgressed in most commercial potato varieties, and potato cyst nematode populations worldwide have been exposed to these resistance genes likely including the ones characterized by [Eves-van den Akker et al., 2016]. So, although these pathotypes share their avirulence concerning the H1 gene, they belong to another *G. rostochiensis* genotype and differ significantly in intra-pathotype variation.

Hence, as a starting point, we used pathotypically characterized inbred lines from which we generated new reference genome sequences. On this basis, complete effector families could be mapped and compared. In essence, the make-up of effector families in lines with distinct pathogenic characteristics could vary because of (1) non-synonymous variants in sequence in a given set of effector genes and/or (2) effector gene loss or gain (3) quantitative variation in expression levels due to SNPs in the promotor region (4) quantitative variation in expression levels due to copy number variation. The balance between these two (dependent) sources of variation varies in a pathogen-dependent manner. The genome-wide comparison of three *Microbotryum* species parasitising distinct *Caryophyllaceae* allowed Beckerson et al. [2019] to define the secretomes of the individual species. Their analyses revealed that host specificity was explained by rapid changes in effector genes rather than by variation in the effector copy numbers. With a similar underlying question, Qutob et al. [2009] investigated two effector genes families of *Phytophthora sojae*, Avr1a and Avr3a in a range of races. The presence of multiple copies of nearly identical genes on the Avr1a and the Avr3a locus was suggested to contribute to the fitness of these races, and races with distinct pathogenicities were characterized by variations in effector gene numbers. These examples demonstrate that both sources of variation can generate differences in pathogenicity among plant pathogens. Here we specifically focused on effector gene loss and gain effects, and observed that both events happen in the avirulent Gr-Line19 as well as the virulent Gr-Line22. Previous studies show that, at least in potato cyst nematodes, single nucleotide polymorphisms are also related to virulence (e.g., [Sacco et al., 2009]), which indicates that both types of genomic variance are relatable with virulence.

In case of the tropical root-knot nematode *Meloidogyne incognita*, Castagnone-Sereno et al. [2019] tried to pinpoint the genetic basis of avirulence and virulence with regard to the

tomato resistance gene Mi-1.2. Genome-wide characterization of two pairs of avirulent and virulent lines revealed 20 gene families that all showed a lower number of copies in both virulent *M. incognita* lines. It is noted that the 20 families included pioneers and household genes, and not known effector families. Hence, although a lower copy number per gene family was associated with virulence, this research did not identify gene loss events that could be causally related to virulence.

We separately considered the dorsal esophageal gland-expressed effectors that are thought to be involved in immune response suppression and feeding site induction, and the subventral gland-expressed effectors that are active during plant penetration. Concerning dorsal gland-expressed effector families, *G. rostochiensis* Gr-Line19 harboured on average 14 genes per effector family, while on average, 19 members were identified per effector family in Gr-Line22, a homozygous virulent line regarding the H1 resistance gene. In our analysis, four effector families showed a higher number of variants in the avirulent Gr-Line 19, and four other families showed a reverse pattern (Fig. 2.3 A). The aforementioned difference in the average number of variants per family is explained by the differences in the extent to which the number of variants had changed in the two lines.

The effector families expressed in the subventral glands that are included in this study showed less expansion than the dorsal gland specific families, with only small differences in copy numbers between the two lines. Strikingly, a substantial number of genes belonging to this category lack a signal peptide presence. Since many of these genes (e.g., glycoside hydrolases, pectate lyases) code for cell wall-degrading or modifying enzymes, the proteins would have to be secreted to make physical contact with the plant in order to perform their function. One hypothesis could be that these genes are, in fact, pseudogenes. However, this seems unlikely as manual inspection showed that most of these SP lacking genes show a RNAseq signal (results not shown). Whereas ample RNAseq data allowed for an accurate prediction of the intron-exon structure, the transcription start site is more difficult to predict without additional experimental data. If transcription start sites were misplaced, we could have missed a preceding signal peptide. Alternatively, it could be that this cyst nematode genuinely harbours effector variants without apparent signal peptide similar to the invertases identified in *Meloidogyne incognita*. These effectors were suggested to be acquired at a late stage during cyst nematode evolution [Flier et al., 2003].

Phylogenetic analysis of effector families as presented here takes along both effector diversification and effector loss and gain. These data clearly demonstrate that the balance between both sources of variation differs per effector family. Whereas effector family 1106 showed overall little copy number variation, SPRYSEC genes were 65% more abundant in Gr-Line22, and in case of the GLAND5 family significant diversification was accompanied by a large difference in copy numbers between both lines. Other population genetic studies on plant-pathogenic fungi and oomycetes showed exclusively low [Talas and McDonald, 2015]

or high [Flier et al., 2003] levels of diversification between effector genes. We are not aware of other plant pathogens for which such drastic contrasts in diversification pattern between effector families were described.

Because of its extreme level of physical clustering of CLE effectors in both *G. rostochiensis* inbred lines we investigated its genomic organisation in more detail. Potato cyst nematodes produce and secrete mimics of plant CLEs. Plant CLEs are signalling components that were shown to be conserved in both *Arabidopsis* and potato roots [Lu et al., 2009]. Among *Globodera* CLE genes two functional classes are distinguished, CLE1 and CLE4. The main difference between these classes is the composition of CLE peptides that are present as small cleavable units separated by small spacers at the protein's C terminus. Mitchum et al. [2012] described a single CLE1 representative, and here a second potential CLE1 variant is identified in both *G. rostochiensis* lines (Fig. 2.8 & Supplemental Figure 1 van Steenbrugge [2022]). This second variant has a domain structure similar to GrCLE1 (Supplemental Figure 1 [van Steenbrugge, 2022]), and the conservation of the domain structure makes it plausible this variant has a CLE1-like function. Notable is the putative duplication event of Gr-CLE4 genes in Gr-Line22. Gr-CLE-4 genes are highly conserved, even between pathotypes and we assume that this duplication event might result in a higher production of GrCLE4 peptides. A dose effect for a nematode effector was previously reported for the 32E03 effector of the beet cyst nematode Heterodera schachtii [Vijayapalani et al., 2018]. So our finding might suggest that the virulent *G. rostochiensis* line 22 might exert a stronger CLE4 peptide-based effects on its host.

2.4 Conclusions

Molecular pathotyping is an essential element in durable disease management. After all, this will allow breeders to use host plant resistances in a targeted way, and it allows farmers to make a more informed decision which potato variety to grow in the field. The existing pathotyping system for *G. rostochiensis* classifies populations into five pathotypes (Ro1-Ro5) on the basis of their relative multiplication rates on a number of *Solanum* differentials, and this systematic was used as starting point for the generation of new pathotyping platform. By generating high quality reference genomes from two pathotypically-distinct inbred lines, we were able to generate broad overviews of effector families including their diversification and spatial organisation. On the basis of a selection of four effector families, dozens of effector variants could be pinpointed that were unique for either of the two inbred lines Gr-Line19 (Ro1) and Gr-line22 (Ro5). Once these data are supplemented by re-sequencing data from well-characterized *G. rostochiensis* field populations, comparative effectoromics would be within reach. Comparative effectoromics will provide a foundation for our understanding of compatible and incompatible host-nematode interactions as well as for a new, biologically insightful pathotyping scheme as a basis for the durable use of host plant resistances.

2.5 Methods

2.5.1 DNA isolation and sequencing

Cysts from two *G. rostochiensis* lines that were previously selected by Janssen et al. [1990] for being fully avirulent Ro1 (Gr-line19) or fully virulent Ro5 (Gr-line22) with regard to the H1 gene were used as starting material for the collection of pre-parasitic second-stage juveniles (J2). J2 nematodes were concentrated, and sucrose centrifugation was used to purify the nematode suspension [Jenkins et al., 1964]. After multiple rounds of washing of the purified nematode suspension in 0.1 M NaCl, nematodes were resuspended in sterilized MQ water. Juveniles were lysed in a standard nematode lysis buffer with proteinase K and beta-mercaptoethanol at 60 °C for 1 h as described by Holterman et al. [2006]. The lysate was mixed with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) (pH 8.0) following a standard DNA purification procedure, and finally, DNA was precipitated with isopropanol. After washing the DNA pellet with 70% ethanol for several times, it was re-suspended in 10mM Tris-HCL (pH 8.0). DNAs of both inbred lines (each 10–20 µg) were sequenced using Pacific biosciences smrt sequencing technology at Bioscience (Wageningen Research, Wageningen, The Netherlands) Gr-line19 was sequenced to a depth of approximately 119X with an average read length of 5,641 bp, whereas Gr-line22 was sequenced 132X with an average read length of 7,469 bp. Depth was calculated based on the assembly sizes. In parallel, a 2x250 bp Illumina NovaSeq run resulted in 188x coverage of paired-end reads per line used to polish the initial assemblies. The raw sequencing reads and the genome assemblies are available under NCBI accession PRJNA695196.

2.5.2 Genome assembly

Raw PacBio reads were first corrected by merging haplotypes with the correction mode of CANU V1.8 [Koren et al., 2017], allowing a corrected error rate of 15% and a corrected coverage of 200. Using long-read assembler WTDGB2 v2.3 [Ruan and Li, 2020], approximately one hundred assemblies were generated per inbred line, optimizing the parameters minimal read length, k-mer size, and minimal read depth. The quality of the initial assemblies was assessed based on whether the assembly size was close to the genome size estimate [Eves-van den Akker et al., 2016]. Completeness of the genome was assessed using BUSCO v3 [Seppey et al., 2019] using the standard library of eukaryotic single copy genes. Based on the criteria mentioned before, the most optimal assembly was then selected for each line and used for post-assembly processing. For Gr-Line19 a minimal read length of 6,000 was used, together with a k-mer size of 20 and a minimal read depth of 6. For Gr-Line22 a minimal read length of 5,000 was used, together with a k-mer size of 15 and a minimal read depth of 6.

After determining the most optimal assembly, remaining unmerged haplotigs were fil-

tered from the assembly using PURGE HAPLOTIGS v1.0.4 [Roach et al., 2018]. The assembly was then tested for contamination using the BLOBTOOLS pipeline v.1.0.1 [Laetsch and Blaxter, 2017]. Contigs were scaffolded with PacBio reads using SSPACE-LONGREAD with a minimum overlap length of 1000 bp and a minimum gap between two contigs of 500 bp [Boetzer and Pirovano, 2014]. The remaining gaps in the scaffolds were then filled using a consensus alignment approach with a minimum coverage per position of 10 reads [van Steenbrugge, 2021]. NovaSeq data were used to polish the resulting assemblies using three iterations of ARROW v2.3.3 at default settings (`https://github.com/PacificBiosciences/GenomicConsensus`) and five iterations of PILON v1.23 [Walker et al., 2014] each. Repeat regions were soft masked using REPEATMODELER v1.0.11 (`https://github.com/Dfam-consortium/RepeatModeler`) and REPEATMASKER v4.0.9 (`http://www.repeatmasker.org/RepeatMasker/`). Gene annotations in gff3 format were predicted for both assemblies using BRAKER v2.1.2 [Brůna et al., 2021]. The prediction of gene models was aided by RNAseq datasets of different life stages of *G. rostochiensis* (NCBI BioProject accessions: PRJEB12075, PRJNA274143). While this data originates from a different *G. rostochiensis* population (JHI-Ro1), addition of this type of data greatly improved the quality of the gene predictions using BRAKER. Sequencing reads from these RNAseq datasets were mapped on both genomes using HISAT v2.2.0 [Kim et al., 2019]. All scripts used for the generation of the genome assemblies including all relevant details are available on GITHUB (`https://github.com/Jorisvansteenbrugge/GROS_genomes`).

2.5.3 Genome Synteny

Genome synteny was determined between the genome assembly of Line19, Line22 and the previous Ro1 reference genome (NCBI BIOPROJECT PRJEB13504) through a progressive genome alignment using MAUVE v2.4.0. The alignment was then visualized in CIRCOS v0.69-9 [Krzywinski et al., 2009], showing only syntenic regions of 3 kb and larger.

2.5.4 Estimating Heterozygosity levels and structural variation

Heterozygosity levels were estimated based on the frequency of heterozygous versus homozygous variants (SNPs and small indels). A comparison was made between Gr-Line19 and the JHI-Ro1 population, using the Gr-Line22 genome assembly as a reference. Illumina reads were mapped with BURROWS-WHEELER ALIGNER [Li and Durbin, 2009] using default settings. For Gr-Line19, a library of Illumina NovaSeq reads (accessions: SRR13560389, SRR13560388) was used, and for JHI-Ro1, Illumina HiSeq reads (accessions: ERR114519) were mapped against the reference. Variants were called with BCFTOOLS v.1.9 [Li, 2011] with multiallelic variant calling enabled, at a maximum depth of 1,000 reads.

The structural variation between the newly generated assemblies of Gr-Line19 and Gr-Line22 was estimated by the frequency of heterozygous versus homozygous structural variants (SVs)

with fragment size > 1 kb. Raw Pacbio reads of Gr-Line19 were mapped against the Gr-Line22 (and vice versa) using NGMLR v0.2.7 [Sedlazeck et al., 2018] with default settings. SVs were then called using SNIFFLES v1.0.10 running the standard settings [Sedlazeck et al., 2018].

2.5.5 Identification of Effector Homologs

Effector genes were identified in both line Gr-line19 and Gr-line22 based on the proteomes predicted by BRAKER2 [Brůna et al., 2021]. PHOBIUS [Käll et al., 2004] was used to check for the presence of a signal peptide for secretion. Homologs for glycoside hydrolase (GH) families 5, 30, 43, 53, Pectate lyase 3, Glutathione Synthetase were identified with HMMER v3.2.1 [Mistry et al., 2013] based on pre-calculated profile HMMs in the PFAM database [Bateman et al., 2004] (entries PF00150, PF02055, PF04616 and PF07745, PF03211, PF03199 respectively). SPRYSEC homologs were identified by testing protein sequences for a SPRY domain (hmm profile PF00622). Arabinogalactan galactosidase homologs were identified with a custom profile HMM-based on UNIPROT sequences (entries O07012, Q65CX5, Q65CX4, D9SM34, P48841, O31529, Q8168, Q5B153, O07013, P83692, P48842, P83691, Q4WJ80, B0XPR3, A1D3T4, Q2UN61, Q0CTQ7, A2RB93, Q9Y7F8, B8NNI2, Q76FP5). CLE-like homologs were identified with a custom profile HMM-based on UNIPROT sequences (D1FNJ7, D1FNK5, D1FNJ9, D1FNK2, D1FNK8, D1FNK3, D1FNK0, D1FNK4). GENBANK peptide sequences JQ912480 to JQ912513 were used to generate a custom profile HMM for the effector family 1106. Based on GENBANK entries KM206198 to KM206272, a custom profile HMM was made for the HYP effector family. Homologs of the *Hetereodera glycines* effector families Hg16B09 (GENBANK: AAO85454) and GLAND1-18 (GENBANK: KJ825712 to KJ825729) were identified with BLASTP, with the following cut-offs: an identity score higher than 35%, a query coverage of at least 50%, and an E-value lower than 0.0001.

2.5.6 Phylogeny

Multiple Sequence Alignments were generated based on the coding sequences of the orthologs per effector family, using MUSCLE v3.8.1551 [Edgar, 2004] using standard options. To test for the best model of DNA substitution, MODELTEST-NG [Darriba et al., 2020] was used. Except for GLAND5, the best model for all effector families was GTRGAMMA. For GLAND5, GTRGAMMAI was marginally better. As the resulting phylogenetic tree was almost identical to the GTRGAMMA, we decided to stick to this model for sake of uniformity. Phylogenetic trees were then generated with RAXML v8.2.12 [Stamatakis, 2014] running a GTRGAMMA model with 100 bootstrap replicates. The resulting trees were visualized and organized in FIGTREE v. 1.4.4.

2.6 Supplemental Data

2.6.1 Supplemental Figures

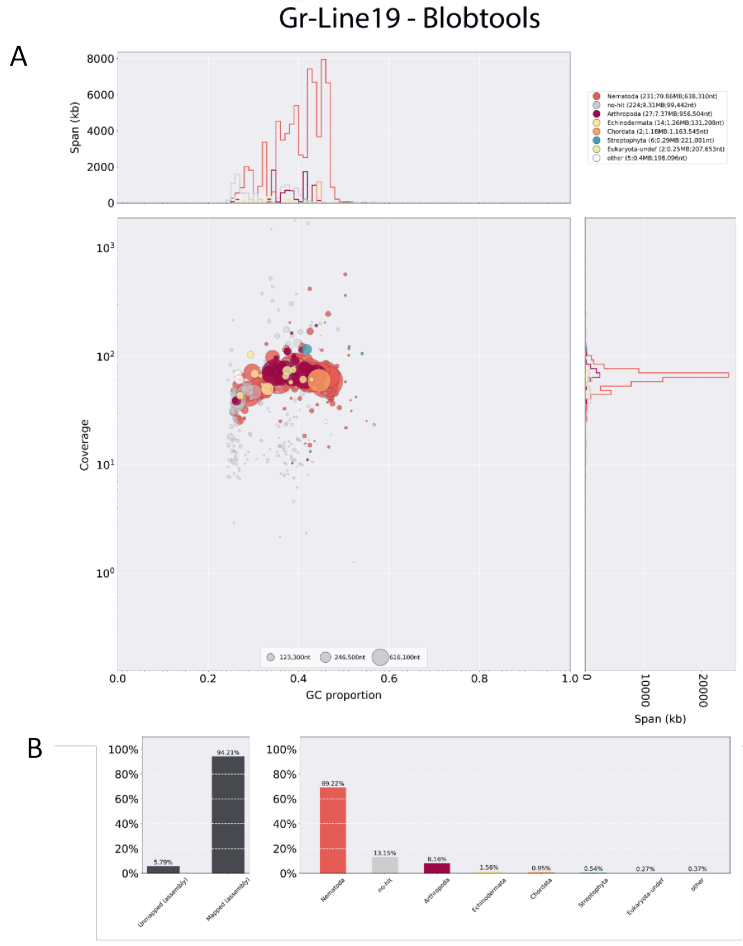


Figure 2.10: BlobTools-based interrogation of genome assembly of Gr-Line 19 to verify for single-taxon origin of the original sequences. Panel A: Each Gr-Line19 scaffold is represented by a single filled circle. Each scaffold is positioned in the main panel based on its GC proportion (x-axis) and coverage by reads from PacBio sequences (y-axis). On the top right the colours of the individual blobs are linked to their taxonomic origin. At the bottom of the main Blobtool figure, the size of the circles is linked to scaffold size. Panel B: on the left the percentage of unmapped versus mapped Gr-Line19 PacBio reads are presented, and the right the taxonomic origin of the reads.

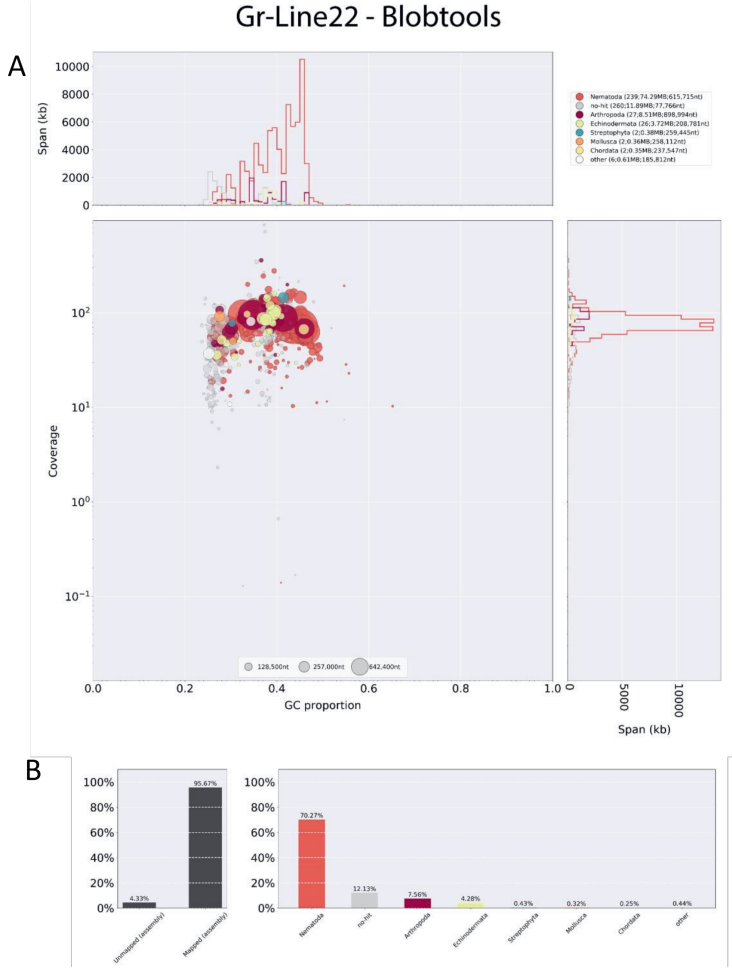


Figure 2.11: BlobTools-based interrogation of genome assembly of Gr-Line 22 to verify for single-taxon origin of the original sequences. Panel A: Each Gr-Line22 scaffold is represented by a single filled circle. Each scaffold is positioned in the main panel based on its GC proportion (x-axis) and coverage by reads from PacBio sequences (y-axis). On the top right the colours of the individual blobs are linked to their taxonomic origin. At the bottom of the main Blobtool figure, the size of the circles is linked to scaffold size. Panel B: on the left the percentage of unmapped versus mapped Gr-Line22 PacBio reads are presented, and the right the taxonomic origin of the reads.

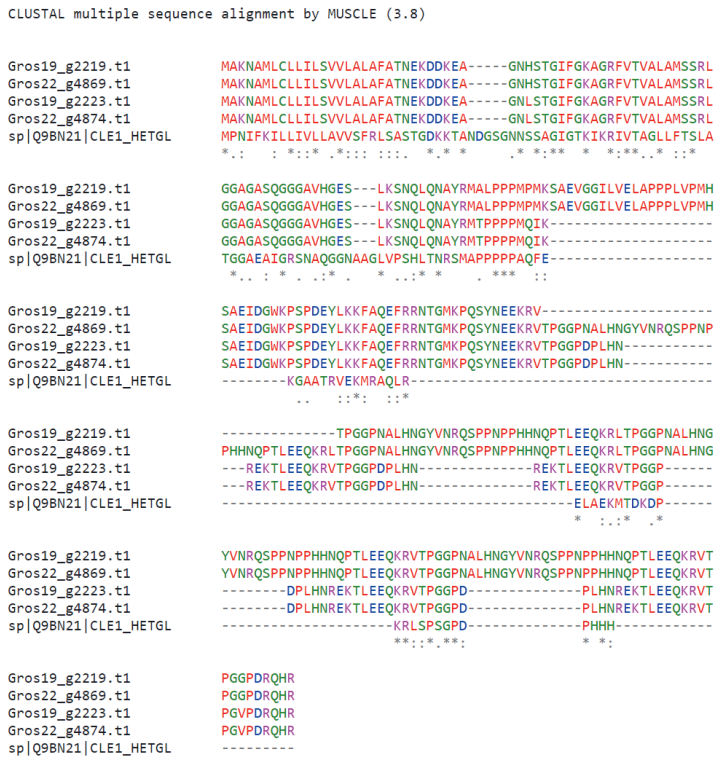


Figure 2.12: Multiple sequence alignment of Gr-CLE-1 protein sequences to verify the conservation of the CLE domain in putative CLE-1 genes. Each gene is represented by a gene identifier. Two genes are included for Gr-Line19 (Gros19_g2219.t1 and Gros19_g2223.t1) and two for Gr-Line22 (Gros22_g4869.t1 and Gros22_g4874.t1). The CLE1 sequence identified in *Heterodera glycines* (Q9BN21) is included as an outgroup.

CHAPTER 3

Genomic reconstruction of the introduction and diversification of *Globodera rostochiensis* populations in Indonesia

Nurul Dwi Handayani[1,2,3], Prabowo Lestari[1,2,3], Wouter van As[2], Martijn Holterman[2,4], Sven van den Elsen[2], Antarjo Dikin[5], Wim Bert[3], Johannes Helder[2], and Joris J.M. Van Steenbrugge[2].

[1] Indonesian Agricultural Quarantine Agency, Ministry of Agriculture, Ragunan, Jakarta, Indonesia
[2] Laboratory of Nematology, Wageningen University and Research, The Netherlands
[3] Nematology Research Unit, Department of Biology, Ghent University, Belgium
[4] Solynta, Wageningen, The Netherlands
[5] Directorate General of Estate Crops, Ministry of Agriculture, Ragunan, Jakarta , Indonesia

This chapter is published in Phytopathology [Handayani et al., 2021].

3.1 Abastract

Potato cyst nematodes (PCNs), the umbrella term for *Globodera rostochiensis* and *G. pallida*, co-evolved with their *Solanaceous* hosts in the Andean Mountain region. From there, PCN proliferated worldwide to virtually all potato production areas. PCN is a major factor limiting the potato production in Indonesia. In our survey, only *G. rostochiensis* was found. Fourteen field populations were collected on Java and Sumatra, and unique variants were called by mapping re-sequencing data on a *G. rostochiensis* reference genome. A phylogenetic tree based on 1.4 million unique variants showed a genotypic separation between the outgroup, a Scottish Ro1 population, and all Indonesian populations. This separation was comparable in size with the genotypic distinction between the Javanese and the Sumatran PCN populations. Next, variants within PCN effector gene families SPRYSEC, 1106, 4D06, and venom allergen-like protein (VAL) that all interfere with the host innate immune system were compared. Distinct selective pressures acted on these effector families; while SPRYSECs (4,341 single-nucleotide polymorphisms/insertions or deletions of bases) behaved like neutral genes, the phylogenetic trees of 1106, 4D06, and VAL proteins (235, 790, and 150 SNPs/indels, respectively) showed deviating topologies. Our data suggest that PCN was introduced on Java not too long after the introduction of potato in the middle of the eighteenth century. Soon thereafter, the pathogen established on Sumatra and started to diversify independently. This scenario was corroborated by diversification patterns of the effector families 1106, 4D06, and VAL. Our data demonstrate how genome re-sequencing data from a non-indigenous pathogen can be used to reconstruct the introduction and diversification process.

3.2 Introduction

Potato is worldwide the fourth most important staple food and the most widely produced non-grain. One of the attractive aspects of potato production is the high average yield per acreage. Whereas corn, wheat, and rice typically yield 10 tonnes or less per hectare ($hectare^{-1}$) annually ($year^{-1}$), a potato crop on average produces ∼20 t ha^{-1} y^{-1}. China and India are the largest potato producers, but a large gap in productivity between developing and developed countries is worth noting. In developing countries, potato yields are in the range of 15 to 20 t ha^{-1} y^{-1}, while more than double this production is commonly reached in northwestern Europe and North America [Faostat, 2015]. With regard to potato consumption, a slight decrease is observed in developed countries, whereas in developing countries potato consumption has doubled between 1960 and 2005 [Jennings et al., 2020].

By far, most potato production is realized outside the center of origin of this crop, the Andean Mountain region in South America. Potato as we know it, *Solanum tuberosum ssp. tuberosum*, is a member of the species-rich genus *Solanum* (1,500 to 2,000 species). Representatives of this genus are found all over the world with concentrations of species diversity in Central and South America and Australia [Hawkes, 1988]. Tuber-bearing *Solanum* species all originate from Central and South America. First reports about human consumption of potatoes outside of its center of origin were found in the account books from 1576 of the Hospital de la Sangre in Seville (Spain; [HAWKES and FRANCISCOORTEGA, 1992]). In the nineteenth century, we see a rapid proliferation of the cultivation of potato as a main crop in Europe.

Large-scale production of a main crop outside its center of diversity is often associated with serious pest and pathogen issues. In the case of potato, this vulnerability is aggravated by the genetic homogeneity of the crop in most production settings. Probably the best illustration of the vulnerability of potato production outside of its native range to pests and diseases is the enormous impact of the oomycete *Phytophthora infestans*, the causal agent of potato late blight, on food production in Ireland in 1845 to 1850. In those years, the potato production collapsed, and in the resulting Irish Potato Famine, the population dropped by ∼50%, mainly because of starvation and emigration (e.g., [Yuen, 2021]).

Potato cyst nematode is the causal agent of a potato disease that co-evolved with potato in the Andean Mountain region and one that has a severe impact on potato production worldwide. Potato cyst nematode is the common name for two related plant-parasitic nematode species, *Globodera rostochiensis* and *G. pallida*. Unlike *P. infestans*, infection with potato cyst nematodes is not immediately clear in the above ground plant parts. These nematodes are soil borne pathogens causing an overall reduction in the growth of the potato plant and an increased vulnerability to other environmental stresses such as drought and opportunistic pathogens. Together with *P. infestans*, potato cyst nematodes belong to the most critical

biotic crop-limiting factors in potato production.

Over the last decades, a worldwide increase in potato production has been accompanied by a focal shift from Europe and North America to Asia. Indonesia is one of the countries in which the steady increase in local production accompanied the increased consumption of potatoes. For smallholder farmers, potato is attractive as potato is one of the most profitable crops. Potato production in Indonesia is not new; the first written report on the cultivation of "Irish potato" was published in 1711 [Lockyer, 1711]. In a historical overview, Boomgaard [2003] reported that potatoes were grown continuously but on a relatively small scale in the highlands of Java and Sumatra from the 1750s onward. According to this overview, potatoes were not grown outside these two main islands till the 1950s.

The potato cyst nematode *G. rostochiensis* was reported for this first time on east Java in 2003 [Indarti et al., 2004]. In a follow-up survey on Java, this potato cyst nematode species was shown to be present in west, central, and east Java in potato fields at an altitude between 1,500 and 2,000 m [Nurjanah et al., 2016]. *G. rostochiensis* was for the first time reported on Sulawesi [Handayani et al., 2020]. In a single report [Suastika et al., 2012], *G. pallida*, the other potato cyst nematode species, was described to be present in central Java. Because of the non-conspicuous nature of potato cyst nematodes and as a consequence, the limited attention that has been paid to this soil borne disease by producers, the cyst nematode has spread enormously and is labeled as a major threat to Indonesian potato production.

Host plant resistances are one of the most successful means to control the potato cyst nematode *G. rostochiensis*. However, potato cyst nematode resistances are invariably only effective against subsets of *G. rostochiensis* populations. These subsets have been grouped into pathotypes, and the scheme proposed by Kort et al. [1977] is most frequently used. Pathotype-specific resistance genes such as H1, a single dominant resistance gene against the *G. rostochiensis* pathotypes Ro1 and Ro4 [Toxopeus and Huijsman, 1952], are highly effective and were shown to be durable. Unfortunately, there are no reliable tests to pinpoint the pathotypic nature of potato cyst nematode in an agronomic setting (see [Folkertsma et al., 1996]). As a consequence, resistant potato varieties are applied in a non-informed manner, thereby potentially disappointing the expectations of farmers and shortening the agronomic life span of these precious resistance genes. Hence, the genotypic characterization of Indonesian *G. rostochiensis* populations would be a first and essential step toward the durable control of this major plant pathogen.

Given the impact of potato cyst nematodes on potato production worldwide, there has been ample research attention for these notorious plant pathogens. Just like for other plant pathogens, effectors play a key role in the interaction between cyst nematodes and their host. Effectors are proteins produced and secreted by pathogens to manipulate their host. In the case of potato cyst nematodes, effectors facilitate host penetration, weaken the

host defense responses, and induce the formation and maintenance of a feeding site (for an overview, [Khan and Khan, 2021]). Effectors are typically encoded by gene families that show some degree of physical genomic clustering. Host plant resistances are usually based on the recognition of a specific effector variant by a resistance protein. In the case of potato cyst nematodes, this can be exemplified by an SPRYSEC variant, RBP-1, that activates the resistance protein GPA2 [Sacco et al., 2009]. Detailed knowledge about the effector repertoire present in field populations of potato cyst nematodes is key to a targeted use of resistant potato varieties.

Whole-genome sequencing has the potential to provide a complete overview of the diversity and genomic organization of effector families. The genomes of the two potato cyst nematodes species have been sequenced [Cotton et al., 2014, Eves-van den Akker et al., 2016], but these genomes are generally spoken of as being too fragmented (4,281 scaffolds in the case of *G. rostochiensis*) to offer a complete picture of the diversification and genomic organization of effector families. van Steenbrugge et al. [2021] generated two novel *G. rostochiensis* genomes that are approximately 25 times less fragmented than the one presented by Eves-van den Akker et al. [2016], and these reference genomes were used as a starting point for the genetic characterization of Indonesian potato cyst nematode populations. Potato cyst nematode populations were collected from major potato production areas in Java and Sumatra. These populations were characterized based on a set of neutral molecular markers; internal transcribed spacer (ITS), mitochondrial cytochrome C oxidase subunit I (COI), and microsatellites. These first two markers were shown to be too conserved to assess diversity among populations, whereas microsatellites allowed for the discrimination between populations from two main islands, Java and Sumatra [Handayani et al., 2020]. Here we used another approach, whole-genome resequencing, to map the genetic diversity of Indonesian *G. rostochiensis* populations and to pinpoint the diversity of effector families. Next to information about genetic diversity, this approach offers clues to reconstruct the introduction of *G. rostochiensis* in Indonesia and its proliferation within the Indonesian archipelago.

3.3 Methods

Soil samples were collected from 14 field sites where potato was grown as a main crop (Table 3.1). Cysts were extracted from a sub sample of 100 ml of dried soil using the Baunacke method (i.e., dried cysts that float in water were decanted and collected on a 250-µm sieve). Then, the cysts were air dried overnight, counted, and stored dry at 4°C.

Table 3.1: Overview of potato cyst nematode collection sites in Indonesia

Province	District	Field	Altitude (m)	Coordinates	
				Latitude	Longitude
North Sumatra	Karo	Cinta Rakyat	1,362	3.16333	98.49555
		Gajah	1,314	3.15333	98.47416
		Guru Singa	1,370	3.19416	98.47861
		Lingga Julu	1,256	3.13555	98.47249
		Suka Ndebi	1,401	3.19888	98.47500
Central Java	Banjarnegara	Bakal Buntu	1,950	-7.28224	109.97797
		Dieng Kulon-1	2,081	-7.20311	109.90293
		Dieng Kulon-2	2,079	-7.20281	109.90215
		Karang Tengah	2,037	-7.31235	109.76660
		Kejajar	1,507	-7.21111	109.92476
	Wonosobo	Patak Banteng	1,983	-7.24371	109.94776
		Tieng	1,766	-7.23916	109.94499
East Java	Batu	Krajan	1,500	-7.86503	112.55813
		Lemah Putih	1,637	-7.77360	112.53643

3.3.1 DNA extraction and purification

A quantity of 15 to 50 cysts per sample were added to 150 µl of Milli-Q water (Millipore Sigma, Burlington, MA) supplemented with two iron beads (Ø = 3 mm) in a 1.5-ml tube, and samples were bead beaten for 80 s at 30 Hz. Homogenized cyst material (without beads) was transferred to a new 2-ml tube prefilled with 150 µl of Milli-Q water. A quantity of 300 µl of 2× nematode lysis buffer [Holterman et al., 2006] was added, and tubes were incubated in a Thermomixer (Eppendorf, Hamburg, Germany) for 1 to 2 h at 60°C at 650 rpm. Once the sample was cooled back to room temperature, 150 µl of 5 M NaCl was added to the lysate. Subsequently, an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) at pH 8.0 (4°C) was added, and the resulting suspension was gently mixed, followed by phase separation by centrifugation (2,500 × g for 10 min). The upper aqueous layer was

transferred to a new tube, and an equal volume of chloroform/isoamyl alcohol (24:1/4°C) was added. After gentle mixing, the DNA was spun down by centrifugation (15,000 × g for 5 min). Thereafter, the supernatant was carefully removed. The DNA pellet was washed with 300 µl of 70% ethanol and spun down again for 1 min at 15,000 × g. The supernatant was removed, and the pellet was dried for 1 min at 60°C. The final, gently dried pellet was resuspended in 50 µl of 10 mM Tris-HCl at pH 8.0. A Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA) was used to quantify the DNA concentrations in the samples, and the purity of the samples was checked on a NanoDrop spectrophotometer (Thermo Fisher Scientific; 260/280 and 260/230 nm). All samples fulfilled the standard requirements for Illumina sequencing. The resulting purified DNA samples were sent out for library preparation and subsequently, sequencing on an Illumina NovaSeq 6000 with an SP flow cell (2 × 250 bp) at Useq (Utrecht, The Netherlands). Sequencing data have been deposited at the NCBI BioProject database under accession number PRJNA730992.

3.3.2 Mapping and variant calling

The sequence data in fastq format of the Indonesian *G. rostochiensis* populations were mapped on the Gr-Ro1-Line19 reference genome sequence (SAMN17613196; van Steenbrugge et al. [2021]) using the BWA-MEM algorithm [Li and Durbin, 2009] at default settings. Sequence variation was determined using the program BCFTOOLS MPILEUP [Li, 2011] with a maximum read depth and insertion or deletion of base (indel) depth of 1,500 and minimum mapping quality of 50. Allelic depth information was saved in the output. Variants were called using the program BCFTOOLS CALL with the built-in alternative model for multi-allelic and rare-variant calling enabled. The resulting variants with a read depth of <10 were pruned using the BCFTOOLS FILTER.

3.3.3 Phylogeny

Phylogeny between the populations was determined based on the variants. First, variants were pruned based on a linear discriminant analysis with a threshold of 0.5. A dissimilarity matrix was calculated on the remaining variants of each population using the snpgdsIBS function in the R package SNPRELATE [Zheng et al., 2012]) based on both autosomal and non-autosomal variants. Hierarchical clustering was then performed on the dissimilarity matrix using the snpgdsHCluster function in the program SNPRELATE. The phylogenetic tree was inferred from the clustering using the snpgdsCutTree function in SNPRELATE at default settings. In our analyses, the genome of a Scottish *G. rostochiensis* Ro1 population [Eves-van den Akker et al., 2016] was used as the outgroup.

3.4 Results

The availability of a high-quality reference genome of the potato cyst nematode *G. rostochiensis* [van Steenbrugge et al., 2021] that allows for the mapping of numerous short reads from Indonesian field populations made it possible to characterize and compare their genomic constitutions. In an attempt to assess the history and the proliferation of Indonesian potato cyst nematode populations over various major production areas, cysts of 14 Indonesian *G. rostochiensis* populations were resequenced. As this plant parasite is strictly amphimictic, cysts harbor a genetically diverse progeny. Cysts are kept together and protected by the hardened cuticle of the dead female, and typically harbor 100 to 400 eggs each. As at least 15 cysts were analyzed per sample, and a minimum of 1,500 genotypes were characterized per population.

An Illumina NovaSeq run generated approximately 38 million paired-end reads (2 × 250 bp) per population. These reads were mapped back to the *G. rostochiensis* Gr-Ro1-Line19 reference genome [van Steenbrugge et al., 2021] using the BWA-MEM algorithm. The coverage per population was 200× on average, ranging from 181× for population Karang Tengah to 234× for population Dieng Kulon. Effector diversification among Indonesian potato cyst nematode populations: The venom allergen-like protein family.

Venom allergen-like proteins (VALs) are effectors for which expression in *G. rostochiensis* coincides with migration in the plant. At least one member (Gr-VAP1) was shown to be expressed in the sub-ventral glands [Postma et al., 2012]. Members of this effector family are secreted in the apoplast where they suppress the plant's basal defense response [Lozano-Torres et al., 2014]. The *G. rostochiensis* line 19 harbors eight VAL variants. In total, 150 variants were detected among the Indonesian *G. rostochiensis* populations. Also here, a high level of dissimilarity was shown between the outgroup (JHI_Ro1) and the Indonesian populations (Fig. 5). Three major subclades can be distinguished: two are island specific, and one comprises the VAL variants that are found in both the central Javanese and Sumatran *G. rostochiensis* populations.

3.4.1 A genome-wide single-nucleotide polymorphism-based comparison of Indonesian potato cyst nematode populations

Phylogenetic trees were rooted with an Ro1 field population from Scotland (United Kingdom) named JHI-Ro1 [Eves-van den Akker et al., 2016]. Keeping the historic trading routes for (seed) potatoes in mind, it seems plausible to assume that *G. rostochiensis* arrived in Indonesia via one of its northwest European trading partners (e.g., The Netherlands, the United Kingdom). Variant calling based on resequencing data mapped to the reference genome Gr-Ro1-Line19 resulted in the identification of 1,404,945 unique variants. Given that the Gr-Ro1-Line19 genome's haploid size is 92 Mb, it implies that these variants cover

~0.76% of the diploid genome.

A phylogenetic tree based on all variants (Fig. 3.1) demonstrated that all populations had distinct and discernible genetic constitutions. This can be illustrated by two closely related populations from east Java: Krajan and Lema Putih (Fig. 3.1). A quantity of 359,618 variants defines the difference in genetic constitution between these two populations (while they share 1,045,327 variants). Moreover, all Indonesian populations were remarkably distinct from the outgroup, JHI-Ro1 from the United Kingdom. Three populations from central Java were positioned at the base of the tree; the remaining central Java populations are in a clade together with the east Java populations. This clade was found to have a sister relation with a distinct clade harboring only Sumatran populations.

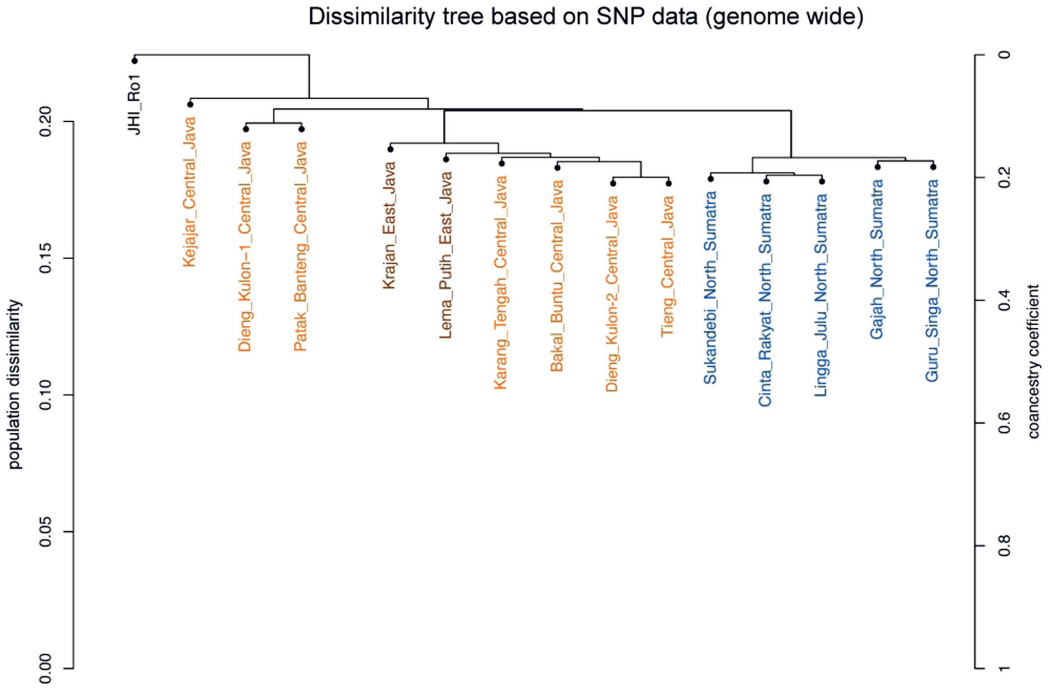


Figure 3.1: Dissimilarity trees based on 1,404,945 unique variants (single-nucleotide polymorphisms [SNPs]/insertions or deletions of bases [indels]). Variants were called by mapping resequencing data from Indonesian field populations on a reference genome of *Globodera rostochiensis* (Gr-Ro1-Line19; van Steenbrugge et al. [2021]). The genome sequence of a selected field population from Scotland (United Kingdom) referred to as JHI-Ro1 [Eves-van den Akker et al., 2016] was used as the outgroup.

3.5 Effector diversification among Indonesian potato cyst nematode populations: The SPRYSEC family

The large majority of the 1.4 million variants that underlie Figure 3.1 is unrelated to pathogenicity. To verify the effect of potato cultivation on pathogenicity, effector families, for which at least one member was shown to be expressed in the dorsal esophageal glands of parasitic juveniles, were investigated. SPRYSEC, a family involved in the disruption of the host's innate immunity [Diaz-Granados et al., 2016], is represented by 60 paralogs in G. rostochiensis Gr-Ro1-Line19. In total 4,341 single-nucleotide polymorphisms/indels were called among SPRYSEC family members represented in the 14 Indonesian G. rostochiensis populations. The topology of the SPRYSEC dissimilarity tree (Fig. 3.2) is remarkably congruent to the tree based on all unique variants (Fig. 3.1). Also here, we observed a clear distinction between the Scottish Ro1 population, and the most closely related Indonesian central Java populations, Kejajar, Dieng Kulon-1, and Patak Banteng. A sister relationship was also observed between all Sumatran potato cyst nematodes populations, and all east Java populations together with central Java populations of Karang Tengah, Bakal Buntu, Dieng Kulon-2, and Tieng. These observations suggest that the SPRYSEC family as a whole behaves as a neutral gene family in the Indonesian potato cultivation areas investigated in this study.

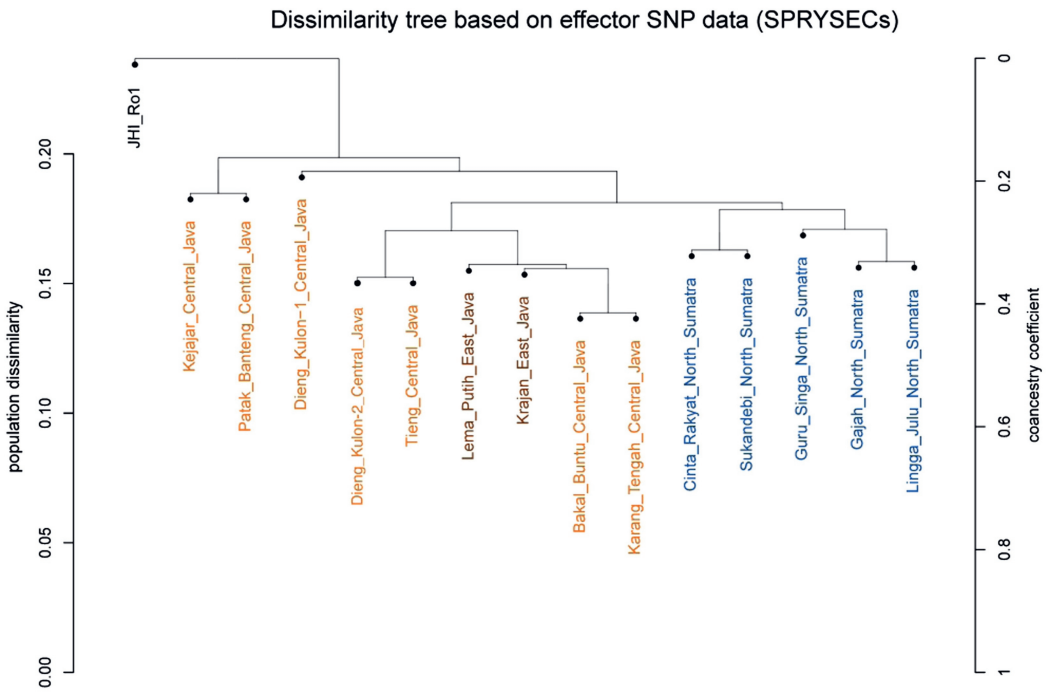


Figure 3.2: Dissimilarity trees based on 4,341 (single-nucleotide polymorphisms [SNPs]/insertions or deletions of bases [indels]) among members of the effector family SPRYSEC as represented in the sampled Indonesian *Globodera rostochiensis* populations. SPRYSEC variants were called by mapping resequencing data from Indonesian field populations on a reference genome of *Globodera rostochiensis* (Gr-Ro1-Line19; van Steenbrugge et al. [2021]). The genome sequence of a selected field population from Scotland (United Kingdom) referred to as JHI-Ro1 [Eves-van den Akker et al., 2016] was used as the outgroup.

