Insight into the genomic architecture of a South African entomopathogenic nematode and its associated bacterial symbiont

by

Stephanie Naidoo

(320265)

Thesis

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Prof. Vincent M. Gray

August 2019

Declaration

I, Stephanie Naidoo (Person number: 320265), am a student registered for the degree of Doctor of Philosophy (PhD) in the academic year 2018.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis/Dissertation/Research Report **HAS NOT** been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature:  **Date:** 01st October 2019

Abstract

Nematoda is one of the most prolific and biologically-diverse of all animal phyla, with species adopting lifestyles ranging from free-living to parasitic. Among the diversity of parasitic nematodes lies a specialized group known as entomopathogenic nematodes that features the ability to rapidly kill insect hosts facilitated through a mutualist association with insect pathogenic bacteria. South Africa offers a range of undisturbed habitats that plays host to a great diversity of plant and insect species. Therein lies an opportunity for the recovery of novel entomopathogenic nematode species with a greater tolerance towards local environments and insects, which could potentially be exploited for biological control. In this study, a small-scale survey was conducted in the Gauteng province of South Africa in 2014, which led to the isolation of a new steinernematid, *Steinernema* sp. HBG28 and its associated bacterial symbiont, *Xenorhabdus khoisanae* strain MCB. It was found that phase I and phase II variants of *X. khoisanae* strain MCB featured the ability to swarm on solid agar. This is the first report of swarming behaviour observed in *Xenorhabdus* bacteria undergoing phenotypic variation. Genomic sequencing was performed on *X. khoisanae* strain MCB, revealing a genome of 4,6 Mbp in length with 3,869 protein-coding genes, of which, fourteen genes were specifically implicated in flagellar motility. Comparative genomic analyses with other *Xenorhabdus* spp. indicated the presence of these flagellar motility genes, with the exception of two chemotaxis genes found only in *X. khoisanae* (*cheA* and *cheY*), suggesting that these genes may be necessary in bacterial swarming during phase variation. Additionally, a draft genome assembly of *Steinernema* sp. HBG28 was produced, which was 97 Mbp in length and consisted of 35,869 predicted protein-coding genes, 1,281 tRNAs and 43 rRNAs. Further genomic characterization of *Steinernema* sp. HBG28 confirmed its mutual role

in insect pathogenicity along with its bacterial symbiont. This study was unique as it represented the first genomic characterization of a new *Steinernema* entomopathogenic nematode species and its associated bacterial symbiont in South Africa. Furthermore, this study paves the way for further research into the biological mechanisms involved in host-parasite and host-symbiont interactions.

In the loving memory of my grandmother, Dhanam Naidoo

1938 – 2006

Acknowledgement

This thesis would not have been possible without the support and encouragement of various funding bodies and individuals. I take great pleasure in acknowledging all those whom have contributed in various ways in the success and completion of this study.

First and foremost, I would like to humbly give thanks to **God Almighty** for granting me every opportunity in this life and being with me every step of the way – through the good and the bad. Without His blessings and favour, this thesis would not have been possible.

I would like to thank the following funding bodies for making this research possible:

- **Financial Aid:**
    - The **National Research Foundation (NRF) Innovation Scholarship** (2016-2017)
    - The **NRF Free-standing Scholarship** (2014-2015)
    - The **University of the Witwatersrand Postgraduate Merit Award** (2015-2016)
    - The **Gauteng Department of Agricultural and Rural Development** Bursary (2014)

I would like to pay tribute to my esteemed supervisor, **Prof. Vincent Myles Gray** for accepting me as a PhD student. This work would not have been seen to completion without his continual support, thoughtful guidance and warm encouragement

throughout various stages of this degree. Under his supervision, I was able to overcome several challenges along the way due to his valuable suggestions, constructive criticism and his extensive discussions around my work. His stimulating and often funny stories are always a pleasure to hear and his open-door policy, regardless of the nature of the visit, has been something that I have always appreciated. Moreover, not only did he serve the role as a supervisor, but he is someone for whom I have the greatest respect and admiration. I believe that the knowledge and life lessons he has imparted has made me into a better individual.

I also owe a great deal of appreciation and gratitude to **Dr. Jonathan Featherston** from the Agricultural and Research Council (ARC) in Onderstepoort, Pretoria for his assistance during the genomic component of this project and for making time to answer all my questions, no matter how trivial they were. His insight into the genome sequencing, assembly and annotation was nothing short of invaluable to this study. Moreover, our shared fondness of 'Star Wars', along with several other common interests, fostered an unlikely kinship built on mutual respect, trust and shared experiences, and for this, I am grateful. I would also be remised if I did not include an apt phrase that he was often fond of using, which has stuck with me throughout the course of this PhD degree, i.e. that '*Science is a cruel mistress*'. I think it perfectly sums up the highs and lows of any scientific study.

I gratefully acknowledge **Dr. Deran Reddy** from the Microscopy and Microanalysis Unit (MMU) for his much valued assistance and advice regarding the application of light microscopy in this study.

The one person that I am exceedingly indebted to, above all else, and without whom, this degree would not have been possible is my partner of thirteen years, **Fabian Franco-Roldan.** Every success I have achieved in my adult life can be attributed to all the sacrifices he has made over the years – both personal and financial, which has enabled me to pursue a tertiary education. Not only am I incredibly grateful for his immense generosity and unwavering faith in me, but I am also thankful for all his tireless efforts – from waking up at the crack of dawn every morning, without fail during my undergraduate degree; to making sure that I was on time for all my early-morning lectures; to ensuring that I lacked for nothing. Not only did he play a supportive role throughout my academic career, but he was someone I was accountable to. It was this very support structure that was conducive towards the completion of this PhD thesis, and for this, I am eternally grateful.

I would like to extend my gratitude to all my teachers during my schooling career, particularly **Mr. K. Govindsamy**, who was one of the best teachers I have ever had in Mathematics. Additionally, I would like to thank **all my lecturers at the University of Witwatersrand**, all of whom have provided me with the necessary skillset in undertaking a PhD degree.

I would like to pay tribute to **Mr. Ivan van Draanen**, an individual whom I briefly worked with many years ago and was instrumental in convincing me of the importance of having a qualification, thus prompting my decision to pursue a tertiary education. I shall be forever grateful for the wise advice he bestowed upon me.

It may be somewhat untoward to include furry companions of the four-legged variety in an acknowledgement section; however, I compelled to do so. I must

acknowledge my cat, **Hummer** who has been my constant companion while writing this PhD thesis. I would also like to honour my late cat, **Sam**, who, in all honesty, got me through my undergraduate degree, by keeping me company during most late nights while studying for exams or completing lab reports.

I would like to acknowledge my dear friend, **Mari Davis** from New Hampshire for lending moral support and encouragement from afar. I am especially grateful for her unwavering belief in me to see this PhD through and just being there when I have needed her the most.

Research Outputs

Original Publications

1. Naidoo, S., Featherston, J. and Gray, V. M. (2015). Draft whole-genome sequence and annotation of the entomopathogenic bacterium *Xenorhabdus khoisanae* strain MCB. *Genome Announcements*. **3:**e00872-15.
2. Naidoo, S., Mothupi, B., Featherston, J., Mpangase, P. T. and Gray, V. M. (2015). Draft genome sequence and assembly of *Photorhabdus heterorhabditis* strain VMG, a bacterial symbiont associated with the entomopathogenic nematode *Heterorhabditis zealandica*. *Genome Announcements*, **3:**e01279-15.

Publications in preparation for journal submission

1. Genome sequence of *Steinernema* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from the grasslands of South Africa. Manuscript in preparation.
2. Genomic characterization of *Xenorhabdus khoisanae* strain MCB focusing on bacterial swarming of phase II variants. Manuscript in preparation.
3. Genomic sequence of *Photorhabdus heterorhabditis* associated to *Heterorhabditis zealandica* entomopathogenic nematodes in South Africa. Manuscript in preparation

Poster Presentations

1. Naidoo, S. and Gray, V. (2014, June). The isolation, molecular identification and foraging behaviour characterization of indigenous entomopathogenic nematodes. Poster presented at the Gauteng Department of Agriculture and

Rural Development (GDARD) 7th Annual Research Symposium, Pretoria, Gauteng.

2. Mothupi, B, Naidoo, S and Gray, V. (2015, June). Isolation and molecular characterization of entomopathogenic nematode species in the Gauteng region. Poster presented at the Gauteng Department of Agriculture and Rural Development GDARD) 8th Annual Research Symposium, Midrand, Gauteng.

Awards

1. Biotech Fundi and Innovation Best Postgraduate Award (2015)
2. Biotech Fundi and Innovation Special Student Recognition Award (2015)

Table of Contents

List of Abbreviations

Abbreviation	Description
BC	before Christ
bp	base pairs
CDS	Coding sequences
CEGMA	Core eukaryotic genes mapping approach
cm	Centimetre
DNA	Deoxyribonucleic acid
EPN(s)	Entomopathogenic nematode(s)
g	Gram
G+C	Guanine +cytosine content
GPCR(s)	G-protein-coupled receptor(s)
Hcp	Haemolysin co-regulated protein
HGT	Horizontal gene transfer
hr(s)	Hour(s)
IJ	Infective juvenile(s)
IGF	Insulin-like growth factor
IPM	Integrated pest management
ITS	Internal transcribed spacer
J1-4	Juvenile stage 1-4
kPA	Kilo Pascal
L	Litre
LCB	Locally collinear block
LINEs	Long interspersed nuclear elements
LTRs	Long terminal repeats

m	meter
Mbp	Mega base pairs
Min	Minute
µl	Microlitre
ml	Millilitre
ML	Maximum Likelihood
MP	Maximum Parsimony
MP	Mate-pair
N50	The size of a contig (or scaffold) in an assembly such that 50% of the actual genome size (if known) is in contigs of that size or larger.
NB	Nutrient broth
NBTA	nutrient broth bromothymol-blue triphenyltetrazolium chloride Agar
NGS	Next-generation sequencing
NCBI	National Centre for Biotechnology Information
ORF	Open reading frame
PE	Paired-end
PCR	Polymerase chain reaction
PGAAP	Prokaryotic Genome Automatic Annotation Pipeline
RAST	Rapid Annotations using Subsystems Technology.
rDNA	Ribosomal deoxyribonucleic acid
rhs	Rearrangement hot spot
rRNA	Ribosomal ribonucleic acid
SINEs	Short interspersed nuclear elements
sp.	Species (singular)

spp.	Species (plural)
SSRs	Simple sequence repeats
TGF-β	Transforming growth factor beta
UN	United Nations
WGS	Whole genome sequencing

List of Figures

List of Tables

Chapter 1

Literature review: An introduction to entomopathogenic nematodes and their symbiotic bacteria

---

1. Introduction

One of the major challenges faced today at a global scale is food security, exacerbated not only by the increase in human population, but also climate change. By 2050, it is estimated that the world's population will reach a projected figure of 9.8 billion, according to the United Nations (UN) (2015). As a result, agricultural production is anticipated to drastically increase to meet the food demand of a rapidly expanding population.