3.6 Effector diversification among Indonesian potato cyst nematode populations: The 1106 family

The 1106 effector family, also referred to as the GLAND4 family in the soybean cyst nematode Heterodera glycines [Noon et al., 2015], interferes with both the basal and the effector triggered immunity response by the host plant [Finkers-Tomczak, 2011]. The reference genome of *G. rostochiensis* Gr-Ro1-Line19 harbors ten 1106 variants. With an antisense DNA probe designed on a conserved part of the 1106 gene family, the 1106 gene family transcripts were shown to localize in the dorsal esophageal gland of infective J2 [Finkers-

Tomczak, 2011]. Despite the apparently similar functions and expression patterns of SPRY-SEC and the 1106 proteins in *G. rostochiensis*, analysis of 235 unique variants in the 1106 gene family members in the Indonesian field populations resulted in a phylogenetic tree with a topology that deviates from the ones shown in Figures 3.1 and 3.2. As shown in Figure 3.3, the 1106 gene effector family variants present in Dieng Kulon 2 from central Java are highly comparable with the diversity present in the JH-Ro1 population. Moreover, the Sumatran populations no longer cluster together. More specifically, four Sumatran populations (Sukandebi, Cinta Rakyat, Lingga Julu, and Gajah) have their own Javanese equivalent, in regard to the composition of the 1106 gene family.

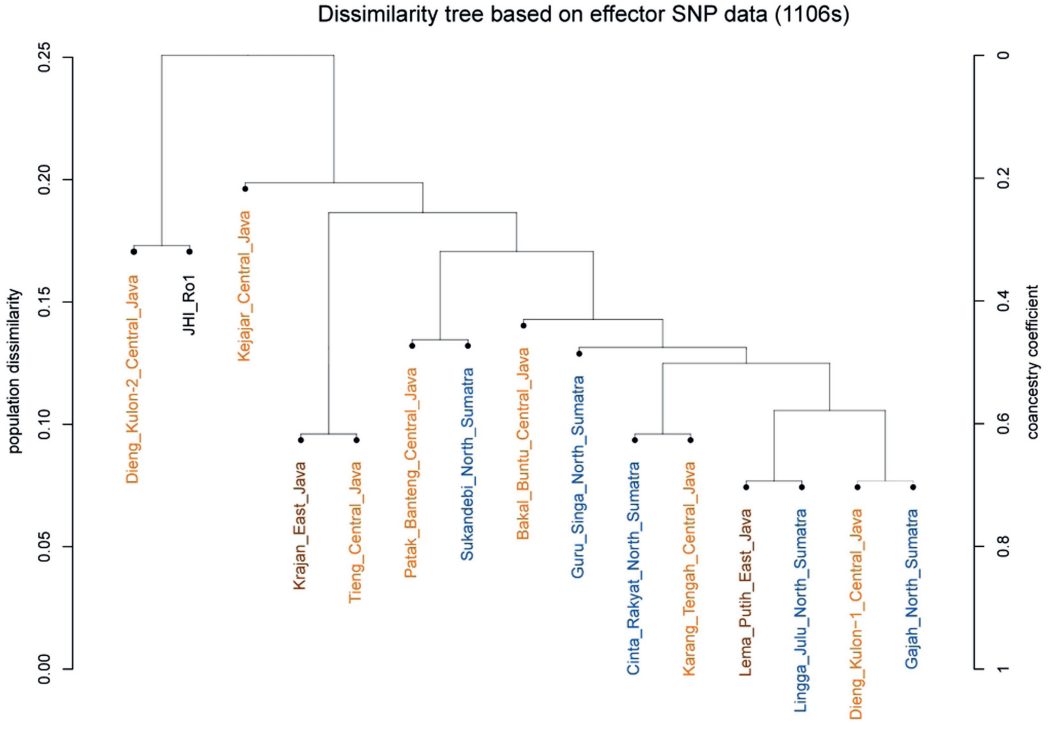


Figure 3.3: Dissimilarity trees based on 235 unique variants among members of the effector family 1106 as represented in the sampled Indonesian *Globodera rostochiensis* populations. The 1106 variants were called by mapping resequencing data from Indonesian field populations on a reference genome of *Globodera rostochiensis* (Gr-Ro1-Line19; van Steenbrugge et al. [2021]). The genome sequence of a selected field population from Scotland (United Kingdom) referred to as JHI-Ro1 [Eves-van den Akker et al., 2016] was used as the outgroup.

3.7 Effector diversification among Indonesian potato cyst nematode populations: The 4D06 family

The 4D06 effector family, which has also been described as GLAND6 in *H. glycines* [Noon et al., 2015], has been characterized as a suppressor of PAMP-triggered immunity [Pogorelko et al., 2020]. In the reference genome Gr-Ro1-Line19 the 4D06 family is represented by 23 variants [van Steenbrugge et al., 2021]. The 4D06 family was shown to be expressed in the dorsal esophageal gland during early stages of plant parasitism in *G. pallida* [Thorpe et al., 2014] and *H. glycines* [Noon et al., 2015], and we assume that members of this effector family are also expressed in the dorsal gland in *G. rostochiensis* during the onset of parasitism. In total, 790 unique variants were identified among the 4D06 family members represented in the Indonesian *G. rostochiensis* populations. The dissimilarity tree generated for this effector family showed a relatively high level of dissimilarity between JHI-Ro1 and the closest Indonesian population, Patak Banteng from central Java (Fig. 3.4). Moreover, we see a clustering of the Sumatran populations interspersed with the Kejajar population from central Java. Hence, unlike the SPRYSEC family that shows a differentiation pattern comparable to the one generated when all unique variants were taken into consideration (Figs. 3.1 and 3.2), variants of the effector families 1106 and 4D06 show distinct diversification patterns (Figs. 3.3 and 3.4) that might reflect the effects of exposure to distinct potato genotypes and agricultural practices.

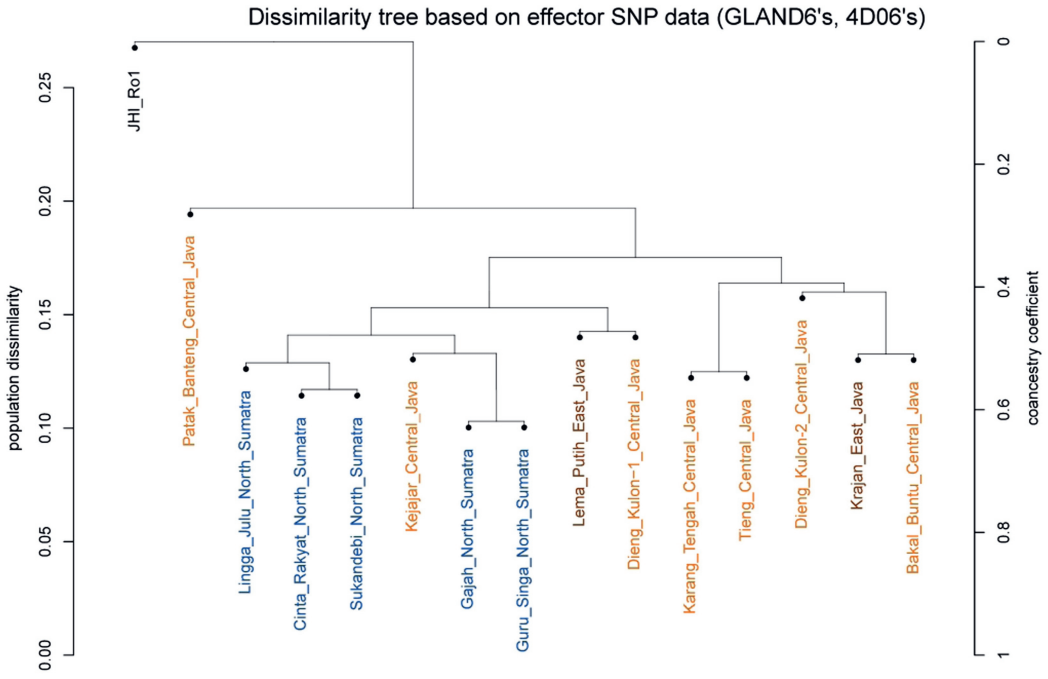


Figure 3.4: Dissimilarity trees based on 235 unique variants among members of the effector family 4D06 as represented in the sampled Indonesian *Globodera rostochiensis* populations. The 4D06 variants were called by mapping resequencing data from Indonesian field populations on a reference genome of *Globodera rostochiensis* (Gr-Ro1-Line19; van Steenbrugge et al. [2021]). The genome sequence of a selected field population from Scotland (United Kingdom) referred to as JHI-Ro1 [Eves-van den Akker et al., 2016] was used as the outgroup.

3.8 Effector diversification among Indonesian potato cyst nematode populations: The venom allergen-like protein family

Venom allergen-like proteins (VALs) are effectors for which expression in *G. rostochiensis* coincides with migration in the plant. At least one member (Gr-VAP1) was shown to be expressed in the subventral glands [Lozano-Torres et al., 2012]. Members of this effector family are secreted in the apoplast where they suppress the plant's basal defense response [Lozano-Torres et al., 2014]. The *G. rostochiensis* line 19 harbors eight VAL variants. In total, 150 variants were detected among the Indonesian *G. rostochiensis* populations. Also

here, a high level of dissimilarity was shown between the outgroup (JHI_Ro1) and the Indonesian populations (Fig. 3.5). Three major subclades can be distinguished: two are island specific, and one comprises the VAL variants that are found in both the central Javanese and Sumatran *G. rostochiensis* populations.

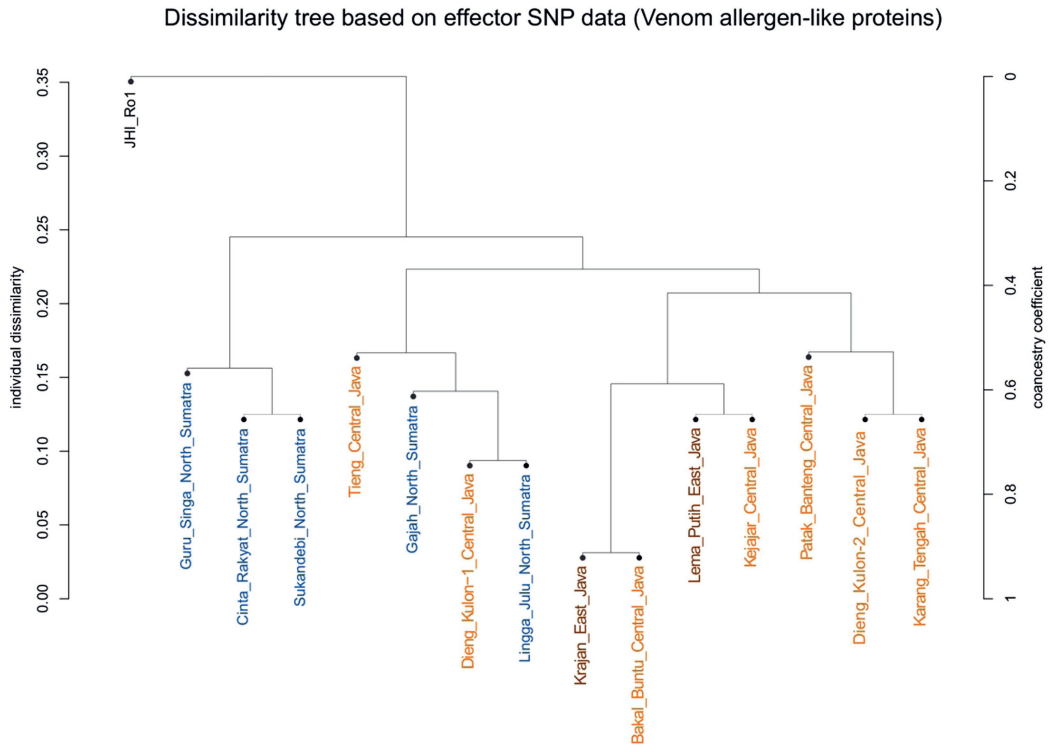


Figure 3.5: Dissimilarity trees based on 150 unique variants among members of the effector family venom allergen-like proteins (VALs) as represented in the sampled Indonesian *Globodera rostochiensis* populations. VAL variants variants were called by mapping resequencing data from Indonesian field populations on a reference genome of *Globodera rostochiensis* (Gr-Ro1-Line19; van Steenbrugge et al. [2021]). The genome sequence of a selected field population from Scotland (United Kingdom) referred to as JHI-Ro1 [Eves-van den Akker et al., 2016] was used as the outgroup.

3.9 Discussion

The potato cyst nematode *G. rostochiensis* co-evolved with its *Solanaceous* hosts in the Andean mountains in South America in or near Bolivia [Boucher et al., 2013], and although this pathogen was first described in Germany in 1923 [Wollenweber, 1923], it seems reasonable to suggest it was introduced in Europe before that time. First reports on the cultivation of potatoes in Indonesia date back to 1711, and it is plausible that this crop was introduced by the Dutch with plant material originating from Europe [Boomgaard, 2003]. Potato as a crop was first introduced on Java, but already in the 1780s potato production was present in Sumatra. Upland farmers grew them as a cash crop for Europeans. It is noted that in those days, this crop was seldom found outside of Java and Sumatra [Boomgaard, 2003]. In our analysis, ∼1.4 million unique variants were called when comparing the Indonesian *G. rostochiensis* population and the European outgroup with the reference genome Gr-Line19. Genome-wide analysis revealed that the distinctiveness between the Javanese and Sumatran *G. rostochiensis* is almost identical to the number of unique variants that distinguish the European Ro1 population from the nearest Javanese population (Kejajar, centra Java) (Fig. 1). For this reason and the aforementioned historical data, we suggest that *G. rostochiensis* was introduced in Indonesia at least two centuries ago, first of all on Java and not that long thereafter, on Sumatra. The suggestion that Sumatran *G. rostochiensis* populations originate from Java, and not from an independent introduction from outside Indonesia, confirms findings based on microsatellite genotyping [Handayani et al., 2020]. Since the time of introduction on Sumatra, *G. rostochiensis* populations have diversified more or less independently on these two main islands. Keeping in mind that the presence of *G. rostochiensis* was reported in 2004 [Indarti et al., 2004], this analysis points at a long unnoticed presence of *G. rostochiensis* in the Indonesian archipelago.

In an alternative scenario, one could suggest that potato cyst nematodes were independently introduced on Java and Sumatra with *G. rostochiensis* populations from Europe. In this scenario, the genotypic constitution of these inocula should have been comparable, otherwise the dissimilarity tree of effector family 1106 would not have given the observed topology in which Sumatran populations are interspersed with populations from east and central Java.

A remarkable finding was the differences in diversification pattern of four effector families, SPRYSEC (60 paralogs in Gr-Lin19), 1106 (10 paralogs), 4D06 (23 paralogs), and VALs (8 paralogs), among the Indonesian PCN populations. Except for the VALs, at least one member of each of these effector families is known to be expressed in the dorsal esophageal gland of potato cyst nematodes at the initial stages of plant parasitism. Moreover, at least one member of each of these effector families was demonstrated to interfere with the host innate immune system, and no other functions of these proteins families are known [Finkers-

Tomczak, 2011, Pogorelko et al., 2020, Postma et al., 2012]. Despite these similar expression patterns and functional parallels, the dissimilarity patterns presented in Figures 2, 3, 4, and 5 differed fundamentally. Whereas the SPRYSEC effector family members diversified under the Indonesian potato production conditions as if it was a neutral gene, the 1106 family showed a completely distinct pattern. We suggest that these distinct patterns may (at least in part) be explainable by the distinct number of paralogs per family. We hypothesize that only a small subset of the 60 SPRYSEC paralogs was directly involved in the interaction with Indonesian potato cultivars, and no selection took place among the non-involved majority. On the other hand, the families with the least number of paralogs, 8 (the VAL effector family) and 10 (the 1106 effector family), behaved as non-neutral genes as demonstrated by the very distinct topologies of Figure 3 (1106 variants only) and Figure 5 (VAL variants only) versus Figure 1 (all unique variants).

In its center of diversity, sympatric speciation resulted in the evolution of two potato cyst nematode species, *G. rostochiensis* and *G. pallida*. Despite a few reports suggesting otherwise [Suastika et al., 2012], 2020 surveys in the Indonesian archipelago suggest that *G. pallida* is absent in this region [Handayani et al., 2020]. It could be hypothesized that *G. pallida* was never introduced in Indonesia. Alternatively, it could be envisaged that the Indonesian environmental conditions are more suitable for *G. rostochiensis* than for *G. pallida*. The latter seems more likely. *G. pallida* performs best at soil temperatures <17.5°C, whereas for *G. rostochiensis* both the number of females per plant and fecundity were highest at soil temperatures between 17.5 and 22.5°C [Jones et al., 2017]. As this survey sampled potato fields at altitudes ranging from 1,300 to 2,100 m (Table 1), the absence of *G. pallida* can be attributed to the unsuitability of the local climate to this sibling species of *G. rostochiensis*.

The exclusive presence of *G. rostochiensis* in the production regions under investigation offers a good perspective for durable management of this major potato pathogen. Multiple resistances against *G. rostochiensis* such as H1 (a single dominant resistance gene; Toxopeus and Huijsman [1952]), the Grp1 locus [Finkers-Tomczak et al., 2009], and Gro1 [Paal et al., 2004] have been introgressed in modern potato cultivars. However, keeping in mind the level of diversification of *G. rostochiensis* between the probable northwest European founder populations and populations as of this writing in the Indonesian archipelago, and the diversification between the Javanese and Sumatran populations, careful selection of potato cultivars with one or multiple resistances against *G. rostochiensis* will be required to durably manage this pathogen in the growing Indonesian potato production sector.

CHAPTER 4

Comparative genomics among cyst nematodes reveals distinct evolutionary histories among effector families and an irregular distribution of effector-associated promoter motifs

Joris J.M. van Steenbrugge[1], Sven van den Elsen[1], Martijn Holterman[1,2], Jose L. Lozano-Torres[1], Vera Putker[1], Peter Thorpe[3], Aska Goverse[1], Mark G. Sterken[1], Geert Smant[1] and Johannes Helder[1]

[1] Laboratory of Nematology, Wageningen University and Research, The Netherlands
[2] Solynta, Wageningen, The Netherlands
[3] School of Medicine, Medical & Biological Sciences, University of St. Andrews, United Kingdom

This chapter is published in Molecular Ecology [van Steenbrugge et al., 2022].

4.1 Abstract

Potato cyst nematodes (PCNs), an umbrella term used for two species, *Globodera pallida* and *G. rostochiensis*, belong worldwide to the most harmful pathogens of potato. Pathotype-specific host plant resistances are essential for PCN control. However, the poor delineation of *G. pallida* pathotypes has hampered the efficient use of available host plant resistances. Long-read sequencing technology allowed us to generate a new reference genome of *G. pallida* population D383 and, as compared to the current reference, the new genome assembly is 42 times less fragmented. For comparison of diversification patterns of six effector families between *G. pallida* and *G. rostochiensis*, an additional reference genome was generated for an outgroup, the beet cyst nematode *Heterodera schachtii* (IRS population). Large evolutionary contrasts in effector family topologies were observed. While VAPs (venom allergen-like proteins) diversified before the split between the three cyst nematode species, the families GLAND5 and GLAND13 only expanded in PCNs after their separation from the genus *Heterodera*. Although DNA motifs in the promoter regions thought to be involved in the orchestration of effector expression ("DOG boxes") were present in all three cyst nematode species, their presence is not a necessity for dorsal gland-produced effectors. Notably, DOG box dosage was only loosely correlated with the expression level of individual effector variants. Comparison of the *G. pallida* genome with those of two other cyst nematodes underlined the fundamental differences in evolutionary history between effector families. Resequencing of PCN populations with different virulence characteristics will allow for the linking of these characteristics to the composition of the effector repertoire as well as for the mapping of PCN diversification patterns resulting from extreme anthropogenic range expansion.

4.2 Introduction

Worldwide, affordable food and feed production depends on large-scale monocropping. For practical and economic reasons crop homogeneity in terms of yield quality and quantity is essential. At the same time, such systems are intrinsically vulnerable to damage by pests and pathogens. The highest susceptibility to biotic stressors is found in genetically homogeneous crops. Potato is the third most important staple food [Birch et al., 2012], and in most production systems clonally propagated seed potatoes are used as starting material. Such production systems need rigorous disease management. Potato cyst nematodes (PCNs), the common name for two species, *Globodera pallida* and *G. rostochiensis*, are among the primary yield-limiting potato pathogens worldwide. PCNs co-evolved with potato in the Andes in South America (see, e.g., [Plantard et al., 2008]), and proliferation of potato as a main crop outside of its native range was unintentionally paralleled by an enormous range expansion of PCNs. For decades, PCN control has mainly been dependent on the application of nematicides. Due to the nonspecific nature of these nematicides, they have a highly negative impact on the environment, and their use is therefore either banned or severely restricted in most parts of the world. Currently, crop rotation and the use of resistant potato varieties are the main means for PCN control. For economic reasons, the use of plant resistances is preferred over crop rotation. However, potato resistance genes such as *H1* [Toxopeus and Huijsman, 1952], *Gro1-4* [Paal et al., 2004], *Gpa2* [Bakker et al., 2003] and *H2* [Strachan et al., 2019] are only effective against specific pathotypes of one of these PCN species. Nevertheless, there is no robust (molecular) pathotyping scheme that would allow for matching of the genetic constitution of field populations with effective host plant resistance genes.

Effectors are proteins secreted by plant pathogens that allow manipulation of the physiology of the host plant and interfere with the host's innate immune response in favour of the invading organism (e.g., [Stergiopoulos and de Wit, 2009]). PCN effectors have some peculiar characteristics. With at least one known exception, HYPer variable proteins (HYPs) [Eves-van den Akker et al., 2014], most effectors are produced in large single-celled glands referred to as the subventral and dorsal esophageal glands. These glands empty into the pharynx lumen, and the lumen is connected to a hollow protrusible stylet with which nematodes pierce plant cell walls. Via the orifice of the stylet, effector proteins are transferred to the apoplast or the cytoplasm of infected host plant cells. Notably, subventral gland effectors are functional during plant penetration. Subsequently, dorsal gland secretions are responsible for feeding site induction and suppression of the host's innate immune system [Smant et al., 2018]. It has been hypothesized that common transcription factors and/or common promoter motifs might facilitate coordinated expression of effectors during the infection process. Such mechanisms have been identified to regulate effector expression in plant pathogenic fungi and oomycetes [Jones et al., 2019, Roy et al., 2013]. Also, among plant-parasitic nematodes,

promotor motifs have been identified upstream of effectors that could contribute to the orchestration of the infection process. In the case of the PCN *G. rostochiensis*, a DOrsal Gland motif ("DOG box") was identified by Eves-van den Akker et al. [2016]. For the pinewood nematode *Bursaphelenchus xylophilus*, a regulatory promotor motif referred to as STATAWAARS was demonstrated to affect effector expression [Espada et al., 2018]. Expression of several effectors of Clade I tropical root-knot nematodes [De Ley et al., 2002b] was suggested to be steered by a putative cis-regulatory motif "Mel-DOG" (*Meloidogyne* DOrsal Gland, [Da Rocha et al., 2021]).

Probably as a reflection of the co-evolution between nematodes and their host(s), effectors are typically encoded by multi-gene families showing family-specific levels of diversification [Masonbrink et al., 2019, Siddique et al., 2022]. Cyst nematodes harbour numerous effector families (e.g., [Pogorelko et al., 2020]), and genome (re-)sequencing is a rigorous approach to generate comprehensive overviews of PCN effector family compositions. The first genomes of *G. pallida* and *G. rostochiensis* were published by Cotton et al. [2014] and Eves-van den Akker et al. [2016]. Although this constituted a major step forward, both genomes are highly fragmented, hampering effector family inventories. Recently, long-read technology allowed for the generation of a less fragmented and more complete reference genome for *G. rostochiensis* with—among other things—a 24-fold reduction in the number of scaffolds as compared to the initial reference genome [van Steenbrugge et al., 2021]. Here, we present a novel reference genome for the other PCN, *G. pallida*, characterized by a 42-fold reduction in the number of scaffolds, together with a reference genome of the beet cyst nematode *Heterodera schachtii*. The *H. schachtii* genome was used to establish the polarity of effectorome contrasts between the two PCN species. Detailed knowledge of the nematode's effector repertoire, a complete overview of variants within effector families and insights in the evolutionary history of individual effector families are essential for a molecular pathotyping scheme. In addition to comparing effector diversification patterns, we investigated DOG box distribution and DOG box dosage (up to 16 DOG boxes were observed per putative promoter region) both within and among effector families. Scrutinizing putative effector promoter regions in three reference genomes allowed us to pinpoint the distribution of this putative regulatory motif among cyst nematode species, as well as among and within effector families. Subsequently, the impact of these new, long-read technology-based reference genomes on ecological PCN diversification in general and on the development of effectorome-based pathotyping systems for PCNs in particular is discussed.

4.3 Results

4.3.1 Use of long read sequence technologies to generate novel reference genomes

The mapping of diversification patterns of effector families requires a high-quality reference genome with preferably a low number of scaffolds and a minimal gap length. Cotton et al. [2014] were the first to present a reference genome of the PCN species *Globodera pallida*. For our specific purpose (i.e., the generation of complete inventories of effector families), this reference genome was too fragmented, and the total gap length was too large. PacBio long-read technology allowed us to generate a new reference genome from the *G. pallida* population D383 with a 42-fold reduction of scaffolds and a 21-fold reduction of the total gap length. As one consequence, the number of predicted genes increased from 16,403 to 18,813, and the level of completeness as estimated by BUSCO increased by more than 10% (Table 4.1).

Table 4.1: Comparative genome statistics of four cyst nematode genome assemblies. In bold, data from the current paper; data on *Globodera pallida* Lindley and *G. rostochiensis* Line 19 genomes were published by respectively Cotton et al. [2014] and van Steenbrugge et al. [2021]

Nematode species population	Size (Mb)	Number of scaffolds	N50 (Mb)	N90 (Mb)	Number of gaps	Gap length	Number of genes	Number of transcripts
Globodera pallida Lindley	124.7	6873	0.122	0.011	6873	19,990,795	16,403	16,403
G. pallida Pa2-D383	113	163	**2.9**	**0.515**	**22,788**	**945,137**	**18,813**	**27,787**
G. rostochiensis Line 19	92	173	1.70	0.582	2733	130,000	17,928	21,037
Heterodera schachtii IRS	190	705	**0.5**	**0.132**	705	**4,285,731**	**29,851**	**31,564**
BUSCO								
G. pallida, Lindley	C:49.0%:[S:44.3%,D:4.7%],F:21.6%,M:29.4%,n:255							
G. pallida, D383	C:59.2%[S:58.4%,D:0.8%], F:19.6%,M:21.1%,n:255							
G. rostochiensis Line 19	C:63.5%[S:62.7%,D:0.8%],F:17.6%,M:18.9%,n:255							
H. schachtii, IRS	C:86.3% [S:80.8%, D:5.5%], F:7.1%, M:6.6%, n:255							

In addition, we assembled the genome sequence of the IRS population of the beet cyst nematode *Heterodera schachtii*. This allowed us to compare effector family diversification among the two PCN species, *G. pallida* and *G. rostochiensis*, and establish the polarity of these contrasts by using *H. schachtii* as an outgroup (both *Globodera* and *Heterodera* belong to the family *Heteroderidae*). The current genome size, 190Mb, is slightly above the genome size estimated by flow cytometry, 160–170Mb [Siddique et al., 2022]. Notably, both the predicted number of genes and transcripts were about 50% higher in *H. schachtii* than in the two *Globodera* species (Table 4.1). These figures largely correspond to a *H. schachtii* genome assembly that was recently published from a German population referred to as "Bonn" with an assembly size of 179 Mb, and 26,700 predicted genes [Siddique et al., 2022].

Two synteny plots were generated based on the alignment of regions >1 and >3 kb to compare the genomic organization of the three cyst nematode species. Not unexpectedly, the two PCN species share numerous >1-kb regions (Figure 4.2a). In the *H. schachtii*

genome, several homologous >1-kb regions cluster together in genomic segments that span over 2 Mb (Figure 4.1a, segments 1–8). The homologous >1-kb regions in these segments have equivalents in both *G. pallida* and *G. rostochiensis*. Alignment of >3-kb fragments severely reduced the number of homologous regions among the three cyst nematode species (Figure 4.2b). Nevertheless, the number of shared >3-kb regions between *G. pallida* and *G. rostochiensis* (N = 76) is considerably higher than the number of shared regions between *H. schachtii* and *G. rostochiensis* (N = 23) (Figure 4.1b).

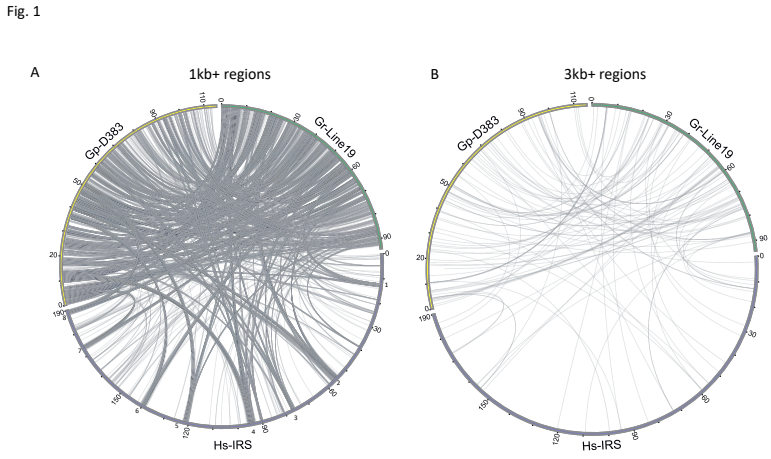


Figure 4.1: Synteny between *Globodera pallida* (population D383), *Globodera rostochiensis* (Gr-Line19) and *Heterodera schachtii* (population IRS) based on a progressive genome alignment in mauve. (a) Syntenic regions larger than 1 kb, (b) syntenic regions larger than 3 kb. In (a), *H. schachtii* genome regions are indicated where multiple syntenic regions cluster together into segments spanning over 2 Mb (segments 1–8). Note that these segments have equivalents in both *G. pallida* and *G. rostochiensis*

4.3.2 Effector family selection

In our comparison between the three cyst nematode species, we concentrated on effectors. Cyst nematodes were shown to harbour numerous effector families. Here we concentrated on six effector families. For four of these families, one or more representatives are known to be involved in the suppression of plant innate immune system: SPRYSEC [Diaz-Granados et al., 2016, Mei et al., 2018], GLAND4 (also referred to as Gr-1106) [Barnes et al., 2018], GLAND5 (also referred to as G11A06) [Yang et al., 2019b] and VAP [Wilbers et al., 2018]. CLE [Wang et al., 2021] is an intriguing effector family involved in feeding site induction, and the GLAND13 [Danchin et al., 2016] family is essential in the hydrolysis of plant sugars

once they are taken up by the nematode.

4.3.3 SPRYSECs

SPRYSEC is an acronym for a family of secreted effectors with an SP1a/RYanodine receptor domain. This family was recently shown to be highly diverged in the PCN *G. rostochiensis* [van Steenbrugge et al., 2021]. Figure 4.2 shows a phylogenetic tree with (supposedly) all SPRYSECs present in the three cyst nematode species under investigation. The number of paralogues in *G. pallida*, *G. rostochiensis* and *H. schachtii* is respectively 24, 60 and 20. Despite the poor backbone resolution of the SPRYSEC tree, two moderately supported SPRYSEC clades (A and B) could be distinguished. Clade A comprises SPRYSEC variants exclusively from the two PCN species, and *G. pallida* SPRYSEC paralogs are interspersed with *G. rostochiensis* SPRYSEC variants. SPRYSEC Clade A is characterized by zero to six DOG box elements. Clade B harbours fewer SPRYSEC paralogues than Clade A (27 vs. 35 in Clade A). Notably, Gr19_g7942 includes a transmembrane domain (and is therefore unlikely to be secreted), whereas three DOG box elements are present in the promoter region directly upstream of this paralogue. Clade B is characterized by a mix of SPRYSEC variants solely originating from *G. pallida* and *G. rostochiensis*. Compared to Clade A, Clade B is typified by an overall higher DOG box dosage (on average, 1.7 and 5.2 DOG boxes per paralogue). Up to 16 DOG box elements were identified in the promoter regions of paralogues in Clade B. The more basal part of the SPRYSEC tree (Figure 4.2, part c) harbours, next to paralogues from the two PCN species, all 20 SPRYSEC variants from *H. schachtii*. Five promoter regions of the 35 SPRYSEC paralogues in this part of the SPRYSEC tree harbour a single DOG box. Both *H. schachtii* and *G. rostochiensis* harbour SPRYSEC paralogues with at least one transmembrane domain (gene ID in italics with lighter colour).

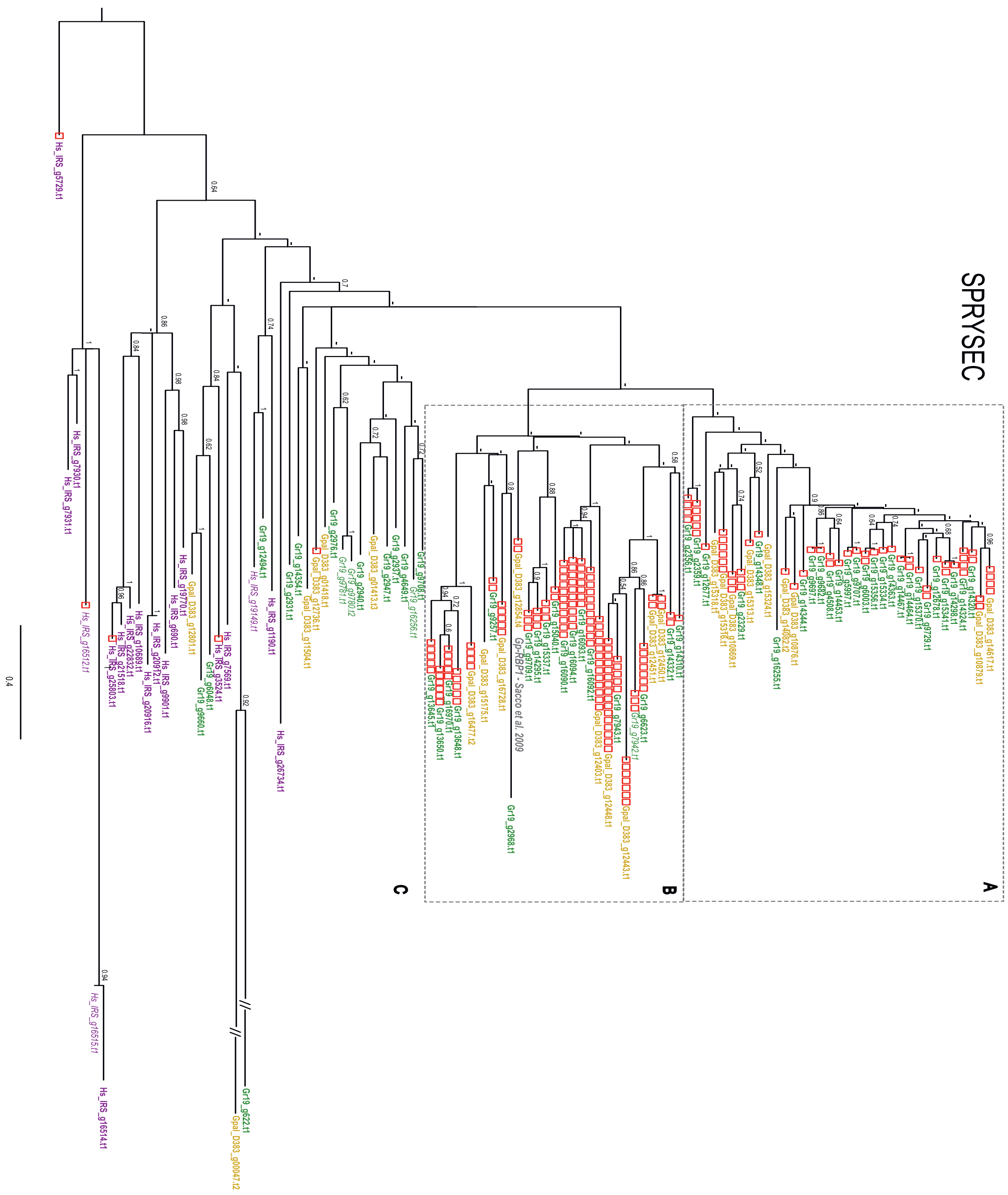


Figure 4.2:

Figure 4.2: Phylogeny of SPRYSEC effector genes (see, e.g., Diaz-Granados et al. [2016] of *Globodera pallida* (population D383) (ochre), *Globodera rostochiensis* (Gr-Line19) (green) and *Heterodera schachtii* (population IRS) (purple). A multiple sequence alignment was made using muscle on the coding sequence. A phylogenetic tree was made using raxml using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a dash. GenBank IDs in italics in lighter shades of ochre, green or purple are used to indicate effector variants with at least one predicted transmembrane domain. Boxed clusters (A) and (B) highlight two moderately well-supported subclades with on average moderate (A) and high DOG box dosages. (C) is the basal part of the SPRYSEC tree

Four *G. pallida* SPRYSEC proteins variants referred to as RBP-1-D383-1, −2, −3 and −4 that were recognized by the potato resistance gene Gpa2 [Sacco et al., 2009] correspond to Gpal_D383_g12854 (Figure 4.2). The four marginally distinct *RBP-1* s presumably include allelic variants of Gpal_D383_g12854. Notably, the closest equivalent of GpSPRY-414-2, a SPRYSEC isolated from the *G. pallida* Pa2/3 population "Lindley" [Mei et al., 2018], was not preceded by a signal peptide for secretion in our genome assembly, and therefore could not be included in Figure 4.2.

4.3.4 GLAND4

The number of GLAND4 (also referred to as Gr-1106) paralogues in Gr-Line19, Gp-D383 and Hs-IRS is respectively 10, nine and 15. The phylogenetic analysis yields a tree with a well-supported backbone (Figure 4.3) showing a clear separation between the outgroup *H. schachtii* and both *Globodera* species. Except for Gpal_D383_g13703, which is positioned at the base of the *G. rostochiensis* cluster, GLAND4 variants end up in separate species-specific clusters. On the other hand, Hs-IRS paralogues show more intra-specific diversification. All but two *G. pallida* paralogues (Gpal_D383_g17346.t1 and Gpal_D383_g13669) contain a signal peptide for secretion. For all but one of the GLAND4 genes in Gr-Line19, the promoter region included a DOG box motif, while promoter regions of only four GLAND4 genes in Gp-D383 and one in Hs-IRS contained such a motif. Previously, over expression of a subset of *G. rostochiensis* GLAND4 (1106) variants in potato resulted in enhanced host plant susceptibility as observed by a significant increase in the number of cysts formed [Finkers-Tomczak, 2011]. The equivalents of two of these variants, E3 and E9, in *G. rostochiensis* line 19 are indicated in Figure 4.3.

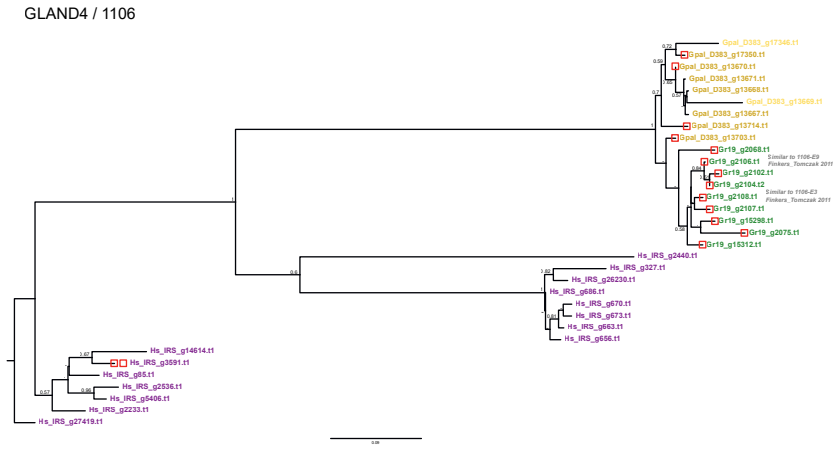


Figure 4.3: Phylogeny of GLAND 4 (equivalent to 1106, see, e.g., Noon et al. [2015]) effector genes of *Globodera pallida* (population D383) (ochre), *Globodera rostochiensis* (Gr-Line19) (green) and *Heterodera schachtii* (population IRS) (purple). A multiple sequence alignment was made using muscle on the coding sequence. A phylogenetic tree was made using Raxml using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a dash. GenBank IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a signal peptide for secretion

4.3.5 GLAND5

With 13 homologues, the GLAND5 effector family (also referred to as G11A06) was significantly less diversified in Gr-Line19 than in Gp-D383 and Hs-IRS with respectively 25 and 27 paralogues. In all three species, the majority of the GLAND5 paralogues harbour a signal peptide for secretion. Note that a relatively high percentage of the GLAND5 paralogues in Gr-Line19 was not preceded by a signal peptide for secretion (23%). In contrast, in *H. schachtii* and *G. pallida*, respectively, 88.9% and 92% of the paralogues comprised a signal peptide. Phylogenetic analysis (Figure 4.4) shows that GLAND5 is a diversified gene family. Several well-supported branching events define a set of subclades that either exclusively comprises *H. schachtii* or contain GLAND5 variants from both PCN species in the more distal branches. Even though the GLAND5 paralogues Gr-Line19 and Gp-D383 occur together in individual subclades, no obvious sets of potential orthologues between the two

species could be identified. In *H. schachtii*, 82% of the paralogues contain at least one DOG box motif in the promoter region. Out of the three GLAND5 paralogues without a signal peptide for secretion, two (Hs-IRS_g6495.t1 and Hs-IRS_g22438.t1) had at least one DOG box motif in their promoter region. DOG boxes were less prominently present among the *G. rostochiensis* and *G. pallida* GLAND5 variants (39% and 20% of the paralogues).

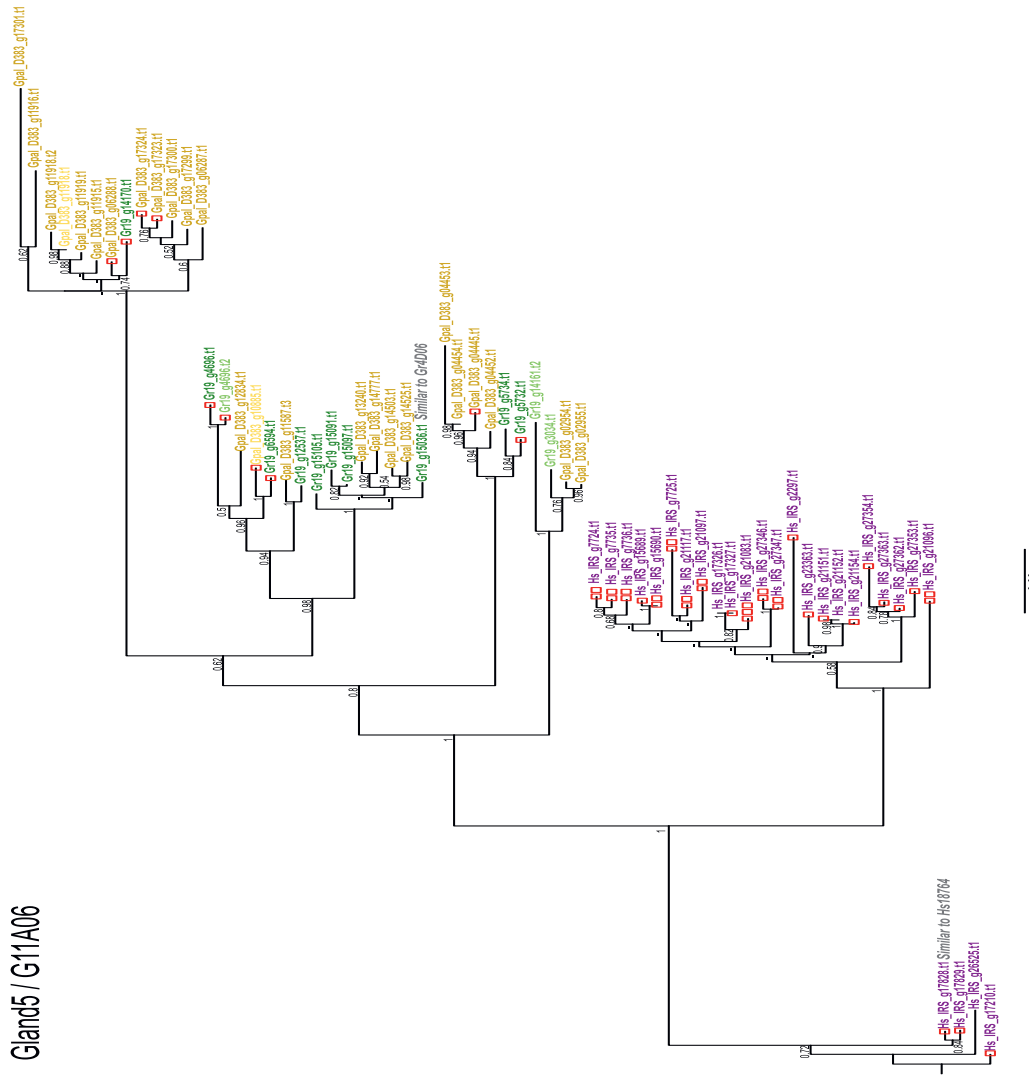


Figure 4.4: Phylogeny of GLAND 5 (equivalent to G11A06, see, e.g., Noon et al. [2015]) effector genes of *Globodera pallida* (population D383) (ochre), *Globodera rostochiensis* (Gr-Line19) (green) and *Heterodera schachtii* (population IRS) (purple). A multiple sequence alignment was made using muscle on the coding sequence. A phylogenetic tree was made using raxml using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a dash. GenBank IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a signal peptide for secretion.

One of the *G. rostochiensis* paralogues (Gr19_g15036) showed similarity with Gr4D06, a GLAND6 effector protein described by Ali et al. [2015a] (Figure 4.4). This effector has an unknown function, and expression in planta did not result in a phenotype. GLAND5 and GLAND6 are related effector families [Noon et al., 2015]. *H. schachtii* GLAND5 variant Hs_IRS_g17828 showed similarity to Hs18764 described by Yang et al. [2019a], and over-expression of Hs18764 in *Arabidopsis thaliana* resulted in an increased susceptibility to *H. schachtii*.

4.3.6 VAP

The levels of diversification in the VAP effector family were highly comparable between the three cyst nematode species. In Gp-D383, Hs-IRS and Gr-Line19, respectively, eight, eight and 10 VAP paralogues were identified. Phylogenetic analysis resulted in a tree with a well-supported backbone (Figure 4.5). The tree contains three clusters (Figure 4.5, boxes A, B and C) with a high level of diversification between the clusters. At the base of the tree, a small cluster of four *H. schachtii* paralogues (Figure 4.5, box C) is present that all lack a signal peptide for secretion. Box B harbours Gp-D383 and Gr-Line19 VAP paralogues, of which all but two lack a signal peptide for secretion. Two sub-clusters are present in Box B: one with exclusively Gr-Line19 variants, a second one with just Gp-D383 variants, and the third with an orthologous pair between Gr-Line19 and Gp-D383. In the largest cluster at the top of the tree (Figure 4.5, box A), VAP paralogues of all three species are present, including the only two secreted VAP variants of Hs-IRS with a DOG box motif in the promoter region.

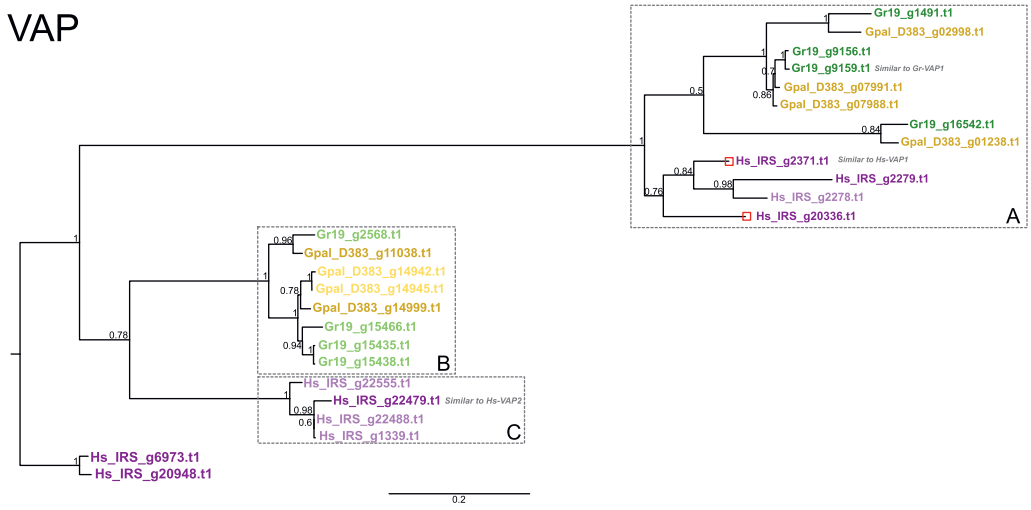


Figure 4.5: Phylogeny of VAP (venom allergen-like protein, e.g., [Wilbers et al., 2018]) effector genes of *Globodera pallida* (population D383) (ochre), *Globodera rostochiensis* (Gr-Line19) (green) and *Heterodera schachtii* (population IRS) (purple). A multiple sequence alignment was made using muscle on the coding sequence. A phylogenetic tree was made using raxml using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a dash. GenBank IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a signal peptide for secretion

Heterologous expression of three VAPs from *G. rostochiensis* (Gr-VAP1) and *H. schachtii* (Hs-VAP1, 2) resulted in the loss of basal immunity to the corresponding cyst nematode species as well as to some other, unrelated pathogens. It was concluded that these VAP variants interfere with the basal immune response of their host plants [Lozano-Torres et al., 2014]. VAP variants highly similar to Gr-VAP1 and Hs-VAP1, 2 are indicated in Figure 4.5.

4.3.7 CLE

With 16 variants, the CLE-like effector family is considerably more diversified in *H. schachtii* than in *G. pallida* and *G. rostochiensis* (respectively, 10 and 11 paralogues). Analysis of the CLE families on the three cyst nematode species resulted in a phylogenetic tree with a reasonably well-resolved backbone (Figure 4.6). Note that nearly all variants are united in species-specific clusters, and in this sense, the CLE diversification patterns resemble the patterns observed for the GLAND4 (Gr-1106) family (Figure 4.3). Whereas Gp-D383 and Gr-

Line19 are characterized by similar-sized, moderately diverged clusters of CLE paralogues, the CLE family is far more diverged in *H. schachtii*.

In *G. rostochiensis*, two functional classes of CLE peptides have been described, CLE-1 and CLE-4 [Lu et al., 2009]. The CLE-1 class (Figure 4.6) comprises two Gr-Line19 paralogues that show only distant homology to *G. pallida* CLEs. Similarly, four *G. rostochiensis* CLEs belonging to functional CLE class 4 (Figure 4.6) do not have clear equivalents in *G. pallida* and *H. schachtii*. Unlike all other effector families investigated so far, all CLE variants from the three cyst nematode species are preceded by a signal peptide for secretion. At the same time, none of them has a DOG box motif in the promoter region.

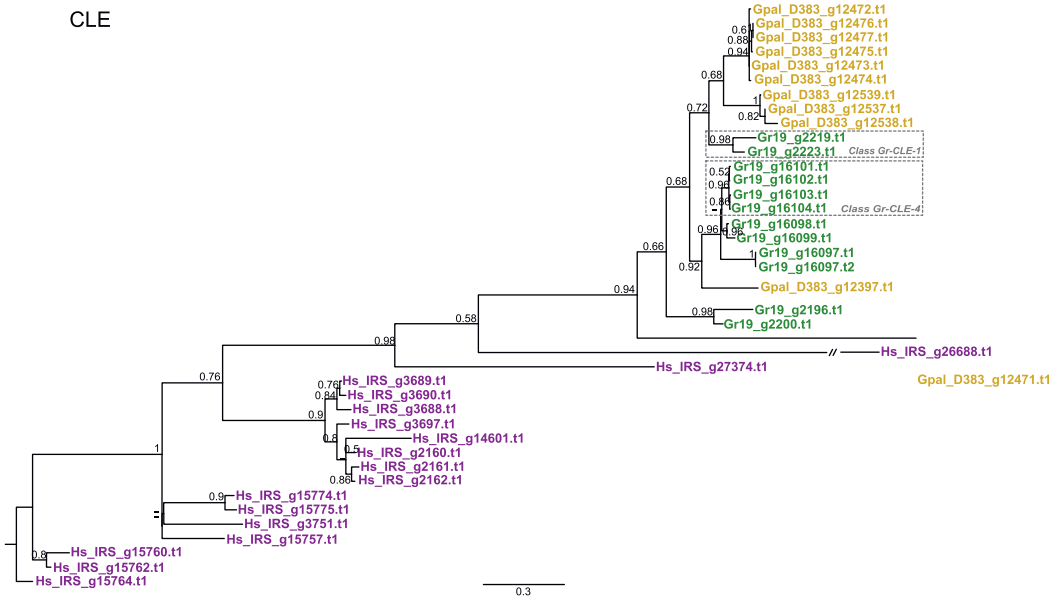


Figure 4.6: Phylogeny of CLE (CLAVATA3/ESR-related peptides, see, e.g., Lu et al. [2009]) effector genes of *Globodera pallida* (population D383) (ochre), *Globodera rostochiensis* (Gr-Line19) (green) and *Heterodera schachtii* (population IRS) (purple). A multiple sequence alignment was made using muscle on the coding sequence. A phylogenetic tree was made using raxml using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a dash. GenBank IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a signal peptide for secretion

4.3.8 GLAND13

GLAND13 effectors investigated so far have been identified in *G. pallida* and coded for invertases belonging to glycosyl hydrolase family 32 (GH32). While these enzymes are not secreted into the plant, they are essential as they catalyse the hydrolysis of the primary type of sugar the nematode takes up from its host, sucrose [Danchin et al., 2016]. This gene family shows a large difference in the number of paralogues present in the three species; while Gr-Line19 and Gp-D383 harbour 10 and seven paralogues, Hs-IRS holds only three copies. In the phylogenetic tree (Figure 4.7), the paralogues in *H. schachtii* are positioned at the tree's base. In Box A, paralogues of Gr-Line19 and Gp-D383 are interspersed, while in Box B, all paralogues except one are from Gr-Line19. In Gr-Line19, 70% of the GLAND13 paralogues comprise a signal peptide for secretion; slightly lower percentages (67% and 57%) were observed in Hs-IRS and Gp-D383. Variants showing high similarity to each of the five GLAND13 paralogues from *G. pallida* (population Lindley; indicated as GPLIN_number) are indicated in Figure 4.7. For more than half of the GLAND13 effector variants in Gr-Line19, a DOG box motif in the promoter region was shown. One gene, Gr19_g13610, contained the motif twice. In *H. schachtii*, this motif was present in two of three genes, while in *G. pallida*, DOG boxes were found in one variant with a signal peptide for secretion (Gpal_D383_g17582), and in one paralogue without such a signal (Gpal_D383_g09388). Note that these DOG box motifs were found in promoter regions of *G. pallida* effectors that are not expressed in the dorsal gland [Danchin et al., 2016]. GLAND13 variants similar to the five invertase variants characterized by [Danchin et al., 2016] including those with the highest expression levels (GPLIN_000950400 and GPLIN_001068900) are shown in Figure 4.7.

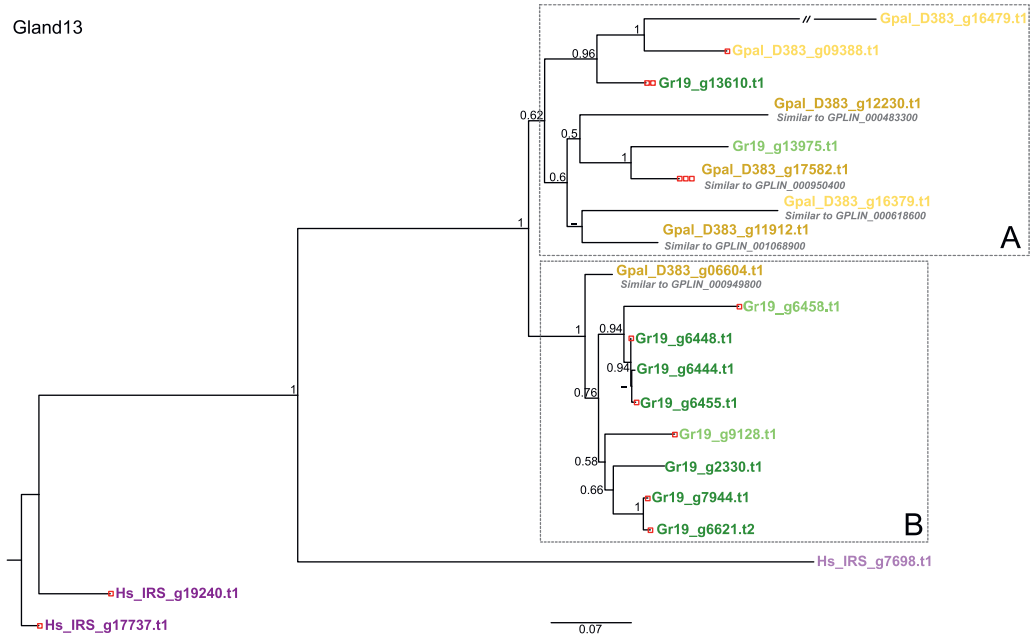


Figure 4.7: Phylogeny of GLAND13 (invertases, e.g., Danchin et al. [2016]) effector genes of *Globodera pallida* (population D383) (ochre), *Globodera rostochiensis* (Gr-Line19) (green) and *Heterodera schachtii* (population IRS) (purple). A multiple sequence alignment was made using muscle on the coding sequence. A phylogenetic tree was made using raxml using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values <50% are indicated by a dash. GenBank IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a signal peptide for secretion

4.3.9 Effects of DOG box dosage on SPRYSEC expression

Although DOG box motifs in the promoter regions of effector variants are present in many effector families, their presence is not a necessity for the functioning of an effector family. For example, none of the variants of the CLE family contained DOG box motifs, regardless of the species (Figure 4.8). Dorsal gland-expressed effectors can therefore be expressed and secreted without the presence of DOG box motifs. This is further illustrated in Figure 8a, which shows DOG box distribution of over the six effector families. For the three cyst nematode species under investigation, it demonstrates that DOG boxes can be entirely absent (CLE), present in some species only (VAP) or present in all species (SPRYSEC, GLAND4, GLAND5, GLAND13). The ample presence of DOG boxes in the diversified SPRYSEC family among the two PCN species prompted us to investigate whether there is a

correlation between the DOG box dosage and SPRYSEC expression levels. Although DOG boxes are present in the putative promoter regions of SPRYSECs from *H. schachtii*, they are relatively rare, and only single motifs were found. In Figure 8b, SPRYSEC effectors from all three cyst nematode species were taken into account. A modest correlation in *G. rostochiensis* and *G. pallida* ($R^2 = .67$ and .62 respectively) between the DOG box dosage and expression levels based on RNA abundance is present for this family. In *G. pallida*, in particular, high expression levels of SPRYSEC variants can be reached in the absence of DOG boxes in its promoter region (Figure 8b). For *H. schachtii*, a slightly negative correlation was found ($R^2 = -0.11$) between DOG box dosage and expression levels.

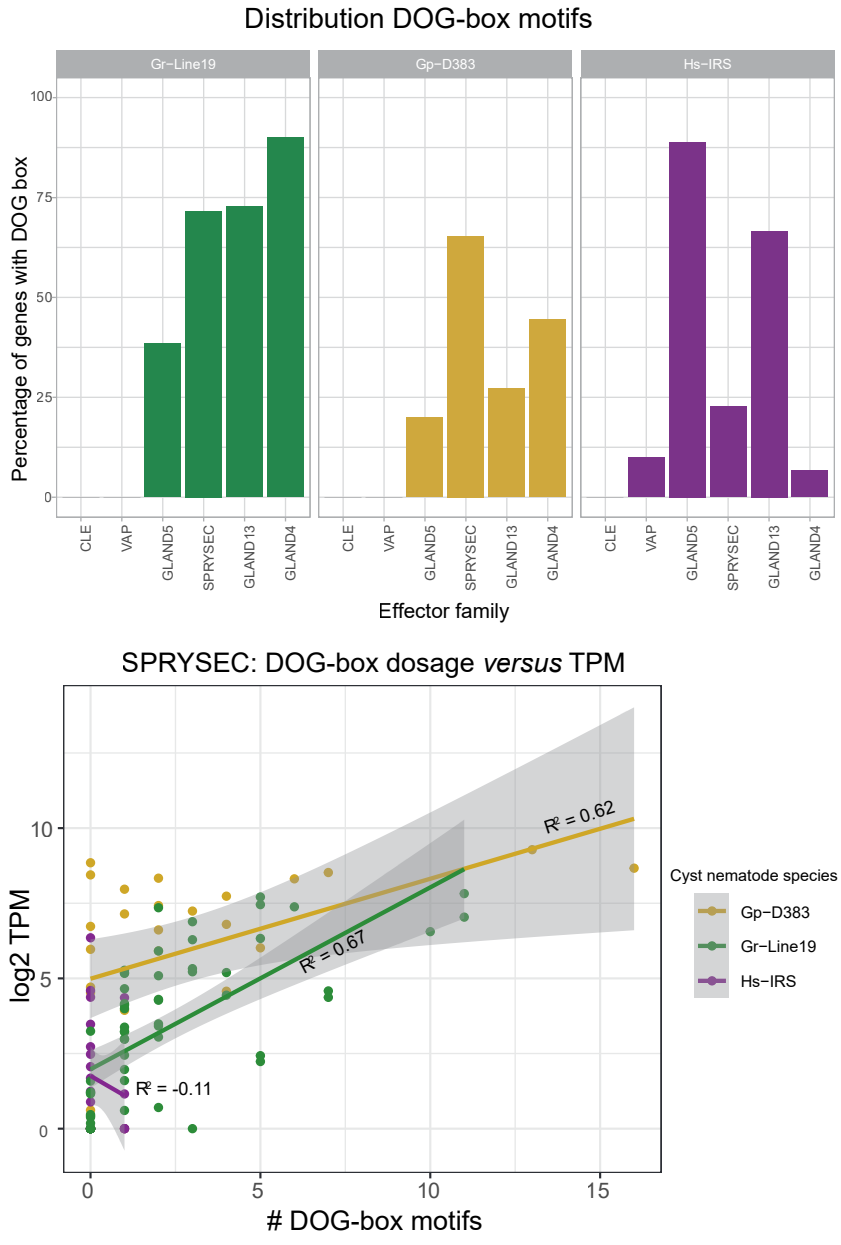


Figure 4.8: This distribution of a DOrsal Gland promoter element motif ("DOG box, Eves-van den Akker et al. [2016]) among a selection of cyst nematode species. (a) Percentages of variants per effector family with one or more a DOG boxes in their promoter region. (b) The relationship between DOG box dosage and expression level (expressed as log2 TPM [transcript count per million]). The relationship is expressed as a Spearman rank-order correlation (R2) per species, for the SPRYSEC gene family

4.4 Discussion

In our attempt to fundamentally understand the interaction between plant-parasitic nematodes and their hosts, the usefulness of high-quality reference genomes of these pathogens is vital. Given the enormous impact of PCNs in all major potato production regions of the world, it is not surprising that high priority was given to the sequencing of both the *Globodera pallida* [Cotton et al., 2014] and the *Globodera rostochiensis* [Eves-van den Akker et al., 2016] genome. This was done before long-read sequencing technologies became available. Although some research questions can be well addressed with these reference genomes, less fragmented genomes are needed for studying effector diversification. Therefore, a new reference genome was generated from *G. pallida* population D383. As compared to the *G. pallida* Lindley genome assembly, this resulted in a 42-fold reduction in the number of scaffolds and a 24-fold increase in N50. In the comparison of the effectoromes of the two PCN species, we included a newly generated genome of the *Heterodera schachtii* population IRS as an outgroup. Note that reference genomes from these obligatory sexually reproducing pathogens are actually population-based consensus genomes. Long read sequencing technologies require DNA from tens of thousands of genetically nonidentical nematodes. While an individual of these diploid species could theoretically carry a maximum of two haplotypes per locus, a population has the potential to carry many more. It is essential to mine these haplotypes and assemble them into a single haploid assembly to generate a proper reference. This is not a trivial process and requires specialized bioinformatics software [Roach et al., 2018]. As the sizes of the current genome assemblies are comparable to the genome sizes assessed by flow cytometry, and as the BUSCO duplication scores are relatively low, we assume that the current genome assemblies are a reasonable reflection of their actual constitution.

4.4.1 Effector diversification

In our analyses we concentrated on six selected effector families, and this selection included relatively widespread effector families such as CLE, GLAND13 and VAP, as well as families that appear to be cyst nematode lineage-specific such as SPRYSEC, GLAND4 and GLAND5. Although the protein architecture is distinct between lineages [Mitchum et al., 2012], the CLE family—a category of effectors involved in feeding site induction—were shown to be present also in root-knot and reniform nematode [Rutter et al., 2014, Wubben et al., 2015]. GLAND13 effectors, members of the glycosyl hydrolase family 32, were shown to be present in a range of root-knot and cyst nematodes species as well as in other plant-parasitic nematodes such as *Nacobbus aberrans* and *Rotylenchus reniformis* [Danchin et al., 2016]. The distribution of VAPs within the phylum Nematoda is even broader [Wilbers et al., 2018]. VAPs were discovered in the animal parasite *Ancylostoma caninum* [Hawdon et al., 1996]. They were later isolated from the root-knot nematode *Meloidogyne incognita* [Ding et al.,

2000], and subsequently in a wide range of obligatory plant-parasitic nematodes including various cyst nematode species. A number of VAP variants were shown to be implicated in the suppression of both PAMP- and effector-triggered immunity [Li et al., 2021] for the burrowing nematode *Radopholus similis*. Our effector family selection also included families that (so far) appear to be specific to the cyst nematode lineages. These include SPRYSEC [Diaz-Granados et al., 2016], GLAND4 (also referred to as Gr-1106) and GLAND5 (also referred to as G11A06). For all of these effector families, at least a subset was shown to be involved in repression of the host plant immune system.

While comparing the overall diversification patterns of the six effector families under investigation, striking differences were observed. In SPRYSEC, GLAND5 (G11A06) and GLAND13 (GH32 members), virtually all *H. schachtii* paralogues appeared to be phylogenetically separate from the *G. pallida* and *G. rostochiensis* effector family variants, while representatives from the two PCN species were present in mixed clusters. These results should be taken with some caution as the backbone resolution of these phylogenetic trees ranges from poor (SPRYSEC) to robust (GLAND5, GLAND13). These patterns suggest that SPRYSEC, GLAND5 and GLAND13 effectors started to diversify after the split between Heterodera and Globodera.

Effector families GLAND4 (Gr-1106) and CLE showed distinct diversification patterns; by far most paralogues are grouped in species-specific clusters. As both effector families show a reasonable backbone resolution, we hypothesize that these effector families might have diverged after the split between *G. pallida* and *G. rostochiensis*.

Phylogenetic analysis of the VAP effector family in the three cyst nematode species revealed an opposite pattern, with almost complete mixtures of representative paralogues from the individual species. VAPs constitute an exceptionally widespread effector family within the phylum Nematoda [Wilbers et al., 2018], and our results indicate diversification of this family before the split between the cyst nematode genera *Globodera* and *Heterodera*.

4.4.2 Regulation of effector gene expression

Various stages of the parasitic life cycle of cyst nematodes such as plant invasion, feeding site induction and feeding site maintenance require the carefully orchestrated expression of distinct blends of effector proteins [Elashry et al., 2020, Cotton et al., 2014]. For some obligatory plant-parasitic nematodes, promoter elements have been identified that were suggested to be involved in this orchestration [Da Rocha et al., 2021, Eves-van den Akker et al., 2016]. For the three cyst nematode species we showed the presence of a short DNA box motif ("DOG box"; ATGCCA) in the promoter region of some members of some of the effector families. The absence of DOG boxes in the CLE family, the scattered presence of DOG boxes in the other five families and the loose correlation between DOG box dosage

and expression level lead us to conclude that DOG boxes might contribute to the orchestration of effector expression, but we see little evidence for a central role of this DNA motif in this process. Further investigation is necessary to elucidate the function of DOG boxes in effector regulation.

In plant-pathogenic fungi, a few transcription factors have been identified that were shown to steer effector expression. SIX Gene Expression 1 (Sge1), a conserved member of the Gti1/Pac2 protein family, was instrumental in the regulation of effector repression in a range of fungal pathogens including *Verticillium dahlia* [Santhanam and Thomma, 2013], *Zymoseptoria tritici* [Gohari et al., 2014] and *Fusarium oxysporum f. sp. cubense* [Hou et al., 2018]. Another example is AbPf2, a zinc cluster transcription factor from the necrotrophic plant pathogen *Alternaria brassicicola*. Via a loss of function approach, this transcription factor was shown to regulate the expression of eight putative effectors [Cho et al., 2013]. Evidently, plant pathogenic fungi are only very distantly related to plant parasitic nematodes, and these examples should only be considered as an illustration of how effector expression is organized in other plant–pathogen systems.

4.5 Conclusions

The PCN *Globodera pallida* and its sibling species *G. rostochiensis* co-evolved with potato in the Andes in South America. These pathogens have been introduced unintentionally in all major potato-growing regions in the world. Currently, PCNs are the most harmful pathogens in potato production systems, and as a result of this extreme anthropogenic range expansion potatoes worldwide cannot be grown without adequate PCN management. For both *G. pallida* and *G. rostochiensis*, host plant resistance are the best way to control these soil pathogens. However, their effectiveness depends on proper matching between the genetic constitution of the PCN field population and the set of host pant resistances present in modern potato varieties. Niere et al. [2014] reported *G. pallida* populations that could no longer be controlled by any of the currently used potato cultivars. This, combined with inherent imperfections of the current *G. pallida* pathotyping system [Phillips and Trudgill, 1983], underlines the need for a new pathotyping system. Such a system will be based on distinctive effector variations present in any given PCN population. The availability of a high-quality reference genome is a prerequisite for the development of such a system. We have demonstrated that the quality of the *G. pallida* genome presented in this paper allows for the mapping of complete effector families. With this resource, resequencing data from pathotypically diverse *G. pallida* populations will provide insight into the ecological diversification of this extreme range expander, and enable the development of a new pathotyping system that will facilitate the targeted and durable use of precious host plant resistances against this notorious plant pathogen.

CHAPTER 5

Comparative genomics among *Meloidogyne chitwoodi* populations reveals remarkably few, spatially concentrated variation hotspots

Joris J.M. van Steenbrugge[1], Stefan van de Ruitenbeek[1] Sven van den Elsen[1], Martijn Holterman[1,2], Hein Overmars[1], Muhammad I. Maulana[1], Geert Smant[1], Mark G. Sterken[1], and Johannes Helder[1]

[1] Laboratory of Nematology, Wageningen University and Research, The Netherlands
[2] Solynta, Wageningen, The Netherlands

5.1 Abstract

Root-knot nematodes (RKN) – *Meloidogyne spp.* – constitute the most impactful group of plant-parasitic nematodes. Subtle host plant penetration as well as the induction of exclusive feeding sites have contributed to the evolutionary success of this genus. In contrast to the tropical RKNs, *Meloidogyne* species from the temperate climate zones have been moderately well characterized. Here we present a highly contiguous genome assembly of a Northwest European population of the Columbia root-knot nematode *Meloidogyne chitwoodi* (NWE-Mc31). This highly polyphagous RKN species is widespread in higher latitudes of both the northern and southern hemisphere. The genomic constitution of NWE-Mc31 was compared with published genome assemblies from three phytopathologically distinct *M. chitwoodi* populations from the Pacific Northwest (PNW) (USA). With only 41,779 variable loci the four *M. chitwoodi* populations are remarkably similar. Most genomic variation was concentrated in five polymorphic regions (PR) spanning in total only 3.6 Mb. Whereas two PRs were variable for all four populations, the remaining PRs showed population-specific levels of variability. Putative effectors were identified in the PRs, and a chorismate mutase and two proteins related to cuticle turnover were found in shared variation hotspots. In contrast, five expansin B variants were found in an essential non variable regions of NWE-Mc31, whereas the same region was highly polymorphic for Race 2, a range expanding PNW population. The current comparison of four *M. chitwoodi* genomes showed a remarkably low overall diversification, while both shared and non-shared variation hotspots were identified. These insights provide a basis for pinpointing the genomic characteristics that underlie pathotype identity and range expansion of this agronomically relevant root-knot nematode.

5.2 Introduction

Plant-parasitic nematodes (PPN) reside among the most impactful categories of pathogens in crop production systems [Jones et al., 2013]. This impact is caused by a small minority (¡ 5%) of the 4,100 plant-parasitic nematode species that have been described in literature [Quist et al., 2015]. Most plant parasites are root hair and rhizodermis feeders causing no detectable damage to their host plants [Qing and Bert, 2019]. Root-knot nematodes (RKN), a common name that refers to all members of the genus Meloidogyne, belong to the aforementioned small minority, and from an evolutionary and agronomic perspective they arguably belong to the most successful (and harmful) plant-parasitic nematodes [Jones et al., 2013, Danchin et al., 2013]. RKNs are obligatory plant parasites, and their common name refers to gall formation on the roots. Galls are the result of disorganized proliferation of roots cells caused by RKN infection. Although there are notable exceptions such as *M. ichinohei*, *M. spartinae*, and *M. sylvestris* [Castillo et al., 2009, Plantard et al., 2007, Araki, 1992], RKNs tend to have broad host ranges. Some of the most well-known RKN species can parasitise hundreds of flowering plant species.

Meloidogyne is a species-rich genus, and distal members are organised into three clades [De Ley et al., 2002a, Tigano et al., 2005]. Clade I consists of RKN species mainly found in the (sub)tropics such as *Meloidogyne incognita*, *M. javanica* and *M. arenaria*. Clade II comprises temperate climate zone-bound species such as *M. hapla*, *M. graminis* and *M. marylandi*. Clade III also includes species that evolved in temperate climate zones, such as *M. chitwoodi*, *M. fallax* and *M. naasi*. Next to their geographic spread, the dominant reproduction modes further characterise these major clades. While Clade I species predominantly multiply by mitotic parthenogenesis, obligatory or facultative meiotic parthenogenesis are the dominant modes of reproduction for Clades II and III [Castagnone-Sereno, 2006].

The Columbia root-knot nematode *Meloidogyne chitwoodi* is a Clade III representative with a broad host range that includes both mono- and dicotylous plant species. It is relatively well characterized because of its economic impact on crops such as potato, tomato, barley and wheat [O'Bannon, 1982]. The species is present in temperate and Mediterranean climate regions in both the northern and southern hemisphere [Wesemael et al., 2011, Onkendi and Moleleki, 2013]. It was hypothesised to originate from the pacific north-west region in the United States as it was first discovered there by [Golden et al., 1980]. However, reanalysis of diseased, well-preserved potato tubers from 1930 in The Netherlands were shown to be infected with *M. chitwoodi* [Brinkman et al., 1996]. Hence, the centre of diversity of *M. chitwoodi* is currently unknown. Further research on the genetic variation between populations of different geographic locations will be required to get insight into the origin of *M. chitwoodi*.

A number of widespread and polyphagous *Meloidogyne* species such as *M. chitwoodi, M.*

fallax and *M. hapla* (race A) multiply via facultative meiotic parthenogenesis (e.g., [Janssen et al., 2017]). This implies that offspring is the product of the fusion between a pronucleus and a second polar body (both haploid) resulting from a single meiosis [Liu et al., 2007]. This mode of reproduction is distinct from asexual reproduction in the very well-characterized model nematode *Caenorhabditis elegans*. In *C. elegans* adult hermaphrodites are essentially females fertilizing oocytes with self-produced sperm. Hence, fertilisation takes place after two meiosis events. Just like *C. elegans*, the RKN species *M. chitwoodi*, *M. fallax* and *M. hapla* (race A) alternate between asexual and sexual modes of reproduction. This phenomenon is referred to as cyclic parthenogenesis and has been described for nematodes such as *Strongyloides* [Viney, 2006] and also for other Ecdysozoa such as *monogonont rotifers*, some *loriciferans* and aphids [Cortés et al., 2008].

To get more insight in the genetic constitution of facultative meiotic parthenogenetic RKNs, it would be relevant to know more about the frequency of alternations in reproduction modes. Sexual reproduction is dominant over asexual reproduction, and takes place when males are formed. Among meiotic parthenogenetic *Meloidogyne* species, sex is environmentally determined, and males will be formed under stressful conditions only [Triantaphyllou, 1973]. Stressful conditions comprise abiotic stress such as drought or high temperatures, as well as biotic factors such as crowding and poor nutritional conditions. Hence, the level of outbreeding of facultative meiotic parthenogenesis RKN populations will depend on the level and the frequency of (a)biotic stress perceived by these populations. We hypothesize that *M. chitwoodi* populations isolated from arable farming production systems perceive little (a)biotic stress and, thus, sexual reproduction will be infrequent among these populations.

The impact of meiotic parthenogenesis on the genetic variation at the genome level has been studied in detail for *M. hapla* (race A) [Liu et al., 2007]. An elegant genetic experiment that used two distinct inbred lines as a starting point revealed unusual marker segregation patterns. These patterns suggested for a preference for four-strand exchanges at similar locations, and such a preference would result in rapid genomic homozygosity. Liu et al. [2007]) hypothesized that these unusual segregation patterns were facilitated by holocentric nature of the chromosomes and very small chromosome size of *M. hapla* (on average 3 Mb). In a more extensive linkage map presented by Opperman et al. [2008], similar segregation patterns were found for most linkage groups, although some genomic regions deviated from this ratio. These regions harboured genes involved in survival and parasitism. As *M. chitwoodi* has holocentric chromosomes [Abad et al., 2008, Castagnone-Sereno et al., 1998], as well as similar small-sized chromosomes, we hypothesize that meiotic parthenogenesis could result in rapid genomic homozygosity in *M. chitwoodi* too.

To assess the impact of facultative meiotic parthenogenesis on the genetic constitution of *M. chitwoodi*, as well as to pinpoint the differences between populations of different geographical origins, the availability of a high-quality reference genome assembly is a prerequisite. Bali

et al. [2021] presented draft genome assemblies of three different pathotypes present in the Pacific Northwest (PNW) region of the United States; PNW-Race 1, PNW-Race 2, and PNW-Race1-Roza. Race 1 and Race 2 populations cannot multiply on hosts harbouring the resistance gene RMc(blb) [Mojtahedi et al., 1995], whereas Race1-Roza can break *S. bulbocastanum SB22* resistance. Race 2 deviates from Race 1 because of its extended host range.

As compared to tropical Clade I RKN representatives, relatively little is known about the effector repertoire of *M. chitwoodi*. Recently Zhang and Gleason [2021] compiled a list of putative effectors in a *M. chitwoodi* Race 1 population based on the comparison between RNA sequence profiles ofpre-parasitic and three parasitic life stages. Most effector genes are under diversifying selection [Carpentier et al., 2012], and we hypothesise that the use of natural selection metrics will contribute to the identification of putative effector genes.

Here, we present a highly contiguous genome assembly of the Northwest European (NWE) reference population Mc-31 [Teklu et al., 2016], and compared it with three draft genome assemblies of *M. chitwoodi* populations from the Pacific Northwest (PNW) (USA); Race1, Race2, Race 1-Roza [Bali et al., 2021]. Host range tests suggested that all Dutch *M. chitwoodi* populations belong to Race 1 [Van der Beek et al., 1999]. By comparing the genomic constitution of four *M. chitwoodi* populations we aimed at pinpointing the level of differentiation between these populations. Next, we investigated the distribution of polymorphic regions within the genomes, while wondering whether differences in variability were present among populations in these overall variable regions. By focusing on putative effector proteins in the polymorphic regions, we investigated whether the nature of the predicted effectors could be linked to the biological and phytopathological characteristic s of these *M. chitwoodi* populations.

5.3 Results

5.3.1 *M. chitwoodi* genome assembly resulted in near-chromosome size scaffolds

To enable the mapping of polymorphic chromosomal regions, we intended to build a high-quality reference genome assembly. To this end we generated a set of long-read PacBio sequencing reads with an average coverage of 228X and an N50 of 28Kb, as well as a dataset of 185 million paired-end 100bp Illumina HiSeq reads at a coverage of 397X. The reference genome produced for the *M. chitwoodi* NWE-Mc31 population has an assembly size of 47.4Mb divided over 19 scaffolds. The NWE-Mc31 assembly contains 972 gaps with a total gap length of 67Kb. A set of gene annotations generated by Braker2 was complemented by RNAseq data from nematode eggs and second-stage juveniles. In total, Braker predicted 12,800 genes and 14,173 transcripts. Together, the assembly and annotations provide a near-chromosome level overview of the NWE-Mc31 genome. The *M. chitwoodi* NWE-Mc31 assembly was compared with three other publicly available *M. chitwoodi* reference genomes, including annotations (Table 5.1). In terms of assembly sizes, the reference genomes are within a 1Mb range from each other. The Mc31 genome is about 1.7 times less fragmented than Race2 and Roza, and 1.3 times less fragmented than Race1. We conclude that the quality of newly generated genome assembly of the Mc31 population is comparable to the *M. chitwoodi* assemblies published by Bali et al. [2021], except that the level of fragmentation that is lower in case of NWE-Mc31. We assessed the completeness score (C) of each assembly using BUSCO, and the four assemblies showed near-identical C scores (64.7%-64.8%), whereas only a small fraction of the complete genes was classified as duplicated (0.8% – 2.4%). BUSCO completion scores tend to be lower for root-knot nematode genome assemblies than for most other eukaryote species. This can be illustrated by representatives of the three major *Meloidogyne* clades [De Ley et al., 2002a]. *M. incognita* (population Morelos, *Meloidogyne* Clade I) had a BUSCO completion score of 71.3% [Abad et al., 2008], *M. hapla* (population VW9, *Meloidogyne* Clade II) had a score of 87.4% [Opperman et al., 2008], and *M. graminicola* (population IARI, *Meloidogyne* Clade III) showed a BUSCO completeness of 82% [Somvanshi et al., 2021].

Table 5.1: Comparative statistics of four *M. chitwoodi* genome assemblies, and corresponding BUSCO (eukaryota$_{odb10}$) scores. Draft genome assemblies from *M. chitwoodi* populations from the Pacific Northwest of the USA ('PNW') were previously published by [Bali et al., 2021]. Mc31 is here referred to as 'NWE – Mc31' to underline its Northwest European origin.

				BUSCO scores (%), n= 255 for each population		
Population	Size (Mb)	# Scaffolds	N50 (Mb)	Complete (Single, Duplicated)	fragmented	missing
NWE - Mc31	47.4	19	2.5	64.7 (63.9, 0.8)	14.5	20.8
PNW - Race1	47.5	30	2.5	64.7 (63.9, 0.8)	14.5	20.8
PNW - Race2	46.9	39	2.3	64.7 (63.9, 0.8)	14.9	20.4
PNW - Race1 Roza	47.7	38	2.4	64.8 (62.4, 2.4)	14.5	20.7

5.3.2 Genome-wide genetic variation is low and mainly located in five divergent regions

The overall phylogenetic relationship between the four populations was determined by clustering analysis (Fig. 5.1). Out of the four populations, the PNW populations Race1 and Race 1 Roza are the most similar. PNW - Race2 shows moderate similarity to the Race1 and Race 1 Roza cluster, while NWE-Mc31 is positioned as an outgroup. This distribution is in line with the geographical origins of the populations; PNW populations cluster together, and are distantly related to NWE-Mc31.

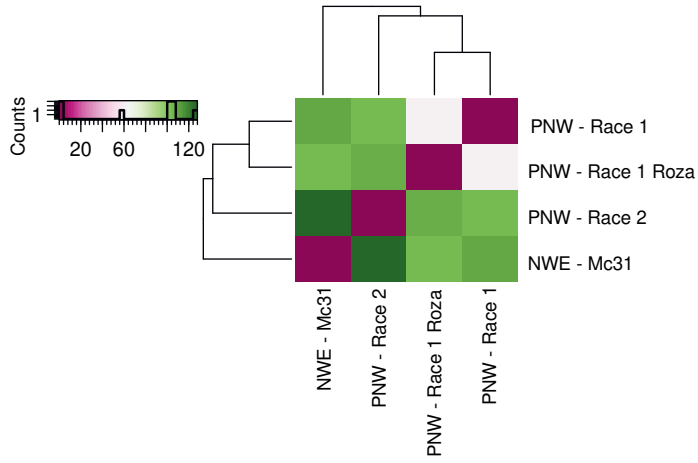


Figure 5.1: Whole-genome heat map showing the pairwise genome conservation distances between *M. chitwoodi* populations. NWE: population originating from Northwestern Europe. PNW: populations originating from the Pacific Northwest of the USA. Colour key (top left) represents the Euclidean distance based on the frequency of alternative alleles within a population in a pairwise comparison between populations.

To characterise the genetic variation in *M. chitwoodi*, Illumina reads of the three PNW populations and the NWE population were mapped on the NWE-Mc31 reference genome followed by variant calling. Overall, only 41,779 quality-filtered variable loci were called in the PNW populations. In addition, we performed a similar analysis by aligning DNA reads of the NWE-Mc31 population on the NWE-Mc31 reference genome and calling variants. This resulted in 7,407 sites that are potentially heterozygous within the NWE-Mc31 population. To further characterise the segregating sites, we compared the distribution of variants across the *M. chitwoodi* NWE-Mc31 genome. We determined the nucleotide diversity by binning the genome in 452 regions of 100 Kb and determining the variants per bin. Each bin contains 120 variants on average. However, most variants are concentrated at specific loci on the genome. The top 5% most diverse bins contained 70% of the variants. We identified five clusters of bins that contain a high number of polymorphisms , McPR1 - McPR5 (Table 5.2, Fig. 5.2). The overall genetic diversity was remarkably low as compared to other (plant-parasitic) nematodes (e.g., Chapter 2).

The distribution of polymorphic regions was visualized comparing the number of variants per 100 kb bin, and distinct variant patterns were observed (Fig. 5.2). Within McPR1 on scaffold 2, a region that variable among all four populations was followed by a region that showing a high number of variants for populations PNW- Race1- Roza and PNW- Race 2.

Table 5.2: Polymorphic Regions (PR) resulting from variant calling using NWE-Mc31 as a reference. Only a relatively low number of variable loci were detected, and most of these loci were clustered on three scaffolds only.

Name	**Scaffold (sc)**	**Polymorphic scaffold regions (start – end, length)**	Mean No. Polymorphisms	No. Potential effectors
Mc-PR1	Mc31_sc02	2,200,000 – 2,700,000 (0.5 Mb)	1,136	44
Mc-PR2	Mc31_sc07	1 - 800,000 (0.8 Mb)	1,527	32
Mc-PR3	Mc31_sc08	2,000,000 - 3,000,000 (1 Mb)	1,406	39
Mc-PR4	Mc31_sc08	3,900,000 - 4,800,000 (0.9 Mb)	1,129	60
Mc-PR5	Mc31_sc14	300,000 - 700,000 (0.4 Mb)	256	26

Scaffold 7 harboured a 0.8 Mb region that showing a high level of polymorphism for all four populations. Within scaffold 8, two regions that were predominantly variable among the PNW populations only were interspersed by a region that was shown to be variable among all four *M. chitwoodi* populations. Within scaffold 14, a genomic region was found that show mainly variation within the Pacific Northwest populations Race 1 - Roza and Race 2.

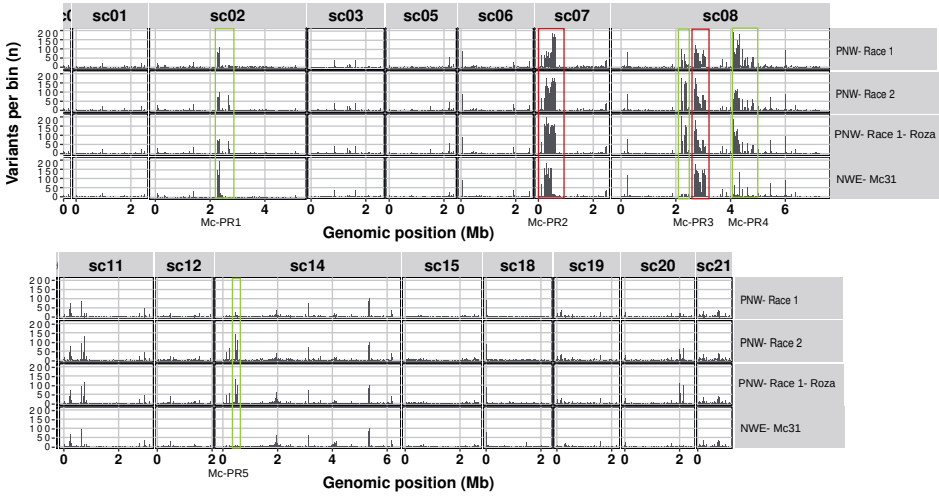


Figure 5.2: Overview of polymorphic regions per scaffold while comparing *M. chitwoodi* population NWE-Mc312 with the PNW populations Race 1, Race 2 and Race 1 Roza. Boxes with a dark red outline are used to delineate regions that are variable for all four populations, boxes with a light green margin refer to regions that are variable for a subset of the *M. chitwoodi* populations only. Scaffolds 04, 10, 13, 16, 17, 22 and 23 were very small (11-73 kb), and collectively merged in Scaffold 0.

5.3.3 Levels of genomic variation in divergent regions among four *M. chitwoodi* populations

The number of polymorphisms in each of the divergent regions presented in Fig. 5.2 is detailed and quantified in Fig. 3. The highest level of polymorphism was found within Mc-PR2 and Mc-PR3 with respectively up to 1,371 and 1,304 variants per bin. Notably, the highest levels of variation were found in genomic regions that were polymorphic for all four *M. chitwoodi* populations. As to be expected, the light green boxed regions show locally disparate levels of genomic variation. In case of scaffold 02, the 2.6 Mb interval was characterized by the virtual absence of variation in PNW-Race 1 and NWE Mc31. A similar pattern was observed for scaffold 14 in the 0.1 - 0.2 and the 0.4 - 0.5 interval. Within scaffold 08, two regions (intervals 2.1 - 2.4 and 4.0 - 4.8 Mb) were identified that were shown

to be considerably more conserved in the Northwest European population than in the three Pacific Northwest populations.

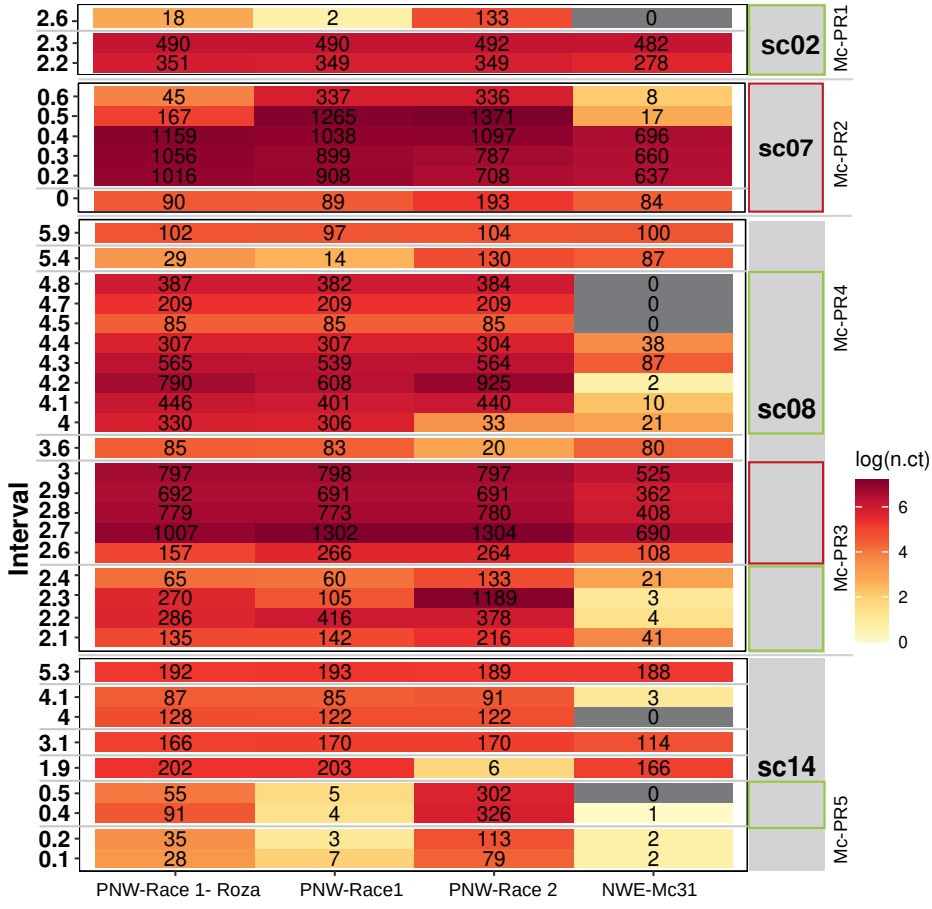


Figure 5.3: Heatmap with number of polymorphisms per *M. chitwoodi* populations in 100 kb segments for scaffolds with relatively large variable-rich regions (scaffolds 2, 7, 8 and 14). Polymorphism-rich fragments defined in Table 5.2, Mc-PR1, 2, 3, 4, 5 (PR= Polymorphic Regions), and visualized as coloured margin around the scaffold identifiers. Colours correspond with colours used in Fig. 5.2.