One approach in safeguarding food security is through the improvement of pest management. Insect pests account for approximately 15% of crop losses worldwide, leading to significant economic losses in agriculture (Maxmen, 2013). Insects also serve as carriers of plant pathogens, which in turn can cause plant disease and result in additional yield loss. Throughout history, various chemicals have been used in efforts to control agricultural pests. The earliest record of insect control in agriculture dates to 2500 BC when Sumerians exploited sulphur compounds. The 19th century saw the use of compounds such as arsenic and Bordeaux mixture, a mixture of copper sulphate and lime, which were effective against most insects (Pretty and Bharucha,

2015). The early 20th century marked the extensive use of dangerous chemicals such as arsenic, mercury and cyanide (Pretty and Bharucha, 2015). However, the true era of chemical control did not commence until 1939, with the advent of the synthetic chemical insecticide, dichlorodiphenyltrichloroethane, more commonly referred to as DDT (Abrol and Shankar, 2012). Following the discovery of DDT, numerous chemical insecticides were later developed.

While there is little doubt that chemical pesticides have played a pivotal role in conventional agriculture, its misuse in crop management has had a serious impact on the environment, human health, as well as insect pests themselves. As some synthetic pesticides are non-biodegradable, they can persist in the environment for extensive periods of time, often contaminating soil, ground water and other vegetation. As such, surrounding wildlife such as birds, fish and beneficial insects are usually affected (Aktar et al., 2009). In addition, long-term exposure to synthetic pesticides, even at low dosages, can affect human health, leading to immune suppression, hormone disruption, neurological conditions, reproductive disorders as well as cancer (Brouwer et al., 1999; Crisp et al., 1998; Sheiner et al., 2003; Hurley et al., 1998).

Over the last few decades, overuse of chemical pesticides has resulted in the emergence of insect resistance in many insect populations. First observed with the use of DDT in the 1940s (Denholm and Devine, 2013), insect resistance has become increasingly common in agriculture. There are over 500 insect species worldwide that have acquired resistance against one or more extensively used insecticide classes (organochlorines, organophosphates, carbonate and pyrethroids) (Hajek, 2004; Denholm and Devine, 2013).

Exploitation of synthetic pesticides for controlling insect pests presents many new and pre-existing challenges. The ease at which once localized insect species can now transfer to different regions, along with shifts in pest and pathogen ranges due to weather pattern and climate changes, has resulted in the emergence of new invasive pests and diseases (Pretty and Bharucha, 2015). Alternative pest control practices, which reduce pest damage, whilst diminishing the harmful effects associated with synthetic pesticide application, are urgently required in insect pest management.

1.1 Biological control

Biological control is an environmentally sound and sustainable approach that reduces or mitigates crop damage, whilst maintaining both quantity and quality, through the exploitation of living organisms and/or their products to control insect populations (Bale et al., 2008). The application of biological control agents includes the use of natural predators of insect pests, parasitoids, parasites, as well as microbial pathogens (Chandler et al., 2011).

Biological control forms an essential component of integrated pest management (IPM), a strategy that can offer long-term insect control, while minimizing the risks of conventional pest control practices to the environment and humans. IPM primarily relies on the deployment of several pest control practices, devised to complement, minimize or replace chemical pesticides (Pretty and Bharucha, 2015), such that the combination of these together overcome the shortcomings of individual practices (Chandler et al., 2011). The main strategies of IPM are listed in Table 1.1, which includes various pest control methods.

A specialized group of nematodes, referred to as entomopathogenic nematodes have demonstrated great potential in controlling various insect pests of agricultural importance. These particular nematodes have evolved unique symbiotic relationships with insect pathogenic bacteria and, in doing so, have acquired the ability to initiate infection in several soil-dwelling insect species, mainly of the orders Lepidoptera and Coleoptera (Lacey and Georgis, 2012). Entomopathogenic nematodes possess specific biological and ecological features, making them excellent candidates as biological control agents for IPM, as well as offering a sustainable alternative to chemical insecticides. Both the entomopathogenic nematode and its bacterial partner exhibit high specificity towards insect targets, posing no threat to the environment. They are also non-toxic to plants, vertebrates and other non-target organisms. Furthermore, due to their highly virulent nature and rapid mode of action, entomopathogenic nematodes and their bacteria can induce insect death within 24-48 hours.

Table 1. 1 Primary IPM prevention and control methods.

Method	Example[1,2]
Modern selective pesticides	Products that are low risk and have high selectivity substitute synthetic pesticides that are highly toxic. For example, the use of biopesticides such as neem.
Cultivation methods	Includes crop rotation, intercropping and/or under-sowing.
Physical methods	Includes mechanical weeding.
Natural compounds	Application of semi-chemicals or biocidal plant extracts
Crop breeding	Breeding crop cultivars to confer partial or complete insect resistance through genetic modification. For example, the insecticidal toxin derived from the bacterium *Bacillus thuringiensis* is expressed in genetically modified *Zea mays* corn, referred to as '*Bt*' corn (Saxena et al., 1999).
Biological control	Exploiting natural enemies to disrupt or minimize insect populations such as predatory insects (e.g. ladybird beetles), parasites and parasitoids (e.g. various species of wasps and flies), and microbial pathogens (e.g. bacteria such as *B. thuringiensis*).

[1]Chandler et. al. (2011)

[2]Pretty and Bharucha (2015)

1.2 Diversity of nematodes

Nematoda is one of the most abundant and biologically-diverse of all the animal phyla, having evolved to inhabit nearly every ecosystem on earth, i.e. terrestrial, freshwater and marine environments (Decraemer et al., 2014). Early members of this ancient phylum are thought to have originated over 500 million years ago during the Precambrian or Cambrian explosion (Wray et al., 1996; Poinar, 2011). Based on their habitat and taxonomic position, nematode species vary in size. Most nematodes are microscopic ranging from 0.5 to 2 mm in length. There are, however, some rare

exceptions of nematodes, which are significantly larger. For instance, *Placentonema gigantissima*, a parasite of sperm whales is the largest nematode species, measuring at 7 m in length (Anderson, 2000).

Generally described as small and non-segmented, nematodes are thread-like in appearance (nema = thread in Greek) (Decraemer et al., 2014); and are often referred to as roundworms due to their distinct tube-within-a-tube morphology. Nematodes possess a simple yet relatively conserved body plan consisting of an external and internal tube, separated by a pseudocoelom. The external tube contains the cuticle, hypodermis, excretory system, nervous system and several muscles, whilst the internal tube consists of the oesophagus, intestines and reproductive organs (in adult nematodes) (Noe, 2007).

With only a small proportion of nematode species having been described (~23,000), half of which being parasitic, nematodes are considered as the most abundant metazoans on earth with an estimated number that ranges from tens of thousands to several million species (Poinar, 1983; Blaxter, 1998). The ubiquitous nature of nematodes as well as their successful adaptation to various ecological environments suggests a high degree of phenotypic plasticity, perhaps best exemplified in differences observed among various life cycle stages, particularly in parasitic nematodes (Viney and Diaz, 2012; Schierenberg and Sommer, 2014).

The phenotype of an individual can signify the difference between living and reproducing, or evolutionary death. The same is holds true for nematodes. Evolutionary fitness of these organisms can be attributed to both its survival and

reproductive success in its environment whilst outcompeting other organisms (Viney and Diaz, 2012).

Nematodes are found to exhibit all modes of reproduction. While most follow an amphimictic (sexual) reproductive strategy, some nematode species can also be hermaphroditic or parthenogenetic (Schierenberg and Sommer, 2014). Amphimixis, often seen in nematode species occupying diverse free-living niches, is regarded as the ancestral mode of reproduction on account of multiple, evolutionary transitions resulting in parthenogenesis and hermaphroditism (Denver et al., 2011). Parthenogenesis, on the other hand, is the second most common mode of reproduction in free-living nematodes (Denver et al., 2011). The other mode of reproduction often seen in nematodes is hermaphrodism, exclusively of the self-fertilizing kind rather than cross-fertilizing hermaphroditism commonly observed in organisms such as annelids and mollusks (Schierenberg and Sommer, 2014). Hermaphroditism has featured several times in nematode evolution. For instance, the genus *Caenorhabditis*, well known for the model organism, *Caenorhabditis elegans*, has undergone at least three independent evolutionary transitions towards hermaphroditism (Denver et al., 2011).

1.3 Importance of nematodes

Nematodes are important in various fields of biology, demonstrating a wide range of unique relationships with several species. Since they are present in nearly every ecosystem, nematodes respond readily to environmental changes. They can, in fact, serve as bioindicators in assessing the quality of an environment, biodiversity and anthropogenic disturbances in biomonitoring (Holt and Miller, 2010). The diversity

and abundance of nematode communities can be used to determine soil quality (Sánchez-Moreno et al., 2006) and even indicate river pollution (Zullini, 1976).