5.3.4 Putative effectors located in variable regions of the *M. chitwoodi* genome.

Within the selected polymorphic regions, putative effectors were identified on the basis of the presence of a signal peptide for secretion and the absence of transmembrane domains. Annotation of 201 putative effectors revealed 23 predicted genes which function directly or indirectly could be related to the interaction between *M. chitwoodi* and its hosts (Table 5.3). It is noted that most putative effectors (88%) lacked a specific annotation (receiving labels such as 'hypothetical protein', 'unnamed protein product', or no annotation), and will not be discussed any further.

Among the putative effectors with informative annotations within regions that are polymorphic among all four *M. chitwoodi* populations, a predicted protein with a chorismate mutase domain was identified (g6083.t1). Chorismate mutases are widespread among plant parasitic nematodes, and at least one chorismate mutase from *Meloidogyne incognita* was found to suppress plant immunity upon nematode infection [Wang et al., 2018]. Next to this, three putative effectors (g3702.t1, g3704.t1, g6054.t1) that are involved in moulting, or functional as component of the cuticle extracellular matrix were discerned in the variable domains of scaffolds 7 and 8. Root-knot nematodes migrate intercellularly in the plant root, and composition and structure of the cuticle will co-determine to what extend the plant innate immune system will be activated. The metallo-endopeptidase encoded by g6224.t1 was functionally linked to defence against Gram-negative bacteria. Inhibition of opportunistic bacteria could easily be envisaged as advantageous for a nematode in the intimate and relatively long-lasting interaction between a parasite and its host. Among the putative effectors with variable levels of polymorphism among the four *M. chitwoodi* populations, the number of putative effectors related to plant cell wall modifications was notable (g694.t1, g4196.t1, g9775.t1 - g9777.t1, and g9787.t1 - g9788.t1). Polygalacturonase were previously identified and characterized in *M. graminicola* [Petitot et al., 2020], cellulases were detected in *M. incognita* and other root- knot nematode species [Ledger et al., 2006], and about 20 candidate expansins were found to be present in the *M. incognita* genome [Abad et al., 2008]. The identification of plant cell wall-modifying proteins might not be unexpected, the fact that they we found them to be located in highly polymorphic regions of the *M. chitwoodi* genome is worth noting. A putative effector localized in Mc-PR4 showing a high level of variation among the PNW populations, and little variation in the European population Mc31 is g3967.t1, coding for a glutathione S-Transferase-resembling protein. Such transferases have been identified previously in *M. incognita*, and suggested to be involved in the protection of the nematode against plant-produced reactive oxygen species [Dubreuil et al., 2007]. Hence, putative effectors residing in genomic regions showing relatively high level of polymorphism amongst *M. chitwoodi* populations were found to be associated with functions such as plant cell wall modification, neutralisation of the host defence response

and, and cuticle formation and modification.

5.4 Discussion

Together with its sibling species *Meloidogyne fallax*, *M. chitwoodi* belongs to the most impactful root-knot nematode species in the temperate climate zones, and detailed knowledge about its genomic constitution could provide us with handles that could be instrumental in the control of this soil borne pathogen. Just like almost all members of *Meloidogyne* Clade III [Holterman et al., 2009], *M. chitwoodi* multiplies by facultative meiotic parthenogenesis, whereas this asexual mode of reproduction is occasionally alternated by amphimixis. It is noted that the formation of males, indicative for a switch towards sexual reproduction, is triggered by stressful environmental conditions. Here a novel reference genome of a Northwest European *M. chitwoodi* population was compared to the genomic constitution of three populations from the Pacific Northwest of the USA [Bali et al., 2021]. As all four population were collected from well-managed arable fields with inherent low levels of environmental stress, we presume that the genomic constitution of all four populations will mainly be a reflection of a meiotic parthenogenetic mode of reproduction. Our analyses revealed that the genomic constitution of *M. chitwoodi* populations showing predominantly asexual reproduction are remarkably similar, despite of their distinct geographical origins, and with 7,407 potentially heterozygous sites the level of heterozygosity the NWE-Mc31 population is remarkably low. To place the low level of diversification observed among *M. chitwoodi* populations in perspective, the inter-population variability at genome level was compared to variation among populations of two other plant parasitic nematodes, the pinewood nematode *Bursaphelenchus xylophilus*, and the soybean cyst nematode *Heterodera glycines* (both obligatory amphimictic). Comparison of genomes of 181 strains *B. xylophilus* (genome size: 77 Mb) from 16 provinces in China resulted in the identification of 7.8 million unique SNPs, and the SNPs were highly associated with the geographic origins of the populations [Ding et al., 2022]. In case of soybean cyst nematodes (genome size: 123 Mb), re-sequencing of 54 populations from its centre of diversity, China, gave rise to the identification 0.8 million SNPs [Lian et al., 2022]. Even if the smaller genome size of *M. chitwoodi* (47 Mb) and the limited number of populations under investigation are taken into consideration, 41,779 variable loci represent a remarkably low level of inter-population diversity.

Putative effectors in genomic regions showing high levels of polymorphisms for all populations. Variation in the *M. chitwoodi* genomes was mainly concentrated in five polymorphic regions distributed over 4 scaffolds. Two of these variable regions (Mc-PR2 and 3' part of Mc-PR3) were shown to be variable in all four populations. One of the putative effectors in Mc-PR2 is a protein that most closely resembles a type II chorismate mutase (CM II) from *M. graminicola*. In *M. incognita* upregulation of this gene promoted the conversion of chorismate to prephenate resulting in a suppression of salicylic acid synthesis [Huang et al., 2005], and thus, plant defence. Among root knot nematodes CMs are distributed between two separate, functionally distinct clades [Wang et al., 2018]. Hence, diversification within

the CM family could relate to optimisation of the suppression of plant defence responses against a wide range of hosts. Two putative effectors (g3702.t1, g6054.t1) related to the cuticle, the outer layer of nematodes that is directly exposed to soil or plant tissues, were shown to be present in generally variable genomic regions. The predicted gene g6054.t1 was expressed mainly in the egg stage and barely in infective second stage juveniles, and is unlikely to play a role in plant-nematode interaction. The g3702.t1 on the other hand was expressed in eggs and in infective juveniles, and it closely resembles galactoside-binding lectins from *M. graminicola* (KAF7635827, KAF7635828, and KAF7626946). An immunohistochemical study of a similar β-galactoside-binding lectin in the bacterivorous nematode *Caenorhabditis elegans* showed this protein was abundantly present in the adult cuticle and in the terminal bulb of the pharynx [Arata et al., 1996]. Currently, no function can be attributed to this protein in relation to its interaction with the host.

Putative effectors in genomic regions showing disparate levels of polymorphisms among populations. Remarkably, genomic regions were identified that showed large difference in the level of variability between the *M. chitwoodi* populations. This can be illustrated by Mc-PR4 in the 4.5 – 4.8 interval, showing a considerable level of variability among the PNW populations, and no variation within the NWE population Mc31 (Fig. 3). One of the putative effectors within this region is g4196.t1 coding for an β-1, 4-endoglucanase. Cellulases are involved in cell wall breakdown allowing the nematode to enter the host plant. Previously, *M. incognita* was shown to secrete 23 cellulases [Gahoi and Gautam, 2017], and we assume *M. chitwoodi* harbours a diversified cellulase family. This can be confirmed by another *M. chitwoodi* cellulase (GenBank AER27807) that showed only 83% identity at amino acid level with the current cellulase. We hypothesize that nematode cellulases might be subject to functional diversification. Cellulases belonging to several glycoside hydrolase families (GHs). For GH45 cellulases from Phytophaga beetles, variants were identified that were able to degrade glucomannan in addition to cellulose, and other GH45 family members lost their ability to degrade amorphous cellulose, and were shown degrade xyloglucan instead [Busch et al., 2019]. We hypothesize that similar functional diversification processes could explain why *M. chitwoodi* cellulases reside in polymorphic regions of the *M. chitwoodi* genome. Polymorphic region Mc-PR4 comprised 5 expansins belonging to the subfamily of the B expansins. Expansins are referred to as molecular lubricants weakening glucan-glucan binding in plant cell walls, and have no catalytic abilities. Functional diversification was shown for cerato-platanin family that are considered to be the functional equivalent of expansins in fungi [Luti et al., 2020]. Among 12 cerato-platanins in the plant-pathogenic fungus *Moniliophthora perniciosa*, individual proteins were shown to have specific binding capacities to fungal cell wall components [De O. Barsottini et al., 2013]. We propose that the five nematode expansins in Mc-PR5 that were shown to diverge mainly in *M. chitwoodi* populations PNW-Race 2, and – to a lesser extent – PNW-Race 1-Roza could be subject to functional diversification. We cannot explain why diversifying selection

among members of the expansin gene family did not take place in all four *M. chitwoodi* populations.

5.5 Conclusion

The generation of a highly contiguous reference genome for a Northwest European population of *M. chitwoodi*, and the comparison of this assembly with three populations from the Pacific Northwest of the USA revealed remarkably little diversification among these populations. Variation was concentrated in five genomic regions, and it is worth noting that only two regions were variable among all four populations, whereas especially NWE-Mc31 and PNW-Race 1 showed very little variation in regions that were highly polymorphic among the two other populations. The *M. chitwoodi* population selection in this research was based on their distinct host ranges (PNW-Race1, PNW-Race2), their disparate responses towards *S. bulbocastanum* SB22 resistance (only PNW-Race 1 - Roza can multiple on potato harbouring the resistance gene RMc(blb)), and the their distinct origins (Northwest Europe versus Pacific Northwest), and it was foreseeable that no molecular features would be found that are causally related to these three variables. Nevertheless, apart from fundamental genomic insights, interesting leads were found. This is exemplified by the variability patterns among the five predicted expansin B proteins in polymorphic region Mc-PR5. Within the 0.4 – 0.5 Mb interval NWE-Mc31 showed in total 1 polymorphism, PNW-Race 1 was characterized by 9 polymorphisms, PNW-Race 1-Roza by 146, and PNW-Race2 by 628 polymorphisms. We hypothesize that functional diversification of expansins as discussed above might be associated with the range expansion that it a key characteristic of PNW-Race 2.

5.6 Methods

5.6.1 DNA and mRNA sequencing

Pre-parasitic J2's from *M. chitwoodi* line NWE-Mc31 were collected from standard susceptible potato plants. J2's were concentrated, and sucrose centrifugation was used to purify the nematode suspension [Jenkins et al., 1964]. After multiple rounds of washing the purified nematode suspension in 0.1 M NaCl, nematodes were resuspended in sterilised MQ water. Juveniles were lysed in a standard nematode lysis buffer with proteinase K and beta-mercaptoethanol at 60°C for 1h [Holterman et al., 2006]. The lysate was mixed with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) (pH 8.0) following a standard DNA purification procedure, and finally, DNA was precipitated with isopropanol. After washing the DNA pellet with 70% ethanol for several times, it was resuspended in 10mM Tris-HCL (pH 8.0). *M. chitwoodi* line NWE-Mc31 (10 - 20 µg) was sequenced using Pacific Biosciences SMRT sequencing technology at Bioscience (Wageningen Research, Wageningen, The Netherlands). For gene annotation, mRNA was isolated from two *M. chitwoodi* life stages, eggs and J2's. The quality of both RNA isolations was RIN 10 (RNA Integrity Number). mRNA sequences were generated using the TruSeq RNA stranded polyA protocol of Useq (Utrecht Sequencing Facility, Utrecht, The Netherlands).

5.6.2 Genome Assembly and Annotation

The genome assembly was generated as described in (Chapter 2). The raw PacBio reads were corrected to an error rate of 15% and a corrected coverage of 200 using Canu v1.8 [Koren et al., 2017]. Initial draft assemblies were created using wtdgb2 v2.3 [Ruan and Li, 2020] to optimise for the parameters minimal read length, k-mer size, and minimal read depth. The most optimal initial assembly was selected based on the completeness, assembly size, and the number of contigs. Completeness was assessed using BUSCO v5.2.2 [Manni et al., 2021] using the eukaryota$_{odb10}$ database. The initial assembly that was generated with wtdbg2 settings minimal read length of 7000, k-mer size of 20 and a minimal read depth of 6 was used for downstream processing. Unmerged haplotigs were purged from the assembly using Purge Haplotigs v1.0.4 [Roach et al., 2018]. Contigs were scaffolded with PacBio reads using SSPACE-Longread [Boetzer and Pirovano, 2014] with a minimum overlap length of 1000bp and a minimum gap length of 500bp. Gaps in the assembly were then filled by taking the consensus of aligned Pacbio reads at a position, allowing a minimum coverage per position of 10 reads. Raw Pacbio reads were then used to polish the assemblies with 3 consecutive runs of Arrow v2.3.3 (`https://github.com/PacificBiosciences/GenomicConsensus`) at default settings, followed by five consecutive runs of Pilon v1.23 [Walker et al., 2014] with the –diploid setting. The resulting assembly was scaffolded an additional time based on three other publicly available *M. chitwoodi* genome assemblies ([Bali et al., 2021],

JACZZP000000000, JACZZO000000000, and JACZZN000000000) using MeDuSa v1.6 [Bosi et al., 2015] at default settings. Additionally, seven scaffolds originally named $Mc31_{Scaffold4}$, $Mc31_{Scaffold10}$, $Mc31_{Scaffold13}$, $Mc31_{Scaffold16}$, $Mc31_{Scaffold17}$, $Mc31_{Scaffold22}$, $Mc31_{Scaffold23}$ that are of sizes between 11Kb and 73Kb were merged in the bucket scaffold $Mc31_{Scaffold0}$. Using Braker2 v2.1.6 [Brůna et al., 2021], gene annotations in gff3 format were predicted, aided by RNAseq data of the egg and J2 life stages of *M. chitwoodi* population Mc31.

5.6.3 Variant calling and identification of polymorphic sites

Raw paired-end Illumina Hiseq reads of *M. chitwoodi* populations Race1, Race2 and Roza provided by Bali et al. [2021] were mapped on the *M. chitwoodi* Mc31 genome assembly using bwa-mem2 [Vasimuddin Md et al., 2019] at default settings. Variants were called using bcftools mpileup with options -d 1500, -L 1500, and -a AD, followed by bcftools call with options -m, and -v enabled. Called variants with a phred quality below 30 were pruned from the final vcf table. To identify hyper-divergent regions in the M. chitwoodi genome, the assembly was divided into bins of sizes 100Kb. For each bin, the nucleotide diversity was calculated using the PopGenome package in R v2.7.5 [Pfeifer et al., 2014]. Divergent bins were defined by a nucleotide diversity above 150. Bins that are adjacent to a divergent bin were also classified as divergent bin. The top five regions with the most diversity were defined as polymorphic sites.

Table 5.3: Overview of putative effector proteins within pre-selected polymorphic regions in genome assemblies from four distinct *M. chitwoodi* populations. SeqNames are presented in dark red for predicted genes from regions that show a high level of variation among all populations, and in light green for genes from regions that show variation for a subset of populations only. Regional variability / impact: whereas all putative effectors are residing in polymorphic regions, the nature of this variability at gene level varied. It is noted that two putative effectors were positioned in a - at individual gene level - non variable region.

Scaffold ID Mc31	Polymorphic region	SeqName	Regional variability - impact	Showing homology with - ontology terms	Length Predicted protein	Expect (e) value	GO IDs
sc02	Mc-PR1	g691.t1	variants impact: 'modifier'	Pept_C1 domain-containing protein - Defence response to Gram-negative bacterium, innate immune response, extracellular region	256	2.00E-152	GO:0050829, GO:0045087, GO:0005576, GO:0050830
		g694.t1	variants impact: 'modifier'	Fibronectin type-III domain-containing protein - Polygalacturonase activity	635	0	GO:0004650, GO:0005975, GO:0005515,
		g784.t1	Divergent region	Chitin-binding type-2 domain-containing protein - chitin binding, extracellular region	687	0	GO:0008061, GO:0005515, GO:0005576, GO:0005886
sc07	Mc-PR2	g6054.t1	variants impact: 'modifier'	Hypothetical protein *M. graminicola*_00007739 - Cuticle development involved in collagen and cuticulin-based cuticle moulting cycle, extracellular space	798	0	GO:0042338, GO:0005615, GO:0016020, GO:0008544
		g6076.t1	Variable region: Untrue	Gut esterase 1 - Carboxylic ester hydrolase activity	590	0	**GO:0052689**
		g6083.t1	Variable region	Chorismate mutase domain-containing protein - Chorismate metabolic process	225	6.00E-101	**GO:0046417**
		g6224.t1	variants impact: 'modifier'	Metallo-endopeptidase - Defence response to Gram-negative bacterium	408	0	GO:0050829, GO:0016239, GO:0007275, GO:0008340, GO:0004806, GO:0019433, GO:0005764
sc08	Mc-PR3	g3428.t1	Variable region: untrue	Positive regulation of macroautophagy, Cellulase domain-containing protein - Carbohydrate metabolic process, extracellular region	333	0	GO:0005975, GO:0004553, GO:0005576
		g3457.t1	Variable region	Cathepsin B - Cysteine-type endopeptidase activity	424	0	GO:0004197, GO:0071480, GO:0005615, GO:0016020, GO:0034644, GO:0051603, GO:0044248, GO:0005764
		g3505.t1	Variable region	Hypothetical protein *M. graminicola* _00005559 - Glycoprotein 6-alpha-L-fucosyltransferase activity	590	2.00E-156	GO:0008424, GO:0016020, GO:0006487, GO:0036071, GO:0044249, GO:0046702
		g3510.t1	Variable region	Saposin B-type domain-containing protein	166	1.00E-92	
		g3613.t1	Variable region	ShKT domain-containing protein, Choriolysin H	418	6.00E-126	
		g3702.t1	Variable region	Hypothetical protein *M. graminicola* _00009648 - Collagen and cuticulin-based cuticle extracellular matrix, galactoside binding, extracellular region	354	7.00E-128	GO:0060102, GO:0005515, GO:0016936, GO:0005576, GO:0030246
		g3704.t1	Variable region	Hypothetical protein *M. graminicola* _00005608 - Collagen and cuticulin-based cuticle moulting cycle	790	0	GO:0018996, GO:0016020
	Mc-PR4	g3948.t1	Variable region untrue	Hypothetical protein *M. graminicola* _00004739 - Collagen and cuticulin-based cuticle extracellular matrix, galactoside binding, extracellular region	114	2.00E-38	GO:0060102, GO:0005515, GO:0016936, GO:0005576, GO:0030246
		g3967.t1	Variable region	Hypothetical protein *M. graminicola* _00009654 GST_C_6 domain-containing protein	643	3.00E-90	GO:0016020; GO:0016021
		g4007.t1	Variable region	Zinc finger BED domain-containing - protein dimerization activity	161	3.00E-07	**GO:0046983**
		g4196.t1	Divergent region	β-1,4-endoglucanase - cellulase activity	220	1.00E-121	GO:0008810, GO:0005975, GO:0016020,
sc14	Mc-PR5	g9775.t1	Divergent region	Expansin B - Nonenzymatic loosing of plant cell wall	131	2.00E-59	GO:0016020
		g9776.t1	Divergent region	Expansin B - Nonenzymatic loosing of plant cell wall	130	9.00E-56	
		g9777.t1	Divergent region	Expansin B - Nonenzymatic loosing of plant cell wall	131	5.00E-64	GO:0016020
		g9787.t1	Variants impact: 'modifier'	Expansin B - Nonenzymatic loosing of plant cell wall	132	1.00E-38	
		g9788.t1	Variants impact: 'modifier'	Expansin B - Nonenzymatic loosing of plant cell wall	136	3.00E-43	

CHAPTER 6

Comparative structural analysis reveals convergently evolved effector proteins in cyst and root-knot nematodes

Joris J.M. van Steenbrugge[1], Geert Smant[1], and Mark G. Sterken[1]

[1] Laboratory of Nematology, Wageningen University and Research, The Netherlands

6.1 Abstract

Plant-parasitic nematodes manipulate host plants by injecting specialised effector proteins into host cells. The ability of cyst- and root-knot nematodes to parasitise plants is believed to have evolved independently. This idea is supported by the observation that only a few effector proteins share sequence similarity between them. However, many evolutionary unrelated effectors are thought to exert similar functions in plants. For example, the suppression of the plant immune system, by converging on common host targets. We hypothesise that convergent evolution on host targets has led to common secondary and tertiary structural motifs in unrelated effectors. To test this hypothesis, we predicted protein structures of all predicted secreted proteins in the cyst nematode *Globodera pallida* and the root-knot nematode *Meloidogyne chitwoodi* using alphafold2. We identified 551 unique *Globodera pallida* proteins and 476 *Meloidogyne chitwoodi* proteins that exhibited structural resemblances, even though the majority of these proteins did not display significant sequence similarity. Using clustering analysis and predicted annotations, we found overlap in predicated effector protein functions, which enabled the functional prediction of 258 previously unannotated proteins. Next, a comparison with protein structures of the full *Caenorhabditis elegans* proteome allowed us to further provide novel annotations for 232 *G. pallida* proteins. In conclusion, our findings give insight into the evolutionary history of secreted proteins between the cyst nematode *G. pallida* and the root-knot nematode *M. chitwoodi*, by distinguishing between evolutionary related proteins, and proteins that may have converged into a similar structure. Furthermore, structure-based comparisons enable functional predictions of putative effectors.

6.2 Introduction

Plant-parasitic nematodes are one of the most harmful plant parasites worldwide. Among the most economically damaging are the cyst- and root-knot nematodes [Jones et al., 2013], which may cause an estimated yield loss of up to 80%-90% without adequate control strategies [Jones et al., 2013]. The plant-parasitic abilities of these two lineages likely evolved independently (i.e., convergent evolution), as have at least three other plant-parasitic lineages (reviewed in [Quist et al., 2015]). Both cyst- and root-knot nematodes are sedentary endoparasites, meaning these parasites establish a permanent feeding site inside the host plant to survive and reproduce. Whereas root-knot nematodes transform plant root cells into a giant multi-nucleated cell [Escobar et al., 2015], cyst-nematodes form a syncytium by fusing multiple host cells [Moens et al., 2018]. Despite the differences in feeding site ontology, cyst- and root-knot nematodes must overcome similar obstacles such as suppressing the plant's immune responses, and reprogramming plant cells to form a nutrient sink [Goverse and Smant, 2014, Rai et al., 2015]. However, the evolutionary adaptations leading to these

similar, yet different, strategies remain unknown. While some of the proteins essential for parasitism by cyst- and root-knot nematodes have been identified, little is known about their specific functions.

In all plant parasitic nematodes the injection of specialised proteins into the host plant plays a role. These proteins, referred to as effectors, are functionally categorised into three groups [Vieira and Gleason, 2019]. The first group contains cell-wall modifying enzymes that dissolve parts of the plant's cell-wall, enabling the nematode to enter the root or puncture the cell. The second group contains immune-suppressing proteins required to survive within the plant tissue. The third group consists of proteins that are involved in influencing host processes including cell differentiation to enable the formation of the feeding site. These effectors are often members of (expanded) gene families (e.g., [van Steenbrugge et al., 2022, Diaz-Granados et al., 2016, Lozano-Torres et al., 2014]), typically contain a signal peptide for secretion [Mitchum et al., 2013], are produced in one of the gland cells, and are mostly excreted via the stylet into the plant. An example of an effector family found in cyst nematodes are SPRYSECs (SPRY-domain containing secreted proteins) [Sacco et al., 2009, Ali et al., 2015b]. Due to the convergently evolved nature of their parasitic abilities, cyst- and root-knot nematodes do not share many homologous effector genes.

While homology in effector genes is rare between cyst- and root-knot nematodes, few conserved families do exist. For example, venom allergen-like proteins (VAP) – involved in the suppression of host-immunity – are present in a broad range of nematode species, including animal parasites [Wilbers et al., 2018]. Variation in VAPs between cyst- and root-knot nematodes are therefore likely the result of divergent evolution. Other examples include various plant cell wall-modifying enzymes including cellulases (GH5, GH43) [Smant et al., 1998, Haegeman et al., 2011] and expansin-like proteins [Qin et al., 2004]. Interestingly, these are thought to be acquired via horizontal gene transfer from bacteria or fungi [Haegeman et al., 2011]. However, for the majority of effector families convergent evolution appears to be the driver for generating functionally similar effector proteins. The identification of effector genes that have converged into a similar function is difficult. Since the primary amino acid sequences of convergently evolved proteins are often different (e.g., [Shafee et al., 2017, Chen et al., 1997, Mackin et al., 2014]), traditional sequence based similarity searches are not effective. An alternative method to identify putative effectors is based on protein structure rather than amino acid sequence, as the protein structure is more conserved [Illergård et al., 2009]. Therefore, we hypothesise that protein structures will enable the identification of functionally similar, convergently evolved, effector proteins between cyst- and root-knot nematodes.

Producing accurate 3D models is essential to for large-scale comparisons between protein structures. Until recently, determining a highly accurate 3D structure of a protein was limited to laboratory techniques such x-ray crystallisation, nuclear magnetic resonance imaging

[Kendrew et al., 1958], or more recently cryo-EM [Bai et al., 2015]. These techniques yield accurate models but the costly, time-consuming and complex process limits high-throughput applications. A commonly used approach to predict a protein structure on a large scale is homology-based modeling, in which the structure of a protein is predicted based on homology with a protein for which the structure has been resolved [Jalily Hasani and Barakat, 2017]. However, the prediction accuracy of homology-based modeling approaches is lower than for laboratory techniques, especially when no similar template structure is available [Nayeem, 2006, Jalily Hasani and Barakat, 2017]. Recently, the protein structure prediction tool ALPHAFOLD2 was published, which predicts highly accurate protein structures, without the requirement of a homologous protein structure [Jumper et al., 2021]. ALPHAFOLD2 utilises a machine learning model that is trained on a large set of known protein structures and corresponding protein sequences. By combining both structures and amino acid sequences, the model has learned how different combinations of amino acids fold, and how changes in amino acids impact protein structure. As a result, ALPHAFOLD2 can predict highly accurate structures based on only amino acid sequences, without the necessity of a template structure like in homology-modeling techniques. ALPHAFOLD2 is therefore more accurate then homology-based modeling, and better scalable than laboratory modeling techniques.

Here, We investigated the structural similarities in the secretomes of the cyst nematode *Globodera pallida* and the root-knot nematode *Meloidogyne chitwoodi* to identify secreted proteins that have evolved convergently. Our hypothesis is that these similarities can reveal overlapping proteins in these evolutionary unrelated plant-parasitic nematodes. We used Alphafold2 to predict protein structures of all predicted protein sequences that contain a signal peptide for secretion. These predicted secretomes contain 1,870 proteins in *G. pallida* and 1,129 proteins in *M. chitwoodi*, and determined pairwise similarity based on structural alignments. In total, we identified 3,146 pairs of protein structures with significant structural similarity, comprising of 551 and 476 unique proteins in *G. pallida* and *M. chitwoodi* respectively. With an average amino acid sequence identity of 8.5%, most of these pairs would not have been identified using sequence based alignments. These similar protein structures contain known both convergently evolved secreted proteins as well as divergently evolved secreted proteins. Structural alignments are therefore capable to infer homology across much larger evolutionary distances than amino acid sequence based methods. Furthermore, structural alignments are capable to distinguish between convergently evolved proteins and divergently evolved proteins.

6.3 Results

6.3.1 Protein Structure is more robust than protein sequence to identify remote homology

We aimed to identify functionally similar effector proteins in evolutionary remotely related plant-parasitic nematode species. We started out with all protein sequences derived from gene predictions of the cyst-nematode *Globodera pallida* containing 18,071 protein-coding genes (Chapter 4), and the root-knot nematode *Meloidogyne chitwoodi* (Chapter 5) which contained 12,800 protein-coding genes. We filtered the full sets of protein sequences to a smaller collection that still likely contains effector proteins, based on the presence of a signal peptide for secretion, and the absence of a transmembrane domain. [Mitchum et al., 2013]. This subset included 1,870 proteins in *G. pallida* and 1,129 proteins in *M. chitwoodi* (Supplemental table 6.1). The most commonly used method to detect homology between proteins is through amino acid sequence alignments. However, when determining sequence similarity between the two effector subsets, we found that these sequences on average share a global sequence identity of 0.8% ($\sigma = 1.9\%$), a local sequence identity of 4.9% ($\sigma = 2.9\%$), and a local sequence similarity of 14% ($\sigma = 5.4\%$) (Supplemental Figure 6.6). We found 691 pairs of homologous proteins ($> 25\%$ global protein sequence identity) through this process, consisting of 271 out of 1,870 (14%) *G. pallida* and 255 out of 1,129 (23%) *M. chitwoodi* protein sequences. Among these homologous genes were VAPs [Lozano-Torres et al., 2014, Wang et al., 2007]. Members of this family share on average only 29% sequence identity between *G. pallida* and *M. chitwoodi*.

Since protein structure is more conserved than protein sequence [Illergård et al., 2009], structural similarities may be a better predictor for similarities in function or resulting from distant homology for evolutionary distinct species. We conducted an analysis to identify functionally similar proteins by predicting structures for each of the proteins in the proteome subsets using Alphafold2. Subsequently, all *G. pallida* and *M. chitwoodi* secreted protein structures were aligned using FATCAT2 [Li et al., 2020]. This analysis yielded 2,111,225 comparisons between protein structures, of which 3,146 were similar (Bonferroni corrected $p < 0.05$) (Fig. 6.1-A). These 3,146 significant alignments consisted of 551 unique *G. pallida* proteins and 476 unique *M. chitwoodi* proteins. While the majority of protein queries matched with a single target protein, many proteins matched with multiple targets (Fig. 6.1-B,C), indicating that many proteins are members of similarly structured protein groups. Since the similarity of these corresponding protein sequences was found to be low, 89% of the similarities captured by the structural alignments would not have been captured by sequence based searches (Fig. 6.1-D,E).

Included in the structurally similar protein structures are members of known effector families. We found that proteins within the same effector gene family show a high degree of structure

similarity, expressed as the residual mean square deviation (RMSD). The RMSD is calculated by comparing the position of atoms in the carbon backbone of one protein to those in the other, where a lower scores indicates a smaller distance and therefore a more similar protein pair [Li et al., 2020]. For example, the eight protein structures belonging to the VAP gene family in *G. pallida* have an RMSD of between 0.31 and 0.82 (mean RMSD: 0.58; mean $p < 1 \times 10^{-8}$) in all versus one pairwise comparisons, using the structure of Gpal_D383_g01238 as a reference. Similarly, the nine protein structures from the more diversified SPRYSEC-b cluster [van Steenbrugge et al., 2022] had an RMSD of between 0.97 and 3.02 (mean RMSD: 1.60, mean $p = 1.6 \times 10^{-5}$) in all versus one pairwise comparisons, using the structure of Gpal_D383_g12450 as a reference. In contrast, when a VAP protein was compared with a SPRYSEC-b protein, an RMSD of 8.5 (p = 0.66) was found (Supplemental Figure 6.7).

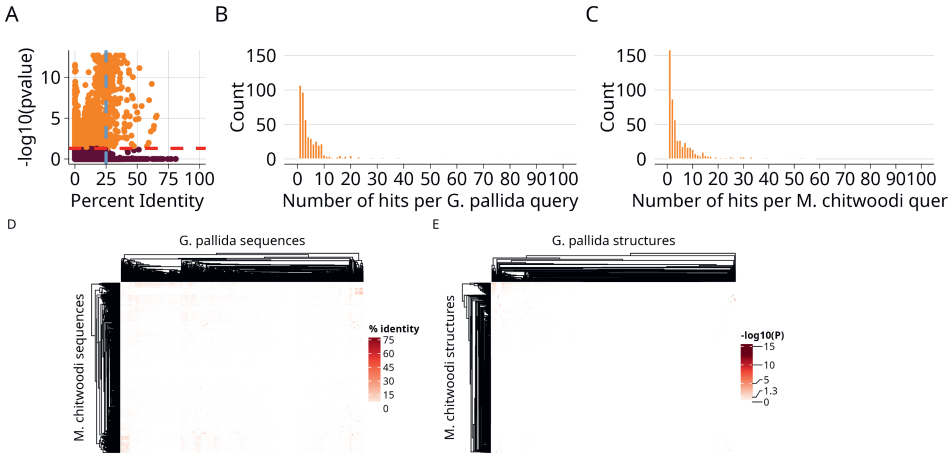


Figure 6.1: **A)** Pairwise comparisons of protein structures between 1,870 *G. pallida* and 1,239 *M. chitwoodi* proteins. Protein pairs are colour coded based on the -log10 Bonferroni corrected p-value (significant: orange, insignificant: black. **B)** Number of structurally similar *M. chitwoodi* protein for each *G. pallida* query. **C)** Number of structurally similar *G. pallida* protein for each *M. chitwoodi* query. **D & E)** Pairwise comparisons of proteins based on sequence alignments (D) and structural alignments (E). Each point represents a unique protein pair. Structural similarity is expressed as the -log10 Bonferroni corrected p-value for each protein pair. Protein sequence similarity is expressed as percent identify.

6.3.2 Protein shape as a predictor of functional similarity in plant-parasitic nematode species

To identify the groups of structurally similar proteins, a cluster analysis using the 3,146 significant pairwise structural comparisons was performed in Cytoscape. We identified 107 clusters of structurally similar proteins (Supplemental Figure 6.8). Out of the 107 clusters, 46 were pairs containing only one *G. pallida* and one *M. chitwoodi* protein. For the remaining 79 clusters the average size was 16.5 proteins, ranging from 3 to 138 (Fig. 6.2-A). On average, *G. pallida* (harmonic mean: 49.7%) contributed more proteins to each cluster than *M. chitwoodi* (harmonic mean: 36.0%) (Fig. 6.2-B). The average pairwise sequence identity between *M. chitwoodi* and *G. pallida* proteins within a cluster (24.3% identity; σ: 17.6%) was much higher than the pairwise identity between any random secreted protein pair (0.8% identity; σ: 2.0%).

To test our hypothesis, that structurally similar proteins were also predicted to have a similar function, we conducted a Gene Ontology analysis. Notably, the majority of clusters had either no common Gene Ontology (GO) terms or only a small proportion of shared ones. However, fifteen clusters share more than 50% of their GO annotations (Fig. 6.2-C). Clusters 5, 6, and 7 (Fig. 6.2-D) overlapped in 46 GO terms including functional annotations (F:GO) such as actin binding (GO:0003779), acyltransferase activity (GO:0016747), and cysteine-type peptidase activity (GO:0008234). Clusters 70 and 71 (Fig. 6.2-E), and similarly Clusters 124 and 63 (Fig. 6.2-H), share functional annotations related to signalling pathways including transmembrane signalling receptor activity (GO:0004888), monoatomic ion channel activity (GO:0005216), extracellular ligand-gated monoatomic ion channel activity (GO:0005230), and chitin binding (GO:0008061). Clusters 4, 37 (Fig. 6.2-F) shared a term related to kinase activity (GO:0016301), whereas clusters 16 and 101 (Fig. 6.2-G) were related to serine-type endopeptidase inhibitor activity (GO:0004867) and peptidase inhibitor activity (GO:0030414).

In the clusters, we identified functional labels for several effector gene families. This includes VAP [Lozano-Torres et al., 2014], and various cell-wall modifying enzyme including endoglucanases, glycoside hydrolases, calreticulins, and pectate lyases [Hamamouch et al., 2012, Smant et al., 1998, Mitreva-Dautova et al., 2006, Popeijus et al., 2000]. For certain effector families, like VAPs, all members are grouped within a single cluster (Fig. 6.3). Contrary to VAP, there are also clusters containing members of multiple effector gene families. For example, Cluster 13 contains multiple cell-wall modifying enzymes belonging to expansins, beta-1,4-endoglucanases, glycosyl hydrolase 30, and glycosyl hydrolase 25 (Fig. 6.4). Notably, the proteins in Cluster 13 form separate sub-clusters that corresponds with their functional annotations. The group of expansin proteins was connected to the beta-1,4-endoglucanases by a shared similarity with only a single unspecified cellulose binding protein. Similarly, GH25 proteins were connected to the endoglucanase group through similarities

with a single predicted GH53 protein. Much stronger was the connection between GH30 and beta-1,4-endoglucanases which were connected by three putative glycoside hydrolases through multiple connections. Additionally, the majority of proteins in multiple clusters (e.g., Cl13, Cl121, Cl24, Cl25, Cl34, Cl74, Cl88) were unannotated or contain annotations such as "putative effector protein" or "hypothetical protein". While it may not be possible to derive a consensus function for these types of clusters, our findings that these groups of proteins are likely to perform similar predicted functions. These results indicate that the shape of proteins can be used as a predictor of functional similarity in independently evolved plant-parasitic nematode species.

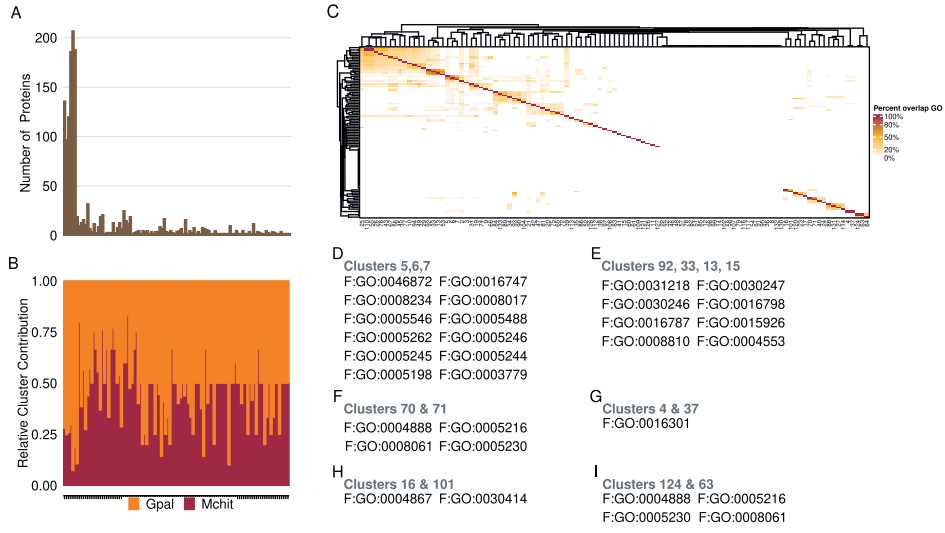


Figure 6.2: Statistics of the cluster analysis, showing the distribution of the number of proteins in a cluster (A), the relative contribution of proteins to each cluster per species (B), The percentage overlap of Gene Ontology (GO) terms between all clusters (C), Overlapping GO terms between different groups of clusters (D-H). To determine overlapping GO terms, only GO terms describing the molecular function were considered.

6.3.3 Protein function prediction based on structural similarity

To further gain insight into the nematode effector proteins, we leveraged the structures to aid in the annotation of uncharacterized proteins. We defined two groups of effector proteins based on their annotation status: (1) unannotated proteins from one species that could be annotated based on structural alignments with annotated proteins from the other species; and (2) unannotated proteins from either species not matching with annotated proteins from

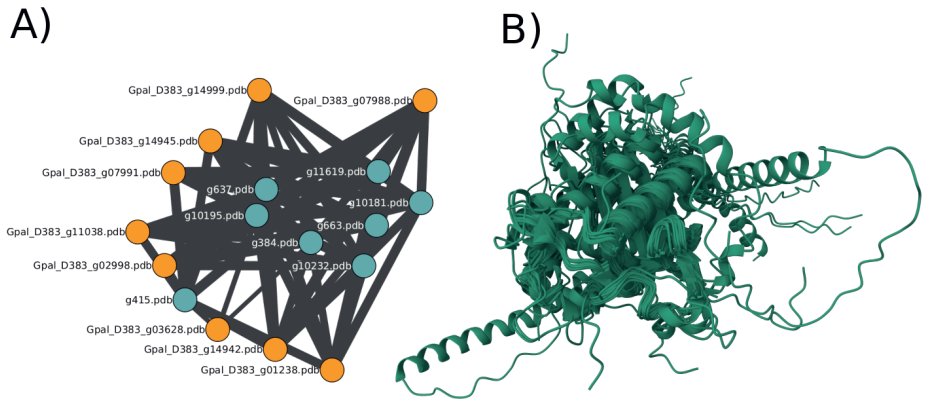


Figure 6.3: **A)** Cluster 66 containing members of the VAP/SCP effector family. *G. pallida* proteins are shown in orange, *M. chitwoodi* proteins are shown in blue. Edges are drawn based on significantly similar protein structures. Edge thickness is based on the pairwise protein sequence identity (higher equals thicker). **B)** Superimposed protein structures of the 17 Venom-allergen like proteins in *G. pallida* and *M. chitwoodi* from Cluster 66.

the other species. In total, there are 346 *G. pallida* and 305 *M. chitwoodi* proteins across all clusters that lack a functional annotation. For each unannotated *G. pallida* protein, we identified the similar *M. chitwoodi* proteins and compared the gene annotations (Supplemental Table 6.3). In G. pallida, 105/346 unannotated proteins (30.3%) shared structure similarity with functionally annotated *M. chitwoodi* proteins. In *M. chitwoodi*, 153/305 (50.2%) unannotated proteins share similarity with functionally annotated *G. pallida* proteins. To provide an example of this first group, we focused on the effector gene family VAP. While members of this family could also be identified through sequence-based methods, it serves as a proof of concept for other effectors that do not share sequence similarity. VAP protein structures were all present in the same cluster (Cluster 66). Although most were already annotated as either VAP or SCP domain-containing proteins, four proteins in the same cluster had a different annotation, such as "putative effector protein". We inferred that these putative effectors could be annotated as VAP proteins using structural comparisons. This inference was supported by the presence of a CAP domain (Supplemental Table 6.2).

The second case where none of the clustering proteins were annotated was more complicated. There we aligned the predicted structures of *G. pallida* proteins with the complete predicted proteome of the model species *Caenorhabditis elegans*. We selected *G. pallida* for the comparison as this species contains the most unannotated proteins (241). The *C. elegans* proteome consists of 19,694 proteins of which 7,588 structurally overlap with one or more of

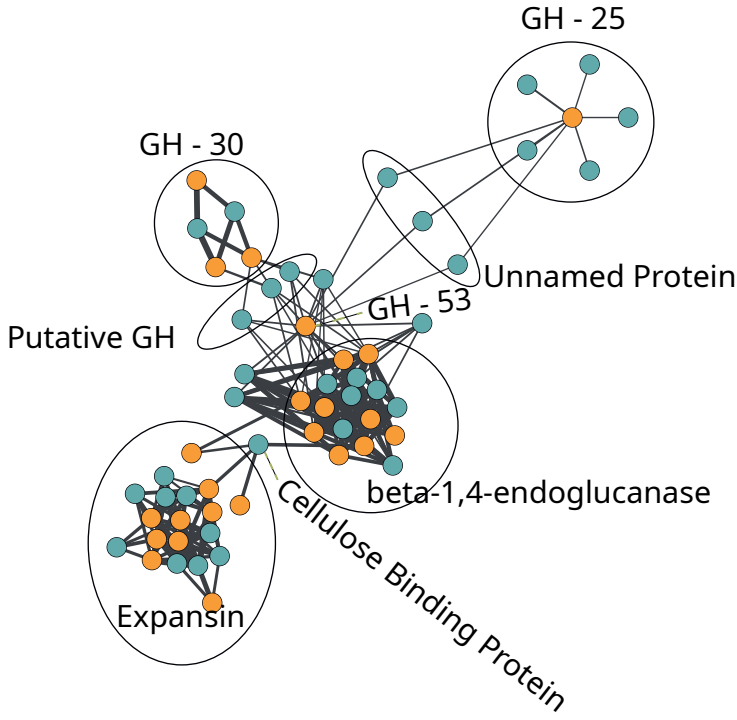


Figure 6.4: Cluster 13 contains members of four functionally related effector families. *G. pallida* proteins are shown in orange, *M. chitwoodi* proteins are shown in blue. Edges are drawn based on significantly similar protein structures. Edge thickness is based on the pairwise protein sequence identity (higher equals thicker).

the secreted *G. pallida* proteins. Of the 241 *G. pallida* proteins, 232 were structurally similar to *C. elegans* proteins. The number of similar *C. elegans* structures for each *G. pallida* protein differed, ranging from 1 to 905 (mean: 10.7, σ: 40.5; Supplemental Tables 6.4, 6.5). For example, Gp_Ce_Cluster_9 (Fig. 5) illustrates the structural similarity between the two *G. pallida* proteins (Gpal_D383_g04161 & Gpal_D383_g13513) and seven Ground-like domain containing proteins in *C. elegans*, a domain that was previously found to only exist in *C. elegans* and *C. briggsae*. The homology between these *C. elegans* proteins and *G. pallida* proteins would not have been identified based on sequence-based searches, as the sequence identities ranges from 9.0% to 24.9% (average: 17.9%). This example underscores the potential of structural comparisons to identify putative homology over much larger evolutionary distances than sequence-based alignments.

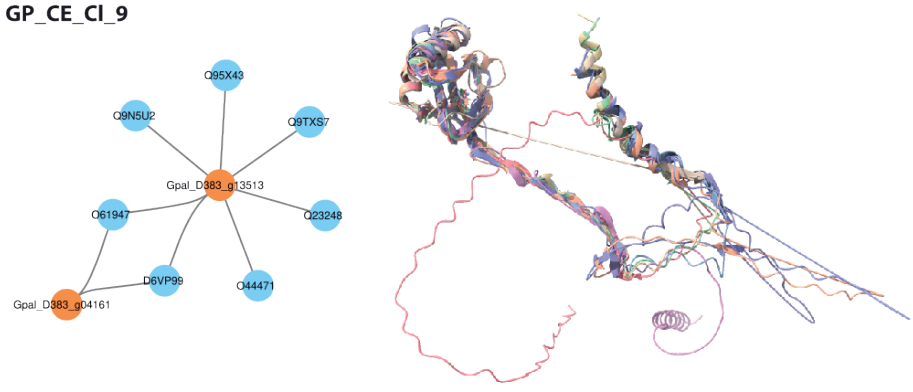


Figure 6.5: Gp_Ce_Cluster 9 of structurally similar *G. pallida* (orange) and *C. elegans* (blue) proteins.

6.4 Discussion

6.4.1 Evolutionary origin of effectors

Plant-parasitic nematode genomes encode for a wide variety of effector proteins, which serve as a crucial tool for their parasitic abilities [Mitchum et al., 2013]. Effectors can be part of highly diverse gene families [Vieira and Gleason, 2019] that are formed through various evolutionary mechanisms, including diversifying selection [van Steenbrugge et al., 2022], neo-functionalisation [Lilley et al., 2018], and horizontal gene transfer [Abad et al., 2008, Haegeman et al., 2011]. The parasitic abilities of cyst- and root-knot nematodes are believed to have evolved independently [Quist et al., 2015], leading to a proportion of effectors that

are evolutionarily distinct. Nevertheless, several effectors from both cyst- and root-knot nematodes have similar functions, including cell-wall modifying enzymes, the modification of cell-differentiation in the host, and the suppression of host immunity. However, also evolution of similar function – despite difference in sequence – has been described for viral RNA structures (for instance the xrn1-stalling structures in flaviviruses [Schnettler et al., 2014, MacFadden et al., 2018]), protein active sites such as disulphide reducing structures, nucleic acid binding motifs, and other functional domains [Graumann and Marahiel, 1996, Gherardini et al., 2007, Kuriyan et al., 1991], as well as whole protein shapes, for instance in defensins [Shafee et al., 2017]. Consequently, we hypothesized that an intersecting set of effectors exists between CN and RKN, possessing similar functions, which have emerged as a product of convergent evolution. To identify this overlap of convergently evolved effectors, we undertook a comparative analysis of 1,870 and 1,129 predicted protein structures in the predicted secretomes of the cyst nematode *G. pallida* and the root-knot nematode *M. chitwoodi*. Before structure prediction methods such as Alphafold2 [Jumper et al., 2021]) became available, only effectors with significantly similar amino acid sequences could be studied, while convergently evolved effectors remained a black box.

Our analysis yielded 3,146 pairs of proteins from the predicted secreteomes of *G. pallida* and *M. chitwoodi* that shared significantly similar protein structures. These similar proteins were organized into 107 clusters based on significant structural similarities with other proteins. The clusters contained proteins of various functions, including both known and unknown effector families. Of the similar proteins, 899 protein pairs could also be identified based on sequence alignments. Therefore, 2,247 of the similarities that we discovered based on protein structure would therefore not have been captured based on sequence similarity searches. We propose that for the 899 pairs with significant sequence similarity, having a common ancestor is the most likely scenario. However, for the 2,247 pairs that were only identified by structural alignments, convergent evolution is a possible scenario. Several previously identified effectors in the clusters exhibited both sequence and structural similarity. These included the cell wall-modifying enzymes from the glycosyl hydrolases family 30, expansins, beta-1,4-endoglucanases, and venom allergen-like proteins. Therefore there appears to be an overlapping core set effectors in *G. pallida* and *M. chitwoodi* that can partially be identified through sequence-based alignment methods, and partially through structure-based alignment methods. To determine whether this core set is indeed present in both types of plant parasitic lineages, more species per lineage need to be included in further analysis.

In the past, homology in molecular research was mainly inferred based on sequence alignments. However, for evolutionary remotely related species, such as cyst and root-knot nematodes, homology can be more difficult to determine due to a generally higher sequence diversity. The amino acid sequences of some, but not all, effector gene families may be especially diversified, due to positive selective pressure (e.g.,[Verhoeven et al., 2023]). Re-

cent insights suggest that for amino acid sequences with less than 35% identity, protein sequence alignments might not accurately represent structural similarities [Rajapaksa et al., 2023]. Sequence alignments with a low percent identity are therefore also unlikely to correctly represent functional similarities, due to misinterpretations on the impact of residue changes and insertions/deletions. Thus, even when remotely related amino acid sequences of effectors are detected, there iss significant uncertainty about their functional relationship. The findings by Rajapaksa et al. [2023] are in line with our results, as not all of the 899 similar amino acid sequences between the secretomes of *G. pallida* and *M. chitwoodi* resulted in significantly similar protein structures. In comparison, we found 3.5 times more similar protein structures than protein sequences, which highlights the increased sensitivity across large evolutionary distances. These findings give a unique insight into the hypothesised independent plant-parasitic evolution of *G. pallida* and *M. chitwoodi*. Due to their enhanced resolution in detecting protein homology, structural protein alignments are suitable for a comprehensive view of shared effectors between root-knot and cyst nematodes, shedding light on the evolutionary mechanisms driving their homology.

6.4.2 Structure alignments improve functional predictions

Plant-parasitic nematode effectors exhibit a wide array of functions, ranging from modifications to the plant cell-wall and the suppression of host immunity, to reprogramming cellular process in the host to establish a feeding site. While some effectors have been functionally characterised (e.g., as reviewed in [Vieira and Gleason, 2019, Mitchum et al., 2013]), many proteins that are predicted as effectors remain uncharacterised. The process of functionally characterising effectors is rather complex as it involves multiple steps, including validating expression of the encoding effector gene in the esophageal gland cells (e.g., [Xie et al., 2016]), the presence of the effector protein in the host tissue (e.g., [Mei et al., 2018]), and measuring a phenotype (e.g., [Ali et al., 2015b]). Because of the complexity involved in the functional validation process, it is infeasible to perform these steps for a large number of putative effectors. To get to a smaller set of putative effectors that can be functionally tested, multiple strategies currently exist to narrow down the search space. Here, we used the most basal filter strategy by pruning protein coding genes that lack a signal peptide for secretion or contain a trans-membrane helix. However other strategies include confirming the encoding gene to be expressed in parasitic life-stages and the presence of specific esophageal gland associated promotor motifs [Vieira and Gleason, 2019]. In addition to these steps, we propose the use of protein structures as an additional method to identify putative effectors by analysing proteins that share structural similarities to known effectors, maximizing the use of our current limited knowledge.

The functional prediction of proteins has been studied for many years (e.g., [Sadowski and Jones, 2009, Laskowski et al., 2005]). Recently, functional prediction tools utilizing pro-

tein structures have surged [Lai and Xu, 2022, Törönen and Holm, 2022, Xia et al., 2022, Zhang et al., 2017], coinciding with the substantial increase in available 3D protein structures, especially with high-throughput prediction tools like Alphafold2 [Jumper et al., 2021]. Compared to the traditional functional prediction approaches based on protein sequences, protein structures contain more information leading to predictions with higher accuracy. For example, not only the presence, but also the localisation of functional domains in the three dimensional structure is important for protein function, including for effectors [Mukhi et al., 2020]. Using sequence-based search methods such as HMMER (`http://hmmer.org`), the presence of a functional domain can be detected, but the correct localisation was previously nearly impossible to determine. We show that secreted proteins with a predicted functional similarity tend to have similar protein structures as well. Therefore, we hypothesise that this method could similarly be applied to functionally annotate putative effector proteins with a previously unknown function, by identifying similarities with other proteins that have functional descriptions.

In addition to functional prediction based on structural similarities with other proteins, an interesting follow-up analysis would be to further investigate the interaction between an effector and the host protein. By studying these interactions in more detail, differences in binding affinity between effector variants and the host protein may help to better understand the impact of these variants in the future. Several interactions between the effector and the host proteins are known, including Gp-RBP-1 and the Gpa-2 immune receptor in potato [Sacco et al., 2009], Gp-RBP-1 and HECT E3 ubiquitin ligase [Diaz-Granados et al., 2020] in potato, and 30D08 in the related cyst nematode *Heterodera schachtii* with SMU2 in *Arabidopsis thaliana* [Verma et al., 2018]. Software exists to investigate these interactions *in silico* through docking approaches [Watson et al., 2023, Roel-Touris et al., 2020, Baek et al., 2021, Schöning-Stierand et al., 2020, Jendele et al., 2019]. The examples of known interactions between the effector and the host protein are prime candidates to study the impact of variations. However, more often, only a general functional descriptions of effectors are available. For instance, various SPRYSEC effectors including as SPRYSEC-4, -5, -8, -15, -18, and -19 [Ali et al., 2015b], the *Meloidogyne spp* effectors *Misp12 [Xie et al., 2016]*, *Mh265* [Gleason et al., 2017], MiMSP32 [Verhoeven et al., 2023], and *Mg01965* [Zhuo et al., 2019] have been found to suppress host immunity. However, only a few specific interactions with the corresponding host target are known. Another new avenue for nematode effector research could therefore be the prediction of host targets. With our current understanding, host target prediction is also computationally feasible, especially for effectors that suppress host immunity as these have been previously shown to only interact with a limited set of host proteins [Mukhtar et al., 2011]. Therefore, the ability to generate hypotheses on protein function makes it possible in the future to screen large sets of proteins in a small time frame.

6.4.3 Potential challenges

The possibility to perform protein structure alignments on complete secreteomes provides a new way to infer homology and predict functional similarities between plant-parasitic nematode effectors. To align protein structures, we used a flexible alignment approach, as implemented in FATCAT2 [Li et al., 2020], which allows *in silico* adjustments of the protein's carbon backbone to improve alignment fit. As a result, the alignments were likely less sensitive to (subtle) variations, which may have a large influence in protein function [Araya and Fowler, 2011, Olivier et al., 2006, Stacey et al., 2008, Breast Cancer Linkage Consortium, 1999]. In plant-parasitic nematodes effectors genetic variation, and therefore variation in the resulting protein structures is abundant [van Steenbrugge et al., 2022]. The most studied examples is a single nucleotide polymorphism in the *rbp-1* gene in *G. pallida* that results in either detection or evasion of the plant immune system [Sacco et al., 2009]. Additional research is therefore necessary to further investigate the accuracy of the molecular function inference based on flexible protein structure alignments.

6.5 Conclusions

Plant-parasitic nematode effectors have evolved through either horizontal gene transfer, divergent evolution, and convergent evolution. Horizontal gene transfer and divergent evolution of effectors have been studied extensively in the past. Due to the poor resolution of sequence-based alignment methods, for low sequence similarities, convergent evolution remained poorly understood. Here, we studied the similarities between all predicted secreted protein structures of the root-knot nematode *M. chitwoodi* and the cyst nematode *G. pallida*. We found that protein structures provide a much higher resolution to infer homology than protein sequences. Furthermore, protein structures can significantly enhance our understanding of functional similarities, enabling us to infer the potential function of putative effectors by aligning their structures with those of well-characterized effectors.

6.6 Methods

6.6.1 Protein sequences

In order to compare protein structures of *G. pallida* and *M. chitwoodi* we obtained the genome assembly of the *G. pallida* D383 population (Chapter 4; NCBI: GCA_020449905.1), and the *M. chitwoodi* Mc31 population (Chapter 5). The corresponding protein sequences were filtered, keeping only the sequences that contain a signal peptide for secretion and that lack a transmembrane domain based on predictions by signalP v6.0 [Teufel et al., 2022]. This resulted in 2,181 secreted protein sequences of *G. pallida* (Supplemental table 1) and

1,509 of *M. chitwoodi* (Supplemental table 2).

The selected *G. pallida* and *M. chitwoodi* protein sequences were functionally annotated using BLAST2GO [Götz et al., 2008], based on all Eukaryota sequences in the NCBI NR database using an E-value cutoff of 1.0E-3.

6.6.2 Protein Models and Pairwise comparisons

For each of the secreted protein sequences the corresponding protein structure was predicted using ALPHAFOLD2 v2.2.0 [Jumper et al., 2021]. ALPHAFOLD2 was ran using a maximum template date of 2022-03-29 for the genetic databases. All ALPHAFOLD predictions are available on Figshare (*M. chitwoodi*: `https://doi.org/10.6084/m9.figshare.22339912` & *G. pallida*: `https://10.6084/m9.figshare.22339903`). Precalculated ALPHAFOLD structure predictions of the *Caenorhabditis elegans* proteome were retrieved from the Alphafold Protein Structure Database using accession UP000001940_6239_CAEEL_v4 [Varadi et al., 2022].

The *G. pallida* and *M. chitwoodi* structures, and *G. pallida* and *C. elegans* protein structures in pdb format were pairwise compared using a flexible alignment approach as implemented in FATCAT v2.0 [Li et al., 2020] in bulk and output in a table using a custom script. The use of flexible alignments reduces the effect of the mentioned sources of uncertainty two-fold. It reduces the impact of errors in gene annotations. For example, if an exon is erroneously added to a gene annotation then the corresponding structure will look different as well. By allowing twist in the backbone of the structure this deviation is less impactful, which similarly implies a reduced impact of inaccuracies in structure predictions.

Probability values produced by FATCAT were corrected using a modified Bonferroni approach, where each comparison is only corrected for the number of subjects (i.e., *M. chitwoodi* or *C. elegans* proteins) that each query (i.e., a *G. pallida* protein) is compared with. Alignments with an adjusted p-value lower than 0.05 are kept for subsequent analyses. This modified method (significant hits: 5,384) is less stringent than a true Bonferroni correction (significant hits: 1,033) but more strict than a False Discovery Rate correction (significant hits: 17,475). As a result, the reported number of structural similarities between *G. pallida* and *M. chitwoodi* is likely an underestimation of the real number of similarities. The modified Bonferroni correction approach was used to minimise noise arising from a potentially higher number of false positives, compared to FDR, while maintaining a relatively low number of true positives, compared to a true Bonferroni correction.

6.6.3 Pairwise statistics

To identify the relation between sequence similarity and protein structure similarity, we additionally performed protein sequence alignments. The global sequence identity between two

protein sequences was calculated using the Needleman-Wunsch algorithm as implemented in EMBOSS [Rice] using default settings, whereas the local sequence identity and similarity was calculated using the Smith-Waterman algorithm as implemented in BLAST [Altschul et al., 1990] using default settings.

6.6.4 Network Analysis

To identify clusters of structurally similar proteins, the pairwise comparisons were exported to CYTOSCAPE [Shannon et al., 2003] to produce a network of significantly similar protein structures. Clusters were defined using the "Fast Greedy" algorithm as implemented in the CYTOSCAPE package clusterMaker2 at default settings. All custom scripts to perform alignments, and all figures and tables are available on `https://github.com/jorisvansteenbrugge/NematodeProtFolding`.

6.7 Supplemental Data

6.7.1 Supplemental Figures

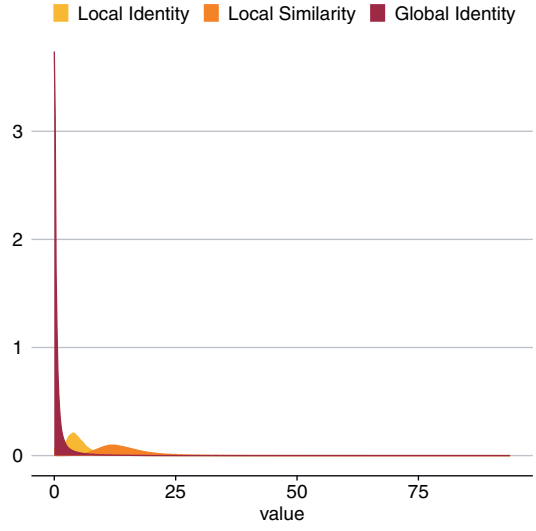


Figure 6.6: Pairwise Sequence identity of significantly similar protein structures.

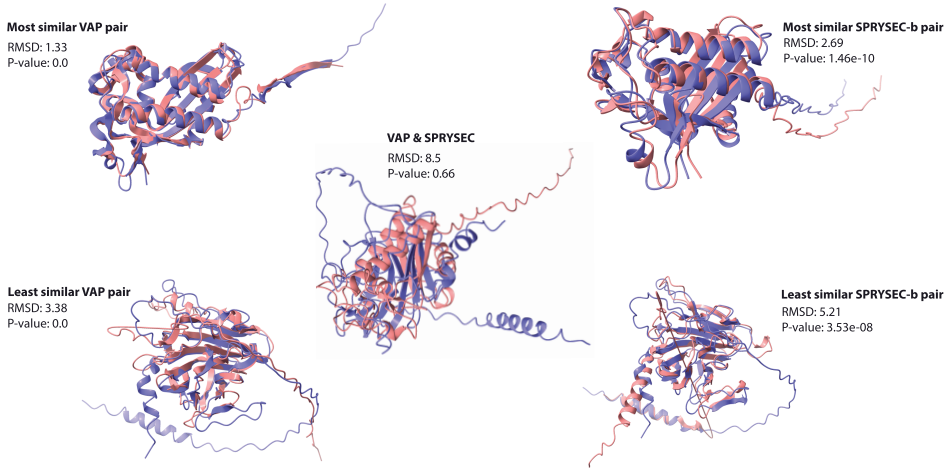


Figure 6.7: RMSD scores compared between the most and least similar two proteins of the VAP and SPRYSEC-b effector families. A comparison between a VAP and a SPRYSEC-b protein pair is included as a negative control.

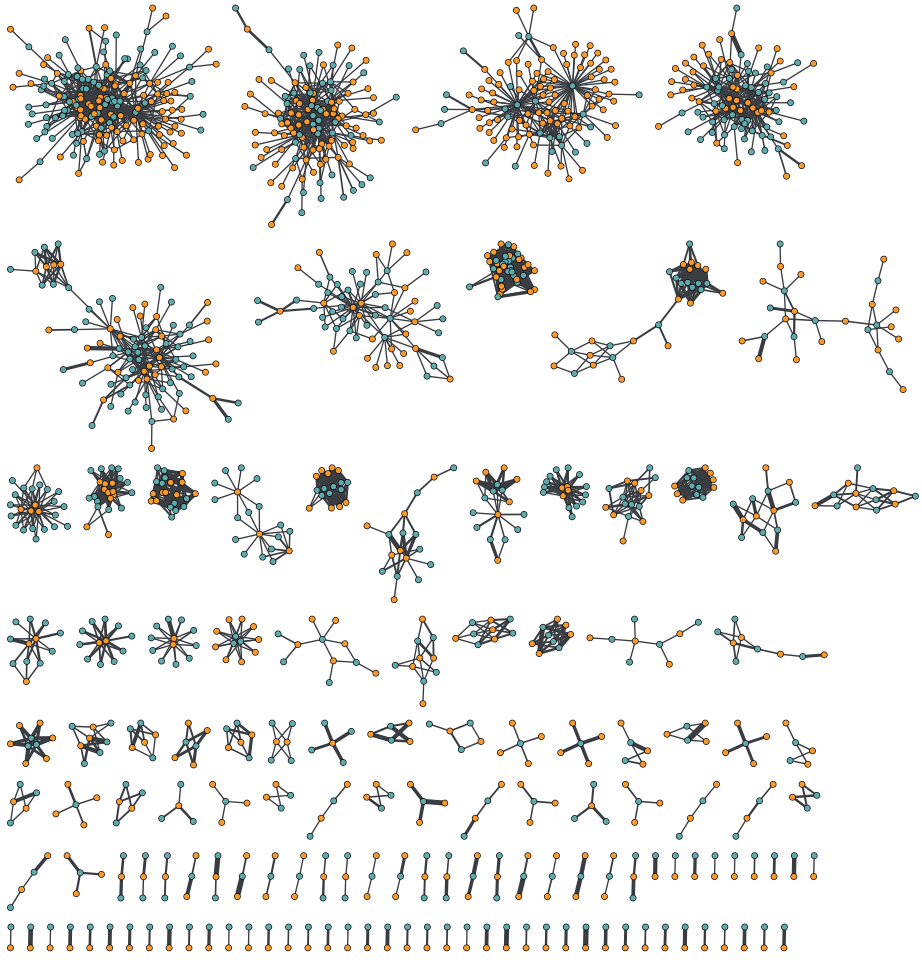


Figure 6.8: Network of structurally similar proteins. *G. pallida* proteins are shown in orange, *M. chitwoodi* proteins are shown in blue. Edges are drawn based on significantly similar protein structures. Edge thickness is based on the pairwise protein sequence identity (higher equals thicker).

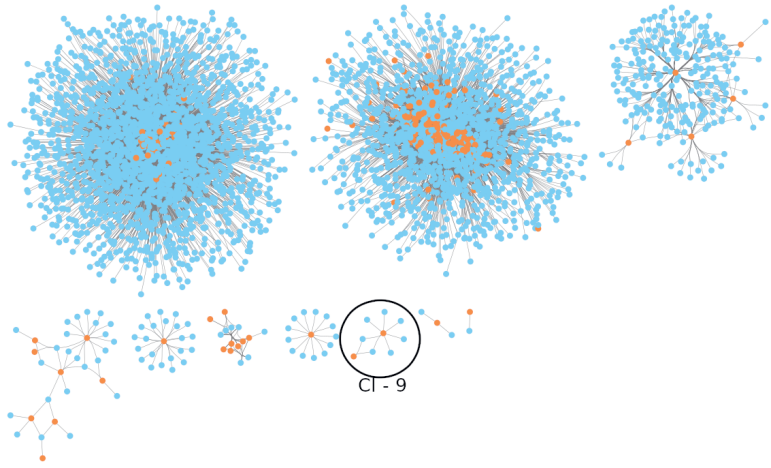


Figure 6.9: Network of structurally similar proteins between *C. elegans* (blue) and *G. pallida* (orange). Edges are drawn based on significantly similar protein structures.

6.7.2 Supplemental Tables

Table 6.1: All significant pairwise comparisons made between *G. pallida* and *M. chitwoodi*, including statistics, cluster assignment, and functional annotations. Only the first rows and a subset of the columns are shown, the full table is available on GitHub (`https://github.com/Jorisvansteenbrugge/nematode_protein_folding`

Query	Subject	P.value	Identity	Similarity	Global Identity	P.adjusted	Cluster	...	Subject_Description	Query_Description
g1002.pdb	Gpal_D383_g07529.pdb	7.77E-16	29.36	42.94	32.5	1.45299E-12	2	...	Cathepsin B-like cysteine proteinase 6	Pept_C1 domain-containing protein
g1002.pdb	Gpal_D383_g07721.pdb	0	36.14	50.3	35.3	0	2	...	cathepsin B	Pept_C1 domain-containing protein
g1002.pdb	Gpal_D383_g07729.pdb	0	32.27	44.09	31.5	0	2	...	cathepsin B isoform X1	Pept_C1 domain-containing protein
g1002.pdb	Gpal_D383_g08427.pdb	0.000000225	17.33	29.07	22	0.00042075	2	...	cathepsin L	Pept_C1 domain-containing protein
g1002.pdb	Gpal_D383_g09608.pdb	5.9E-10	18.89	33.33	16.8	0.0000011033	2	...	Proteinase inhibitor I25 and Proteinase inhibitor I29 ...	Pept_C1 domain-containing protein
g1002.pdb	Gpal_D383_g09916.pdb	3.06E-14	25.62	40.3	27.3	5.7222E-11	2	...	cathepsin B	Pept_C1 domain-containing protein

Table 6.2: Functional domain annotations of all *G. pallida* and *M. chitwoodi* proteins in Cluster 15, annotated with Interproscan. All protein sequences, including the 'putative effectors' contain the functional CAP domain. Only the first rows and a subset of the columns are shown, the full table is available on GitHub (`https://github.com/Jorisvansteenbrugge/nematode_protein_folding`

Query	Subject	Annotation	IPR ID	Description
g10195.t1	PR00837	Allergen V5/Tpx-1 family signature	IPR001283	Cysteine-rich secretory protein-related
g10195.t1	PR00837	Allergen V5/Tpx-1 family signature	IPR001283	Cysteine-rich secretory protein-related
g10195.t1	PR00837	Allergen V5/Tpx-1 family signature	IPR001283	Cysteine-rich secretory protein-related
g10195.t1	SSF55797	PR-1-like	IPR035940	CAP superfamily
g10195.t1	cd05380	CAP_euk	-	-
g10195.t1	PF00188	Cysteine-rich secretory protein family	IPR014044	CAP domain
g10195.t1	SignalP-noTM	SignalP-noTM	-	-
g10195.t1	PR00838	Venom allergen 5 signature	IPR002413	Venom allergen 5-like
g10195.t1	PR00838	Venom allergen 5 signature	IPR002413	Venom allergen 5-like
g10195.t1	PR00838	Venom allergen 5 signature	IPR002413	Venom allergen 5-like

Table 6.3: Unannotated *G. pallida* proteins with structurally similar annotated *M. chitwoodi* proteins. Only the first rows and a subset of the columns are shown, the full table is available on GitHub (`https://github.com/Jorisvansteenbrugge/nematode_protein_folding`

Cluster	Species	Unannotated_gene	Target	P-value	Annotation
2	Gpal	Gpal_D383_g01135.t1	g9832.t1	0.0000201	Protein CBG11905
2	Gpal	Gpal_D383_g01135.t1	Gpal_D383_g01135.t1	NA	—NA—
2	Gpal	Gpal_D383_g01796.t1	g8382.t1	0.0000204	proline-rich transmembrane protein 1
2	Gpal	Gpal_D383_g01796.t1	Gpal_D383_g01796.t1	NA	—NA—
2	Gpal	Gpal_D383_g02790.t1	g276.t1	0.0000019	CBN-JUD-4 protein
2	Gpal	Gpal_D383_g02790.t1	g10728.t1	0.0000019	EGF-like domain-containing protein
2	Gpal	Gpal_D383_g02790.t1	g11244.t1	0.0000081	emp24 gp25L p24 domain containing protein
2	Gpal	Gpal_D383_g02790.t1	g5956.t1	0.0000133	Pribosyltran_N domain-containing protein
2	Gpal	Gpal_D383_g02790.t1	g8382.t1	0.0000145	proline-rich transmembrane protein 1

Table 6.4: Network table of structurally similar proteins between *G. pallida* and *C. elegans*. Only the first rows and a subset of the columns are shown, the full table is available on GitHub (`https://github.com/Jorisvansteenbrugge/nematode_protein_folding`

Cluster	Name
GP_CE_Cl_3	A0A060Q5Z9
GP_CE_Cl_3	Gpal_D383_g00989
GP_CE_Cl_3	Gpal_D383_g06424
GP_CE_Cl_3	Gpal_D383_g10528
GP_CE_Cl_3	Gpal_D383_g17246
GP_CE_Cl_3	A0A060Q608
GP_CE_Cl_3	Gpal_D383_g02913
GP_CE_Cl_3	Gpal_D383_g11988
GP_CE_Cl_3	A0A061AD30

Table 6.5: Functional mapping table of unannotated *G. pallida* proteins against structurally similar, and annotated, *C. elegans* proteins. Only the first rows and a subset of the columns are shown, the full table is available on GitHub (`https://github.com/Jorisvansteenbrugge/nematode_protein_folding`

Gpal_ID	Wormbase ref	Protein Name	Gene Name	Uniprot_ID
Gpal_D383_g00009	B0348.10;	Uncharacterized protein	B0348.10 CELE_B0348.10	A0A0K3ARQ1
Gpal_D383_g00009	C49F5.13;	Uncharacterized protein	C49F5.13 CELE_C49F5.13	A0A0S4XR61
Gpal_D383_g00009	F11G11.15;	Uncharacterized protein	CELE_F11G11.15 F11G11.15	A0A2C9C2N5
Gpal_D383_g00009	C32E8.13;	Uncharacterized protein	C32E8.13 CELE_C32E8.13	A0A2K5ATM1
Gpal_D383_g00009	F53G2.12a;	Complex1_LYR_dom domain-containing protein	CELE_F53G2.12 F53G2.12	A0A2R8F5A8
Gpal_D383_g00009	C40C9.4a;	Uncharacterized protein	C40C9.4 CELE_C40C9.4	A0A3B1E543
Gpal_D383_g00009	R160.13;	Uncharacterized protein	CELE_R160.13 R160.13	A0A5E4M2K2
Gpal_D383_g00009	Y80D4G.2;	Uncharacterized protein	CELE_Y80D4G.2 Y80D4G.2	A0A5S9MMR3
Gpal_D383_g00009	F07C6.6;	Uncharacterized protein	CELE_F07C6.6 F07C6.6	A0FLR5

CHAPTER 7

General Discussion

7.1 Introduction

Plant-parasitic nematodes are a major pest in crop production worldwide, which makes them an important research topic. Plant-parasitic nematodes have been studied for centuries, with the first description of a plant-parasitic species dating back to 1743 [Needham, 1743], however, little is known about what causes their virulence. Over the years, significant effort has been invested in characterising virulence of plant-parasitic nematodes, with a particular focus on factors determining host range (e.g., [Santo, 1980, Stone and Roberts, 1981]) and defining pathotypes (e.g., [Kort et al., 1977, Andersen and Andersen, 1982]). However, because the first genome assembly of a plant-parasitic nematode was only published in 2008 [Abad et al., 2008], the underlying genetic architecture of virulence, and how genetic diversity in virulence genes can explain differences in host range and pathotypes, has remained poorly understood.

An important bottleneck for understanding virulence has been the lack of high-quality reference genome assemblies. In this thesis, I describe the generation of five high-quality genome assemblies of lineages of plant-parasitic nematodes, and used these assemblies to study genetic diversity in virulence in these nematodes. For the potato cyst nematode *Globodera rostochiensis*, I found that virulence is correlated with copy number variation and allelic variation in effector gene families. (**Chapter 2**). In a comparison among fourteen Indonesian *G. rostochiensis* populations of whole-genome genetic diversity versus effector family-specific diversity, I found that there likely has been a single introduction event of *G. rostochiensis* in Indonesia, and that the genetic diversity is correlated with the geographic distribution of the populations (**Chapter 3**). In a comparison between the related cyst nematode species *G. rostochiensis*, *G. pallida*, and *Heterodera schachtii*, I describe the diversity and evolutionary history of six effector gene families. Furthermore, I found that there is no clear correlation between a previously described dorsal gland specific promotor motif, that is thought to orchestrate effector expression, and gene expression of these six effector gene families (**Chapter 4**). In the Columbia root-knot nematode *Meloidogyne chitwoodi*, I observed that the overall genetic diversity is remarkably low, with most variation concentrated in five polymorphic regions. Remarkably, these polymorphic regions encode homologs of known nematode effectors, which might aid in the adaptation of host range and pathotypes (**Chapter 5**). The endoparasitic abilities of cyst and root-knot nematodes have evolved independently [Quist et al., 2015]. While the evolution of many cyst and root-knot nematode effectors have converged into related functionalities, only a small number of effector genes seem to share a common ancestor. However, using AI prediction tools we found much more structural similarity between effector families than previously possible using sequence-based alignment methods. (**Chapter 6**). In this chapter, I will discuss how the quality of the genome assemblies described in this thesis compares to other plant-parasitic nematode genome assemblies, how the quality of genome assemblies can be more objectively assessed,

and how assembly quality impacts our understanding of virulence of nematodes.

7.2 Plant-parasitic nematode reference genomes

In recent years, there has been a notable increase in the number of available nematode genome assemblies, with 31 plant-parasitic nematode genomes currently submitted to WormBase ParaSite (version: WBPS18; [Howe et al., 2017]). These genome assemblies represent 16 different species of plant-parasitic nematodes (Table 1). The assemblies can be categorised into three generations based on the sequence technology used to generate the reads. The first generation comprises only two assemblies, the *Meloidogyne hapla* Freeze_1 assembly [Opperman et al., 2008] and the original *M. incognita* Morelos assembly [Abad et al., 2008]. Both assemblies were constructed using Sanger sequencing data of BAC/YAC clones. This technology can produce accurate reads of moderate lengths (i.e., 500 - 600 bp) resulting in highly contiguous genome assemblies (e.g., [The C. elegans Sequencing Consortium, 1998]). However, it is costly to achieve sufficient coverage with Sanger sequencing. The second generation consists of genomes assemblies that were produced with short-read sequencing data (i.e., 50 - 300 bp), mostly with Illumina sequencing technology (i.e., HiSeq). Assemblies generated with short reads are typically fragmented but have a high single base accuracy. The third generation consists of genome assemblies based on long-read data (i.e., 15 - 20 kb) generated with Oxford Nanopore Technology or PacBio complemented with short-read sequencing to improve the accuracy. The combination of long-read and short-read sequencing has resulted in more contiguous assemblies with higher single base accuracy. Due to the differences in assembly contiguity and accuracy among the sequencing technologies, it is important to consider how these variations impact our findings on plant-parasitic nematode virulence.

Research on the genomic organisation of plant-parasitic nematode effectors is valuable as it provides insights into the evolutionary history of parasitism and virulence. However, investigating the genomic organisation of effectors has proven to be difficult when working with genome assemblies generated using short-read sequencing data. For instance, the first genome assemblies of *Globodera pallida* (GPAL001; Table 7.1; [Cotton et al., 2014]) and *G. rostochiensis* (nGr; Table 7.1; [Eves-van den Akker et al., 2016]) were each generated based only on short-read sequencing data. Consequently, both assemblies are highly fragmented as compared to newer long-read based assemblies (e.g., *G. rostochiensis* assemblies WUR_GloGros_L19 (Chapter 2) and WUR_GloGros_L22 (Chapter 2), and *G. pallida* assemblies D383 (Chapter 4) and SA_Gpal_Newton PRJNA702104). For the earlier genome assemblies of *G. pallida* and *G. rostochiensis*, the authors predicted gene models and annotated these models based on sequence homology with previously identified effectors. However, due to the high degree of fragmentation, the genomic organisation of effector families

in plant-parasitic nematode could not be resolved accurately. To my knowledge, the only study applying a short-read based genome assembly to explore effector gene organisation was based on the *G. rostochiensis*-nGr assembly [Eves-van den Akker et al., 2016]. The authors studied the genetic organisation of previously reported effectors by determining their positions on the genome. They observed that most of these effector genes were located in gene dense regions. This finding was in line with earlier observations that many effector families have diversified through gene duplication events [Kikuchi et al., 2017].

Long-read sequencing has opened up new avenues for genome research on plant-parasitic nematode effectors. First and foremost, the higher structural integrity of long-read assemblies has enabled more reliable inferences of the genomic organisation of effector genes, as exemplified in Chapter 2. Second, it allows to assess the impact of diversifying selection and its distribution across genomes, as shown in Chapter 5 and in more detail by Lee et al. [2021] Lee et al. [2021]. A more reliable inference of the genomic organisation of effectors also enables further studies on the regulatory mechanisms driving the expression of effector genes. Previous work has identified three conserved cis-regulatory motifs that are believed to influence effector gene expression in the dorsal gland of various nematode species [Eves-van den Akker et al., 2016, Espada et al., 2018, Da Rocha et al., 2021]. A causal relationship between presence of these cis-regulatory motifs and gene expression can be investigated further by identifying transcription factors that bind to these motifs. A physical interaction between such transcription factors and these motifs can be established using high-throughput chromosome conformation capture (Hi-C) [Tomás-Daza et al., 2023].

In addition to studying the genomic organisation of effector genes, long-read sequencing offers more accurate identification of structural variation within nematode genomes [Huddleston et al., 2017, Chaisson et al., 2015, English et al., 2015]. Structural variants are abundant in eukaryotic genomes and are known to have an impact on numerous phenotypes including on disease susceptibility [Conrad et al., 2010] in humans. In plant-parasitic nematodes, structural variants play an important role in the development of their effector repertoires through gene loss or gene duplication events [Bird et al., 2015, Kikuchi et al., 2017, Castagnone-Sereno et al., 2019].

Table 7.1: Plant-parasitic nematode genome assemblies as published in WormBase ParaSite (version WBPS18), grouped by their primary sequencing technology.

Species Name	Assembly	NCBI ID	Genome Size (Mb)	Number of Scaffolds	Publication Year
		Moderate Read Length			
Meloidogyne hapla	Freeze_1	PRJNA29083	53	3452	2008
Meloidogyne incognita	Morelos	CABB01000001–CABB01009538	86	2817	2008
		Short Read Length			
Acrobeloides nanus	v1	PRJEB26554	248	30759	2018
Bursaphelenchus xylophilus	ASM23113v1_submitted	PRJEA64437	75	5527	2011
Globodera pallida	GPAL001	PRJEB123	124	6873	2014
Globodera rostochiensis	nGr	PRJEB13504	96	4281	2016
Meloidogyne arenaria	ASM90000398v1	PRJEB8714	258	26196	2017
Meloidogyne arenaria	ASM369356v1	PRJNA340324	164	46436	2017
Meloidogyne enterolobii	ASM369367v1	PRJNA340324	163	42008	2017
Meloidogyne floridensis	nMf_1_1	PRJEB6016	97	58696	2014
Meloidogyne floridensis	ASM369360v1	PRJNA340324	75	8887	2017
Meloidogyne graminicola	Mgraminicola_V1	PRJNA411966	38	4304	2018
Meloidogyne incognita	Meloidogyne_incognita_V3	PRJEB8714	183	12091	2017
Meloidogyne incognita	ASM369364v1	PRJNA340324	122	33351	2017
Meloidogyne javanica	ASM90000394v1	PRJEB8714	236	31341	2017
Meloidogyne javanica	ASM369362v1	PRJNA340324	150	34316	2017
		Long Read Length			
Aphelenchoides besseyi	AORJ.fasta	PRJNA834627	46	32	2023
Aphelenchoides besseyi	AORT.fasta	PRJNA834627	47	39	2023
Aphelenchoides besseyi	APFT.fasta	PRJNA834627	45	80	2023
Aphelenchoides besseyi	APVT.fasta	PRJNA834627	45	28	2023
Bursaphelenchus xylophilus	BXYJ5	PRJEB40022	78	11	2020
Ditylenchus destructor	ASM157970v1	PRJNA312427	111	1761	2016
Ditylenchus dipsaci	D.dipsaci.v1.0	PRJNA498219	227	1394	2019
Globodera pallida	SA_Gpal_Newton	PRJNA702104	120	173	2023
Globodera pallida	D383_Gpal_1.0	PRJNA764088	113	163	2023
Globodera rostochiensis	WUR_GloGros_L19	PRJNA695196	93	88	2021
Globodera rostochiensis	WUR_GloGros_L22	PRJNA695196	102	135	2021
Heterodera glycines	ASM414822v2	PRJNA381081	158	9	2021
Heterodera schachtii	Hsch_Bonn_V1.0	PRJNA722882	179	395	2022
Heterodera schachtii	WUR_Hsch_IRS_1.0	PRJNA767548	190	705	2023
Meloidogyne arenaria	ASM313380v1	PRJNA438575	284	2224	2018
Meloidogyne chitwoodi	ASM1518303v1	PRJNA666745	47	30	2021

7.3 Plant-parasitic Nematode Genome Assembly

In this thesis, the genome assemblies are presented of five nematode lineages: *Globodera rostochiensis* Line-19 and *G. rostochiensis* Line-22 (**Chapter 2**), *G. pallida* D383 and *Heterodera schachtii* IRS (**Chapter 4**), and *Meloidogyne chitwoodi* Mc31 (**Chapter 5**). Compared to second generation genome assemblies (e.g., [Cotton et al., 2014, Eves-van den Akker et al., 2016]), a significantly higher degree of contiguity was achieved for these five assemblies. This improvement was primarily due to the use of long-read sequencing technologies. The publication of the genome assembly of *Ditylenchus destructor* in 2016 was the first genome assembly of a plant-parasitic nematode generated using long-read data [Zheng et al., 2016]. While this latter assembly used similar sequence technology as the assemblies in this thesis (i.e., PacBio), it is an order of magnitude more fragmented. Since the assembly size of *D. destructor* is not significantly larger, this discrepancy can be caused by either improvements in PacBio chemistry [Rhoads and Au, 2015], differences in the assem-

bly protocol, or a combination of both. Notably, a more traditional assembly procedure was used for the *D. destructor* genome. Contigs were produced using the short-read assembler ALLPATHS-LG [Gnerre et al., 2011] based on short-read and mate-pair libraries. These contigs were then scaffolded using a PacBio RSII long-read library, which was sequenced at low coverage [Zheng et al., 2016]. One possible explanation for this high number of scaffolds is a low coverage of PacBio sequencing data. Alternatively, the order in which both long-read and short-read data was used in the assembly process could also cause a lower level of contiguity. Previous work suggests that short-read derived contigs misrepresent the full complexity of repeats and transposons [Alkan et al., 2011]. It is therefore possible that the contigs in the *D. destructor* assembly contain missing or incorrectly sized repeats, resulting in mismatches with the long reads that were used for scaffolding. In this thesis, all initial assemblies were produced using long-read data and subsequently polished using short-read data, which resulted in much more contiguous assemblies.

It is not yet possible to isolate sufficient high molecular weight DNA from individual microscopically small plant-parasitic nematodes for high-coverage sequencing. Instead, high molecular weight DNA is isolated from a large number of individual nematodes (e.g., 10,000 juveniles) from a population rather than a single nematode. Consequently, the template DNA used for sequencing consists of a diverse pool of haplotypes. This pool of haplotypes can represent high levels of heterozygosity in diploid, sexually reproducing species due to DNA crossovers during meiosis (**Chapter 2**). Genome assembly programs are typically designed to assemble sequence reads from a single haplotype, instead of merging multiple haplotypes into a consensus genome sequence. It was, therefore, necessary to adapt the standard assembly protocols to handle high levels of heterozygosity in the DNA pools isolated from plant-parasitic nematodes in this thesis.

The assembly protocol used in this thesis dealt with the heterozygosity in two different ways. First, a rigorous correction step was performed on the PacBio or Oxford Nanopore Technology long-reads using the error correction mode in the assembly program Canu [Koren et al., 2017]. The error correction step essentially detects the overlap of the sequencing reads and merges overlaps that have a lower error rate than the threshold. For example, the authors of Canu opted for a default error rate for PacBio reads of 4.5% [Koren et al., 2017]. Therefore, reads with 95.5% or more identical base calls are merged. When multiple haplotypes of the same locus are sequenced, these reads could potentially be less similar than the default threshold and remain unmerged. During the genome assembly phase, the assembler may identify a proportional amount of these haplotypes as separate loci, introducing spurious contigs, or so called haplotigs. We raised this threshold from 4.5% to 15%, to allow Canu to merge haplotypes of the same locus. As a second step to address heterozygosity, remaining haplotigs were removed from the assembly using the Purge Haplotigs program [Roach et al., 2018]. Especially the first correction step has the risk to incorrectly prune parts

of the genome by collapsing low-complexity regions, in addition to collapsing the intended haplotypes. We validated the assemblies for this in two ways. First, we ensured that these steps produced genome assemblies that are close to the expected assembly sizes based on flow cytometry estimates. We additionally assessed completeness using BUSCO [Seppey et al., 2019] (**Chapters 2, 4, 5**). Second, we mapped the original uncorrected reads to the assemblies and called structural variants. The number of homozygous structural variants, which are indicative for miss-assembled regions of the genome, was very low in both *G. rostochiensis*-Line19 and *G. rostochiensis*-Line22 genome assemblies (**Chapter 2**). Altogether, this suggests that the steps in our assembly protocol adequately dealt with the high levels of heterozygosity.

To improve the contiguity of the assemblies even further in the future, it would be helpful to incorporate experimental data from optical mapping technologies. A method like Hi-C could enable (near) chromosome resolution by resolving very large repeats, for which long-reads are often still too short [Yuan et al., 2020]. Genome scaffolding using Hi-C data was first introduced in 2013 to assembly *Homo sapiens*, *Mus musculus*, and *Drosophila melanogaster* genomes [Burton et al., 2013]. It has since then been applied in numerous other genome assembly projects, including the mosquito *Aedes aegypti* [Dudchenko et al., 2017], sweet cherry (*Prunus avium L.*) [Wang et al., 2020], the nematode *Caenorhabditis remanei [Teterina et al., 2020]*, the pinewood nematode *Bursaphelenchus xylophilus* [Dayi et al., 2020], and the soybean cyst nematode *Heterodera glycines* [Masonbrink et al., 2021], to produce highly contiguous assemblies, up to the chromosome level.

7.4 Evaluation of Assembly Quality

Genome researchers often rate a genome assembly with a qualitative label, such as high-quality. In 2009, when next/second generation sequencing was more widely adopted, the number of available genome assemblies increased rapidly. At the same time, large differences in the quality of assemblies became evident, ranging from draft assemblies with minimal manual or automatic improvements (e.g., [Ling et al., 2013, Franzén et al., 2009, Chan et al., 2010]) to (nearly) finished assemblies (e.g., [The C. elegans Sequencing Consortium, 1998, Liu et al., 2009, Church et al., 2009]). To facilitate the categorisation of assembly quality, Chain et al. [2009] proposed a standardised set of assembly quality labels, Standard Draft, High-Quality Draft, Improved High-Quality Draft, and Finished. These labels remain valuable descriptors to date, as they are technology independent. The assemblies generated in this thesis (**Chapters 2, 4**, and **5**) meet the criteria of an *Improved High-Quality Draft* by Chain et al. [2009]. They cover at least 90% of the genome, with attempts made to purge contaminations, scaffold contigs in the correct order, and resolve gaps. This qualification still leaves room for misassemblies in repetitive regions, and base call errors in low-quality

regions.

While the standards set by Chain et al. [2009] aid in determining whether the quality of an assembly aligns with project goals, the authors did not make recommendations on how to validate an assembly. The current description of the labels is still quite ambiguous. For example, the description of an Improved High-Quality Draft includes requirements for scaffolding and gap closing (i.e., assembly contiguity), misassemblies (i.e., assembly correctness), and genome coverage (i.e., assembly completeness). However, there is no golden standard or metric to valid assemblies. The absence of such an evaluation standard can be explained by the complexity and nuances involved evaluating a genome assembly. Nevertheless, the quality of assemblies is currently most often assessed based on three properties, namely contiguity, correctness, and completeness, which are recently coined as the *3C criterion* [Molina-Mora et al., 2020]. I will elaborate on each of these properties in the subsequent sections of this chapter.

7.4.1 Contiguity

When producing a genome assembly, most effort is often spent on the contiguity and its relation to the total assembly size. The assembly size indicates potential missing or erroneously included sequences. contiguity is furthermore important for gene prediction. Long stretches of contiguous sequences will increase the number, and the accuracy, of gene predictions, as genes that span across two contigs will not be (accurately) predicted by gene prediction programs [Yandell and Ence, 2012]. contiguous genome assemblies therefore facilitate studies on the genomic organisation of genes and genetic variation within these genes (e.g., this thesis, [Lee et al., 2021]).

contiguity is commonly assessed using the N50 statistic, and the number of contigs. The N50 statistic is calculated by ordering all contigs based on size in descending order. Starting with the largest contig, their size is iteratively summed until the cumulative size reaches 50% of the total assembly size. The N50 is then the size of the contig that exceeds this 50% mark. This metric is an effective descriptor for assemblies with a gradual size distribution, however, it is biased against contigs with a highly unequal size distribution [Bradnam et al., 2013]. This could occur after extensive scaffolding, when the assembly contains a small number of very large contigs and many extremely small contigs. Such an assembly is then in fact highly contiguous, but the N50 would be misleading. In this same situation, the number of contigs would likewise indicate a fragmented assembly. More informative than the N50 would therefore be a contig size distribution plot, sorted from largest to smallest contigs. Furthermore, considering the genome size and karyotype could make the N50 metric more informative when conducting cross-species comparisons. An interesting addition to a size distribution would therefore be the contig/chromosome ratio [Wang and Wang, 2023]. This ratio represents the number of contigs relative to the number of chromosome pairs, and

together with the N50 and contig/chromosome ratio, would be a more robust and species independent way to assess the contiguity of an assembly.