While most nematodes species are free-living in nature, many have developed a variety of parasitic lifestyles, resulting in several human diseases as well as severe loss of crops and livestock. According to the World Health Organization (WHO), parasitic nematode account for the infections of 1.5 billion people worldwide. Some parasitic nematodes known as helminths are directly responsible for causing several well-known diseases in humans, such as river blindness, elephantiasis (filariasis) and ascariasis (Taylor et al., 2010). A more indirect effect on human health is through zoonotic infections brought about by parasitic nematodes affecting livestock and domestic pets. For example, the zoonotic nematode, *Trichinella spiralis* affects livestock and can easily be transmitted to humans upon ingesting infected meat (Robinson and Dalton, 2009). *Dirofilaria immitis* and *D. repens* can infect feline, canine and human hosts, resulting in cardiopulmonary dirofilariasis (Simón et al., 2012).

Some nematodes, such as the root-knot (*Meloidogyne*) and cyst (*Heterodera* and *Globodera*) nematodes, are obligate plant parasites (Jones et al., 2013) and have had devastating effects on agriculture. With over 4,100 plant-parasitic species having been described (Decraemer and Hunt, 2006), the damage inflicted by these nematode parasites can be devastating, often resulting in severe economic losses to agriculture, i.e. an annual loss estimated at $US80 billion from crop yield loss (Nicol et al., 2011).

While many nematode species are regarded as pests owing to the diseases they can inflict on both humans and animals, as well as the considerable impact they have on several agricultural products, there are also a significant number of nematodes

that are beneficial. Some nematodes, for instance, are insect parasites such as the entomopathogenic nematodes and they can be exploited as biological control agents against problematic insect pests.

1.4 Entomopathogenic nematodes

Among the diversity of insect pathogens extensively studied, entomopathogenic nematodes have aroused considerable interest over the years due to their unique ability of controlling a wide range of insect hosts. Having adapted specific mechanisms to associate with bacterial pathogens of either the genus *Xenorhabdus* or *Photorhabdus*, entomopathogenic nematodes of the families Steinernematidae and Heterorhabditidae, respectively, share remarkably similar life strategies, despite being phylogenetically unrelated (Dillman et al., 2012). The term 'entomopathogenic nematode' was first used to describe a group of nematodes, capable of inducing rapid insect death within 48 hours of infection (Gaugler and Kaya, 1990).

There are various forms of nematode-insect associations, ranging from beneficial to antagonistic, i.e. phoresy, necromeny, parasitism and entomopathogeny (Dillman et al., 2012). Entomopathogeny holds the most potential for the biological control of insects. The key features of this type of nematode-insect association include: i) the transmission of lethal insect-killing bacteria amongst susceptible insect hosts by infective juveniles; ii) location and invasion of insect hosts by infective juveniles; iii) the release of the bacteria into the haemolymph of the insect host; iv) bacterial proliferation subject to cadaver-nutrient utilization leading to infection and the subsequent demise of the insect host; v) development of nematodes and re-association of the bacterial symbionts with infective juveniles; and vi) emergence of

newly formed infective juveniles from the nutrient-depleted insect cadaver (Kaya and Gaugler, 1993; Dillman et al., 2012).

Until recently, other nematode species have also demonstrated characteristic hallmarks of entomopathogeny in their use of bacterial pathogens to infect insect hosts (Dillman et al., 2012), suggesting that this insect-parasitic lifestyle is not exclusive to steinernematid and heterorhabditid nematodes alone. Several *Oscheius* species have been identified as insect pathogens and are now recognized as entomopathogenic nematodes (Abebe et al., 2011; Stock et al., 2011). *Oscheius* nematodes have been shown to associate exclusively with members of the bacterial genus, *Serratia* (Zhang et al., 2008).

1.4.1 Taxonomy and history of entomopathogenic nematodes

The taxonomic classification of entomopathogenic nematode families, Steinernematidae and Heterorhabditidae, are presented below (Table 1.2). The first entomopathogenic nematode, *Aplectana kraussei*, which later became known as *Steinernema kraussei* was described by Steiner in 1923. The name, *Steinernema* was later adopted (Travassos, 1927). In 1929, another genus and species, *Neoaplectana glaseri* (now *S. glaseri*) (Steiner, 1929) was isolated from the Japanese beetle grub *Popillia japonica* by Glaser and Fox (1930). Chitwood and Chitwood (1937) proposed the family Steinernematidae. After re-examination, it was found that both species *S. kraussei* and *N. glaseri* belonged to the same genus (Wouts et al., 1982).

The genus *Heterorhabditis* was first reported by Poinar (1976). A striking feature of this nematode genus was its association with bacterial symbionts that were able to fluoresce (Poinar et al., 1980). Like with *Steinernema*, there exists various

geographical species and strains of *Heterorhabditis* (Poinar 1990; Stock and Hunt, 2005). The global distribution of *Steinernema* and *Heterorhabditis* nematodes suggests the presence of the lineages when all land masses were combined as the supercontinent, Pangaea.

Table 1. 2 Taxonomy of Steinernematidae and Heterorhabditidae[1]

Phylum: Nematoda
Class: Chromadorea
Subclass: Chromadoria
Order: Rhabditida
Suborder: Tylenchina
Infraorder: Panagrolaimomorpha
Superfamily: Strongyloidoidea
Family: Steinernematidae
Suborder: Rhabditina
Infraorder: Rhabditomorpha
Superfamily: Strongloidea
Family: Heterorhabditidae

[1]Stock and Hunt (2005)

1.4.2 Origin of entomopathogenic nematodes

Entomopathogenic nematodes can be traced back to the mid-Palaeozoic period, about 375 million years ago (Poinar, 1983, 1993). Based on morphological data, it is believed that *Steinernema* spp. evolved from a 'proto-Rhabditonema' ancestor and *Heterorhabditis* spp. arose from a '*Pellioditis*-like' ancestor in a marine environment (Poinar, 1983). This theory is consistent with the findings of a phylogenetic study performed on the Nematoda phylum, based on the highly conserved 18S ribosomal

DNA (rDNA) (Blaxter et al., 1998). Although both genera are not closely related phylogenetically, they do share similar life histories and morphology, having undergone independent, convergent evolution with enteric bacteria and arthropods (Poinar, 1993). It has also been shown that *Steinernema* is most phylogenetically related to Panagrolaimoidea and Strongyloididae; whilst *Heterorhabditis* is related to a suborder of vertebrate parasites known as Strongylida (Blaxter et al.,1998), indicating that a free-living *Rhabditis* group gave rise to both groups (Kiontke el al., 2007).

According to Blaxter et al. (1998), nematode parasitism (plant, invertebrate and vertebrate) has evolved from free-living nematodes at least five times. It is then plausible that the ancestral species of entomopathogenic nematodes originated from a free-living, microbivorous nematode, which ultimately led to the acquisition of its insect-pathogenic bacterial partner. This is supported by the fact that entomopathogenic nematodes have retained their ancestral microbivorous traits in that they depend on their symbiotic bacteria for their nutritional and developmental needs. Over time, the ancestral nematode-bacterium association evolved, becoming highly-specific and virulent in nature, evitability leading to insect parasitism. Other evolving associations between nematodes and insect-pathogenic bacteria include *Oscheius* spp. and *Caenorhabditis briggsae* with *Serratia* bacteria (Torres-Barragan et al., 2011; Abebe et al., 2011), which supports the idea that bacteria-feeding nematodes can evolve into vectors for insect-pathogenic bacteria.

1.4.3 Life cycle of entomopathogenic nematodes

Many nematodes can undergo a developmentally-arrested dauer stage in response to harsh environmental conditions. In fact, the dauer stage can be considered as an equivalent to the infective stage demonstrated by most parasitic nematodes including entomopathogenic nematodes. The third larval or juvenile stage (L3 or J3), referred to as the infective juvenile, is the only infective stage of entomopathogenic nematodes, allowing for survival, dispersal as well as host seeking (Figure 1.1). In fact, infective juveniles are the only free-living form of entomopathogenic nematodes, while all other developmental stages can only occur within an infected insect host (Dillman et al., 2012).

The non-feeding infective juveniles can persist in the soil for weeks or even months at a time, relying on lipid reserves as their sole source of energy (Selvan et al., 1993). The presence of a moulted cuticle retained as a sheath from their second-stage offers some form of protection against desiccation and adverse environmental conditions (Kaya and Gaugler, 1993).

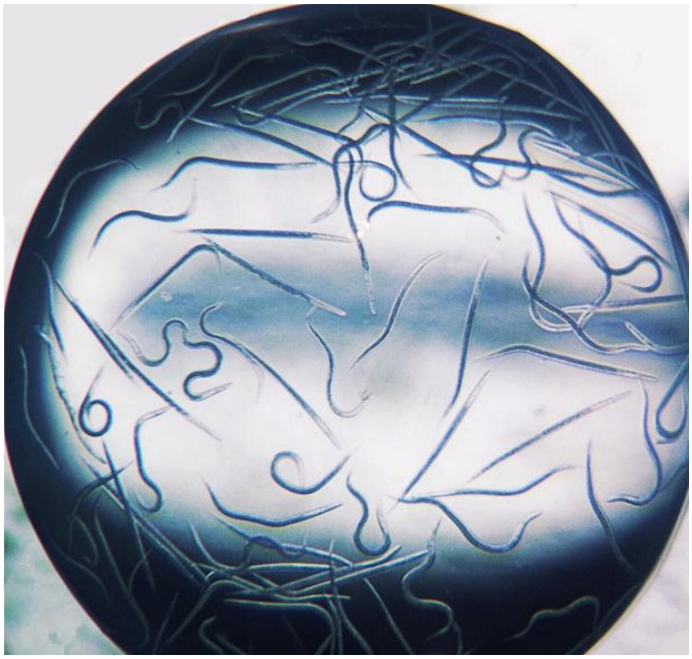


Figure 1.1. Third-stage entomopathogenic nematodes, referred to as infective juveniles. Image of infective juveniles in a microlitre (μL) of water, viewed under a dissecting microscope. Live infective juveniles may either adopt a "J"-shape at resting position or an "S" pattern whilst moving (Lewis, 1999). Straight, motionless infective juveniles are non-viable and typically, are dead. Image taken with a Samsung S3 camera phone, with no scale bar included.

Upon locating a suitable insect host, infective juveniles gain entry into the haemocoel either through natural openings such as the mouth, anus or spiracles, or through the cuticle (Kaya and Gaugler, 1993). *Steinernema* nematodes make use of the former route, entering the body cavity of an insect host through natural openings. In contrast, infective juveniles of *Heterorhabditis* are able to penetrate directly through the thin cuticle of the intersegmental membrane using a dorsal tooth – a morphological feature absent in *Steinernema* nematodes (Bedding and Molyneux, 1982).

Once inside an insect host, the bacterial symbionts are released in the haemolymph by the infective juveniles, where the bacteria rapidly proliferate,

producing an array of enzymes and antimicrobial agents (Poinar, 1990), as shown in Figure 1.2. The accumulation of bacterial toxins results in insect death through septicaemia (Adams and Nguyen, 2002). The production of bacterial enzymes leads to the breakdown of host tissue into a nutrient 'soup', which is then utilized by both bacteria and nematode. The antimicrobial activity of the associated bacteria also suppresses the invasion of other competing microorganisms in the soil. The invading nematodes exit the infective juvenile stage, moulting into fourth stage (J4) juveniles. The J4 juveniles mature into adults (Griffin et al., 2001). In the case of *Steinernema* nematodes, they can undergo one to three generations, before nutrients are depleted. When this occurs, egg laying ceases and the female can undergo a process referred to as *endotokia matricida*. This process is crucial for infective juvenile development in both *Steinernema* and *Heterorhabditis* nematodes (Goodrich-Blair and Clark, 2007). *Endotokia matricida* ensures nematode survival by permitting their development in a nutrient-restricted environment (Johnigk and Ehlers, 1999). Newly formed infective juveniles then re-associate with their bacterial symbiont, before emerging from a nutrient-depleted insect cadaver.

The symbiosis between the infective juveniles and its bacterial counterpart has served as a beneficial and necessary requirement in the evolution of insect pathogenicity. The nematode offers a protective, competitor-free environment for its bacterial symbionts, whilst serving as its vector (Forst et al., 1997). By comparison, the bacteria serve as a nutrient source for nematode growth and development, in addition to suppressing the immune response of the insect host, as well as preventing other competing microorganisms from invading the host cadaver (Forst et al., 1997).

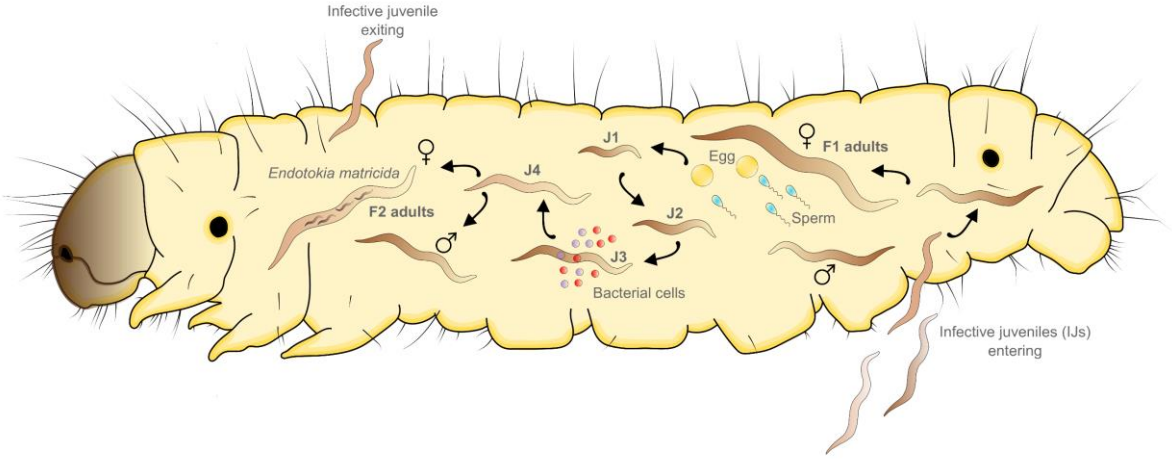


Figure 1.2. The life cycle of *Steinernema* entomopathogenic nematodes. Upon entry into the haemocoel of the insect host, the infective juveniles (IJs) release their symbiotic bacteria, which rapidly multiply and induce septicaemia in the insect. The insect succumbs to infection and dies within 24-48 hrs. The infective juveniles then develop by moulting into fourth stage (J4) juveniles, which the matures into adults. The nematodes can undergo one to three generations, before nutrient depletion. The unique infective juveniles (J3) take up the symbiotic bacteria before vacating the insect cadaver and emerge as infective juveniles once more.

1.4.4 Host range and safety

Entomopathogenic nematodes are unique parasites due to their symbiotic relationship with entomopathogenic bacteria, as well as their ability to kill insect hosts quickly after infection. Though considered as obligate parasites or pathogens of insects (Poinar, 1979), entomopathogenic nematodes do not adhere to the close, highly-adapted relationships distinctive of other host-parasite infections (Kaya and Gaugler, 1993). Their ability to induce rapid insect mortality quickly allows these nematodes to target a broad range of hosts, spanning most insect orders (Kaya and Gaugler, 1993; Poinar, 1979). Whilst many entomopathogenic nematode species have been shown to be highly infective towards numerous insect hosts, particularly under optimum laboratory conditions, this is rarely observed in the field, whereby

entomopathogenic nematodes tend to target a narrower host range (Kaya and Gaugler, 1993; Adams et al., 2007).

As entomopathogenic nematodes are adapted to a soil environment, their host range extends to species that develop or dwell in the upper layers of the soil for a certain period during their development (Bathon, 1996). The survival and pathogenicity of entomopathogenic nematodes are often subject to external factors, such as environmental conditions (Kaya and Gaugler, 1993), ecological conditions (Kung et. al. 1991), energy constraints (Vanninen (1990), morphological barriers of insect hosts (Eidt and Thurston, 1995), as well as evasive behaviour sometimes exhibited by these insects (Gaugler et al., 1994). For instance, some insect larvae, such as white grubs, display grooming behaviour (scratching and biting) to circumvent infection entirely when the presence of infective juveniles is detected on their cuticle (Gaugler et al., 1994).

There have been extensive field studies conducted with varied success, which have demonstrated the biocontrol potential of entomopathogenic nematodes to infect various insect hosts. Infective juveniles of *S. carpocapsae*, for example, have been shown to control field populations of fungus gnats, strawberry root weevils, citrus root weevils and cutworms (Kaya, 1993). The exploitation of entomopathogenic nematodes in biological pest control has often raised safety concerns, especially with regards to their effects, if any, on populations of non-target species. Apart from tadpoles (Poinar, 1989), there have not been any reports on acute or chronic toxicity of entomopathogenic nematodes on humans or other vertebrates, nor have there been any lasting effects on non-target invertebrate populations. Numerous studies have demonstrated that vertebrates cannot be included in the host range of

entomopathogenic nematodes, nor do these nematodes pose any safety hazard as biological control agents (Akhurst and Smith, 2002; Bathon,1996; Wang et al., 1983; Wang and Liu, 1983; Wang et al., 1984).

1.4.5 Behaviour

Host seeking is an intricate behaviour, involving chemosensory, thermosensory and hygrosensory cues (Ashton et al., 1999; Haas, 2003; Haas et al., 2005 and Torr et al., 2004). Entomopathogenic nematodes engage in a wide range of host-seeking behaviour, detecting various insect host cues, such as carbon dioxide (Gaugler et al., 1980), host faeces (Grewal et al., 1993a), host gut contents (Grewal et al., 1993b) and host odour (Hallem et al., 2011). Adoption of a specific foraging strategy is influenced by certain aspects such as parasite ecology, physiology, anatomy and behaviour, all of which determine parasite-host interactions (Campbell and Lewis, 2002). Often, host selection by parasitic nematodes adheres to a hierarchical sequence of steps, i.e. host seeking, host recognition and host suitability (Doutt, 1964).

When seeking out insect hosts, entomopathogenic nematodes may utilize one of three foraging strategies, including ambushing, cruising or an intermediate approach of both strategies. Ambush foragers typically adopt an energy-conserving approach when searching for insect targets, exploiting a unique behaviour referred to as nictation. Nematodes, exhibiting this nictating-behaviour, balance on the ends of their tails, erecting themselves in an upright manner, whilst swaying their heads in a wave-like motion (Campbell and Gaugler, 1993). Nictation is a mechanism, which facilitates the capture of mobile insects closer to the soil surface. By contrast, cruiser foragers are best adapted for seeking out relatively sedentary targets, making use of various host cues during their search.

Search behaviour of entomopathogenic nematodes is also influenced by energy constraints (Kaya and Gaugler, 1993). Since infective juveniles exist in a perpetual, non-feeding state in nature when searching for a new food source, they are incapable of adjusting their energy intake to compensate for their energy expenditure, as with many foraging organism (Wolf and Hainsworth, 1977). Studies conducted by Vanninen (1990) have demonstrated the significance of search cost in infective juvenile nematodes with depleted lipid reserves. While these infective juveniles were successful at infecting stationary *Tenebrio molitor* hosts, nematode infectivity declined when active host search was required for mobile insect targets.