7.4.2 Correctness

There are multiple sources of errors during genome assembly, ranging from base-level errors that can result from sequencing errors to structural errors. Multiple methods exist to assess the *correctness* of an assembly, which can be classified as either reference-based, or reference-free approaches. Reference-based approaches such as QUAST-LG [Mikheenko et al., 2018] assesses correctness by identifying small (i.e., ranging from 200bp to 1000kb) and larger (1000kb+) structural errors based on comparisons between the *de novo* assembly, the raw sequencing data, and an existing reference genome assembly. The disadvantage of reference-based assessment is that it requires a (nearly) finished reference genome sequence. This limitation is probably the reason why the authors of QUAST-LG demonstrated their software only on model species with nearly finished reference genome sequences [Mikheenko et al., 2018]. In many cases, including plant-parasitic nematodes, reference genome assemblies of such high quality are not available.

The most commonly used reference-free correctness assessment approaches for *ab initio* assembled genomes are K-mer based. Software such as Merqury [Rhie et al., 2020] or KAT [Mapleson et al., 2017] compare the K-mer distribution of the assembly with the K-mer distribution of raw Illumina sequencing reads, to estimate base level correctness, and identify possible contaminations. Both of these methods are designed to be performed on genome assemblies of diploid organisms, and require a high sequencing depth of accurate sequencing reads [Mapleson et al., 2017, Rhie et al., 2020]. In this thesis, the nematodes of which the genomes were assembled, are indeed all diploid organisms. However, since DNA was extracted from a heterogenous population, rather than an individual nematode, the sequence pool contains highly heterozygous reads. Because in this case the assembly is a consensus representation of the population, many reads that map on the same locus will therefore not be identical to the assembly. This heterozygosity will impact the K-mer distribution and likely results in a higher number of reported errors.

Since the introduction of high-throughput long-read sequencing, various programs have been developed to resolve structural complexity in genome assembly (e.g., [Koren et al., 2017, Ruan and Li, 2020]). However, there is only a limited number of methods available to detect structural errors in an existing assembly. One example is the program Inspector, that detects both larger structural- and small-scale errors in an assembly using long-read data [Chen et al., 2021]. Inspector reports on both large and small-scale assembly errors and reports a normalised Phred Quality Value (QV) score that incorporates all error-types. A genome assembly with a QV score of 40 (i.e., 99.99% correct) might be considered finished [Schmutz et al., 2004]. However, for the assemblies in this thesis a lower QV score suffices, as a finished

assembly is not strictly necessary for the characterisation of the genomic organisation of effector genes, and their genetic diversity. Only a limited research effort has been dedicated to structural variants within and between plant-parasitic nematode populations. Structural variants are abundantly present within human populations and impact a large proportion of the genome [Sudmant et al., 2015, Feuk et al., 2006]. Structural variants are therefore also expected to occur in these highly heterozygous plant-parasitic nematode populations. The majority of structural variants will likely occur in low frequencies [Abel et al., 2020] and might therefore remain undetected due to a low sequencing coverage. These variants present in the population will therefore likely have a small impact on the QV score. Future research is necessary to determine an appropriate QV score threshold for high-quality plant-parasitic nematode assemblies.

7.4.3 Completeness

As measures of assembly completeness, the most commonly used metrics are the genome assembly size and quantitative indicators of completeness algorithms such as CEGMA and BUSCO [Seppey et al., 2019, Parra et al., 2007]. A much smaller assembly size than expected based on flow cytometry estimations [DOLEŽEL and BARTOŠ, 2005, Vinogradov, 1998]) is indicative of missing genomic regions, and therefore an incomplete assembly. However, the assembly size as a completeness index is not sensitive towards erroneous duplications that may inflate the assembly size and therefore hide unrepresented regions, and should therefore be used in conjunction with other completeness estimates.

CEGMA and BUSCO use a similar principle, which involves identifying a set of universal genes in a taxonomic group (e.g., eukaryotes), mapping these genes to the assembly, and then reporting the percentage of genes found, missing, and fragmented. The main difference between the two programs is in the composition of the gene sets that are used to measure completeness. The CEGMA gene sets (COGs) are based on clusters of orthologous genes found in only seven (nearly) finished genomes of eukaryotic model species [Parra et al., 2007], which may not be representative for all eukaryotic species [Seppey et al., 2019]. CEGMA furthermore was discontinued in 2015 and therefore does not receive gene-set updates anymore. Surprisingly, despite this shortcoming, it is still being used to report measure of completeness (e.g., [Chen et al., 2023, Yan et al., 2023, Song et al., 2023]). BUSCO on the other hand uses curated sets (BUSCOs) for different taxonomic groups of universal single copy genes (i.e., present in $> 90\%$ of the species) based on OrthoDB [Kuznetsov et al., 2023]. BUSCO has two potential shortcomings when used as measure of completeness for genome assemblies of plant-parasitic nematodes. First, likely due to the small number of finished genomes, the phylum Nematoda is significantly underrepresented in OrthoDB. In the current release at the time of writing of this thesis (i.e., release 11) only eight nematode species are included, five of which are animal-parasitic nematodes species and three are free-living nematode species

(i.e., members of the genus *Caenorhabditis*). As a result, the set of universal genes that occur in 90% of these nematode species, do not represent plant-parasitic nematodes species, which are evolutionary distinct from free-living and animal-parasitic nematodes.

Secondly, BUSCO maps the genes in a gene set to an assembly by first predicting gene models using Augustus [Stanke et al., 2006]. The assembly completeness score reported by BUSCO is therefore not only dependent on the completeness of the assembly, but also on the accuracy of gene prediction by Augustus. Since other gene annotation pipelines exist that outperform gene models that are predicted by only Augustus (e.g., [Brůna et al., 2021, Campbell et al., 2014]), it would be more informative to calculate BUSCO completeness scores based on these more accurate gene predictions [Seppey et al., 2019]). I would therefore recommend publishing both the BUSCO scores of the assembly, as well as the accompanying gene models. Furthermore, if in a future update of OrthoDB more nematode species would be included, BUSCO would provide an even more robust completeness index.

7.5 Future perspectives

Effector repertoires deployed by plant-parasitic nematodes display significant diversity, resulting from mutations, gene duplications, diversifying selection, and horizontal gene transfers [Lilley et al., 2018, Haegeman et al., 2011, Vieira and Gleason, 2019]. This thesis sheds light on the effector diversity in economically important nematode species by examining the diversification of previously reported effector gene families across different populations. However, it is essential to recognise that the effector gene families explored in **Chapters 2 and 4** represent only a few examples of a much larger effector pool. Given the varying levels of diversification among effector gene families, capturing a broader spectrum of this diversity is crucial to fully understand plant-parasitic nematode virulence.

In spite of the public availability of numerous plant-parasitic nematode effector sequences, the absence of a centralized nematode effector repository requires extensive literature studies to gain a complete overview of the available information. Furthermore, the same nematode effectors may be redundantly published under different names, for example Gr-1106 [Finkers-Tomczak, 2011] and Hg-GLAND4 [Noon et al., 2015], which requires additional manual curation. To make effector sequences and their metadata more accessible, a complete overview of all known plant-parasitic nematode effectors in a database is therefore necessary. Such a database would provide a standardised and non-redundant resource, encompassing the full spectrum of known effectors. This information can then be used to easily compare the effector repertoire between nematode species. Using commonly used search tools like BLAST [Altschul et al., 1990] or HMMER [Eddy, 1998] it can be determined which effectors are present in each of the available genomes, and store this information in a table. These tables could be continuously updated as new genome assemblies becomes available,

or when new effectors are identified. Furthermore, the presence-absence information could be enriched with additional details, including copy number variations and allelic variation. This database could be compiled from various sources, including published data, and could also be enriched using predictions based on multiple sources of experimental evidence. To contribute to this enrichment, an initial step for identifying potential effector candidates for inclusion may involve utilizing gene models and applying common criteria such as the presence of a signal peptide for secretion and the absence of a transmembrane domain [Mitchum et al., 2013]. This prediction step can serve as a preliminary screening to screen candidates, which can then be further validated through subsequent experimental evidence. As spatial and temporal gene expression have an important role in differentiating functional roles of proteins, establishing the expression of putative effectors in specific tissues or life stages, particularly during host interaction, is valuable for accurate prediction and validation. For example, requiring the coding gene of a predicted effector to be expressed in parasitic life-stages [Mitchum et al., 2012] provides a layer of validation, ensuring the predicted effectors are relevant to parasitism. Newer techniques, such as single-cell sequencing, could further enhance confidence in these predictions in the future, specifically if the putative effector is expressed in either the dorsal or sub-ventral esophageal gland cells, where the majority of effector proteins are produced [Mitchum et al., 2012]. Similar to how the UniProt database distinguishes between predicted proteins (TrEMBL) and curated proteins (Swiss-Prot), the effector database could discriminate between predicted effectors curated annotations [The UniProt Consortium et al., 2023, O'Donovan, 2002]. An intuitive approach would be to implement a score similar to the UniProt "annotation score," calculated based on available experimental data. Putative effectors of which the prediction is only based on the presence of a signal peptide and absence of a transmembrane domain would receive lower scores, while those with functional validations at the protein level would score higher. This scoring system would enable researchers to assess the reliability of effector predictions more effectively.

Another major challenge in plant-parasitic nematode effector research, is the functional characterisation of effectors, especially at the protein level. While there are instances where effectors have been rigorously analysed for function - including confirming their interaction with proteins of their host (e.g., [Sacco et al., 2009, Wang et al., 2011, Verhoeven et al., 2023]) - we find that many effectors have not been equally explored in terms of their functionality. There is a recognized need to functionally characterize a larger set of effectors, but due to the complexity of the methods involved, analyzing a large number of putative effectors at once is not feasible. Hence, leveraging *in silico* predictions is an important initial step to narrow down the selection, and focus experimental efforts on a more manageable subset of proteins. With the recent innovations in artificial intelligence prediction tools such as the protein structure prediction tool Alphafold2 [Jumper et al., 2021], it is now possible to predict protein structures at the proteome level (**Chapter 6**). These structural predic-

tions may then be used to infer protein function using structural alignments with proteins with a known function, the prediction of interactions with host-proteins, and active site predictions. Therefore, employing artificial intelligence tools will improve the accuracy of functional predictions, which will enable a more reliable identification of potential effectors for functional screening.

In conclusion, this thesis contributes five high quality genome assemblies, and has provided insights into the diversification, and genomic organisation of nematode effector genes. These differences in effectors have mainly been characterised from three angles. First, genetic variation was studied between different populations of a potato cyst nematode (**Chapter 2, and 3**). Second, I described the genetic variation between three different cyst nematode species (**Chapter 4**). Third, the similarities in putative effectors between a cyst nematode and a root-knot nematode was described (**Chapter 6**). These findings, combined with future research, are important in enhancing our understanding of how plant-parasitic nematodes gain their virulence and ability to adapt to changes of their host. The improved knowledge will allow for a better informed and more targeted application of control strategies in the future.

Bibliography

Pierre Abad, Jérôme Gouzy, Jean-Marc Aury, Philippe Castagnone-Sereno, Etienne G. J. Danchin, Emeline Deleury, Laetitia Perfus-Barbeoch, Véronique Anthouard, François Artiguenave, Vivian C. Blok, Marie-Cécile Caillaud, Pedro M. Coutinho, Corinne Dasilva, Francesca De Luca, Florence Deau, Magali Esquibet, Timothé Flutre, Jared V. Goldstone, Noureddine Hamamouch, Tarek Hewezi, Olivier Jaillon, Claire Jubin, Paola Leonetti, Marc Magliano, Tom R. Maier, Gabriel V. Markov, Paul McVeigh, Graziano Pesole, Julie Poulain, Marc Robinson-Rechavi, Erika Sallet, Béatrice Ségurens, Delphine Steinbach, Tom Tytgat, Edgardo Ugarte, Cyril van Ghelder, Pasqua Veronico, Thomas J. Baum, Mark Blaxter, Teresa Bleve-Zacheo, Eric L. Davis, Jonathan J. Ewbank, Bruno Favery, Eric Grenier, Bernard Henrissat, John T. Jones, Vincent Laudet, Aaron G. Maule, Hadi Quesneville, Marie-Noëlle Rosso, Thomas Schiex, Geert Smant, Jean Weissenbach, and Patrick Wincker. Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. *Nature Biotechnology*, 26(8):909–915, August 2008. ISSN 1546-1696. doi: 10.1038/nbt.1482.

Haley J. Abel, David E. Larson, Allison A. Regier, Colby Chiang, Indraniel Das, Krishna L. Kanchi, Ryan M. Layer, Benjamin M. Neale, William J. Salerno, Catherine Reeves, Steven Buyske, Tara C. Matise, Donna M. Muzny, Michael C. Zody, Eric S. Lander, Susan K. Dutcher, Nathan O. Stitziel, and Ira M. Hall. Mapping and characterization of structural variation in 17,795 human genomes. *Nature*, 583(7814):83–89, July 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2371-0.

Shawkat Ali, Maxime Magne, Shiyan Chen, Natasa Obradovic, Lubna Jamshaid, Xiaohong Wang, Guy Bélair, and Peter Moffett. Analysis of globodera rostochiensis effectors reveals conserved functions of sprysec proteins in suppressing and eliciting plant immune responses. *Frontiers in plant science*, 6:623, 2015a.

Shawkat Ali, Maxime Magne, Shiyan Chen, Natasa Obradovic, Lubna Jamshaid, Xiaohong Wang, Guy Bélair, and Peter Moffett. Analysis of Globodera rostochiensis effectors reveals conserved functions of SPRYSEC proteins in suppressing and eliciting plant immune responses. *Frontiers in Plant Science*, 6, August 2015b. ISSN 1664-462X. doi: 10.3389/fpls.2015.00623.

Can Alkan, Saba Sajjadian, and Evan E Eichler. Limitations of next-generation genome sequence assembly. *Nature Methods*, 8(1):61–65, January 2011. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.1527.

Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, October 1990. ISSN 00222836. doi: 10.1016/S0022-2836(05)80360-2.

S. Andersen and K. Andersen. Suggestions for Determination and Terminology of Pathotypes and Genes for Resistance in Cyst-forming Nematodes, especially Heterodera avenae1. *EPPO Bulletin*, 12(4):379–386, 1982. ISSN 1365-2338. doi: 10.1111/j.1365-2338.1982.tb01819.x.

Dmitry Antipov, Anton Korobeynikov, Jeffrey S. McLean, and Pavel A. Pevzner. hybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics*, 32(7):1009–1015, April 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv688.

M Araki. Description of Meloidogyne ichinohei n. sp. (Nematoda: Meloigynidae) from Iris laevigata in Japan. *Japanese Journal of Nematology*, 22(11-19), 1992.

Yoichiro Arata, Yoshihiro Akimoto, Jun Hirabayashi, Ken-Ichi Kasai, and Hiroshi Hirano. An immunohistochemical study of the 32-kDa galectin (β-galactoside-binding lectin) in the nematodeCaenorhabditis elegans. *The Histochemical Journal*, 28(3):201–207, March 1996. ISSN 0018-2214, 1573-6865. doi: 10.1007/BF02331444.

Carlos L. Araya and Douglas M. Fowler. Deep mutational scanning: Assessing protein function on a massive scale. *Trends in Biotechnology*, 29(9):435–442, September 2011. ISSN 01677799. doi: 10.1016/j.tibtech.2011.04.003.

Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, August 2021. doi: 10.1126/science.abj8754.

Xiao-chen Bai, Greg McMullan, and Sjors H. W. Scheres. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40(1):49–57, January 2015. ISSN 0968-0004. doi: 10.1016/j.tibs.2014.10.005.

E. Bakker, P. Butterbach, J. Rouppe van der Voort, E. van der Vossen, J. van Vliet, J. Bakker, and A. Goverse. Genetic and physical mapping of homologues of the virus resistance gene rx1 and the cyst nematode resistance gene gpa2 in potato. *THEORETICAL AND APPLIED GENETICS*, 106(8):1524–1531, MAY 2003. ISSN 0040-5752. doi: 10.1007/s00122-003-1213-4.

Sapinder Bali, Shengwei Hu, Kelly Vining, Chuck Brown, Hassan Mojtahedi, Lei Zhang, Cynthia Gleason, and Vidyasagar Sathuvalli. Nematode Genome Announcement: Draft Genome of *Meloidogyne chitwoodi* , an Economically Important Pest of Potato in the Pacific Northwest. *Molecular Plant-Microbe Interactions®*, pages MPMI–12–20–0337, September 2021. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-12-20-0337-A.

Stacey N. Barnes, Catherine L. Wram, Melissa G. Mitchum, and Thomas J. Baum. The plant-parasitic cyst nematode effector GLAND4 is a DNA-binding protein. *Molecular Plant Pathology*, 19(10):2263–2276, 2018. ISSN 1364-3703. doi: 10.1111/mpp.12697.

A Bateman, L Coin, R Durbin, RD Finn, V Hollich, S Griffiths-Jones, A Khanna, M Marshall, S Moxon, ELL Sonnhammer, DJ Studholme, C Yeats, and SR Eddy. The pfam protein families database. *NUCLEIC ACIDS RESEARCH*, 32(SI):D138–D141, JAN 1 2004. ISSN 0305-1048. doi: 10.1093/nar/gkh121.

William C. Beckerson, Ricardo C. Rodriguez de la Vega, Fanny E. Hartmann, Marine Duhamel, Tatiana Giraud, and Michael H. Perlin. Cause and effectors: Whole-genome comparisons reveal shared but rapidly evolving effector sets among host-specific plant-castrating fungi. *MBIO*, 10(6), NOV-DEC 2019. ISSN 2150-7511. doi: 10.1128/mBio.02391-19.

Sadia Bekal, JP Craig, ME Hudson, TL Niblack, LL Domier, and KN Lambert. Genomic dna sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing. *Molecular Genetics and Genomics*, 279 (5):535–543, 2008.

Stéphane Bellafiore, Zhouxin Shen, Marie-Noelle Rosso, Pierre Abad, Patrick Shih, and Steven P. Briggs. Direct Identification of the Meloidogyne incognita Secretome Reveals Proteins with Host Cell Reprogramming Potential. *PLoS Pathogens*, 4(10):e1000192, October 2008. ISSN 1553-7374. doi: 10.1371/journal.ppat.1000192.

Paul R. J. Birch, Glenn Bryan, Brian Fenton, Eleanor M. Gilroy, Ingo Hein, John T. Jones, Ankush Prashar, Mark A. Taylor, Lesley Torrance, and Ian K. Toth. Crops that feed

the world 8: Potato: are the trends of increased global production sustainable? *FOOD SECURITY*, 4(4):477–508, DEC 2012. ISSN 1876-4517. doi: 10.1007/s12571-012-0220-1.

David McK. Bird, John T. Jones, Charles H. Opperman, Taisei Kikuchi, and Etienne G. J. Danchin. Signatures of adaptation to plant parasitism in nematode genomes. *Parasitology*, 142(S1):S71–S84, February 2015. ISSN 0031-1820, 1469-8161. doi: 10.1017/S0031182013002163.

Marten Boetzer and Walter Pirovano. SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, 15(1):211, December 2014. ISSN 1471-2105. doi: 10.1186/1471-2105-15-211.

P Boomgaard. In the shadow of rice - roots and tubers in indonesian history, 1500-1950. *AGRICULTURAL HISTORY*, 77(4):582–610, FAL 2003. ISSN 0002-1482. doi: 10.1525/ah.2003.77.4.582.

Emanuele Bosi, Beatrice Donati, Marco Galardini, Sara Brunetti, Marie-France Sagot, Pietro Lió, Pierluigi Crescenzi, Renato Fani, and Marco Fondi. MeDuSa: A multi-draft based scaffolder. *Bioinformatics*, 31(15):2443–2451, August 2015. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btv171.

Annie Christine Boucher, Benjamin Mimee, Josselin Montarry, Sylvie Bardou-Valette, Guy Belair, Peter Moffett, and Eric Grenier. Genetic diversity of the golden potato cyst nematode globodera rostochiensis and determination of the origin of populations in quebec, canada. *MOLECULAR PHYLOGENETICS AND EVOLUTION*, 69(1):75–82, OCT 2013. ISSN 1055-7903. doi: 10.1016/j.ympev.2013.05.020.

Keith R Bradnam, Joseph N Fass, Anton Alexandrov, Paul Baranay, Michael Bechner, Inanç Birol, Sébastien Boisvert, Jarrod A Chapman, Guillaume Chapuis, Rayan Chikhi, Hamidreza Chitsaz, Wen-Chi Chou, Jacques Corbeil, Cristian Del Fabbro, T Roderick Docking, Richard Durbin, Dent Earl, Scott Emrich, Pavel Fedotov, Nuno A Fonseca, Ganeshkumar Ganapathy, Richard A Gibbs, Sante Gnerre, Élénie Godzaridis, Steve Goldstein, Matthias Haimel, Giles Hall, David Haussler, Joseph B Hiatt, Isaac Y Ho, Jason Howard, Martin Hunt, Shaun D Jackman, David B Jaffe, Erich D Jarvis, Huaiyang Jiang, Sergey Kazakov, Paul J Kersey, Jacob O Kitzman, James R Knight, Sergey Koren, Tak-Wah Lam, Dominique Lavenier, François Laviolette, Yingrui Li, Zhenyu Li, Binghang Liu, Yue Liu, Ruibang Luo, Iain MacCallum, Matthew D MacManes, Nicolas Maillet, Sergey Melnikov, Delphine Naquin, Zemin Ning, Thomas D Otto, Benedict Paten, Octávio S Paulo, Adam M Phillippy, Francisco Pina-Martins, Michael Place, Dariusz Przybylski, Xiang Qin, Carson Qu, Filipe J Ribeiro, Stephen Richards, Daniel S Rokhsar, J Graham Ruby, Simone Scalabrin, Michael C Schatz, David C Schwartz, Alexey Sergushichev, Ted

Sharpe, Timothy I Shaw, Jay Shendure, Yujian Shi, Jared T Simpson, Henry Song, Fedor Tsarev, Francesco Vezzi, Riccardo Vicedomini, Bruno M Vieira, Jun Wang, Kim C Worley, Shuangye Yin, Siu-Ming Yiu, Jianying Yuan, Guojie Zhang, Hao Zhang, Shiguo Zhou, and Ian F Korf. Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*, 2(1):2047–217X–2–10, December 2013. ISSN 2047-217X. doi: 10.1186/2047-217X-2-10.

T. Breast Cancer Linkage Consortium. Cancer Risks in BRCA2 Mutation Carriers. *JNCI Journal of the National Cancer Institute*, 91(15):1310–1316, August 1999. ISSN 0027-8874, 1460-2105. doi: 10.1093/jnci/91.15.1310.

H. Brinkman, J. J. M. Goossens, and H. R. Riel. Comparative host suitability of selected crop plants toMeloidogyne chitwoodi Golden et al. 1980 andM. fallax Karssen 1996. *Anzeiger für Schädlingskunde Pflanzenschutz Umweltschutz*, 69(6):127–129, August 1996. ISSN 0340-7330, 1612-4766. doi: 10.1007/BF01904713.

Tomáš Brůna, Katharina J Hoff, Alexandre Lomsadze, Mario Stanke, and Mark Borodovsky. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1):lqaa108, March 2021. ISSN 2631-9268. doi: 10.1093/nargab/lqaa108.

Ola Brynildsrud, Snorre Gulla, Edward J. Feil, Simen Foyn Norstebo, and Linda D. Rhodes. Identifying copy number variation of the dominant virulence factors msa and p22 within genomes of the fish pathogen renibacterium salmoninarum. *MICROBIAL GENOMICS*, 2(4), APR 2016. ISSN 2057-5858. doi: 10.1099/mgen.0.000055.

D.G. Bullock. Crop rotation. https://www.tandfonline.com/doi/epdf/10.1080/07352689209382349?nee 1992.

Joshua N. Burton, Andrew Adey, Rupali P. Patwardhan, Ruolan Qiu, Jacob O. Kitzman, and Jay Shendure. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, 31(12):1119–1125, December 2013. ISSN 1546-1696. doi: 10.1038/nbt.2727.

André Busch, Etienne G. J. Danchin, and Yannick Pauchet. Functional diversification of horizontally acquired glycoside hydrolase family 45 (GH45) proteins in Phytophaga beetles. *BMC Evolutionary Biology*, 19(1):100, December 2019. ISSN 1471-2148. doi: 10.1186/s12862-019-1429-9.

Michael S. Campbell, Carson Holt, Barry Moore, and Mark Yandell. Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, 48 (1):4.11.1–4.11.39, 2014. ISSN 1934-340X. doi: 10.1002/0471250953.bi0411s48.

Jean Carpentier, Magali Esquibet, Didier Fouville, Maria J. Manzanares-Dauleux, Marie-Claire Kerlan, and Eric Grenier. The evolution of the *Gp-Rbp-1* gene in *Globodera pallida* includes multiple selective replacements. *Molecular Plant Pathology*, 13(6):546–555, August 2012. ISSN 1464-6722, 1364-3703. doi: 10.1111/j.1364-3703.2011.00769.x.

P Castagnone-Sereno. Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity*, 96(4):282–289, April 2006. ISSN 0018-067X, 1365-2540. doi: 10.1038/sj.hdy.6800794.

Philippe Castagnone-Sereno. Genetic variability of nematodes: A threat to the durability of plant resistance genes? *Euphytica*, 124(2):193–199, March 2002. ISSN 1573-5060. doi: 10.1023/A:1015682500495.

Philippe Castagnone-Sereno, Hélène Leroy, Jean-Philippe Semblat, Frédéric Leroy, Pierre Abad, and Carolien Zijlstra. Unusual and Strongly Structured Sequence Variation in a Complex Satellite DNA Family from the Nematode Meloidogyne chitwoodi. *Journal of Molecular Evolution*, 46(2):225–233, February 1998. ISSN 0022-2844. doi: 10.1007/PL00006297.

Philippe Castagnone-Sereno, Karine Mulet, Etienne G. J. Danchin, Georgios D. Koutsovoulos, Mégane Karaulic, Martine Da Rocha, Marc Bailly-Bechet, Loris Pratx, Laetitia Perfus-Barbeoch, and Pierre Abad. Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita*. *Molecular Ecology*, 28(10):2559–2572, May 2019. ISSN 0962-1083, 1365-294X. doi: 10.1111/mec.15095.

P. Castillo, N. Vovlas, A. Troccoli, G. Liébanas, J. E. Palomares Rius, and B. B. Landa. A new root-knot nematode, *Meloidogyne silvestris* n. sp. (Nematoda: Meloidogynidae), parasitizing European holly in northern Spain. *Plant Pathology*, 58(3):606–619, June 2009. ISSN 0032-0862, 1365-3059. doi: 10.1111/j.1365-3059.2008.01991.x.

P. S. G. Chain, D. V. Grafham, R. S. Fulton, M. G. FitzGerald, J. Hostetler, D. Muzny, J. Ali, B. Birren, D. C. Bruce, C. Buhay, J. R. Cole, Y. Ding, S. Dugan, D. Field, G. M. Garrity, R. Gibbs, T. Graves, C. S. Han, S. H. Harrison, S. Highlander, P. Hugenholtz, H. M. Khouri, C. D. Kodira, E. Kolker, N. C. Kyrpides, D. Lang, A. Lapidus, S. A. Malfatti, V. Markowitz, T. Metha, K. E. Nelson, J. Parkhill, S. Pitluck, X. Qin, T. D. Read, J. Schmutz, S. Sozhamannan, P. Sterk, R. L. Strausberg, G. Sutton, N. R. Thomson, J. M. Tiedje, G. Weinstock, A. Wollam, Genomic Standards Consortium Human Microbiome Project Jumpstart Consortium, and J. C. Detter. Genome Project Standards in a New Era of Sequencing. *Science*, 326(5950):236–237, October 2009. doi: 10.1126/science.1180614.

Mark J. P. Chaisson, John Huddleston, Megan Y. Dennis, Peter H. Sudmant, Maika Malig, Fereydoun Hormozdiari, Francesca Antonacci, Urvashi Surti, Richard Sandstrom, Matthew Boitano, Jane M. Landolin, John A. Stamatoyannopoulos, Michael W. Hunkapiller, Jonas Korlach, and Evan E. Eichler. Resolving the complexity of the human genome using single-molecule sequencing. *Nature*, 517(7536):608–611, January 2015. ISSN 1476-4687. doi: 10.1038/nature13907.

Agnes P. Chan, Jonathan Crabtree, Qi Zhao, Hernan Lorenzi, Joshua Orvis, Daniela Puiu, Admasu Melake-Berhan, Kristine M. Jones, Julia Redman, Grace Chen, Edgar B. Cahoon, Melaku Gedil, Mario Stanke, Brian J. Haas, Jennifer R. Wortman, Claire M. Fraser-Liggett, Jacques Ravel, and Pablo D. Rabinowicz. Draft genome sequence of the oilseed species Ricinus communis. *Nature Biotechnology*, 28(9):951–956, September 2010. ISSN 1546-1696. doi: 10.1038/nbt.1674.

Hongyu Chen, Miaoxian Guo, Shuting Dong, Xinling Wu, Guobin Zhang, Liu He, Yuannian Jiao, Shilin Chen, Li Li, and Hongmei Luo. A chromosome-scale genome assembly of Artemisia argyi reveals unbiased subgenome evolution and key contributions of gene duplication to volatile terpenoid diversity. *Plant Communications*, 4(3):100516, May 2023. ISSN 2590-3462. doi: 10.1016/j.xplc.2023.100516.

Liangbiao Chen, Arthur L. DeVries, and Chi-Hing C. Cheng. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proceedings of the National Academy of Sciences*, 94(8):3817–3822, April 1997. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.94.8.3817.

Yu Chen, Yixin Zhang, Amy Y. Wang, Min Gao, and Zechen Chong. Accurate long-read de novo assembly evaluation with Inspector. *Genome Biology*, 22(1):312, November 2021. ISSN 1474-760X. doi: 10.1186/s13059-021-02527-4.

Yangrae Cho, Robin A. Ohm, Igor V. Grigoriev, and Akhil Srivastava. Fungal-specific transcription factor abpf2 activates pathogenicity in alternaria brassicicola. *PLANT JOURNAL*, 75(3):498–514, AUG 2013. ISSN 0960-7412. doi: 10.1111/tpj.12217.

Deanna M. Church, Leo Goodstadt, LaDeana W. Hillier, Michael C. Zody, Steve Goldstein, Xinwe She, Carol J. Bult, Richa Agarwala, Joshua L. Cherry, Michael DiCuccio, Wratko Hlavina, Yuri Kapustin, Peter Meric, Donna Maglott, Zoë Birtle, Ana C. Marques, Tina Graves, Shiguo Zhou, Brian Teague, Konstantinos Potamousis, Christopher Churas, Michael Place, Jill Herschleb, Ron Runnheim, Daniel Forrest, James Amos-Landgraf, David C. Schwartz, Ze Cheng, Kerstin Lindblad-Toh, Evan E. Eichler, Chris P. Ponting, and The Mouse Genome Sequencing Consortium. Lineage-Specific Biology Revealed by a Finished Genome Assembly of the Mouse. *PLOS Biology*, 7(5):e1000112, May 2009. ISSN 1545-7885. doi: 10.1371/journal.pbio.1000112.

Martin Barry Cole, Mary Ann Augustin, Michael John Robertson, and John Michael Manners. The science of food security. *npj Science of Food*, 2(1):14, August 2018. ISSN 2396-8370. doi: 10.1038/s41538-018-0021-9.

European Commission. Pioneering proposals to restore Europe's nature by 2050. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_3746, 2022.

Donald F. Conrad, Dalila Pinto, Richard Redon, Lars Feuk, Omer Gokcumen, Yujun Zhang, Jan Aerts, T. Daniel Andrews, Chris Barnes, Peter Campbell, Tomas Fitzgerald, Min Hu, Chun Hwa Ihm, Kati Kristiansson, Daniel G. MacArthur, Jeffrey R. MacDonald, Ifejinelo Onyiah, Andy Wing Chun Pang, Sam Robson, Kathy Stirrups, Armand Valsesia, Klaudia Walter, John Wei, Chris Tyler-Smith, Nigel P. Carter, Charles Lee, Stephen W. Scherer, and Matthew E. Hurles. Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289):704–712, April 2010. ISSN 1476-4687. doi: 10.1038/nature08516.

T. Cortés, D. Tagu, J.C. Simon, A. Moya, and D. Martínez-Torres. Sex versus parthenogenesis: A transcriptomic approach of photoperiod response in the model aphid Acyrthosiphon pisum (Hemiptera: Aphididae). *Gene*, 408(1-2):146–156, January 2008. ISSN 03781119. doi: 10.1016/j.gene.2007.10.030.

James A Cotton, Catherine J Lilley, Laura M Jones, Taisei Kikuchi, Adam J Reid, Peter Thorpe, Isheng J Tsai, Helen Beasley, Vivian Blok, Peter J Cock, Sebastian Eves-van den Akker, Nancy Holroyd, Martin Hunt, Sophie Mantelin, Hardeep Naghra, Arnab Pain, Juan E Palomares-Rius, Magdalena Zarowiecki, Matthew Berriman, John T Jones, and Peter E Urwin. The genome and life-stage specific transcriptomes of Globodera pallida elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology*, 15(3): R43, 2014. ISSN 1465-6906. doi: 10.1186/gb-2014-15-3-r43.

Martine Da Rocha, Caroline Bournaud, Julie Dazenière, Peter Thorpe, Marc Bailly-Bechet, Clément Pellegrin, Arthur Péré, Priscila Grynberg, Laetitia Perfus-Barbeoch, Sebastian Eves-van den Akker, and Etienne G. J. Danchin. Genome Expression Dynamics Reveal the Parasitism Regulatory Landscape of the Root-Knot Nematode Meloidogyne incognita and a Promoter Motif Associated with Effector Genes. *Genes*, 12(5):771, May 2021. ISSN 2073-4425. doi: 10.3390/genes12050771.

Etienne G. J. Danchin, Marie-Jeanne Arguel, Amandine Campan-Fournier, Laetitia Perfus-Barbeoch, Marc Magliano, Marie-Noëlle Rosso, Martine Da Rocha, Corinne Da Silva, Nicolas Nottet, Karine Labadie, Julie Guy, François Artiguenave, and Pierre Abad. Identification of Novel Target Genes for Safer and More Specific Control of Root-Knot Nematodes from a Pan-Genome Mining. *PLoS Pathogens*, 9(10):e1003745, October 2013. ISSN 1553-7374. doi: 10.1371/journal.ppat.1003745.

Etienne G. J. Danchin, Elena A. Guzeeva, Sophie Mantelin, Adokiye Berepiki, and John T. Jones. Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode globodera pallida to feed on host-derived sucrose. *MOLECULAR BIOLOGY AND EVOLUTION*, 33(6):1571–1579, JUN 2016. ISSN 0737-4038. doi: 10.1093/molbev/msw041.

Diego Darriba, David Posada, Alexey M. Kozlov, Alexandros Stamatakis, Benoit Morel, and Tomas Flouri. Modeltest-ng: A new and scalable tool for the selection of dna and protein evolutionary models. *MOLECULAR BIOLOGY AND EVOLUTION*, 37(1):291–294, JAN 2020. ISSN 0737-4038. doi: 10.1093/molbev/msz189.

Laura J. Davies and Axel A. Elling. Resistance genes against plant-parasitic nematodes: A durable control strategy? *Nematology*, 17(3):249–263, 2015. ISSN 1388-5545, 1568-5411. doi: 10.1163/15685411-00002877.

Mehmet Dayi, Simo Sun, Yasunobu Maeda, Ryusei Tanaka, Akemi Yoshida, Isheng Jason Tsai, and Taisei Kikuchi. Nearly Complete Genome Assembly of the Pinewood Nematode Bursaphelenchus xylophilus Strain Ka4C1. *Microbiology Resource Announcements*, 9(42): e01002–20, October 2020. ISSN 2576-098X. doi: 10.1128/MRA.01002-20.

Irma Tandingan De Ley, Paul De Ley, Andy Vierstraete, Gerrit Karssen, Maurice Moens, and Jacques Vanfleteren. Phylogenetic Analyses of Meloidogyne Small Subunit rDNA. *Journal of Nematology*, 34(4):319–327, December 2002a. ISSN 0022-300X.

IT De Ley, P De Ley, A Vierstraete, G Karssen, M Moens, and J Vanfleteren. Phylogenetic analyses of meloidogyne small subunit rdna. *JOURNAL OF NEMATOLOGY*, 34(4): 319–327, DEC 2002b. ISSN 0022-300X.

Mario R. De O. Barsottini, Juliana F. De Oliveira, Douglas Adamoski, Paulo J. P. L. Teixeira, Paula F. V. Do Prado, Henrique O. Tiezzi, Mauricio L. Sforça, Alexandre Cassago, Rodrigo V. Portugal, Paulo S. L. De Oliveira, Ana C. De M. Zeri, Sandra M. G. Dias, Gonçalo A. G. Pereira, and Andre L. B. Ambrosio. Functional Diversification of Cerato-Platanins in *Moniliophthora perniciosa* as Seen by Differential Expression and Protein Function Specialization. *Molecular Plant-Microbe Interactions®*, 26(11):1281–1293, November 2013. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-05-13-0148-R.

Johan Desaeger, Catherine Wram, and Inga Zasada. New reduced-risk agricultural nematicides - rationale and review. *Journal of Nematology*, 52(1):1–16, January 2020. doi: 10.21307/jofnem-2020-091.

Stephane Deschamps, Yun Zhang, Victor Llaca, Liang Ye, Abhijit Sanyal, Matthew King, Gregory May, and Haining Lin. A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping. *NATURE COMMUNICATIONS*, 9, NOV 19 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-07271-1.

Viraj Deshpande, Eric D. K. Fung, Son Pham, and Vineet Bafna. Cerulean: A Hybrid Assembly Using High Throughput Short and Long Reads. In Aaron Darling and Jens Stoye, editors, *Algorithms in Bioinformatics*, Lecture Notes in Computer Science, pages 349–363, Berlin, Heidelberg, 2013. Springer. ISBN 978-3-642-40453-5. doi: 10.1007/978-3-642-40453-5_27.

Amalia Diaz-Granados, Andrei-José Petrescu, Aska Goverse, and Geert Smant. SPRYSEC Effectors: A Versatile Protein-Binding Platform to Disrupt Plant Innate Immunity. *Frontiers in Plant Science*, 7, October 2016. ISSN 1664-462X. doi: 10.3389/fpls.2016.01575.

Amalia Diaz-Granados, Mark G. Sterken, Hein Overmars, Roel Ariaans, Martijn Holterman, Somnath S. Pokhare, Yulin Yuan, Rikus Pomp, Anna Finkers-Tomczak, Jan Roosien, Erik Slootweg, Abdenaser Elashry, Florian M.W. Grundler, Fangming Xiao, Aska Goverse, and Geert Smant. The effector GpRbp-1 of *Globodera pallida* targets a nuclear HECT E3 ubiquitin ligase to modulate gene expression in the host. *Molecular Plant Pathology*, 21 (1):66–82, January 2020. ISSN 1464-6722, 1364-3703. doi: 10.1111/mpp.12880.

X Ding, J Shields, R Allen, and RS Hussey. Molecular cloning and characterisation of a venom allergen ag5-like cdna from meloidogyne incognita. *INTERNATIONAL JOURNAL FOR PARASITOLOGY*, 30(1):77–81, JAN 2000. ISSN 0020-7519. doi: 10.1016/S0020-7519(99)00165-4.

Xiaolei Ding, Yunfei Guo, Jianren Ye, Xiaoqin Wu, Sixi Lin, Fengmao Chen, Lihua Zhu, Lin Huang, Xiaofeng Song, Yi Zhang, Ling Dai, Xiaotong Xi, Jinsi Huang, Kai Wang, Ben Fan, and De-Wei Li. Population differentiation and epidemic tracking of *Bursaphelenchus xylophilus* in China based on chromosome-level assembly and whole-genome sequencing data. *Pest Management Science*, 78(3):1213–1226, March 2022. ISSN 1526-498X, 1526-4998. doi: 10.1002/ps.6738.

JAROSLAV DOLEŽEL and JAN BARTOŠ. Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size. *Annals of Botany*, 95(1):99–110, January 2005. ISSN 0305-7364. doi: 10.1093/aob/mci005.

G. Dubreuil, M. Magliano, E. Deleury, P. Abad, and M. N. Rosso. Transcriptome analysis of root-knot nematode functions induced in the early stages of parasitism*. *New Phytologist*, 176(2):426–436, October 2007. ISSN 0028-646X, 1469-8137. doi: 10.1111/j.1469-8137.2007.02181.x.

Olga Dudchenko, Sanjit S. Batra, Arina D. Omer, Sarah K. Nyquist, Marie Hoeger, Neva C. Durand, Muhammad S. Shamim, Ido Machol, Eric S. Lander, Aviva Presser Aiden, and Erez Lieberman Aiden. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*, 356(6333):92–95, April 2017. doi: 10.1126/science.aal3327.

S R Eddy. Profile hidden Markov models. *Bioinformatics*, 14(9):755–763, January 1998. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/14.9.755.

Robert C Edgar. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–1797, 2004.

Abdelnaser M. Elashry, Samer S. Habash, Paramasivan Vijayapalani, Nahal Brocke-Ahmadinejad, Roman Bluemel, Arun Seetharam, Heiko Schoof, and Florian M. W. Grundler. Transcriptome and parasitome analysis of beet cyst nematode heterodera schachtii. *SCIENTIFIC REPORTS*, 10(1), FEB 24 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-60186-0.

Adam C. English, William J. Salerno, Oliver A. Hampton, Claudia Gonzaga-Jauregui, Shruthi Ambreth, Deborah I. Ritter, Christine R. Beck, Caleb F. Davis, Mahmoud Dahdouli, Singer Ma, Andrew Carroll, Narayanan Veeraraghavan, Jeremy Bruestle, Becky Drees, Alex Hastie, Ernest T. Lam, Simon White, Pamela Mishra, Min Wang, Yi Han, Feng Zhang, Pawel Stankiewicz, David A. Wheeler, Jeffrey G. Reid, Donna M. Muzny, Jeffrey Rogers, Aniko Sabo, Kim C. Worley, James R. Lupski, Eric Boerwinkle, and Richard A. Gibbs. Assessing structural variation in a personal genome—towards a human reference diploid genome. *BMC Genomics*, 16(1):286, April 2015. ISSN 1471-2164. doi: 10.1186/s12864-015-1479-3.

Carolina Escobar, Marta Barcala, Javier Cabrera, and Carmen Fenoll. Overview of Root-Knot Nematodes and Giant Cells. In *Advances in Botanical Research*, volume 73, pages 1–32. Elsevier, 2015. ISBN 978-0-12-417161-9. doi: 10.1016/bs.abr.2015.01.001.

Margarida Espada, Sebastian Eves-van den Akker, Tom Maier, Paramasivan Vijayapalani, Thomas Baum, Manuel Mota, and John T. Jones. STATAWAARS: A promoter motif associated with spatial expression in the major effector-producing tissues of the plant-parasitic nematode Bursaphelenchus xylophilus. *BMC Genomics*, 19(1):553, July 2018. ISSN 1471-2164. doi: 10.1186/s12864-018-4908-2.

K. Evans and A. R. Stone. A Review of the Distribution and Biology of the Potato Cyst-Nematodes Globodera rostochiensis and G. pallida. *PANS*, 23(2):178–189, June 1977. ISSN 0309-7943. doi: 10.1080/09670877709412426.

Sebastian Eves-van den Akker, Catherine J. Lilley, John T. Jones, and Peter E. Urwin. Identification and characterisation of a hyper-variable apoplastic effector gene family of the potato cyst nematodes. *PLOS PATHOGENS*, 10(9), SEP 2014. ISSN 1553-7366. doi: 10.1371/journal.ppat.1004391.

Sebastian Eves-van den Akker, Dominik R. Laetsch, Peter Thorpe, Catherine J. Lilley, Etienne G. J. Danchin, Martine Da Rocha, Corinne Rancurel, Nancy E. Holroyd, James A.

Cotton, Amir Szitenberg, Eric Grenier, Josselin Montarry, Benjamin Mimee, Marc-Olivier Duceppe, Ian Boyes, Jessica M. C. Marvin, Laura M. Jones, Hazijah B. Yusup, Joël Lafond-Lapalme, Magali Esquibet, Michael Sabeh, Michael Rott, Hein Overmars, Anna Finkers-Tomczak, Geert Smant, Georgios Koutsovoulos, Vivian Blok, Sophie Mantelin, Peter J. A. Cock, Wendy Phillips, Bernard Henrissat, Peter E. Urwin, Mark Blaxter, and John T. Jones. The genome of the yellow potato cyst nematode, Globodera rostochiensis, reveals insights into the basis of parasitism and virulence. *Genome Biology*, 17(1):124, December 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-0985-1.

FAO Faostat. Food and agriculture organization of the united nations: Rome, 2015.

Lars Feuk, Andrew R. Carson, and Stephen W. Scherer. Structural variation in the human genome. *Nature Reviews Genetics*, 7(2):85–97, February 2006. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg1767.

Matt A. Field, Benjamin D. Rosen, Olga Dudchenko, Eva K. F. Chan, Andre E. Minoche, Richard J. Edwards, Kirston Barton, Ruth J. Lyons, Daniel Enosi Tuipulotu, Vanessa M. Hayes, Arina D. Omer, Zane Colaric, Jens Keilwagen, Ksenia Skvortsova, Ozren Bogdanovic, Martin A. Smith, Erez Lieberman Aiden, Timothy P. L. Smith, Robert A. Zammit, and J. William O. Ballard. Canfam_gsd: De novo chromosome-length genome assembly of the german shepherd dog (canis lupus familiaris) using a combination of long reads, optical mapping, and hi-c. *GIGASCIENCE*, 9(4), APR 2020. ISSN 2047-217X. doi: 10.1093/gigascience/giaa027.

Anna Finkers-Tomczak, Sarah Danan, Thijs van Dijk, Amelework Beyene, Liesbeth Bouwman, Hein Overmars, Herman van Eck, Aska Goverse, Jaap Bakker, and Erin Bakker. A high-resolution map of the grp1 locus on chromosome v of potato harbouring broad-spectrum resistance to the cyst nematode species globodera pallida and globodera rostochiensis. *THEORETICAL AND APPLIED GENETICS*, 119(1):165–173, JUN 2009. ISSN 0040-5752. doi: 10.1007/s00122-009-1026-1.

Anna M. Finkers-Tomczak. *Co-Evolution between Globodera Rostochiensis and Potato Driving Sequence Diversity of NB-LRR Resistance Loci and Nematode Suppressors of Plant Immunity*. 2011. ISBN 978-90-8585-949-9.