Interestingly, some entomopathogenic nematode species have even been shown to exhibit scavenging behaviour, preferring dead insect hosts to living ones (San-Blas et al. 2008). It has been suggested that this alternative life strategy may serve to bypass defence mechanisms of an insect host, according to experiments performed by San-Blas and Gowen (2008).

1.4.6 Entomopathogenic nematodes as biological control agents

Biological control provides an alternative method for controlling problematic insect pests, whilst minimizing the adverse effects on human health and the environment. Entomopathogenic nematodes can serve as an important part of the spectrum of biocontrol agents (Ehlers, 1996), and have already been used to target insect pests of high-value crops and can potentially be used on a larger scale in IPM, organic farming and sustainable agriculture to control a variety of economically-important insect pests.

Regardless of how well-suited entomopathogenic nematodes are in targeting insect hosts, their implementation may prove unsuccessful if the nematodes are not delivered in an effective manner that ensures adequate contact with and infection of the intended insect targets. There are several important considerations when applying entomopathogenic nematodes in the field, and the choice of application equipment and way these nematodes are delivered can have a significant impact of their efficacy in pest control.

Generally, conventional equipment designed for pesticides, fertilizers and irrigation delivery can be used in entomopathogenic nematodes application, including spray equipment, such as pressurized sprayers, mist blowers, and electrostatic sprayers (Georgis, 1990). The most common way of nematode application is spraying the soil surface. For field application against soil-borne insects, a rate of 2.5 billion infective juveniles/ha is considered to be effective, though this quantity may vary according to crop type, target insect, and application technology (Nyugen and Smart, 1996).

According to Poinar (1986) and Georgis (1990), tips with openings no smaller than 0.05 mm can be used during nematode delivery, though allowances must be made for larger nematode species. While most agricultural pumping systems do not generate sufficient pressure that can cause shearing damage to nematodes, infective juvenile nematodes can withstand pressures of up to 2000 kPA (Georgis, 1990).

Several environmental factors can affect nematode survival, movement, reproduction and development, thereby influencing the success of entomopathogenic

nematode application in the soil. These factors include, but are not limited to, ultraviolet radiation, moisture content and soil type (Kaya, 1990). Ultraviolet radiation is especially harmful to entomopathogenic nematodes (Gaugler and Boush, 1978), therefore if infective juveniles are to be applied to the surface of the soil, it is advisable that application be carried out when sun exposure is at its minimum (Sharpiro-Ilan et al., 2006). Adequate moisture levels in the soil are also crucial for nematode motility, as well as their survival. Nematodes are essentially aquatic organisms, requiring a thin film of water to facilitate their movement (Norton, 1978); however, too much moisture may restrict movement altogether (Kaya, 1990). Optimal moisture levels will differ according to the species of nematode. Dry conditions can also have an adverse effect on nematode motility in soil, in turn affecting their efficacy (Womersley, 1990). Under severe environmental stress, such as the absence of water, some nematodes can become quiescent, a response involving the reduction in metabolic activity (Glazer, 1996). Following a gradual loss of water, nematodes can sometimes enter a state of anhydrobiosis (a type of cryptobiosis) (Crowe and Madin, 1975), forming tight coils when subjected to desiccation (Wharton, 1986; Womersley, 1987). Soil texture also plays a crucial role in nematode activity, having been shown to influence their motility, survival and efficacy (Kung et al., 1990). Generally, clay soils are found to be restrictive to nematode movement, whilst sandy-loam soils provide a medium through which nematodes move easily (Kaya, 1990).

Often, the application of biological control agents in integrated pest control involves the simultaneous use of agrochemicals, which can vary in its effects on infective juvenile nematodes. Most fertilizers, for example, have little to no effect on the efficacy of entomopathogenic nematodes, when applied at recommended rates; however, high levels can adversely affect nematode activity and survival (Bednarek

and Gaugler, 1997; Shapiro et al., 1996). Nematode species vary in their susceptibility to chemical pesticides. Some chemical pesticides can be extremely toxic to entomopathogenic nematodes, such as dodine, methomyl and parathion. Others (e.g., chlorpyrifos and endosulfan) can be compatible with nematode application, while other pesticides act synergistically with entomopathogenic nematodes, even improving their effectiveness in pest control (Alumai and Grewal, 2004; Koppenhöfer, 2000; Koppenhöfer and Kaya, 1998; Mannion et al., 2000; Nishimatsu and Jackson, 1998; Rovesti and Deseö, 1991).

Like with other organisms, entomopathogenic nematodes can be susceptible to infection and have their own suite of natural predators, including some phages, bacteria, protozoan parasites, nematophagous fungi, mites and other nematodes, (Kaya, 2002). In addition, it has been reported that some entomopathogenic nematodes have synergistic interactions with other entomopathogens, i.e. *Bacillus thuringiensis* (Koppenhöfer and Kaya, 1997). It has been shown that *S. carpocapsae and H. bacteriophora* were capable of infecting insect hosts exposed to *B. thuringiensis*, though the time at which *B. thuringiensis* exposure occurred was crucial for nematode development (Bednarek, 1986; Kaya and Burlando, 1989; Poinar et al., 1990). Other studies on the interactions between the entomopathogenic fungus, *Beauveria bassiana* and entomopathogenic nematodes on the same insect hosts, have demonstrated an antagonistic effect for the most part (Brinkman and Gardner, 2000).

1.5 Pathogenicity of *Photorhabdus* and *Xenorhabdus*

Entomopathogenic nematodes of the families Steinernematidae and Heterorhabditidae have forged symbiotic associations with bacteria of the genera

Xenorhabdus (Thomas and Poinar, 1979) and *Photorhabdus* (Boemare et al., 1993, respectively. Members of *Xenorhabdus* and *Photorhabdus* belong to the family *Enterobacteriaceae* in the γ-subclass of Proteobacteria and share characteristics of being Gram-negative, non-fermentative rods that are motile and oxidase-negative (Brenner and Farmer, 2005). Unlike other genera within the family, *Xenorhabdus* and *Photorhabdus* feature unique characteristics, both phenotypically and genotypically (Holt el al., 1994; Brenner and Farmer, 2005).

Xenorhabdus and *Photorhabdus* have adopted distinct yet similar lifestyles both in parasitism and mutualism as a result of convergent evolution (Goodrich-Blair and Clarke, 2007; Poinar, 1993). The symbiotic bacteria are obligate insect pathogens, lethal to various insects, although one species, *Photorhabdus asymbiotica* tends to be an opportunistic pathogen to humans as well (Costa et. al., 2009). *Xenorhabdus* and *Photorhabdus* bacteria are also mutualists of entomopathogenic nematodes, engaging in symbiotic interactions that are beneficial to both partners. In fact, free-living forms of the symbiotic bacteria are rarely isolated outside of their nematode host (Forst et al., 1997). Infective juvenile nematodes serve as vectors for the transmission of *Xenorhabdus* and *Photorhabdus* spp. to various insect hosts, whilst offering a protective, competitor-free environment in which the bacterial symbionts can thrive. The bacterial symbionts, in turn, contribute to the symbiotic nature of the relationship by creating conditions that are favourable for nematode growth and development (Boemare et al., 1997).

Entomopathogenic nematodes and their symbiotic bacteria play a vital role in the control of soil food webs. Together, they make a potent insecticidal complex, serving as a natural enemy to several soil-borne insects. *Xenorhabdus* and *Photorhabdus* have

received considerable interest over the years, representing a potential source of pharmaceutical and agrochemical agents (Webster et al., 2002).

1.5.1 Life cycle of *Xenorhabdus* and *Photorhabdus*

The life cycle of *Xenorhabdus* and *Photorhabdus* spp. share similar mechanisms – commencing and ending with intestinal colonization of the third-stage infective juvenile nematodes (Goodrich-Blair and Clarke, 2007). The bacterial symbionts are maintained in the gut of third-stage infective juveniles. *Xenorhabdus* colonizes a specialized intestinal vesicle in *Steinernema* nematodes (Bird and Akhurst, 1983). By comparison, *Photorhabdus* mainly colonizes the proximal region of the intestine of *Heterorhabditis* spp. (ffrench-Constant et al., 2003), though the bacterial cells can be found throughout the intestine. Upon locating and successfully entering an insect host, infective juvenile nematodes make their way to the haemolymph where their bacterial symbionts are released. Within the haemolymph, the bacteria rapidly multiply, evading the host immune response and inducing insect death using a variety of toxic mechanisms (Hinchliffe et al., 2010).

Photorhabdus infections are characterized by the colonization of the midgut in insect hosts, where the secretion of endo- and exotoxins results in the destruction of the midgut epithelium (Hinchliffe et al., 2010). In contrast, *Xenorhabdus* bacteria initially proliferates within the haemolymph before migrating to the midgut where they colonize the connective tissues surrounding the muscle fibres, trachea and other organs of insect hosts. The bacteria metabolize the host tissues, while producing several antimicrobial molecules that inhibit the growth of other competing microorganisms within the insect cadaver (Akhurst, 1982; Forst and Nealson, 1996). Developing nematodes feed on the symbiont biomass and metabolized host tissues,

which allows them to undergo one to three rounds of reproduction (Boemare, 2002). Once all nutrients have been depleted, the bacterial symbiont recolonizes the intestinal tract of the nematode progeny and a new generation of infective juveniles emerge from the insect cadaver in search of new insect hosts.

The bacterial cells of *Xenorhabdus* and *Photorhabdus* spp. feature the ability to spontaneously undergo phase variation, occurring as two forms. This phenotypic switching can be attributed to changes in the environment or their growth conditions (van der Woude and Baumler, 2004). Typically, phase I bacterial variants are retained by infective juvenile nematodes, contributing to nematode development and pathogenicity. In contrast, phase II bacteria can arise spontaneously, usually under artificial conditions (Volgyi et al., 1998). Additionally, phase II variants do not sufficiently support nematode growth and development and are associated to reduced virulence (Akhurst, 1980; Akhurst and Boemare, 1990; Forst et al., 1997).

1.5.2 Pathogenicity of *Xenorhabdus* and *Photorhabdus*

Xenorhabdus and *Photorhabdus* are pathogenic to a wide range of insect larvae spanning most insect orders, including Lepidoptera, Coleopteran, Hymenoptera and Dictyoptera (Bowen and Ensign, 1998; Morgan et al., 2001). The symbiotic bacteria, together with their nematode partner can induce insect death within 24-48 hours – a striking feature, given that insects have developed a strong innate immune response that involves a combination humoral and cellular defences, similar to the immune responses of mammals (Leulier et al., 2003).

Upon infection, the presence of the nematode and its bacterial symbiont does not go undetected by the host insect. Several recognition and antimicrobial proteins are

released that recognize and bind to specific sites found on the surface of these invading microorganisms – elicited as part of the humoral response. In particular, circulating haemocytes (blood cells), which forms part of the cellular innate immunity response in insects, are involved in the recognition of foreign particles. The recruitment of accessory haemocytes called lamellocytes lead to the encapsulation of the invading organism(s) (Goodrich-Blair and Clarke, 2007), which are then melanized through the enzymatic action of phenoloxidase and eliminated through phagocytosis (Brennan and Anderson, 2004; Kanost et al., 2004). The bacterial symbionts have evolved various strategies to evade the host immune responses. It has been shown that both *Xenorhabdus* and *Photorhabdus* spp. suppress the enzymatic activity of phospholipase A2 – an important insect enzyme involved in the eicosanoid biosynthesis pathway, which prevents haemocyte aggregation and nodulation (Kim et al., 2005). Some species of *Photorhabdus* encode a type III secretion system (TTSS), which allows for the delivery of an effector protein, LopT, responsible for suppressing phagocytosis and nodule reduction entirely within insects (Brugirard-Ricaud et al., 2004; 2005).

Humoral innate immunity of insects, on the other hand, involves the synthesis of various cationic antimicrobial peptides (CAMP) in response to infection, which specifically targets bacterial membranes (Nappi and Ottaviani, 2000). *Photorhabdus* spp. can evade this insect innate immune response entirely through the modification of their lipopolysaccharides, whilst *Xenorhabdus* spp. only hinder CAMP production without membrane alteration (Goodrich-Blair and Clarke, 2007).

1.6 Genome sequencing

1.6.1 Introduction and overview

The genetic information of an organism is encoded within its genome. Genome sequences are made up of protein-coding regions, RNA genes, functional intergenic and non-coding regions and non-functional DNA. The functional non-coding DNA comprises promoters, enhancers, regulators, transcription factors and telomeres, while the non-functional DNA includes transposons, repetitive sequences and pseudogenes. Genomes can vary in size, ranging from 580-kilobases (kb) for the smallest bacterial genome, *Mycoplasma genitalium* (Fraser et al., 1995) to 32-gigabases (Gb) for the Mexican salamander, *Ambystoma mexicanum*, which is the largest genome described to date (Nowoshilow et al., 2018).

The first complete genomes of free-living organisms were sequenced in 1995, belonging to the bacteria, *Haemophilus influenzae* (Fleischmann et al., 1995) and *M. genitalium* (Fraser et al., 1995). Several other genome sequences soon followed, such as the first eukaryotic genome, *Saccharomyces cerevisiae* (Goffeau, 1996); the first multicellular animal genome, *Caenorhabditis elegans* (The C. elegans Sequencing Consortium, 1998) and the first plant genome, *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000). Over the last decade, the number of sequenced genomes has risen considerably, and genomics is fast evolving into one of the most recognized disciplines in science. To date, a total of 15,242 sequencing projects has been completed (https://gold.jgi.doe.gov/). Studying these genomes has provided insight into the important differences in genome organization between prokaryotes and eukaryotes, as well highly conserved gene and genome structures across phyla and domains.

1.6.2 Gene structure of eukaryotic genomes

The genome of eukaryotes consists of a haploid set of chromosomes. A typical eukaryotic gene contains regions that encode for proteins referred to as exons, which are split by non-coding intragenic regions referred to as introns (Castellana and Bafna, 2010). Though introns are transcribed, they are not translated into protein, and are subsequently removed out of the final mRNA product through RNA splicing. A general process of gene expression in eukaryotes includes the transcription of a gene from DNA to pre-mRNA (Zhang, 2002). Through RNA processing, mRNA is produced through the addition of a 5' cap and 3' poly-A tail at the start and end of the RNA transcript, respectively, and RNA splicing taking place. The transcript is then exported to the cytoplasm from the nucleus, where translation is carried out. Eukaryotic genomes have several features that are not found in prokaryotes. In addition to the presence of introns and mRNA splicing, eukaryotes are not limited by space constraints in terms of genome size, resulting in the accumulation of large amounts of repetitive DNA and non-coding sequences such as transposons within the genome.

1.6.3 Genome organization of prokaryotes

The genomes of prokaryotic organisms are very different to those of eukaryotes. In prokaryotes, the genome is usually made up of a single circular chromosome. This is certainly true for *Escherichia coli* and several other commonly researched bacteria, though recent studies have indicated that some prokaryotes may have as many as four chromosomes, with DNA molecules that are linear in nature (Griswold, 2008), such as

in *Borrelia* (Ferdows and Barbour, 1989) and *Streptomyces* (Casjens, 1998; Bentley et al., 2002).

Similar to eukaryotic chromosomes, a prokaryotic genome has to fit into a substantially small space. The genomic organization of prokaryotic genomes can be regarded as highly efficient, as many key genes are located on a single chromosome, in spite of space limitations. There is also minimal non-coding DNA within bacterial genomes, which accounts for an average of 12% of the prokaryotic genomes, in contrast to the 98% in eukaryotes (Ahnert et al., 2008).

Much of the knowledge known about DNA organization within prokaryotic genomes has been derived from studies involving *E. coli*. For *E. coli*, the unfolded circumference of its circular chromosome is roughly 1.6 mm in length (Pettijohn, 1996), with dimensions of roughly 1.0 x 2.0 μm, suggesting that considerable compaction (i.e. nearly 1000-fold) is necessary to fit the entire bacterial chromosome into the nucleoid (Holmes and Cozzarelli, 2000). One way prokaryotes achieve this is through supercoiling. Supercoiling arises when DNA twists to form tiny coils, which twist even further to form additional folds, thus resulting in a condensed ball. When additional twists are introduced into bacterial DNA in the direction of the double helix, this results in positive supercoiling, whilst the removal of twists in the direction opposite to the double helix results in negative supercoiling, often favoured in most bacterial genomes (Griswold, 2008).

Another genomic feature found in *E. coli*, which is also observed in other bacterial groups, is the occurrence of mobile elements called plasmids. Prokaryotes can carry one or more plasmids, which are small pieces of extrachromosomal DNA, often

circular, but not always, that are capable of self-replicating, integrating into chromosomes and migrating between cells. While plasmids can be useful, especially as they contain genes that encode for proteins that confer antibiotic resistance and virulence, their presence in a prokaryotic cell is not contingent to bacterial survival.

1.6.4 Sequencing technologies

DNA sequencing technologies have existed since the late 1970s. At the time, two pioneering sequencing methods were developed almost simultaneously, which included the Chemical Cleavage (Maxam and Gilbert, 1977) and Sanger dideoxy (Sanger et al., 1977) sequencing methods. Although both methods were suitable for sequencing DNA, the use of radioactive labelling and toxic chemicals limited the use of the chemical cleavage method in research. Improvements in Sanger sequencing led to the development of automated sequence analysers, making it the favoured method of DNA sequencing.

Sanger sequencing involves the separation of double-stranded DNA by denaturation to obtain a single-stranded DNA template. With a primer bound to the template strand, DNA polymerase synthesizes a complementary strand by using deoxyribonucleotide-triphosphates (dNTPs). Also included in the reaction mix are chemically modified nucleotides called di-deoxynucleotide-triphosphates (ddNTPs), which prevent the addition of new nucleotides when incorporated into the DNA strand, due to the lack of a 3' hydroxyl group, thereby terminating the DNA chain. The presence of both dNTP and ddNTP leads to DNA fragments of various lengths, which can then be radioactively or fluorescently labelled, separated by size using electrophoresis and visualised by autoradiography or ultraviolet light.

For some time, Sanger sequencing played an important role in the analysis of genomic sequences and remained the gold standard when it came to DNA sequencing. The demand for new sequencing strategies, capable of generating large-scale sequence data in a fast and affordable manner, culminated in the development of the Next-generation Sequencing (NGS) in the early 2000s. Similar to the conventional Sanger sequencing method, most widely-used NGS technologies rely on amplification steps, without the need of cloning DNA fragments into plasmid vectors (Anderson, 1981), including Roche 454 pyrosequencing (454 Roche Applied Science) which has since been discontinued, Illumina (Illumina), Ion Torrent (ThermoFisher) and SOLiD platform (Life Technologies). Briefly, Sanger sequencing amplify and sequence one DNA fragment at a time. In contrast, NGS methods employed the same basic principle of clonal amplification by generating millions of DNA fragments in parallel and simultaneously producing millions of sequence reads from these fragments. This, of course, resulting in high-throughput sequencing technologies producing large volumes of genomic sequence data in the form of reads within a short timeframe and at low cost. Alternative sequencing technologies have also been developed that focuses solely on single molecule sequencing without the need for amplification, such as Helicos BioSciences, Pacific Bioscience and Oxford Nanopore. NGS technologies has led to the rapid increase in the number of whole genome sequences that have become accessible on various public databases.