Wilbert G Flier, Niklaus J Grünwald, Laurens PNM Kroon, Anne K Sturbaum, Trudy BM van den Bosch, Edith Garay-Serrano, Hector Lozoya-Saldaña, William E Fry, and Lod J Turkensteen. The population structure of phytophthora infestans from the toluca valley of central mexico suggests genetic differentiation between populations from cultivated potato and wild solanum spp. *Phytopathology*, 93(4):382–390, 2003.

Rolf T Folkertsma, JNAM Rouppe Van Der Voort, Koen E De Groot, Peter M Van Zandvoort, Arjen Schots, FJ Gommers, and J Bakker. Gene pool similarities of potato cyst

nematode populations assessed by aflp analysis. *MPMI-Molecular Plant Microbe Interactions*, 9(1):47–54, 1996.

Oscar Franzén, Jon Jerlström-Hultqvist, Elsie Castro, Ellen Sherwood, Johan Ankarklev, David S. Reiner, Daniel Palm, Jan O. Andersson, Björn Andersson, and Staffan G. Svärd. Draft Genome Sequencing of Giardia intestinalis Assemblage B Isolate GS: Is Human Giardiasis Caused by Two Different Species? *PLOS Pathogens*, 5(8):e1000560, August 2009. ISSN 1553-7374. doi: 10.1371/journal.ppat.1000560.

JM Fuller and HW Howard. Breeding for resistance to the white potato cyst-nematode, heterodera pallida. *Annals of Applied Biology*, 77(2):121–128, 1974.

Shachi Gahoi and Budhayash Gautam. Genome-wide analysis of Excretory/Secretory proteins in root-knot nematode, Meloidogyne incognita provides potential targets for parasite control. *Computational Biology and Chemistry*, 67:225–233, April 2017. ISSN 14769271. doi: 10.1016/j.compbiolchem.2017.01.014.

Bingli Gao, R Allen, Tom Maier, Eric L Davis, Thomas J Baum, and Richard S Hussey. The parasitome of the phytonematode heterodera glycines. *Molecular Plant-Microbe Interactions*, 16(8):720–726, 2003.

Pier Federico Gherardini, Mark N. Wass, Manuela Helmer-Citterich, and Michael J.E. Sternberg. Convergent Evolution of Enzyme Active Sites Is not a Rare Phenomenon. *Journal of Molecular Biology*, 372(3):817–845, September 2007. ISSN 00222836. doi: 10.1016/j.jmb.2007.06.017.

Godelieve Gheysen and Melissa G. Mitchum. Phytoparasitic nematode control of plant hormone pathways. *PLANT PHYSIOLOGY*, 179(4):1212–1226, APR 2019. ISSN 0032-0889. doi: 10.1104/pp.18.01067.

Cynthia Gleason, Frederik Polzin, Samer S. Habash, Lei Zhang, Jan Utermark, Florian M. W. Grundler, and Abdelnaser Elashry. Identification of Two *Meloidogyne hapla* Genes and an Investigation of Their Roles in the Plant-Nematode Interaction. *Molecular Plant-Microbe Interactions®*, 30(2):101–112, February 2017. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-06-16-0107-R.

Sante Gnerre, Iain MacCallum, Dariusz Przybylski, Filipe J. Ribeiro, Joshua N. Burton, Bruce J. Walker, Ted Sharpe, Giles Hall, Terrance P. Shea, Sean Sykes, Aaron M. Berlin, Daniel Aird, Maura Costello, Riza Daza, Louise Williams, Robert Nicol, Andreas Gnirke, Chad Nusbaum, Eric S. Lander, and David B. Jaffe. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, 108(4):1513–1518, January 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1017351108.

Amir Mirzadi Gohari, Rahim Mehrabi, Olivier Robert, Ikbal Agah Ince, Sjef Boeren, Martin Schuster, Gero Steinberg, Pierre J. G. M. De Wit, and Gert H. J. Kema. Molecular characterization and functional analyses of ztwor1, a transcriptional regulator of the fungal wheat pathogen zymoseptoria tritici. *MOLECULAR PLANT PATHOLOGY*, 15(4):394–405, MAY 2014. ISSN 1464-6722. doi: 10.1111/mpp.12102.

A. M. Golden, J. H. O'Bannon, G. S. Santo, and A. M. Finley. Description and SEM Observations of Meloidogyne chitwoodi n. sp. (Meloidogynidae), a Root-knot Nematode on Potato in the Pacific Northwest. *Journal of Nematology*, 12(4):319–327, October 1980. ISSN 0022-300X.

Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Tim D. Williams, Shivashankar H. Nagaraj, María José Nueda, Montserrat Robles, Manuel Talón, Joaquín Dopazo, and Ana Conesa. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36(10):3420–3435, June 2008. ISSN 0305-1048. doi: 10.1093/nar/gkn176.

Aska Goverse and Geert Smant. The Activation and Suppression of Plant Innate Immunity by Parasitic Nematodes. *Annual Review of Phytopathology*, 52(1):243–265, August 2014. ISSN 0066-4286, 1545-2107. doi: 10.1146/annurev-phyto-102313-050118.

Peter Graumann and Moharned A. Marahiel. A case of convergent evolution of nucleic acid binding modules. *BioEssays*, 18(4):309–315, April 1996. ISSN 0265-9247, 1521-1878. doi: 10.1002/bies.950180409.

E Grisi, Paul R Burrows, Roland N Perry, and William M Hominick. The genome size and chromosome complement of the potato cyst nematode globodera pallida. *Fundamental and applied Nematology*, 18(1):67–70, 1995.

Annelies Haegeman, John T. Jones, and Etienne G. J. Danchin. Horizontal Gene Transfer in Nematodes: A Catalyst for Plant Parasitism? *Molecular Plant-Microbe Interactions®*, 24 (8):879–887, August 2011. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-03-11-0055.

Noureddine Hamamouch, Chunying Li, Tarek Hewezi, Thomas J. Baum, Melissa G. Mitchum, Richard S. Hussey, Lila O. Vodkin, and Eric L. Davis. 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(10): 3683–3695, June 2012. ISSN 1460-2431, 0022-0957. doi: 10.1093/jxb/ers058.

Nurul Dwi Handayani, Magali Esquibet, Josselin Montarry, Prabowo Lestari, Marjolein Couvreur, Antarjo Dikin, Johannes Helder, Eric Grenier, and Wim Bert. Distribution, dna barcoding and genetic diversity of potato cyst nematodes in indonesia. *European Journal of Plant Pathology*, 158(2):363–380, 2020.

Nurul Dwi Handayani, Prabowo Lestari, Wouter Van As, Martijn Holterman, Sven Van den Elsen, Antarjo Dikin, Wim Bert, Johannes Helder, and Joris J.M. Van Steenbrugge. Genomic reconstruction of the introduction and diversification of golden potato cyst nematode populations in Indonesia. *Phytopathology®*, pages PHYTO–04–21–0150–R, June 2021. ISSN 0031-949X, 1943-7684. doi: 10.1094/PHYTO-04-21-0150-R.

JM Hawdon, BF Jones, DR Hoffman, and PJ Hotez. Cloning and characterization of ancylostoma-secreted protein - a novel protein associated with the transition to parasitism by infective hookworm larvae. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 271 (12):6672–6678, MAR 22 1996. ISSN 0021-9258. doi: 10.1074/jbc.271.12.6672.

JG HAWKES and J FRANCISCOORTEGA. The potato in spain during the late 16th-century. *ECONOMIC BOTANY*, 46(1):86–97, JAN-MAR 1992. ISSN 0013-0001. doi: 10.1007/BF02985257.

John G Hawkes. The evolution of cultivated potatoes and their tuber-bearing wild relatives. *Die Kulturpflanze*, 36(1):189–208, 1988.

Tarek Hewezi, Peter Howe, Tom R Maier, Richard S Hussey, Melissa Goellner Mitchum, Eric L Davis, and Thomas J Baum. Cellulose binding protein from the parasitic nematode heterodera schachtii interacts with arabidopsis pectin methylesterase: cooperative cell wall modification during parasitism. *The Plant Cell*, 20(11):3080–3093, 2008.

M. Holterman, G. Karssen, S. Van Den Elsen, H. Van Megen, J. Bakker, and J. Helder. Small Subunit rDNA-Based Phylogeny of the Tylenchida Sheds Light on Relationships Among Some High-Impact Plant-Parasitic Nematodes and the Evolution of Plant Feeding. *Phytopathology®*, 99(3):227–235, March 2009. ISSN 0031-949X, 1943-7684. doi: 10.1094/PHYTO-99-3-0227.

Martijn Holterman, Andre van der Wurff, Sven van den Elsen, Hanny van Megen, Tom Bongers, Oleksandr Holovachov, Jaap Bakker, and Johannes Helder. Phylum-wide analysis of ssu rdna reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Molecular biology and evolution*, 23(9):1792–1800, 2006.

Xingrong Hou, Bang An, Qiannan Wang, Yunfeng Guo, Hongli Luo, and Chaozu He. Sge1 is involved in conidiation and pathogenicity of fusarium oxysporum f.sp cubense. *CANADIAN JOURNAL OF MICROBIOLOGY*, 64(5):349–357, MAY 2018. ISSN 0008-4166. doi: 10.1139/cjm-2017-0638.

Kevin L. Howe, Bruce J. Bolt, Myriam Shafie, Paul Kersey, and Matthew Berriman. WormBase ParaSite - a comprehensive resource for helminth genomics. *Molecular and Biochemical Parasitology*, 215:2–10, July 2017. ISSN 01666851. doi: 10.1016/j.molbiopara.2016.11.005.

Yanfeng Hu, Jia You, Chunjie Li, Fengjuan Pan, and Congli Wang. The heterodera glycines effector hg16b09 is required for nematode parasitism and suppresses plant defense response. *PLANT SCIENCE*, 289, DEC 2019. ISSN 0168-9452. doi: 10.1016/j.plantsci.2019.110271.

Guozhong Huang, Ruihua Dong, Rex Allen, Eric L. Davis, Thomas J. Baum, and Richard S. Hussey. Two chorismate mutase genes from the root-knot nematode *Meloidogyne incognita*. *Molecular Plant Pathology*, 6(1):23–30, January 2005. ISSN 1464-6722, 1364-3703. doi: 10.1111/j.1364-3703.2004.00257.x.

John Huddleston, Mark J.P. Chaisson, Karyn Meltz Steinberg, Wes Warren, Kendra Hoekzema, David Gordon, Tina A. Graves-Lindsay, Katherine M. Munson, Zev N. Kronenberg, Laura Vives, Paul Peluso, Matthew Boitano, Chen-Shin Chin, Jonas Korlach, Richard K. Wilson, and Evan E. Eichler. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Research*, 27(5):677–685, May 2017. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.214007.116.

Kristoffer Illergård, David H. Ardell, and Arne Elofsson. Structure is three to ten times more conserved than sequence-A study of structural response in protein cores. *Proteins: Structure, Function, and Bioinformatics*, 77(3):499–508, November 2009. ISSN 08873585, 10970134. doi: 10.1002/prot.22458.

S Indarti, RTP Bambang, Mulyadi, and B Triman. First record of potato cyst nematode globodera rostochiensis in indonesia. *AUSTRALASIAN PLANT PATHOLOGY*, 33(2): 325–326, 2004. ISSN 0815-3191. doi: 10.1071/AP04018.

Horia Jalily Hasani and Khaled Barakat. Homology Modeling: An Overview of Fundamentals and Tools. *International Review on Modelling and Simulations (IREMOS)*, 10(2): 129, April 2017. ISSN 1974-983X, 1974-9821. doi: 10.15866/iremos.v10i2.11412.

Pierre-Loup Jan, Cecile Gracianne, Sylvain Fournet, Eric Olivier, Jean-Francois Arnaud, Catherine Porte, Sylvie Bardou-Valette, Marie-Christine Denis, and Eric J. Petit. Temporal sampling helps unravel the genetic structure of naturally occurring populations of a phytoparasitic nematode. 1. insights from the estimation of effective population sizes. *EVOLUTIONARY APPLICATIONS*, 9(3):489–501, MAR 2016. ISSN 1752-4571. doi: 10.1111/eva.12352.

Richard Janssen, Jaap BAKKER, and Fred J Gommers. Selection of virulent and avirulent lines of globodera rostochiensis for the h, resistance gene in solanurn tuberosurn ssp. andigena cpc 1673. *Revue Nérnatol*, 13(3):265–268, 1990.

Toon Janssen, Gerrit Karssen, Olivera Topalović, Danny Coyne, and Wim Bert. Integrative taxonomy of root-knot nematodes reveals multiple independent origins of mi-

totic parthenogenesis. *PLOS ONE*, 12(3):e0172190, March 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0172190.

Lukas Jendele, Radoslav Krivak, Petr Skoda, Marian Novotny, and David Hoksza. PrankWeb: A web server for ligand binding site prediction and visualization. *Nucleic Acids Research*, 47(W1):W345–W349, July 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkz424.

WRB Jenkins et al. A rapid centrifugal-flotation technique for separating nematodes from soil. *Plant disease reporter*, 48(9), 1964.

Stewart A. Jennings, Ann-Kristin Koehler, Kathryn J. Nicklin, Chetan Deva, Steven M. Sait, and Andrew J. Challinor. Global potato yields increase under climate change with adaptation and co2 fertilisation. *FRONTIERS IN SUSTAINABLE FOOD SYSTEMS*, 4, DEC 15 2020. doi: 10.3389/fsufs.2020.519324.

Darcy A. B. Jones, Evan John, Kasia Rybak, Huyen T. T. Phan, Karam B. Singh, Shao-Yu Lin, Peter S. Solomon, Richard P. Oliver, and Kar-Chun Tan. A specific fungal transcription factor controls effector gene expression and orchestrates the establishment of the necrotrophic pathogen lifestyle on wheat. *SCIENTIFIC REPORTS*, 9, NOV 4 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-52444-7.

John T. Jones, Annelies Haegeman, Etienne G. J. Danchin, Hari S. Gaur, Johannes Helder, Michael G. K. Jones, Taisei Kikuchi, Rosa Manzanilla-López, Juan E. Palomares-Rius, Wim M. L. Wesemael, and Roland N. Perry. Top 10 plant-parasitic nematodes in molecular plant pathology: Top 10 plant-parasitic nematodes. *Molecular Plant Pathology*, 14(9): 946–961, December 2013. ISSN 14646722. doi: 10.1111/mpp.12057.

Laura M. Jones, Ann-Kristin Koehler, Mirek Trnka, Jan Balek, Andrew J. Challinor, Howard J. Atkinson, and Peter E. Urwin. Climate change is predicted to alter the current pest status of globodera pallida and g.rostochiensis in the united kingdom. *GLOBAL CHANGE BIOLOGY*, 23(11):4497–4507, NOV 2017. ISSN 1354-1013. doi: 10.1111/gcb.13676.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03819-2.

Lukas Käll, Anders Krogh, and Erik LL Sonnhammer. A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology*, 338(5):1027–1036, 2004.

Noel T. Keen and Philip A. Roberts. Plant parasitic nematodes: Digesting a page from the microbe book. *Proceedings of the National Academy of Sciences*, 95(9):4789–4790, April 1998. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.95.9.4789.

J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666, March 1958. ISSN 0028-0836. doi: 10.1038/181662a0.

Masudulla Khan and Azhar U. Khan. Plant parasitic nematodes effectors and their crosstalk with defense response of host plants: A battle underground. *RHIZOSPHERE*, 17, MAR 2021. doi: 10.1016/j.rhisph.2020.100288.

Taisei Kikuchi, Hajime Shibuya, Takuya Aikawa, and John T. Jones. Cloning and Characterization of Pectate Lyases Expressed in the Esophageal Gland of the Pine Wood Nematode Bursaphelenchus xylophilus. *Molecular Plant-Microbe Interactions®*, 19(3): 280–287, March 2006. ISSN 0894-0282. doi: 10.1094/MPMI-19-0280.

Taisei Kikuchi, Sebastian Eves-van Den Akker, and John T. Jones. Genome Evolution of Plant-Parasitic Nematodes. *Annual Review of Phytopathology*, 55(1):333–354, August 2017. ISSN 0066-4286, 1545-2107. doi: 10.1146/annurev-phyto-080516-035434.

Daehwan Kim, Joseph M. Paggi, Chanhee Park, Christopher Bennett, and Steven L. Salzberg. Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *NATURE BIOTECHNOLOGY*, 37(8):907+, AUG 2019. ISSN 1087-0156. doi: 10.1038/s41587-019-0201-4.

Sergey Koren, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. Canu: Scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. *Genome Research*, 27(5):722–736, May 2017. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.215087.116.

J. Kort, H. Ross, A. R. Stone, and H. J. Rumpenhorst. An International Scheme for Identifying and Classifying Pathotypes of Potato Cyst-Nematodes Globodera Rostochiensis and G. Pallida. *Nematologica*, 23(3), 1977.

Martin Krzywinski, Jacqueline Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. Circos: An information aesthetic for comparative genomics. *GENOME RESEARCH*, 19(9):1639–1645, SEP 2009. ISSN 1088-9051. doi: 10.1101/gr.092759.109.

John Kuriyan, T. S. R. Krishna, Lim Wong, Brian Guenther, Arno Pahler, Charles H. Williams, and Peter Model. Convergent evolution of similar function in two structurally divergent enzymes. *Nature*, 352(6331):172–174, July 1991. ISSN 0028-0836, 1476-4687. doi: 10.1038/352172a0.

Dmitry Kuznetsov, Fredrik Tegenfeldt, Mosè Manni, Mathieu Seppey, Matthew Berkeley, Evgenia V Kriventseva, and Evgeny M Zdobnov. OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*, 51(D1):D445–D451, January 2023. ISSN 0305-1048. doi: 10.1093/nar/gkac998.

Dominik R Laetsch and Mark L Blaxter. Blobtools: Interrogation of genome assemblies. *F1000Research*, 6(1287):1287, 2017.

Boqiao Lai and Jinbo Xu. Accurate protein function prediction via graph attention networks with predicted structure information. *Briefings in Bioinformatics*, 23(1):bbab502, January 2022. ISSN 1467-5463, 1477-4054. doi: 10.1093/bib/bbab502.

Roman A. Laskowski, James D. Watson, and Janet M. Thornton. Protein Function Prediction Using Local 3D Templates. *Journal of Molecular Biology*, 351(3):614–626, August 2005. ISSN 00222836. doi: 10.1016/j.jmb.2005.05.067.

Terence Neil Ledger, Stéphanie Jaubert, Nathalie Bosselut, Pierre Abad, and Marie-Noëlle Rosso. Characterization of a new β-1,4-endoglucanase gene from the root-knot nematode Meloidogyne incognita and evolutionary scheme for phytonematode family 5 glycosyl hydrolases. *Gene*, 382:121–128, November 2006. ISSN 0378-1119. doi: 10.1016/j.gene.2006.06.023.

Daehan Lee, Stefan Zdraljevic, Lewis Stevens, Ye Wang, Robyn E. Tanny, Timothy A. Crombie, Daniel E. Cook, Amy K. Webster, Rojin Chirakar, L. Ryan Baugh, Mark G. Sterken, Christian Braendle, Marie-Anne Félix, Matthew V. Rockman, and Erik C. Andersen. Balancing selection maintains hyper-divergent haplotypes in Caenorhabditis elegans. *Nature Ecology & Evolution*, 5(6):794–807, June 2021. ISSN 2397-334X. doi: 10.1038/s41559-021-01435-x.

Heng Li. A statistical framework for snp calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27 (21):2987–2993, 2011.

Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows–wheeler transform. *bioinformatics*, 25(14):1754–1760, 2009.

Junyi Li, Chunling Xu, Sihua Yang, Chun Chen, Shiqiao Tang, Jiafeng Wang, and Hui Xie. A venom allergen-like protein, rsvap, the first discovered effector protein of radopholus similis

that inhibits plant defense and facilitates parasitism. *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES*, 22(9), MAY 2021. doi: 10.3390/ijms22094782.

Zhanwen Li, Lukasz Jaroszewski, Mallika Iyer, Mayya Sedova, and Adam Godzik. FATCAT 2.0: Towards a better understanding of the structural diversity of proteins. *Nucleic Acids Research*, 48(W1):W60–W64, July 2020. ISSN 0305-1048. doi: 10.1093/nar/gkaa443.

Yun Lian, Georg Koch, Dexin Bo, Jinshe Wang, Henry T. Nguyen, Chun Li, and Weiguo Lu. The Spatial Distribution and Genetic Diversity of the Soybean Cyst Nematode, Heterodera glycines, in China: It Is Time to Take Measures to Control Soybean Cyst Nematode. *Frontiers in Plant Science*, 13:927773, June 2022. ISSN 1664-462X. doi: 10.3389/fpls.2022.927773.

Catherine J. Lilley, Abbas Maqbool, Duqing Wu, Hazijah B. Yusup, Laura M. Jones, Paul R. J. Birch, Mark J. Banfield, Peter E. Urwin, and Sebastian Eves-van den Akker. Effector gene birth in plant parasitic nematodes: Neofunctionalization of a housekeeping glutathione synthetase gene. *PLOS Genetics*, 14(4):e1007310, April 2018. ISSN 1553-7404. doi: 10.1371/journal.pgen.1007310.

Hong-Qing Ling, Shancen Zhao, Dongcheng Liu, Junyi Wang, Hua Sun, Chi Zhang, Huajie Fan, Dong Li, Lingli Dong, Yong Tao, Chuan Gao, Huilan Wu, Yiwen Li, Yan Cui, Xiaosen Guo, Shusong Zheng, Biao Wang, Kang Yu, Qinsi Liang, Wenlong Yang, Xueyuan Lou, Jie Chen, Mingji Feng, Jianbo Jian, Xiaofei Zhang, Guangbin Luo, Ying Jiang, Junjie Liu, Zhaobao Wang, Yuhui Sha, Bairu Zhang, Huajun Wu, Dingzhong Tang, Qianhua Shen, Pengya Xue, Shenhao Zou, Xiujie Wang, Xin Liu, Famin Wang, Yanping Yang, Xueli An, Zhenying Dong, Kunpu Zhang, Xiangqi Zhang, Ming-Cheng Luo, Jan Dvorak, Yiping Tong, Jian Wang, Huanming Yang, Zhensheng Li, Daowen Wang, Aimin Zhang, and Jun Wang. Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature*, 496(7443):87–90, April 2013. ISSN 1476-4687. doi: 10.1038/nature11997.

Qingli L Liu, Varghese P Thomas, and Valerie M Williamson. Meiotic Parthenogenesis in a Root-Knot Nematode Results in Rapid Genomic Homozygosity. *Genetics*, 176(3): 1483–1490, July 2007. ISSN 1943-2631. doi: 10.1534/genetics.107.071134.

Yue Liu, Xiang Qin, Xing-Zhi Henry Song, Huaiyang Jiang, Yufeng Shen, K. James Durbin, Sigbjørn Lien, Matthew Peter Kent, Marte Sodeland, Yanru Ren, Lan Zhang, Erica Sodergren, Paul Havlak, Kim C. Worley, George M. Weinstock, and Richard A. Gibbs. Bos taurus genome assembly. *BMC Genomics*, 10(1):180, April 2009. ISSN 1471-2164. doi: 10.1186/1471-2164-10-180.

Charles Lockyer. *An Account of the Trade in India: Containing Rules for Good Government in Trade, Price Courants, and Tables, with Descriptions of Fort St. George, Acheen,*

Malacca, Condore, Canton, Anjengo, Muskat, Gombroon, Surat, Goa, Carwar, Telichery, Panola, Calicut, the Cape of Good-Hope, and St. Helena: Their Inhabitants, Customs, Religion, Government, Animals, Fruits, &c.: to which is Added, An Account of the Management of the Dutch in Their Affairs in India. Number 4707. author, and sold, 1711.

Jose L. Lozano-Torres, Ruud H. P. Wilbers, Piotr Gawronski, Jordi C. Boshoven, Anna Finkers-Tomczak, Jan H. G. Cordewener, Antoine H. P. America, Hein A. Overmars, John W. Van 't Klooster, Lukasz Baranowski, Miroslaw Sobczak, Muhammad Ilyas, Renier A. L. van der Hoorn, Arjen Schots, Pierre J. G. M. de Wit, Jaap Bakker, Aska Goverse, and Geert Smant. Dual disease resistance mediated by the immune receptor cf-2 in tomato requires a common virulence target of a fungus and a nematode. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 109(25):10119–10124, JUN 19 2012. ISSN 0027-8424. doi: 10.1073/pnas.1202867109.

Jose L. Lozano-Torres, Ruud H. P. Wilbers, Sonja Warmerdam, Anna Finkers-Tomczak, Amalia Diaz-Granados, Casper C. van Schaik, Johannes Helder, Jaap Bakker, Aska Goverse, Arjen Schots, and Geert Smant. Apoplastic Venom Allergen-like Proteins of Cyst Nematodes Modulate the Activation of Basal Plant Innate Immunity by Cell Surface Receptors. *PLoS Pathogens*, 10(12):e1004569, December 2014. ISSN 1553-7374. doi: 10.1371/journal.ppat.1004569.

Shun-Wen Lu, Shiyan Chen, Jianying Wang, Hang Yu, Demosthenis Chronis, Melissa G. Mitchum, and Xiaohong Wang. Structural and functional diversity of clavata3/esr (cle)-like genes from the potato cyst nematode globodera rostochiensis. *MOLECULAR PLANT-MICROBE INTERACTIONS*, 22(9):1128–1142, SEP 2009. ISSN 0894-0282. doi: 10.1094/MPMI-22-9-1128.

Simone Luti, Luca Sella, Alessandra Quarantin, Luigia Pazzagli, and Ivan Baccelli. Twenty years of research on cerato-platanin family proteins: Clues, conclusions, and unsolved issues. *Fungal Biology Reviews*, 34(1):13–24, March 2020. ISSN 17494613. doi: 10.1016/j.fbr.2019.10.001.

Andrea MacFadden, Zoe O'Donoghue, Patricia A. G. C. Silva, Erich G. Chapman, René C. Olsthoorn, Mark G. Sterken, Gorben P. Pijlman, Peter J. Bredenbeek, and Jeffrey S. Kieft. Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. *Nature Communications*, 9(1):119, January 2018. ISSN 2041-1723. doi: 10.1038/s41467-017-02604-y.

Kristine A. Mackin, Richard A. Roy, and Douglas L. Theobald. An Empirical Test of Convergent Evolution in Rhodopsins. *Molecular Biology and Evolution*, 31(1):85–95, January 2014. ISSN 1537-1719, 0737-4038. doi: 10.1093/molbev/mst171.

Mosè Manni, Matthew R Berkeley, Mathieu Seppey, Felipe A Simão, and Evgeny M Zdobnov. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, 38(10):4647–4654, September 2021. ISSN 1537-1719. doi: 10.1093/molbev/msab199.

Daniel Mapleson, Gonzalo Garcia Accinelli, George Kettleborough, Jonathan Wright, and Bernardo J Clavijo. KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, 33(4):574–576, February 2017. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw663.

Rick Masonbrink, Tom R. Maier, Usha Muppirala, Arun S. Seetharam, Etienne Lord, Parijat S. Juvale, Jeremy Schmutz, Nathan T. Johnson, Dmitry Korkin, Melissa G. Mitchum, Benjamin Mimee, Sebastian Eves-van den Akker, Matthew Hudson, Andrew J. Severin, and Thomas J. Baum. The genome of the soybean cyst nematode (heterodera glycines) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC GENOMICS*, 20, FEB 7 2019. ISSN 1471-2164. doi: 10.1186/s12864-019-5485-8.

Rick E. Masonbrink, Tom R. Maier, Matthew Hudson, Andrew Severin, and Thomas Baum. A chromosomal assembly of the soybean cyst nematode genome. *Molecular Ecology Resources*, 21(7):2407–2422, October 2021. ISSN 1755-098X, 1755-0998. doi: 10.1111/1755-0998.13432.

Bruce A. McDonald and Celeste Linde. The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica*, 124(2):163–180, March 2002. ISSN 1573-5060. doi: 10.1023/A:1015678432355.

Yuanyuan Mei, Kathryn M. Wright, Annelies Haegeman, Lander Bauters, Amalia Diaz-Granados, Aska Goverse, Godelieve Gheysen, John T. Jones, and Sophie Mantelin. The Globodera pallida SPRYSEC Effector GpSPRY-414-2 That Suppresses Plant Defenses Targets a Regulatory Component of the Dynamic Microtubule Network. *Frontiers in Plant Science*, 9:1019, July 2018. ISSN 1664-462X. doi: 10.3389/fpls.2018.01019.

Alla Mikheenko, Andrey Prjibelski, Vladislav Saveliev, Dmitry Antipov, and Alexey Gurevich. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, 34(13): i142–i150, July 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty266.

Stephen B. Milligan, John Bodeau, Jafar Yaghoobi, Isgouhi Kaloshian, Pim Zabel, and Valerie M. Williamson. The Root Knot Nematode Resistance Gene Mi from Tomato Is a Member of the Leucine Zipper, Nucleotide Binding, Leucine-Rich Repeat Family of Plant Genes. *The Plant Cell*, 10(8):1307–1319, August 1998. ISSN 1040-4651. doi: 10.1105/tpc.10.8.1307.

Benjamin Mimee, Marc-Olivier Duceppe, Pierre-Yves Veronneau, Joel Lafond-Lapalme, Martine Jean, Francois Belzile, and Guy Belair. A new method for studying population genetics of cyst nematodes based on pool-seq and genomewide allele frequency analysis. *MOLECULAR ECOLOGY RESOURCES*, 15(6):1356–1365, NOV 2015. ISSN 1755-098X. doi: 10.1111/1755-0998.12412.

Jaina Mistry, Robert D Finn, Sean R Eddy, Alex Bateman, and Marco Punta. Challenges in homology search: Hmmer3 and convergent evolution of coiled-coil regions. *Nucleic acids research*, 41(12):e121–e121, 2013.

Melissa G. Mitchum, Xiaohong Wang, Jianying Wang, and Eric L. Davis. Role of Nematode Peptides and Other Small Molecules in Plant Parasitism. *Annual Review of Phytopathology*, 50(1):175–195, September 2012. ISSN 0066-4286, 1545-2107. doi: 10.1146/annurev-phyto-081211-173008.

Melissa G. Mitchum, Richard S. Hussey, Thomas J. Baum, Xiaohong Wang, Axel A. Elling, Martin Wubben, and Eric L. Davis. Nematode effector proteins: An emerging paradigm of parasitism. *New Phytologist*, 199(4):879–894, September 2013. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.12323.

Makedonka Mitreva-Dautova, Erwin Roze, Hein Overmars, Leo De Graaff, Arjen Schots, Johannes Helder, Aska Goverse, Jaap Bakker, and Geert Smant. A Symbiont-Independent Endo-1,4-β-Xylanase from the Plant-Parasitic Nematode *Meloidogyne incognita*. *Molecular Plant-Microbe Interactions®*, 19(5):521–529, May 2006. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-19-0521.

M. Moens, R. N. Perry, and J. T. Jones. Cyst nematodes - life cycle and economic importance. In R. N. Perry, M. Moens, and J. T. Jones, editors, *Cyst Nematodes*, pages 1–26. CAB International, UK, 1 edition, January 2018. ISBN 978-1-78639-083-7 978-1-78639-084-4. doi: 10.1079/9781786390837.0001.

H. Mojtahedi, C. R. Brown, and G. S. Santo. Characterization of Resistance in a Somatic Hybrid of Solanum bulbocastanum and S. tuberosum, to Meloidogyne chitwoodi. *Journal of Nematology*, 27(1):86–93, March 1995. ISSN 0022-300X.

José Arturo Molina-Mora, Rebeca Campos-Sánchez, César Rodríguez, Leming Shi, and Fernando García. High quality 3C de novo assembly and annotation of a multidrug resistant ST-111 Pseudomonas aeruginosa genome: Benchmark of hybrid and non-hybrid assemblers. *Scientific Reports*, 10(1):1392, January 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-58319-6.

Josselin Montarry, Sylvie Bardou-Valette, Romain Mabon, Pierre-Loup Jan, Sylvain Fournet, Eric Grenier, and Eric J. Petit. Exploring the causes of small effective population

sizes in cyst nematodes using artificial globodera pallida populations. *PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 286(1894), JAN 2 2019. ISSN 0962-8452. doi: 10.1098/rspb.2018.2359.

Nitika Mukhi, Danylo Gorenkin, and Mark J. Banfield. Exploring folds, evolution and host interactions: Understanding effector structure/function in disease and immunity. *New Phytologist*, 227(2):326–333, July 2020. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16563.

M. Shahid Mukhtar, Anne-Ruxandra Carvunis, Matija Dreze, Petra Epple, Jens Steinbrenner, Jonathan Moore, Murat Tasan, Mary Galli, Tong Hao, Marc T. Nishimura, Samuel J. Pevzner, Susan E. Donovan, Lila Ghamsari, Balaji Santhanam, Viviana Romero, Matthew M. Poulin, Fana Gebreab, Bryan J. Gutierrez, Stanley Tam, Dario Monachello, Mike Boxem, Christopher J. Harbort, Nathan McDonald, Lantian Gai, Huaming Chen, Yijian He, European Union Effectoromics Consortium, Jean Vandenhaute, Frederick P. Roth, David E. Hill, Joseph R. Ecker, Marc Vidal, Jim Beynon, Pascal Braun, and Jeffery L. Dangl. Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. *Science*, 333(6042):596–601, July 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1203659.

B. A. Mullin and B. B. Brodie. Effects of Host Resistance on Second-stage Juveniles and Adult Males of Globodera rostochiensis. *Journal of Nematology*, 20(3):335–339, July 1988. ISSN 0022-300X.

A. Nayeem. A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. *Protein Science*, 15(4):808–824, March 2006. ISSN 0961-8368, 1469-896X. doi: 10.1110/ps.051892906.

Turbevil Needham. XVI. A letter from Mr. Turbevil Needham, to the President; concerning certain chalky tubulous concretions, called malm; with some microscopical observations on the farina of the red lily, and of worms discovered in smutty corn. *Philosophical Transactions of the Royal Society of London*, 42(471):634–641, December 1743. ISSN 0261-0523, 2053-9223. doi: 10.1098/rstl.1742.0101.

TL Niblack, PR Arelli, GR Noel, CH Opperman, JH Orf, DP Schmitt, JG Shannon, and GL2620582 Tylka. A revised classification scheme for genetically diverse populations of heterodera glycines. *Journal of nematology*, 34(4):279, 2002.

J. M. Nicol, S. J. Turner, D. L. Coyne, L. den Nijs, S. Hockland, and Z. Tahna Maafi. Current Nematode Threats to World Agriculture. In John Jones, Godelieve Gheysen, and Carmen Fenoll, editors, *Genomics and Molecular Genetics of Plant-Nematode Interactions*, pages 21–43. Springer Netherlands, Dordrecht, 2011. ISBN 978-94-007-0434-3. doi: 10.1007/978-94-007-0434-3_2.

B Niere, S Krüssel, and K Osmers. Auftreten einer außergewöhnlich virulenten population der kartoffelzystennematoden. *Journal für Kulturpflanzen*, 66(12):426–427, 2014.

Jason B. Noon, Tarek Hewezi, Thomas R. Maier, Carl Simmons, Jun-Zhi Wei, Gusui Wu, Victor Llaca, Stéphane Deschamps, Eric L. Davis, Melissa G. Mitchum, Richard S. Hussey, and Thomas J. Baum. Eighteen New Candidate Effectors of the Phytonematode *Heterodera glycines* Produced Specifically in the Secretory Esophageal Gland Cells During Parasitism. *Phytopathology®*, 105(10):1362–1372, October 2015. ISSN 0031-949X, 1943-7684. doi: 10.1094/PHYTO-02-15-0049-R.

Nurjanah, Y Andi Trisyono, Siwi Indarti, and Sedyo Hartono. Identification, distribution and genetic diversity of the golden potato cyst nematode (globodera rostochiensis) in java indonesia. In *AIP Conference Proceedings*, volume 1755, page 130006. AIP Publishing LLC, 2016.

J. H. O'Bannon. Host Range of the Columbia Root-Knot Nematode. *Plant Disease*, 66(1): 1045, 1982. ISSN 0191-2917. doi: 10.1094/PD-66-1045.

C. O'Donovan. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Briefings in Bioinformatics*, 3(3):275–284, January 2002. ISSN 1467-5463, 1477-4054. doi: 10.1093/bib/3.3.275.

Magali Olivier, Anita Langer$\surd\prod$d, Patrizia Carrieri, Jonas Bergh, Sigrid Klaar, Jorunn Eyfjord, Charles Theillet, Carmen Rodriguez, Rosette Lidereau, Ivan Bi$\surd$che, Jennifer Varley, Yves Bignon, Nancy Uhrhammer, Robert Winqvist, Arja Jukkola-Vuorinen, Dieter Niederacher, Shunsuke Kato, Chikashi Ishioka, Pierre Hainaut, and Anne-Lise B$\surd\prod$rresen-Dale. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clinical Cancer Research*, 12(4):1157–1167, February 2006. ISSN 1078-0432, 1557-3265. doi: 10.1158/1078-0432.CCR-05-1029.

E. M. Onkendi and L. N. Moleleki. Distribution and genetic diversity of root-knot nematodes (M *eloidogyne* spp.) in potatoes from South Africa. *Plant Pathology*, 62(5):1184–1192, October 2013. ISSN 0032-0862, 1365-3059. doi: 10.1111/ppa.12035.

Charles H. Opperman, David M. Bird, Valerie M. Williamson, Dan S. Rokhsar, Mark Burke, Jonathan Cohn, John Cromer, Steve Diener, Jim Gajan, Steve Graham, T. D. Houfek, Qingli Liu, Therese Mitros, Jennifer Schaff, Reenah Schaffer, Elizabeth Scholl, Bryon R. Sosinski, Varghese P. Thomas, and Eric Windham. Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39):14802–14807, September 2008. ISSN 1091-6490. doi: 10.1073/pnas.0805946105.

J Paal, H Henselewski, J Muth, K Meksem, CM Menendez, F Salamini, A Ballvora, and C Gebhardt. 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. *PLANT JOURNAL*, 38(2):285–297, APR 2004. ISSN 0960-7412. doi: 10.1111/j.1365-313X.2004.02047.x.

Genis Parra, Keith Bradnam, and Ian Korf. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23(9):1061–1067, May 2007. ISSN 1367-4803. doi: 10.1093/bioinformatics/btm071.

Alexander Payne, Nadine Holmes, Vardhman Rakyan, and Matthew Loose. BulkVis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*, 35(13):2193–2198, July 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty841.

Anne-Sophie Petitot, Alexis Dereeper, Corinne Da Silva, Julie Guy, and Diana Fernandez. Analyses of the Root-Knot Nematode (Meloidogyne graminicola) Transcriptome during Host Infection Highlight Specific Gene Expression Profiling in Resistant Rice Plants. *Pathogens*, 9(8):644, August 2020. ISSN 2076-0817. doi: 10.3390/pathogens9080644.

Bastian Pfeifer, Ulrich Wittelsbürger, Sebastian E. Ramos-Onsins, and Martin J. Lercher. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology and Evolution*, 31(7):1929–1936, July 2014. ISSN 1537-1719, 0737-4038. doi: 10.1093/molbev/msu136.

MS Phillips and DL Trudgill. Variations in the ability of globodera pallida to produce females on potato clones bred from solanum vernei or s. tuberosum ssp. andigena cpc 2802. *Nematologica*, 29(2):217–226, 1983.

Olivier Plantard, Sylvie Valette, and Michael F. Gross. The Root-Knot Nematode Producing Galls on Spartina alterniflora Belongs to the Genus Meloidogyne: Rejection of Hypsoperine and Spartonema spp. *Journal of Nematology*, 39(2):127–132, June 2007. ISSN 0022-300X.

Origin Plantard, Damien Picard, S Valette, M Scurrah, Eric Grenier, and Didier Mugniery. Origin and genetic diversity of western european populations of the potato cyst nematode (globodera pallida) inferred from mitochondrial sequences and microsatellite loci. *Molecular Ecology*, 17(9):2208–2218, 2008.

Gennady Pogorelko, Jianying Wang, Parijat S. Juvale, Melissa G. Mitchum, and Thomas J. Baum. Screening soybean cyst nematode effectors for their ability to suppress plant immunity. *MOLECULAR PLANT PATHOLOGY*, 21(9):1240–1247, SEP 2020. ISSN 1464-6722. doi: 10.1111/mpp.12972.

Herman Popeijus, Hein Overmars, John Jones, Vivian Blok, Aska Goverse, Johannes Helder, Arjen Schots, Jaap Bakker, and Geert Smant. Degradation of plant cell walls by a nematode. *Nature*, 406(6791):36–37, July 2000. ISSN 0028-0836, 1476-4687. doi: 10.1038/35017641.

Wiebe J. Postma, Erik J. Slootweg, Sajid Rehman, Anna Finkers-Tomczak, Tom O. G. Tytgat, Kasper van Gelderen, Jose L. Lozano-Torres, Jan Roosien, Rikus Pomp, Casper van Schaik, Jaap Bakker, Aska Goverse, and Geert Smant. The effector sprysec-19 of globodera rostochiensis suppresses cc-nb-lrr-mediated disease resistance in plants. *PLANT PHYSIOLOGY*, 160(2):944–954, OCT 2012. ISSN 0032-0889. doi: 10.1104/pp.112.200188.

James A. Price, Danny Coyne, Vivian C. Blok, and John T. Jones. Potato cyst nematodes Globodera rostochiensis and G. pallida. *Molecular Plant Pathology*, 22(5):495–507, 2021. ISSN 1364-3703. doi: 10.1111/mpp.13047.

Ling Qin, Urszula Kudla, Erwin H. A. Roze, Aska Goverse, Herman Popeijus, Jeroen Nieuwland, Hein Overmars, John T. Jones, Arjen Schots, Geert Smant, Jaap Bakker, and Johannes Helder. A nematode expansin acting on plants. *Nature*, 427(6969):30–30, January 2004. ISSN 0028-0836, 1476-4687. doi: 10.1038/427030a.

Xue Qing and Wim Bert. Family Tylenchidae (Nematoda): An overview and perspectives. *Organisms Diversity & Evolution*, 19(3):391–408, September 2019. ISSN 1439-6092, 1618-1077. doi: 10.1007/s13127-019-00404-4.

Casper W. Quist, Geert Smant, and Johannes Helder. Evolution of Plant Parasitism in the Phylum Nematoda. *Annual Review of Phytopathology*, 53(1):289–310, August 2015. ISSN 0066-4286, 1545-2107. doi: 10.1146/annurev-phyto-080614-120057.

Dinah Qutob, Jennifer Tedman-Jones, Suomeng Dong, Kuflom Kuflu, Hai Pham, Yuanchao Wang, Daolong Dou, Shiv D Kale, Felipe D Arredondo, Brett M Tyler, et al. Copy number variation and transcriptional polymorphisms of phytophthora sojae rxlr effector genes avr1a and avr3a. *PloS one*, 4(4):e5066, 2009.

Krishan Mohan Rai, Vimal Kumar Balasubramanian, Cassie Marie Welker, Mingxiong Pang, Mei Mei Hii, and Venugopal Mendu. Genome wide comprehensive analysis and web resource development on cell wall degrading enzymes from phyto-parasitic nematodes. *BMC Plant Biology*, 15(1):187, December 2015. ISSN 1471-2229. doi: 10.1186/s12870-015-0576-4.

Sandun Rajapaksa, Arun S. Konagurthu, and Arthur M. Lesk. Sequence and structure alignments in post-AlphaFold era. *Current Opinion in Structural Biology*, 79:102539, April 2023. ISSN 0959440X. doi: 10.1016/j.sbi.2023.102539.

Sajid Rehman, Wiebe Postma, Tom Tytgat, Pjotr Prins, Ling Qin, Hein Overmars, Jack Vossen, Laurentiu-N Spiridon, Andrei-J Petrescu, Aska Goverse, et al. A secreted spry domain-containing protein (sprysec) from the plant-parasitic nematode globodera rostochiensis interacts with a cc-nb-lrr protein from a susceptible tomato. *Molecular plant-microbe interactions*, 22(3):330–340, 2009.

Arang Rhie, Brian P. Walenz, Sergey Koren, and Adam M. Phillippy. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, 21(1):245, September 2020. ISSN 1474-760X. doi: 10.1186/s13059-020-02134-9.

Anthony Rhoads and Kin Fai Au. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, 13(5):278–289, October 2015. ISSN 16720229. doi: 10.1016/j.gpb.2015.08.002.

Peter Rice. EMBOSS: The European Molecular Biology Open Software Suite.

Michael J. Roach, Simon A. Schmidt, and Anthony R. Borneman. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19(1): 460, December 2018. ISSN 1471-2105. doi: 10.1186/s12859-018-2485-7.

Jorge Roel-Touris, Alexandre M J J Bonvin, and Brian Jiménez-García. LightDock goes information-driven. *Bioinformatics*, 36(3):950–952, February 2020. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btz642.

Sourav Roy, Meenakshi Kagda, and Howard S. Judelson. Genome-wide prediction and functional validation of promoter motifs regulating gene expression in spore and infection stages of phytophthora infestans. *PLOS PATHOGENS*, 9(3), MAR 2013. ISSN 1553-7366. doi: 10.1371/journal.ppat.1003182.

Jue Ruan and Heng Li. Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, 17(2):155–158, February 2020. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-019-0669-3.

William B. Rutter, Tarek Hewezi, Tom R. Maier, Melissa G. Mitchum, Eric L. Davis, Richard S. Hussey, and Thomas J. Baum. Members of the meloidogyne avirulence protein family contain multiple plant ligand-like motifs. *PHYTOPATHOLOGY*, 104(8):879–885, AUG 2014. ISSN 0031-949X. doi: 10.1094/PHYTO-11-13-0326-R.

Melanie Ann Sacco, Kamila Koropacka, Eric Grenier, Marianne J. Jaubert, Alexandra Blanchard, Aska Goverse, Geert Smant, and Peter Moffett. The Cyst Nematode SPRYSEC Protein RBP-1 Elicits Gpa2- and RanGAP2-Dependent Plant Cell Death. *PLoS Pathogens*, 5(8):e1000564, August 2009. ISSN 1553-7374. doi: 10.1371/journal.ppat.1000564.

M I Sadowski and D T Jones. The sequence–structure relationship and protein function prediction. *Current Opinion in Structural Biology*, 19(3):357–362, June 2009. ISSN 0959440X. doi: 10.1016/j.sbi.2009.03.008.

Parthasarathy Santhanam and Bart P. H. J. Thomma. Verticillium dahliae sge1 differentially regulates expression of candidate effector genes. *MOLECULAR PLANT-MICROBE INTERACTIONS*, 26(2):249–256, FEB 2013. ISSN 0894-0282. doi: 10.1094/MPMI-08-12-0198-R.

G. S. Santo. Occurrence and Host Range of a New Root-Knot Nematode (*Meloidogyne chitwoodi*) in the Pacific Northwest. *Plant Disease*, 64(10):951, 1980. ISSN 0191-2917. doi: 10.1094/PD-64-951.