1.7 Project rationale

The study of nematode genomes has largely relied on the concerted efforts made on the model organism, *Caenorhabditis elegans* (Dillman et al., 2012). *C. elegans* has paved the way for genetic research as being the first metazoan to be sequenced using

Sanger sequencing (*C. elegans* Sequencing Consortium, 1998). The genome of *C. elegans* is 100.2 MB, with 21,000 protein-coding genes (Blaxter et al., 2012). Strikingly, *C. elegans* is very similar to humans in terms of genome complexity and the number of protein encoding genes, i.e. 20,000 genes in humans.

Entomopathogenic nematodes can serve as excellent candidates for whole genome sequencing due to their biological relevance in controlling insect pests. Despite years of intensive research, a great deal of information remains unknown about these organisms, such as their behaviour, distribution and diversity, host specialization and their symbiotic and parasitic nature. Genome sequencing can be used as a valuable tool in understanding pathogenicity, parasitism and mutualism in entomopathogenic nematodes and their bacterial symbionts, thereby allowing for the discovery of useful genes and gene products of agricultural importance.

There is also tremendous scope for the present study, in that genome sequencing of entomopathogenic nematodes and their associated symbiotic bacteria is still in its infancy in South Africa. The genomic data held within these genomes may offer insight into understanding genetic mechanisms that trigger nematode and bacterial virulence, bacterial antimicrobial production, insect-host specificity, nematode host-seeking and stress tolerance behaviour.

Also, entomopathogenic nematodes share similar lifestyles to other parasitic nematodes and could therefore serve as model systems for understanding parasitic nematode interactions, with respect to host seeking and foraging behaviour. The knowledge gained from behavioural responses of entomopathogenic nematodes may also culminate in a better understanding of human-nematode interactions.

Furthermore, comparative genomic studies of entomopathogenic nematode species and their symbiotic bacteria at an intra- or inter-species level may also provide insight into their convergent and divergent evolution, regulatory mechanisms and virulence factors.

1.8 Aims and objectives

The main aim of this research project, after having isolated and molecularly characterized indigenous entomopathogenic nematode species during a small-scale survey of the North West Province and Gauteng in South Africa, was to assemble and annotate the genome of an entomopathogenic nematode, preferably one that had not been previously sequenced, making use of Illumina short-read sequencing technology. In doing so, the genomic architecture of the selected nematode of interest could then be explored. The second aim was to sequence the bacterial genome of the symbiont associated with the nematode of interest; and, thereby, perform comparative genomic analyses using other entomopathogenic bacterial genomes.

The objectives of this research project are listed as follows:

1. To isolate and molecularly identify entomopathogenic nematodes from indigenous soil samples collected during a small-scale survey of the North West Province and Gauteng of South Africa by sequencing the internal transcribed spacer (ITS) regions of the 18S rDNA and performing phylogenetic analyses.
2. To morphologically and morphometrically characterize any potentially novel entomopathogenic species recovered during this study.

3. To isolate and molecularly identify the bacterial symbionts associated with the nematode isolates recovered from soil samples collected during a small-survey in the North West Province and Gauteng by sequencing the internal transcribed spacer (ITS) regions of the 16S rDNA and performing phylogenetic analyses.
4. To sequence, assemble and annotate the genome of the bacterial symbiont associated with the nematode of interest and perform comparative genomic analyses with the bacterial symbiont genome.
5. To perform genomic sequencing and assembly on an entomopathogenic nematode species that had not been sequenced previously in order to produce a draft genome that could be used for subsequent downstream genome analyses.
6. To perform genomic annotation on the assembled genome of an entomopathogenic nematode species.
7. To characterize the predicted protein sequences of nematode of interest, as well as identify putative gene products in the nematode genome related to pathogenicity and host seeking behaviour.
8. To explore the genomic architecture of the nematode of interest and compare these genomic features to those of other previously sequenced nematode genomes.

1.9 References

1. Abebe, E., Bonner, K., Gray, V. and Thomas, W. K. (2011). Response to 'Bugs don't make worms kill'. *Journal of Experimental Biology*, **214:**1053–1054.
2. Abebe, E., Abebe-Akele, F., Morrison, J., Cooper, V., et al. (2011). An insect pathogenic symbiosis between a *Caenorhabditis* and *Serratia*. *Virulence*, **2:**158–161.
3. Abrol, D. P. and Shankar, U. (2012). History, Overview and Principles of Ecologically-based Pest Management. In: Abrol, D. P. and Shankar, U. (Eds.), *Integrated Pest Management: Principles and Practice*. CABI. pp. 1–26.
4. Adams, B. J. and Nguyen, K. B. (2002). Taxonomy and systematics. In: Gaugler, R. (Eds.), *Entomopathogenic nematology*. CABI Publishing, Wallingford, UK, pp. 1–33.
5. Adams, B. J., Peat, S. M. and Dillman, A. R. (2007). Phylogeny and evolution. In: Nguyen, K. B. and Hunt, D. J. (Eds) *Entomopathogenic nematodes: systematics, phylogeny and bacterial symbionts*. Nematology Monographs and Perspectives. Brill, Leiden, The Netherlands, pp. 693–733.
6. Ahnert, S. E., Fink, T. M. and Zinovyev, A. (2008). How much non-coding DNA do eukaryotes require? *Journal of Theoretical Biology*, **252:**587–592.
7. Akhurst, R. and Smith, K. (2002). Regulations and safety. In: Gaugler, R. (Eds.) *Entomopathogenic nematology*. CABI Publishing, Wallingford, UK, pp. 331–332.
8. Akhurst, R. J. (1980). Morphological and functional dimorphism in *Xenorhabdus* spp., bacteria symbiotically associated with the insect pathogenic nematodes *Neoaplectana* and *Heterorhabditis*. *Journal of General Microbiology*, **121:**303–309.

9. Akhurst, R. J. (1982). Antibiotic activity of *Xenorhabdus* spp., bacteria symbiotically associated with insect pathogenic nematodes of the families Heterorhabditidae and Steinernematidae. *Journal of General Microbiology*, **128:**3061–3065.

10. Akhurst, R. J. and Boemare, N. E. (1990). Biology and taxonomy of Xenorhabdus. In: Gaugler, R. and Kaya, H. K. (Eds.), *Entomopathogenic Nematodes in Biological Control*. Boca Raton, FL: CRC Press, pp. 79–90.

11. Aktar, W., Dwaipaya, S. and Chowdhury, A. (2009). Impact of pesticides use in agriculture: their benefits and hazards. *Interdisciplinary Toxicology*, **22:**1–12.

12. Alumai, A. and Grewal, P. S. (2004). Tank-mix compatibility of the entomopathogenic nematodes, *Heterorhabditis bacteriophora* and *Steinernema carpocapsae*, with selected chemical pesticides used in turfgrass. *Biocontrol Science and Technology*, **14:**725–730.

13. Anderson, R. C. (2000). Nematode parasites of vertebrates: their development and transmission, 2nd ed. New York: CABI.

14. Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature,* **408:**796–815.

15. Ashton, F. T., Li, J. and Schad, G. A. (1999). Chemo- and thermosensory neurons: Structure and function in animal parasitic nematodes. *Veterinary Parasitology*, **84:**297–316.

16. Bale, J. S., van Lenteren, J. C. and Bigler, F. (2008). Biological control and sustainable food production. *Philosophical Transactions of the Royal Society*, **363:**761–776.

17. Bathon, H. (1996). Impact of entomopathogenic nematodes on nontarget hosts. *Biocontrol Science and Technology*, **6:**421–434.

18. Bedding, R. A., and Molyneux, A. S. (1982). Penetration of Insect Cuticle by Infective Juveniles of *Heterorhabditis* spp. (Heterorhabditidae, Nematoda). *Nematologica*, **28:**354–359.

19. Bednarek, A. (1986). Development of the *Steinernema feltiae* (Filipjev) entomogenous nematode (Steinernematidae) in the conditions of occurrence in the insect's body cavity of other pathogens. *Annals of Warsaw Agricultural University SGGW-AR Animal Science*, **20:**69–74.

20. Bell, W. J. (1991). Searching behaviour: The behavioural ecology of finding resources. London: Chapman and Hall Animal Behaviour Series.

21. Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417:**141–147.

22. Bird, A. F., and Akhurst, R. J. (1983). The nature of the intestinal vesicle in nematodes of the family Steinernematidae. *International Journal for Parasitology*, **13:**599–606.

23. Blaxter, M. and Koutsovoulos, G. (2015). The evolution of parasitism in Nematoda. *Parasitology*, **142:**S26–S39.

24. Blaxter, M. L. (1998). *Caenorhabditis elegans* is a nematode. *Science*, **282:**2041–2046.

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

26. Blaxter, M., Kumar, S., Kaur, G., Koutsovoulos, G., et al. (2012). Genomics and Transciptomics across the diversity of Nematoda. *Parasite Immunology*, **34:**108–120.

27. Boemare, N. E. (2002). Biology, taxonomy, and systematics of *Photorhabdus* and *Xenorhabdus*. In: Gaugler, R., editor. Entomopathogenic Nematology; p. 35–56.

28. Boemare, N. E., Akhurst, R. J. and Mourant, R. G. (1993). DNA relatedness between *Xenorhabdus* spp. (*Enterobacteriaceae*), symbiotic bacteria of entomopathogenic nematodes, and a proposal to transfer *Xenorhabdus luminescens* to a new genus, *Photorhabdus* gen. nov. *International Journal of Systematic Bacteriology*, **43:**249–255.

29. Boemare, N. E., Givaudan, A., Brehelin, M. and Laumond, C. (1997). Symbiosis and pathogenicity of nematode-bacterium complexes. *Symbiosis,* **22:**21–45.

30. Brennan, C. A., and Anderson, K. V. (2004). *Drosophila*: the genetics of innate immune recognition and response. *Annual Review of Immunology*, **22:**457–483.

31. Brenner, D. J. and Farmer, J. L. (2005). Family 1. *Enterobacteriaceae*. In: Garrity, G. M., Brenner, D. J., Krieg, N. R. and Staley, J. T. (Eds.), *Bergey's Manual of Systematic Bacteriology, Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria*, 2nd ed., Springer, New York, USA, pp. 587–607.