Jeremy Schmutz, Jeremy Wheeler, Jane Grimwood, Mark Dickson, Joan Yang, Chenier Caoile, Eva Bajorek, Stacey Black, Yee Man Chan, Mirian Denys, Julio Escobar, Dave Flowers, Dea Fotopulos, Carmen Garcia, Maria Gomez, Eidelyn Gonzales, Lauren Haydu, Frederick Lopez, Lucia Ramirez, James Retterer, Alex Rodriguez, Stephanie Rogers, Angelica Salazar, Ming Tsai, and Richard M. Myers. Quality assessment of the human genome sequence. *Nature*, 429(6990):365–368, May 2004. ISSN 1476-4687. doi: 10.1038/nature02390.

Esther Schnettler, Hana Tykalová, Mick Watson, Mayuri Sharma, Mark G. Sterken, Darren J. Obbard, Samuel H. Lewis, Melanie McFarlane, Lesley Bell-Sakyi, Gerald Barry, Sabine Weisheit, Sonja M. Best, Richard J. Kuhn, Gorben P. Pijlman, Margo E. Chase-Topping, Ernest A. Gould, Libor Grubhoffer, John K. Fazakerley, and Alain Kohl. Induction and suppression of tick cell antiviral RNAi responses by tick-borne flaviviruses. *Nucleic Acids Research*, 42(14):9436–9446, August 2014. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gku657.

Katrin Schöning-Stierand, Konrad Diedrich, Rainer Fährrolfes, Florian Flachsenberg, Agnes Meyder, Eva Nittinger, Ruben Steinegger, and Matthias Rarey. ProteinsPlus: Interactive analysis of protein–ligand binding interfaces. *Nucleic Acids Research*, 48(W1):W48–W53, July 2020. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkaa235.

Fritz J. Sedlazeck, Philipp Rescheneder, Moritz Smolka, Han Fang, Maria Nattestad, Arndt von Haeseler, and Michael C. Schatz. Accurate detection of complex structural variations using single-molecule sequencing. *NATURE METHODS*, 15(6):461+, JUN 2018. ISSN 1548-7091. doi: 10.1038/s41592-018-0001-7.

Mathieu Seppey, Mosè Manni, and Evgeny M. Zdobnov. BUSCO: Assessing Genome Assembly and Annotation Completeness. In Martin Kollmar, editor, *Gene Prediction: Methods and Protocols*, Methods in Molecular Biology, pages 227–245. Springer, New York, NY, 2019. ISBN 978-1-4939-9173-0. doi: 10.1007/978-1-4939-9173-0_14.

Thomas M. A. Shafee, Fung T. Lay, Thanh Kha Phan, Marilyn A. Anderson, and Mark D. Hulett. Convergent evolution of defensin sequence, structure and function. *Cellular and Molecular Life Sciences*, 74(4):663–682, February 2017. ISSN 1420-682X, 1420-9071. doi: 10.1007/s00018-016-2344-5.

Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, November 2003. ISSN 1088-9051. doi: 10.1101/gr.1239303.

Shahid Siddique, Zoran S. Radakovic, Clarissa Hiltl, Clement Pellegrin, Thomas J. Baum, Helen Beasley, Andrew F. Bent, Oliver Chitambo, Divykriti Chopra, Etienne G. J. Danchin, Eric Grenier, Samer S. Habash, M. Shamim Hasan, Johannes Helder, Tarek Hewezi, Julia Holbein, Martijn Holterman, Sławomir Janakowski, Georgios D. Koutsovoulos, Olaf P. Kranse, Jose L. Lozano-Torres, Tom R. Maier, Rick E. Masonbrink, Badou Mendy, Esther Riemer, Mirosław Sobczak, Unnati Sonawala, Mark G. Sterken, Peter Thorpe, Joris J. M. Van Steenbrugge, Nageena Zahid, Florian Grundler, and Sebastian Eves-van Den Akker. The genome and lifestage-specific transcriptomes of a plant-parasitic nematode and its host reveal susceptibility genes involved in trans-kingdom synthesis of vitamin B5. *Nature Communications*, 13(1):6190, October 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-33769-w.

Satyandra Singh, Bijendra Singh, and A. P. Singh. Nematodes: A Threat to Sustainability of Agriculture. *Procedia Environmental Sciences*, 29:215–216, January 2015. ISSN 1878-0296. doi: 10.1016/j.proenv.2015.07.270.

Geert Smant, Jack P. W. G. Stokkermans, Yitang Yan, Jan M. De Boer, Thomas J. Baum, Xiaohong Wang, Richard S. Hussey, Fred J. Gommers, Bernard Henrissat, Eric L. Davis, Johannes Helder, Arjen Schots, and Jaap Bakker. Endogenous cellulases in animals: Isolation of β-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences*, 95(9):4906–4911, April 1998. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.95.9.4906.

Geert Smant, Johannes Helder, and Aska Goverse. Parallel adaptations and common host cell responses enabling feeding of obligate and facultative plant parasitic nematodes. *PLANT JOURNAL*, 93(4):686–702, FEB 2018. ISSN 0960-7412. doi: 10.1111/tpj.13811.

Vishal Singh Somvanshi, Manoranjan Dash, Chaitra G. Bhat, Roli Budhwar, Jeffrey Godwin, Rohit N. Shukla, Andrea Patrignani, Ralph Schlapbach, and Uma Rao. An improved draft genome assembly of Meloidogyne graminicola IARI strain using long-read sequencing. *Gene*, 793:145748, August 2021. ISSN 03781119. doi: 10.1016/j.gene.2021.145748.

Aiping Song, Jiangshuo Su, Haibin Wang, Zhongren Zhang, Xingtan Zhang, Yves Van de Peer, Fei Chen, Weimin Fang, Zhiyong Guan, Fei Zhang, Zhenxing Wang, Likai Wang, Baoqing Ding, Shuang Zhao, Lian Ding, Ye Liu, Lijie Zhou, Jun He, Diwen Jia, Jiali Zhang, Chuwen Chen, Zhongyu Yu, Daojin Sun, Jiafu Jiang, Sumei Chen, and Fadi Chen. Analyses of a chromosome-scale genome assembly reveal the origin and evolution of cultivated chrysanthemum. *Nature Communications*, 14(1):2021, April 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-37730-3.

Minviluz G. Stacey, Ami Patel, William E. McClain, Melanie Mathieu, Melissa Remley, Elizabeth E. Rogers, Walter Gassmann, Dale G. Blevins, and Gary Stacey. The Arabidopsis AtOPT3 Protein Functions in Metal Homeostasis and Movement of Iron to Developing Seeds. *Plant Physiology*, 146(2):323–324, February 2008. ISSN 1532-2548. doi: 10.1104/pp.107.108183.

Alexandros Stamatakis. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *BIOINFORMATICS*, 30(9):1312–1313, MAY 1 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu033.

Mario Stanke, Oliver Keller, Irfan Gunduz, Alec Hayes, Stephan Waack, and Burkhard Morgenstern. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, 34(suppl_2):W435–W439, July 2006. ISSN 0305-1048. doi: 10.1093/nar/gkl200.

Dave T. Ste-Croix, Anne-Frederique Gendron St-Marseille, Etienne Lord, Richard R. Belanger, Jacques Brodeur, and Benjamin Mimee. Genomic profiling of virulence in the soybean cyst nematode using single-nematode sequencing. *PHYTOPATHOLOGY*, 111 (1):137–148, JAN 2021. ISSN 0031-949X. doi: 10.1094/PHYTO-08-20-0348-FI.

Ioannis Stergiopoulos and Pierre J. G. M. de Wit. Fungal effector proteins. *ANNUAL REVIEW OF PHYTOPATHOLOGY*, 47:233–263, 2009. ISSN 0066-4286. doi: 10.1146/annurev.phyto.112408.132637.

A.R. Stone and P.A. Roberts. Host Ranges of Globodera Species Within Solanum Subgenus Leptostemonum. *Nematologica*, 27(2):172–189, 1981. ISSN 0028-2596, 1875-2926. doi: 10.1163/187529281X00232.

Shona M. Strachan, Miles R. Armstrong, Amanpreet Kaur, Kathryn M. Wright, Tze Yin Lim, Katie Baker, John Jones, Glenn Bryan, Vivian Blok, and Ingo Hein. Mapping the h2 resistance effective against globodera pallida pathotype pa1 in tetraploid potato. *THEORETICAL AND APPLIED GENETICS*, 132(4):1283–1294, APR 2019. ISSN 0040-5752. doi: 10.1007/s00122-019-03278-4.

Gede Suastika et al. Identification of potato cyst nematode in indonesia by polymerase chain reaction. *Australasian Plant Disease Notes*, 7(1):133–135, 2012.

Peter H. Sudmant, Tobias Rausch, Eugene J. Gardner, Robert E. Handsaker, Alexej Abyzov, John Huddleston, Yan Zhang, Kai Ye, Goo Jun, Markus Hsi-Yang Fritz, Miriam K. Konkel, Ankit Malhotra, Adrian M. Stütz, Xinghua Shi, Francesco Paolo Casale, Jieming Chen, Fereydoun Hormozdiari, Gargi Dayama, Ken Chen, Maika Malig, Mark J. P. Chaisson, Klaudia Walter, Sascha Meiers, Seva Kashin, Erik Garrison, Adam Auton, Hugo Y. K. Lam, Xinmeng Jasmine Mu, Can Alkan, Danny Antaki, Taejeong Bae, Eliza Cerveira, Peter Chines, Zechen Chong, Laura Clarke, Elif Dal, Li Ding, Sarah Emery, Xian Fan, Madhusudan Gujral, Fatma Kahveci, Jeffrey M. Kidd, Yu Kong, Eric-Wubbo Lameijer, Shane McCarthy, Paul Flicek, Richard A. Gibbs, Gabor Marth, Christopher E. Mason, Androniki Menelaou, Donna M. Muzny, Bradley J. Nelson, Amina Noor, Nicholas F. Parrish, Matthew Pendleton, Andrew Quitadamo, Benjamin Raeder, Eric E. Schadt, Mallory Romanovitch, Andreas Schlattl, Robert Sebra, Andrey A. Shabalin, Andreas Untergasser, Jerilyn A. Walker, Min Wang, Fuli Yu, Chengsheng Zhang, Jing Zhang, Xiangqun Zheng-Bradley, Wanding Zhou, Thomas Zichner, Jonathan Sebat, Mark A. Batzer, Steven A. McCarroll, 1000 Genomes Project Consortium, Ryan E. Mills, Mark B. Gerstein, Ali Bashir, Oliver Stegle, Scott E. Devine, Charles Lee, Evan E. Eichler, and Jan O. Korbel. An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571): 75–81, October 2015. ISSN 1476-4687. doi: 10.1038/nature15394.

Firas Talas and Bruce A. McDonald. Genome-wide analysis of fusarium graminearum field populations reveals hotspots of recombination. *BMC GENOMICS*, 16, NOV 24 2015. ISSN 1471-2164. doi: 10.1186/s12864-015-2166-0.

Misghina G. Teklu, Corrie H. Schomaker, Thomas H. Been, and Leendert P.G. Molendijk. A routine test for the relative susceptibility of potato genotypes with resistance to Meloidogyne chitwoodi. *Nematology*, 18(9):1079–1094, 2016. ISSN 1388-5545, 1568-5411. doi: 10.1163/15685411-00003016.

Anastasia A Teterina, John H Willis, and Patrick C Phillips. Chromosome-Level Assembly of the Caenorhabditis remanei Genome Reveals Conserved Patterns of Nematode Genome Organization. *Genetics*, 214(4):769–780, April 2020. ISSN 1943-2631. doi: 10.1534/genetics.119.303018.

Felix Teufel, José Juan Almagro Armenteros, Alexander Rosenberg Johansen, Magnús Halldór Gíslason, Silas Irby Pihl, Konstantinos D. Tsirigos, Ole Winther, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, 40(7):1023–1025, July 2022. ISSN 1546-1696. doi: 10.1038/s41587-021-01156-3.

The C. elegans Sequencing Consortium. Genome Sequence of the Nematode *C. elegans* : A Platform for Investigating Biology. *Science*, 282(5396):2012–2018, December 1998. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.282.5396.2012.

The UniProt Consortium, Alex Bateman, Maria-Jesus Martin, Sandra Orchard, Michele Magrane, Shadab Ahmad, Emanuele Alpi, Emily H Bowler-Barnett, Ramona Britto, Hema Bye-A-Jee, Austra Cukura, Paul Denny, Tunca Dogan, ThankGod Ebenezer, Jun Fan, Penelope Garmiri, Leonardo Jose Da Costa Gonzales, Emma Hatton-Ellis, Abdulrahman Hussein, Alexandr Ignatchenko, Giuseppe Insana, Rizwan Ishtiaq, Vishal Joshi, Dushyanth Jyothi, Swaathi Kandasaamy, Antonia Lock, Aurelien Luciani, Marija Lugaric, Jie Luo, Yvonne Lussi, Alistair MacDougall, Fabio Madeira, Mahdi Mahmoudy, Alok Mishra, Katie Moulang, Andrew Nightingale, Sangya Pundir, Guoying Qi, Shriya Raj, Pedro Raposo, Daniel L Rice, Rabie Saidi, Rafael Santos, Elena Speretta, James Stephenson, Prabhat Totoo, Edward Turner, Nidhi Tyagi, Preethi Vasudev, Kate Warner, Xavier Watkins, Rossana Zaru, Hermann Zellner, Alan J Bridge, Lucila Aimo, Ghislaine Argoud-Puy, Andrea H Auchincloss, Kristian B Axelsen, Parit Bansal, Delphine Baratin, Teresa M Batista Neto, Marie-Claude Blatter, Jerven T Bolleman, Emmanuel Boutet, Lionel Breuza, Blanca Cabrera Gil, Cristina Casals-Casas, Kamal Chikh Echioukh, Elisabeth Coudert, Beatrice Cuche, Edouard De Castro, Anne Estreicher, Maria L Famiglietti, Marc Feuermann, Elisabeth Gasteiger, Pascale Gaudet, Sebastien Gehant, Vivienne Gerritsen, Arnaud Gos, Nadine Gruaz, Chantal Hulo, Nevila Hyka-Nouspikel, Florence Jungo, Arnaud Kerhornou, Philippe Le Mercier, Damien Lieberherr, Patrick Masson, Anne Morgat, Venkatesh Muthukrishnan, Salvo Paesano, Ivo Pedruzzi, Sandrine Pilbout, Lucille Pourcel, Sylvain Poux, Monica Pozzato, Manuela Pruess, Nicole Redaschi, Catherine Rivoire, Christian J A Sigrist, Karin Sonesson, Shyamala Sundaram, Cathy H Wu, Cecilia N Arighi, Leslie Arminski, Chuming Chen, Yongxing Chen, Hongzhan Huang, Kati Laiho, Peter McGarvey, Darren A Natale, Karen Ross, C R Vinayaka, Qinghua Wang, Yuqi Wang, and Jian Zhang. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1):D523–D531, January 2023. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkac1052.

Romain Thevenoux, Laurent Folcher, Magali Esquibet, Didier Fouville, Josselin Montarry, and Eric Grenier. The hidden diversity of the potato cyst nematode globodera pallida in the south of peru. *EVOLUTIONARY APPLICATIONS*, 13(4):727–737, APR 2020. ISSN 1752-4571. doi: 10.1111/eva.12896.

Peter Thorpe, Sophie Mantelin, Peter Ja Cock, Vivian C. Blok, Mirela C. Coke, Sebastian Eves-van den Akker, Elena Guzeeva, Catherine J. Lilley, Geert Smant, Adam J. Reid, Kathryn M. Wright, Peter E. Urwin, and John T. Jones. Genomic characterisation of the effector complement of the potato cyst nematode globodera pallida. *BMC GENOMICS*, 15, OCT 23 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-923.

M.S. Tigano, R.M. Carneiro, A Jeyaprakash, D.W. Dickson, and B.J. Adams. Phylogeny of Meloidogyne spp. based on 18S rDNA and the intergenic region of mitochondrial DNA sequences. *Nematology*, 7(6):851–862, 2005.

Laureano Tomás-Daza, Llorenç Rovirosa, Paula López-Martí, Andrea Nieto-Aliseda, François Serra, Ainoa Planas-Riverola, Oscar Molina, Rebecca McDonald, Cedric Ghevaert, Esther Cuatrecasas, Dolors Costa, Mireia Camós, Clara Bueno, Pablo Menéndez, Alfonso Valencia, and Biola M. Javierre. Low input capture Hi-C (liCHi-C) identifies promoter-enhancer interactions at high-resolution. *Nature Communications*, 14(1):268, January 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-35911-8.

Petri Törönen and Liisa Holm. PANNZER —A practical tool for protein function prediction. *Protein Science*, 31(1):118–128, January 2022. ISSN 0961-8368, 1469-896X. doi: 10.1002/pro.4193.

HJ Toxopeus and CA Huijsman. Genotypical background of resistance to heterodera rostochiensis in solanum tuberosum, var. andigenum. *Nature*, 170(4337):1016–1016, 1952.

A C Triantaphyllou. Environmental Sex Differentiation of Nematodes in Relation to Pest Management. *Annual Review of Phytopathology*, 11(1):441–462, September 1973. ISSN 0066-4286, 1545-2107. doi: 10.1146/annurev.py.11.090173.002301.

DL Trudgill. Potato cyst nematodes: A critical review of the current pathotyping scheme. *EPPO Bulletin*, 15(3):273–279, 1985.

Emmanuel A. Tzortzakakis, Mohamed A. M. Adam, Vivian C. Blok, Charalampos Paraskevopoulos, and Kostas Bourtzis. Occurrence of Resistance-breaking Populations of Root-knot Nematodes on Tomato in Greece. *European Journal of Plant Pathology*, 113(1): 101–105, September 2005. ISSN 0929-1873, 1573-8469. doi: 10.1007/s10658-005-1228-6.

J. G. Van der Beek, P. W. Maas, G. J. Janssen, C. Zijlstra, and C. H. Van Silfhout. A Pathotype System to Describe Intraspecific Variation in Pathogenicity of Meloidogyne chitwoodi. *Journal of Nematology*, 31(4):386–392, December 1999. ISSN 0022-300X.

JNAM Rouppe van der Voort, ELJG Van Enckevort, LP Pijnacker, J Helder, FJ Gommers, and J Bakker. Chromosome number of the potato cyst nematode globodera rostochiensis. *Fundamental and applied nematology*, 19:369–374, 1996.

JJM van Steenbrugge. Jorisvansteenbrugge/gapfiller: Gros assembly version. *Zenodo. doi: http://doi. org/10.5281/zenodo*, 4627096, 2021.

Joris J. M. van Steenbrugge, Sven Elsen, Martijn Holterman, Jose L. Lozano-Torres, Vera Putker, Peter Thorpe, Aska Goverse, Mark G. Sterken, Geert Smant, and Johannes Helder. Comparative genomics among cyst nematodes reveals distinct evolutionary histories among effector families and an irregular distribution of effector-associated promoter motifs. *Molecular Ecology*, page mec.16505, June 2022. ISSN 0962-1083, 1365-294X. doi: 10.1111/mec.16505.

Joris J.M. van Steenbrugge. Chapter 2 supplemental figure 1. *Figshare. doi: 10.6084/m9.figshare.21252273.v1*, 2022. URL `https://figshare.com/articles/figure/H2_supplemental_figure_1/21252273`.

Joris J.M. van Steenbrugge, Sven van den Elsen, Martijn Holterman, Mark G. Sterken, Peter Thorpe, Aska Goverse, Geert Smant, and Johannes Helder. Comparative genomics of two inbred lines of the potato cyst nematode Globodera rostochiensis reveals disparate effector family-specific diversification patterns. *BMC Genomics*, 22(1):611, December 2021. ISSN 1471-2164. doi: 10.1186/s12864-021-07914-6.

Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Žídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, January 2022. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkab1061.

Vasimuddin Md, Sanchit Misra, Heng Li, and Srinivas Aluru. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *arXiv:1907.12931 [cs, q-bio]*, July 2019.

Ava Verhoeven, Anna Finkers-Tomczak, Pjotr Prins, Debbie R. Valkenburg-van Raaij, Casper C. van Schaik, Hein Overmars, Joris J. M. van Steenbrugge, Wannes Tacken, Koen Varossieau, Erik J. Slootweg, Iris F. Kappers, Michaël Quentin, Aska Goverse, Mark G. Sterken, and Geert Smant. The root-knot nematode effector MiMSP32 targets host 12-oxophytodienoate reductase 2 to regulate plant susceptibility. *New Phytologist*, 237(6):2360–2374, 2023. ISSN 1469-8137. doi: 10.1111/nph.18653.

Anju Verma, Chris Lee, Stephanie Morriss, Fiona Odu, Charlotte Kenning, Nancy Rizzo, William G. Spollen, Marriam Lin, Amanda G. McRae, Scott A. Givan, Tarek Hewezi, Richard Hussey, Eric L. Davis, Thomas J. Baum, and Melissa G. Mitchum. The novel cyst nematode effector protein 30D08 targets host nuclear functions to alter gene expression in feeding sites. *New Phytologist*, 219(2):697–713, July 2018. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15179.

Paulo Vieira and Cynthia Gleason. Plant-parasitic nematode effectors — insights into their diversity and new tools for their identification. *Current Opinion in Plant Biology*, 50: 37–43, August 2019. ISSN 13695266. doi: 10.1016/j.pbi.2019.02.007.

Paramasivan Vijayapalani, Tarek Hewezi, Frederic Pontvianne, and Thomas J. Baum. An effector from the cyst nematode heterodera schachtii derepresses host rrna genes by altering histone acetylation. *PLANT CELL*, 30(11):2795–2812, NOV 2018. ISSN 1040-4651. doi: 10.1105/tpc.18.00570.

M.E. Viney. The biology and genomics of Strongyloides. *Medical Microbiology and Immunology*, 195(2):49–54, June 2006. ISSN 0300-8584, 1432-1831. doi: 10.1007/s00430-006-0013-2.

Alexander E. Vinogradov. Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. *Cytometry*, 31(2):100–109, 1998. ISSN 1097-0320. doi: 10.1002/(SICI)1097-0320(19980201)31:2⟨100::AID-CYTO5⟩3.0.CO;2-Q.

Pieter Vos, Rene Hogers, Marjo Bleeker, Martin Reijans, Theo van de Lee, Miranda Hornes, Adrie Friters, Jerina Pot, Johan Paleman, Martin Kuiper, et al. Aflp: a new technique for dna fingerprinting. *Nucleic acids research*, 23(21):4407–4414, 1995.

Bruce J. Walker, Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, Qiandong Zeng, Jennifer Wortman, Sarah K. Young, and Ashlee M. Earl. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE*, 9(11):e112963, November 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0112963.

Gaofeng Wang, Deliang Peng, Bingli Gao, Wenkun Huang, Lingan Kong, Haibo Long, Huan Peng, and Heng Jian. Comparative transcriptome analysis of two races of heterodera glycines at different developmental stages. *PLOS ONE*, 9(3), MAR 24 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0091634.

Jianying Wang, Amy Replogle, Richard Hussey, Thomas Baum, Xiaohong Wang, Eric L. Davis, and Melissa G. Mitchum. Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode Heterodera schachtii: Beet cyst nematode CLEs. *Molecular Plant Pathology*, 12(2):177–186, February 2011. ISSN 14646722. doi: 10.1111/j.1364-3703.2010.00660.x.

Jianying Wang, Andi Dhroso, Xunliang Liu, Thomas J. Baum, Richard S. Hussey, Eric L. Davis, Xiaohong Wang, Dmitry Korkin, and Melissa G. Mitchum. Phytonematode peptide effectors exploit a host post-translational trafficking mechanism to the er using a novel translocation signal. *NEW PHYTOLOGIST*, 229(1, SI):563–574, JAN 2021. ISSN 0028-646X. doi: 10.1111/nph.16765.

Jiawei Wang, Weizhen Liu, Dongzi Zhu, Po Hong, Shizhong Zhang, Shijun Xiao, Yue Tan, Xin Chen, Li Xu, Xiaojuan Zong, Lisi Zhang, Hairong Wei, Xiaohui Yuan, and Qingzhong Liu. Chromosome-scale genome assembly of sweet cherry (Prunus avium L.) cv. Tieton

obtained using long-read and Hi-C sequencing. *Horticulture Research*, 7:122, January 2020. ISSN 2052-7276. doi: 10.1038/s41438-020-00343-8.

Peng Wang and Fei Wang. A proposed metric set for evaluation of genome assembly quality. *Trends in Genetics*, 39(3):175–186, March 2023. ISSN 01689525. doi: 10.1016/j.tig.2022.10.005.

X. Wang, B. Xue, J. Dai, X. Qin, L. Liu, Y. Chi, J. T. Jones, and H. Li. A novel *Meloidogyne incognita* chorismate mutase effector suppresses plant immunity by manipulating the salicylic acid pathway and functions mainly during the early stages of nematode parasitism. *Plant Pathology*, 67(6):1436–1448, August 2018. ISSN 0032-0862, 1365-3059. doi: 10.1111/ppa.12841.

Xuan Wang, Hongmei Li, Yongjian Hu, Peng Fu, and Jianhua Xu. Molecular cloning and analysis of a new venom allergen-like protein gene from the root-knot nematode Meloidogyne incognita. *Experimental Parasitology*, 117(2):133–140, October 2007. ISSN 0014-4894. doi: 10.1016/j.exppara.2007.03.017.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion. *Nature*, July 2023. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-023-06415-8.

Wim Wesemael, Nicole Viaene, and Maurice Moens. Root-knot nematodes (Meloidogyne spp.) in Europe. *Nematology*, 13(1):3–16, 2011. ISSN 1388-5545, 1568-5411. doi: 10.1163/138855410X526831.

Ruud H. P. Wilbers, Roger Schneiter, Martijn H. M. Holterman, Claire Drurey, Geert Smant, Oluwatoyin A. Asojo, Rick M. Maizels, and Jose L. Lozano-Torres. Secreted venom allergen-like proteins of helminths: Conserved modulators of host responses in animals and plants. *PLOS Pathogens*, 14(10):e1007300, October 2018. ISSN 1553-7374. doi: 10.1371/journal.ppat.1007300.

V. M. Williamson, J. Y. Ho, F. F. Wu, N. Miller, and I. Kaloshian. A PCR-based marker tightly linked to the nematode resistance gene, Mi, in tomato. *Theoretical and Applied Genetics*, 87(7):757–763, February 1994. ISSN 0040-5752, 1432-2242. doi: 10.1007/BF00221126.

Valerie M. Williamson and Amar Kumar. Nematode resistance in plants: The battle underground. *Trends in Genetics*, 22(7):396–403, July 2006. ISSN 0168-9525. doi: 10.1016/j.tig.2006.05.003.

HW Wollenweber. Krankheiten und beschädigungen der kartoffel. arbeiten des forschungsinstitutes für kartoffelbau, heft 7, 1923.

World Meteorological Organization. Scientific assessment of ozone depletion: 2002 Global Ozone Research Monitoring Project Report #47. Technical report, 2003.

Martin J. Wubben, Lily Gavilano, Thomas J. Baum, and Eric L. Davis. Sequence and spatiotemporal expression analysis of cle-motif containing genes from the reniform nematode (rotylenchulus reniformis linford & oliveira). *JOURNAL OF NEMATOLOGY*, 47 (2):159–165, JUN 2015. ISSN 0022-300X.

Weiqi Xia, Lingyan Zheng, Jiebin Fang, Fengcheng Li, Ying Zhou, Zhenyu Zeng, Bing Zhang, Zhaorong Li, Honglin Li, and Feng Zhu. PFmulDL: A novel strategy enabling multiclass and multi-label protein function annotation by integrating diverse deep learning methods. *Computers in Biology and Medicine*, 145:105465, June 2022. ISSN 00104825. doi: 10.1016/j.compbiomed.2022.105465.

Jialian Xie, Shaojun Li, Chenmi Mo, Gaofeng Wang, Xueqiong Xiao, and Yannong Xiao. A Novel Meloidogyne incognita Effector Misp12 Suppresses Plant Defense Response at Latter Stages of Nematode Parasitism. *Frontiers in Plant Science*, 7, June 2016. ISSN 1664-462X. doi: 10.3389/fpls.2016.00964.

Haidong Yan, Min Sun, Zhongren Zhang, Yarong Jin, Ailing Zhang, Chuang Lin, Bingchao Wu, Min He, Bin Xu, Jing Wang, Peng Qin, John Pablo Mendieta, Gang Nie, Jianping Wang, Chris S. Jones, Guangyan Feng, Rakesh K. Srivastava, Xinquan Zhang, Aureliano Bombarely, Dan Luo, Long Jin, Yuanying Peng, Xiaoshan Wang, Yang Ji, Shilin Tian, and Linkai Huang. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nature Genetics*, 55(3):507–518, March 2023. ISSN 1546-1718. doi: 10.1038/s41588-023-01302-4.

Mark Yandell and Daniel Ence. A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics*, 13(5):329–342, May 2012. ISSN 1471-0064. doi: 10.1038/nrg3174.

Shanshan Yang, Yiran Dai, Yongpan Chen, Jun Yang, Dan Yang, Qian Liu, and Heng Jian. A novel g16b09-like effector from heterodera avenae suppresses plant defenses and promotes parasitism. *FRONTIERS IN PLANT SCIENCE*, 10, FEB 8 2019a. ISSN 1664-462X. doi: 10.3389/fpls.2019.00066.

Shanshan Yang, Lingling Pan, Yongpan Chen, Dan Yang, Qian Liu, and Heng Jian. Heterodera avenae gland5 effector interacts with pyruvate dehydrogenase subunit of plant to promote nematode parasitism. *FRONTIERS IN MICROBIOLOGY*, 10, JUN 4 2019b. doi: 10.3389/fmicb.2019.01241.

Yuxuan Yuan, Claire Yik-Lok Chung, and Ting-Fung Chan. Advances in optical mapping for genomic research. *Computational and Structural Biotechnology Journal*, 18:2051–2062, January 2020. ISSN 2001-0370. doi: 10.1016/j.csbj.2020.07.018.

Jonathan Yuen. Pathogens which threaten food security: Phytophthora infestans, the potato late blight pathogen. *FOOD SECURITY*, 13(2):247–253, APR 2021. ISSN 1876-4517. doi: 10.1007/s12571-021-01141-3.

Chengxin Zhang, Peter L. Freddolino, and Yang Zhang. COFACTOR: Improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic Acids Research*, 45(W1):W291–W299, July 2017. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkx366.

Lei Zhang and Cynthia Gleason. Transcriptome Analyses of Pre-parasitic and Parasitic Meloidogyne Chitwoodi Race 1 to Identify Putative Effector Genes. *Journal of Nematology*, 53:e2021–84, 2021. ISSN 0022-300X. doi: 10.21307/jofnem-2021-084.

Shu Zhao and John G. Gibbons. A population genomic characterization of copy number variation in the opportunistic fungal pathogen aspergillus fumigatus. *PLOS ONE*, 13(8), AUG 2 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0201611.

Jinshui Zheng, Donghai Peng, Ling Chen, Hualin Liu, Feng Chen, Mengci Xu, Shouyong Ju, Lifang Ruan, and Ming Sun. The Ditylenchus destructor genome provides new insights into the evolution of plant parasitic nematodes. *Proceedings of the Royal Society B: Biological Sciences*, 283(1835):20160942, July 2016. doi: 10.1098/rspb.2016.0942.

Xiuwen Zheng, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S. Weir. A high-performance computing toolset for relatedness and principal component analysis of snp data. *BIOINFORMATICS*, 28(24):3326–3328, DEC 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts606.

Kan Zhuo, Diana Naalden, Silke Nowak, Nguyen Xuan Huy, Lander Bauters, and Godelieve Gheysen. A *Meloidogyne graminicola* C-type lectin, Mg01965, is secreted into the host apoplast to suppress plant defence and promote parasitism: Meloidogyne graminicola C-type lectin effector. *Molecular Plant Pathology*, 20(3):346–355, March 2019. ISSN 14646722. doi: 10.1111/mpp.12759.

English Summary

Plant-parasitic nematodes are among the most economically damaging plant-parasites globally. From an evolutionary perspective, multiple clades of plant-parasitic nematodes exist, including the economically relevant cyst and root-knot nematodes. Nematode species belonging to these clades parasitise plants using a mixture of specialised secreted proteins, commonly referred to as effectors. These effectors manipulate the host in multiple ways, such as modification of the host cell-wall, suppression of host immunity, and hijacking cellular processes in the host to enable the establishment of a permanent feeding site. The composition of the mixture of effector proteins ultimately determines virulence. However, the effector repertoire is highly variable and may differ between different species, but also between individual nematodes of the same species. These differences can be explained by the encoding genes for effectors, which are often part of diverse gene families. In this thesis, the genetic diversity within key effector gene families is characterised by examining copy number variations and sequence variations, both inter and intra species.

Chapter 2 investigates the genetic diversification in effector gene families between two inbred lines of the potato cyst nematode *Globodera rostochiensis*. One of the primary control strategies against *G. rostochiensis* is the H1 resistance gene in potato. However, H1 resistance is not effective against all pathotypes of *G. rostochiensis*. The two *G. rostochiensis* lines studied in this chapter, Gr-Line19 and Gr-Line22, were derived from different field populations with a distinct pathotype. Gr-Line19, derived from a Ro1 pathotype, is avirulent to the H1 resistance gene, while Gr-Line22 originated from a Ro5 population that is predominantly virulent towards the same H1 resistance. It should be emphasised that although a population may exhibit a predisposition for virulence against the H1 resistance gene, genetic variability among individuals within that population can lead to varying degrees of virulence. To ensure maximum uniformity within Gr-Line19 and Gr-Line22, both

lines were initiated from a single pair of male and female parents and then inbred. As a result, Gr-Line19 exhibits complete avirulence, while Gr-Line22 displays complete virulence against potato carrying H1 resistance. To study the genetic differences between these populations, a high-quality reference genome assembly was generated for each inbred line using a combination of long-read and short-read sequencing to obtain assemblies with both a high structural and single-base accuracy. These assembled genomes were subsequently utilised to investigate the copy number variation of 19 previously reported effector gene families. Additionally, the phylogeny of four of these effector gene families was examined in greater detail. These analyses resulted in multiple copy number variants and sequence variants that were unique for either of the two inbred lines. In the future, these insights could be supplemented by resequencing data of other populations to develop pathotype-specific molecular markers.

Chapter 3 investigates the introduction of the potato cyst nematode *Globodera rostochiensis* in Indonesia. While potato cultivation in Indonesia was first documented in 1711, the species *G. rostochiensis* was not discovered there until 2003, almost 300 years later. In this chapter, a genomic reconstruction of the introduction of *G. rostochiensis* was performed. This analysis was performed based on genomic sequencing data of fourteen *G. rostochiensis* field populations collected on Java and Sumatra, and a Scottish population as an outgroup. Using the Gr-Line19 genome assembly as a reference, sequence variants were called for each population and compared, both genome-wide as well as for four effector gene families. Based on the deviation patterns in comparison with the outgroup, *G. rostochiensis* was most likely introduced on Java in the middle of the eighteenth century, followed by an introduction on Sumatra shortly after. The populations on Sumatra thereafter diversified independently from the populations on Java.

To manage the damages caused by plant-parasitic nematodes, host-resistance is for environmental and economic reasons an important control strategy. However, plant resistance genes are often only effective against specific pathotypes of a single species. Since no molecular-based pathotyping schemes are available, host resistance genes are not always used in an informed manner. To work towards molecular pathotyping schemes, it is necessary to obtain detailed knowledge of the nematode's effector repertoire, an overview of variants within effector gene families and insights into the evolutionary history of these gene families. **Chapter 4** therefore studies the genetic diversification in effector gene families between three cyst nematode species, *Globodera rostochiensis*, *Globodera pallida*, and *Heterodera schachtii*. Thereto, high-quality reference genome assemblies were generated for the *G. pallida* D383 population and the *H. schachtii* IRS population. For *G. rostochiensis*, the Gr-Line19 assembly from Chapter 2 was used. In total five effector gene families were studied in detail. This analysis revealed distinct evolutionary histories, including effector gene families that diversified after the split between the genera *Globodera* and *Heterodera*, effector gene families that diversified

after the split between the species *G. rostochiensis* and *G. pallida*, and an effector gene family that diversified before the split between the genera *Globodera* and *Heterodera*. **Chapter 4** also investigated the role of the promoter motif DOG box. It is hypothesized that this motif enhances gene expression within the dorsal gland, a secretory gland cell where effectors are produced. However only a modest correlation was found between DOG box dosage and dorsal gland gene expression in *G. rostochiensis* and *G. pallida* based on RNA sequencing data. In *H. schachtii* no correlation was found. Therefore, the DOG box appears to be highly specific, and might not be required for gene expression of effectors produced in the dorsal gland.

Chapter 5 presents a highly contiguous genome assembly of the Columbia root-knot nematode *Meloidogyne chitwoodi* and examines the genomic diversity in the species. Root-knot nematodes constitute the most impactful group of plant-parasitic nematodes. Subtle host plant penetration as well as the induction of exclusive feeding sites have contributed to the evolutionary success of this genus. In contrast to tropical RKNs, *Meloidogyne* species from temperate climate zones, such as *M. chitwoodi* were moderately well characterised. The reproduction mode of *M. chitwoodi* was previously characterised as facultative meiotic parthenogenetic. To study the impact of this reproduction mode on the genetic constitution of *M. chitwoodi*, a reference genome assembly was produced of the Dutch Mc31 population that is remarkably contiguous, achieving near chromosome-level resolution. In a comparison with three other *M. chitwoodi* populations from the United States, an incredibly low number of sequence variants were discovered, which were centred into separate polymorphic regions on the genome, which included known effector genes. These genomic characteristics suggest that, although *M. chitwoodi* is capable of sexual reproduction, it likely reproduces predominantly through asexual means. Furthermore, the effector genes that are located in the polymorphic regions provide basis for further research to study pathotype identity and host-range expansion.

Chapter 6 explores the evolutionary history of effectors between the cyst nematode *G. pallida* and the root-knot nematode *M. chitwoodi* by investigating structural similarities between secreted and putative effector proteins. The plant-parasitic abilities of cyst and root-knot nematodes have evolved independently. As a result, most effectors that are deployed by cyst and root-knot nematodes are not evolutionary related. However, effectors involved in the modification of the plant cell-wall, suppression of host immunity and influencing host processes to enable formation of a feeding site, are similar in function and have therefore convergently evolved. Traditionally, protein homology is determined through alignments of amino acid sequences. However, proteins that have evolved convergently do not share sequence similarity, necessitating an alternative approach to determine homology. In this chapter, similarities between convergently evolved proteins were therefore determined based on protein structure alignments. The protein structures of the secretomes of

G. pallida and *M. chitwoodi* were predicted using AlphaFold2, and a significant overlap in structurally similar proteins was identified. Most of these structurally similar proteins would have been overlooked by sequence alignment methods. These findings therefore provide a novel perspective into the evolutionary history of plant-parasitic nematode effectors.

The research presented in this thesis provides new insights into the genetic diversity, genomic organisation and evolutionary history of plant-parasitic nematode effectors. These findings are important in enhancing our understanding of how plant-parasitic nematodes gain their virulence and ability to adapt to changes of their host. The improved insights will allow for a better informed and more targeted application of control strategies in the future.

Acknowledgements

Doing a PhD is often portrayed as a journey of hardships, filled with many frustrating and negative experiences. After seeing all the memes online about stubborn supervisors and those dreadful failed experiments, one might argue that someone would be somewhat foolish to pursue this. Well, actually, no. While most of the stereotypical complaints are somewhat valid (except the ones about the supervisors reading this, of course), my overall experience was great. This would not have been possible without my wonderful colleagues, friends, and family.

First of all, I extend my deepest gratitude to my supervisors for their invaluable contributions to my PhD journey. To **Geert**, for his academic support and keen interest in my well-being, particularly during the critical final stages of my project. To **Hans** for his personal and friendly mentorship, his hands-on approach and positivity, which made my PhD that much more enjoyable. And to **Mark** for his incredible mentorship, down-to-earth demeanour, and for being my fellow nerd. Their combined guidance has made a big impact on both my professional and personal growth.

I also want to thank my two paranymphs, who both supported me immensely. I can't thank you both enough. **Vera**, you are one of the people that made my time at Nematology so great. I truly enjoyed the many things we did together, including our cycling rounds, goofing around, and complaining about work things. To be continued! **Marijke**, while we never actually did any work together, we are a good team and great friends. Thank you for all your support, countless talks about cats, and all the fun things we did together!

Paula (Pipa), I remember you immediately stood out during my very first day at Nematology (this will not come as a surprise if you know Paula, but I mean it in a good way). Not only did you bring cake that day, but we also had a fun chemistry that led to numerous

collaborations and a friendship I value a lot. For this, you deserve special mention. Thank you for all our fun over the years and endless support whenever I ranted about work.

Over the years, many other colleagues made a positive impact, including my fellow PhD candidates. **Robbert**, you joined Nematology as my office mate, and I liked you immediately. Thank you for all the silly talks during the coffee breaks, the chats about computer and sequencing, and the countless board game/Mario party nights (and thank you, **Jenny**!). **Dennie**, up until not that long ago, you were the only PhD besides me doing real bioinformatics (instead of writing spaghetti code R scripts like the others ;-)), and you are a great person to be around. I will always remember our nerd talks about computers and, of course, emacs (#RMS). **Nina**, you were one of the two PhDs already there when I started more than five years ago. While we never worked together, I loved having you around, and you helped me a lot by sharing experiences, having fun talks during coffee breaks and cooking risotto for us! **Jaap-Jan**, as part of the furniture at Nematology, you were the other PhD already there when I started (and likely long after). You were always one of the conversation starters at the coffee break, which made them a lot of fun, whether by making cringy jokes, calling Arno a clown or anything else. Thank you so much for all your help brainstorming ideas and support while writing my thesis. **Sara**, thank you for all the helpful talks. I really enjoyed working together! **Myrna**, we also did not really work together, but I loved having you around. You were always interested to talk about anything fun or by complaining about something annoying, haha. **Alex**, I enjoyed all the breaks and, of course, the drinks! **Joliese**, thank you for all the good times. You are a great person, and I'd wish we did some projects together.

Furthermore, I would like to thank everyone who helped me during work meetings, including **Arno**, who lives up every room he enters, **Iqbal**, who always showed interest in how I was doing (whether it was good, ok, alright, or great), and **Daan** and **Marije**, who joined the team recently. And all the other great PhDs, **Jorge**, it's too bad you spent so much time in Germany. I really enjoyed it when you were at Nema. **Alejandro**, or Señor protisto, I will never forget when Stefan made you do shots of some weird Chinese liquor on your first day. I'm glad it did not scare you away because that would have been a big loss. **Anna**, thank you for the nice years. I will surely come to you when I need any board game recommendations! **Yuhao**, I enjoyed talking about bioinformatics with you. **Mijke** & **Geartsje**, unfortunately for me, you both joined Nema during my last month, so we did not get the chance to get to know each other much. I wish you both the best of luck! A special thank you to my 'bonus' PhD colleague **Laura**. We met while you were doing a master's thesis at Nematology. Since then, we have become friends and (for a short while) colleagues. Thank you for all the fun we had, the nice coffee breaks, walks and for your support!

Nematology is a lovely place with many more great colleagues besides PhD candidates. **Jaap**, unfortunately, we did not meet often during my time at Nematology, even though

you were my promoter before retiring. As most people will know, this was usually because you were away ice skating, which I understand entirely. I want to thank you for organising the yearly barbecue in your garden, those were always great! **Jet**, a special thanks goes to you because I would not have gotten this position without you. Who would have thought that an ACT project about something as weird as the artificial colouring of tuna would lead to something as weird as a PhD in small worms? Thank you! **Lisa**, when I started, you were also working on your PhD, and since then, you kept coming back, lucky for us! To me, you have always felt like a beacon of support and positivity (Team Positievo). I want to thank you for all your support and fun times. **Jessica**, I especially enjoyed talking to you during the breaks, and I still chuckle occasionally about the South African word for chameleon that you told me. **André**, my Brazilian friend, your warm personality and friendliness will always stay with me. **Jan**, when you interviewed me as part of the job application process, I thought you were pretty stern. However, I was utterly wrong. You are super friendly, and I appreciated your unique perspectives on all the topics we discussed. Another well-deserved thank you goes to all the lab managers/technicians. Even though I never did any work in the lab, the sequencing data I used so often was all made by you, so thank you! **Casper** or Kareltje, I enjoyed the many breaks together and all the silly jokes. It was a lot of fun! **Sven**, I always enjoyed listening to your stories, even though I heard some of them more than 10 times by now. **Joost**, I used to think you were the most 'normal' person working at Nematology, but that was only until I got to know you better. You are a very nice person to be around, and I want to thank you for all the years. **Hein**, thank you for all the sequencing data and your overall goofiness. **Stefan**, when you started, I got you up to speed in bioinformatics, which was a lot of fun. You are one of the friendliest people that I know; never change! And, of course, a thank you for making my stay at Nematology better to everyone else, especially **Arjen**, **Ruud**, **Pieter**, **Stefan**, and **Jose**.

Furthermore, I would like to thank my wife **Sacha** for all her help and endless support. You were always interested in my work, even though you didn't always understand everything. I greatly appreciate you, and I am not sure I would have finished this PhD without you. My parents **Henk** & **Hannie**, brother **Tom** and sister **Eefje** also deserve a thank you for their support. I tried not to bore you too much with my talk about these strange worms, but I am thankful you cheered me on, especially towards the end of the writing phase.

Last but not least, I want to thank my best friends **Jarno**, **Stephan**, and **Remon** (and their partners **Jis**, **Stefanne**, and **Camiela**) for keeping me grounded by repeatedly telling me that doing a PhD is actually quite dumb. Still, I am glad we spent so much time together playing games, drinking beers, and especially talking about things other than work. That really helped me through the sometimes frustrating periods.

About the author

Joris Johannes Matheus van Steenbrugge was born on April 16, 1995, in Breda, The Netherlands. His academic journey in bioinformatics started in 2012 at the Hogeschool van Arnhem en Nijmegen, driven by his interests in biology and technologies. Joris developped these interests further during his internships at RadboudUMC and the Radboud Institute for Molecular Life Sciences (RIMLS), where he worked on a classification database for disease-causing genetic variants in humans, and a pipeline for enhancing automatic gene annotations.

After completing his Bachelor's degree in 2016, he pursued a Master's in Bioinformatics at Wageningen University. His master's thesis focused on identifying traits in microbial communities using time-series meta-transcriptomics data, a project he further developed during his internship at the Netherlands Institute of Ecology (NIOO-KNAW).

Joris began his PhD research at the Laboratory of Nematology at Wageningen University in 2018, where he engaged in studying the genomes of plant-parasitic nematodes and the diversification of their effector gene families. Looking ahead, Joris aims to keep contributing to the field of bioinformatics and hopes to apply his findings in practical, real-world scenarios.

The research described in this thesis was financially supported by the Netherlands Organization for Scientific Research (NWO) as part of the Applied and Engineering Sciences domain (AES) under grant no. 14708.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Bregje Jaspers | proefschriftontwerp.nl
Printed by ProefschriftMaken