32. Brinkman, M. A. and Gardner, W. A. (2000). Possible antagonistic activity of two entomopathogens infecting workers of the red imported fire ant (Hymenoptera: Formicidae). *Journal of Entomological Science*, **35:**205–207.

33. Brouwer, A., Longnecker, M. P., Birnbaum, L. S., Cogliano, J., et al. (1999). Characterization of potential endocrine-related health effects at low-dose levels of exposure to PCBs. *Environmental health perspectives*, **107:**639–649.

34. Brugirard-Ricaud, K., Duchaud, E., Givaudan, A., Girard, P. A., et al. (2005). Site-specific antiphagocytic function of the *Photorhabdus luminescens* type III

secretion system during insect colonization. *Cellular Microbiology*, **7:**363–371.

35. Brugirard-Ricaud, K., Givaudan, A., Parkhill, J., Boemare, N., et al. (2004). Variation in the effectors of the type III secretion system among *Photorhabdus* species as revealed by genomic analysis. *Journal of Bacteriology*, **186:**4376–4381.

36. Campbell, J. F. and Gaugler, R. (1993). Nictation behaviour and its ecological implications in the host search strategies of entomopathogenic nematodes. *Behaviour*. **126:**155–169.

37. Casjens, S. (1998). The diverse and dynamic structure of bacterial genomes. *Annual Review of Genetics*, **32:**339–377.

38. Castelletto, M. L., Gang, S. S., Okubo, R. P., Tselikova, A. A., et al. (2014). Diverse host-seeking behaviors of skin-penetrating nematodes. *PLoS Pathogens*, **10:**e1004305.

39. Castellana, N. and Bafna, V. (2010). Proteogenomics to discover the full coding content of genomes: A computational perspective. *Journal of Proteomics*, **73:**2124–2135.

40. Chandler, D., Bailey, A. S., Tatchel, G. M., Davidson, G., et al. (2011). The development, regulation and use of biopesticides for integrated pest management. *Philosophical Transactions of the Royal Society*, **366:**1987–1998.

41. Costa., S. C. P., Girard, P. A., Brehélin, M. and Zumbihl, R. (2009). The Emerging Human Pathogen *Photorhabdus asymbiotica* Is a Facultative Intracellular Bacterium and Induces Apoptosis of Macrophage-Like Cells. *Infection and Immunity*, **77:**1022–1030.

42. Crisp, T. M., Clegg, E. D., Cooper, R. L., Wood, W. P., et al. (1998). Environmental endocrine disruption: an effects assessment and analysis. *Environmental health perspectives*, **106:**11–56.

43. Decraemer, W. and Hunt, D. J. (2006). Structure and classification. In: Perry, R. N. and Moens, M., (Eds.). *Plant Nematology*. Wallingford, Oxfordshire: CAB International. pp. 3–32.

44. Decraemer, W., Coomans, A. and Baldwin, J. (2014). Morphology of Nematoda. In: Schmidt-Rhaesa, A. (ed.). Handbook of zoology: Gastrotricha, Cycloneuralia and Gnathifera. *Nematoda* vol. 2. Berlin: De Gruyter.

45. Denholm, I. and Devine, G. J. (2013). Insect resistance. In: S. A. Levin, ed. *Encyclopedia of Biodiversity*. Waltham, MA: Academic Press.

46. Denver, D. R., Clark, K. A. and Raboin, M. J. (2011). Reproductive mode evolution in nematodes: insight from molecular phylogenies and recently discovered species. *Molecular Phylogenetics and Evolution*, **61:**584–592.

47. Dillman, A. R., Mortazavi, A. and Sternberg, P. W. (2012). Incorporating Genomics into the Toolkit of Nematology. *Journal of Nematology*, **44:**191–205.

48. Ehler, R. U. (1996). Current and Future Use of Nematodes in Biocontrol: Practice and Commercial Aspects with Regard to Regulatory Policy Issues. *Biocontrol Science and Technology*, **6:**303–316.

49. Ferdows, M. S. and Barbour, A. G. (1989). Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. *Proceedings of the National Academy of Sciences of the United States of America*, **86:**5969–5973.

50. Ferris, H., Minoshima, H., Sánchez-Moreno, S., and Jackson, L. (2006). Linking soil properties and nematode community composition: effects of soil management on soil food webs. *Nematology*, **8:**703–715.

51. ffrench-Constant, R., Waterfield, N., Daborn, P., Joyce, S., et al. (2003). *Photorhabdus*: towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiology Reviews*, **26:**433–456.

52. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269:**496–512.

53. Forst S. and Nealson K. (1996). Molecular biology of the symbiotic-pathogenic bacteria *Xenorhabdus* spp. and *Photorhabdus* spp.. *Microbiology Reviews*, **60:**21–43.

54. Forst, S., Dowds, B., Boemare, N. and Stackebrandt, E. (1997). *Xenorhabdus* spp. and *Photorhabdus* spp.: bugs that kill bugs. *Annual Review of Microbiology,* **51:**47–72.

55. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science*, **20:**397–404.

56. Gaugler, R. and Boush, G.M. (1978). Effects of ultraviolet radiation and sunlight on the entomogenous nematode *Neoaplectana carpocapsae*. *Journal of Invertebrate Pathology*, **32:**291–296.

57. Gaugler, R. and Kaya, H. K. (1990). Entomopathogenic nematodes in biological control. Boca Raton: CRC Press.

58. Gaugler, R., Lebeck, L., Nakagaki, B. et. al. (1980). Orientation of the entomopathogenic nematodes *Neoplectana carpocapsa*e to carbon dioxide. *Environmental Entomology*, **9:**649-659.

59. Gaugler, R., Wang, Y. and Campbell, J. F. (1994). Aggressive and evasive behaviours in *Popillia japonica* (Coleoptera: Scarabaeidae) larvae: defenses against entomopathogenic nematode attack. *Journal of Invertebrate Pathology*, **64:**193–199.

60. Georgis, R. (1990). Formulation and application technology, In: Gaugler, R. and Kaya, H. K., (Eds.), *Entomopathogenic Nematodes in Biological Control*. CRC Press, Boca Raton, FL, pp. 173–191.

61. Georgis, R. and Hague, N. G. M. (1981). A neoaplectanid nematode in the larch sawfly *Cephalcia lariciphila* (Hymenoptera: Pamphiliidae). *Annals of Applied Biology*, **99:**171–177.

62. Glaser, R. W. and Fox, H. (1930). A nematode parasite of the Japanese beetle (*Popillia japonica* Newm.). *Science*, **71:**16–17.

63. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., et al. (1996). Life with 6000 genes. *Science*, **27:**546–567.

64. Goodrich-Blair, H. and Clarke, D. J. (2007). Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination. *Molecular Microbiology*, **64:**260–268.

65. Grewal, P. S., Gaugler, R. and Campbell, J. F. (1993c). Host finding behaviour as a predictor of foraging strategy in entomopathogenic nematodes. *Parasitology*, **108:**207–215.

66. Grewal, P. S., Gaugler, R. and Lewis, E. E. (1993b). Host recognition behaviour by entomopathogenic nematodes during contact with insect gut contents. *Journal of Parasitology*, **79:**495–503.

67. Grewal, P. S., Gaugler, R. and Selvan, S. (1993a). Host Recognition by entomopathogenic nematodes: Behavioural response to contact with host feces. *Journal of Chemical Ecology*, **19:**1219-1231.

68. Griffin, C. T., Callaghan, K. M. and Dix. I. (2001). A self-fertile species of *Steinernema* from Indonesia: further evidence of convergent evolution amongst entomopathogenic nematodes? *Parasitology*, **122:**181-186.

69. Griswold, A. (2008). Genome packaging in prokaryotes: the circular chromosome of *E. coli*. *Nature Education*, **1:**57.

70. Haas, W. (2003). Parasitic worms: Strategies of host finding, recognition and invasion. *Zoology (Jena)*, **106:**349–364.

71. Haas, W., Haberi, B., Syafruddin, I., Idris, I., et al. (2005). Behavioural strategies used by the hookworms *Necator americanus* and *Ancylostoma duodenale* to find, recognize and invade the human host. *Parasitology Research*, **95:**30–39.

72. Hajek, A. (2004). Natural enemies: an introduction to biological control. Cambridge, UK: Cambridge University Press.

73. Hallen, E. A., Dillman, A. R., Hong, A. V., Zhang, Y., et al. (2011). A Sensory Code for Host Seeking in Parasitic Nematodes. *Current Biology*, **21:**377–383.

74. Heong, K. L., Tan, K. H., Garcia, C. P. F., Fabellar, L. T. and Lu, Z. ed. (2011). Insecticidal toxicology. In Heong, K. L., Tan, K. H., Garcia, C. P. F., Fabellar, L. T. and Lu, Z. (ed.). *Research Methods in Toxicology and Insecticide Resistance Monitoring of Rice Planthoppers*. Los Baños: International Rice Research Institute. pg. 29.

75. Holt, E. A. and Miller, S. W. (2010). Bioindicators: Using Organisms to Measure Environmental Impacts. *Nature Education Knowledge*, **3:**8.

76. Holt, J. G., Krieg, N. R., Sneath, P. H., Staley, J. T., et al. (1994). Bergey's manual of determinative bacteriology, 9th ed. 188–189.

77. Holmes, V. F. and N. R. Cozzarelli. (2000). Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proceedings of the National Academy of Sciences of the United States of America*. 97:1322–1324.

78. Hurley, P. M. (1998). Mode of carcinogenic action of pesticides inducing thyroid follicular cell tumors in rodents. *Environmental Health Perspectives*, **106:**437–445.

79. Johnigk, S. A. and Ehlers, R. (1999). *Endotokia matricida* in hermaphrodites of *Heterorhabditis* spp. and the effect of food supply. *Nematology*, **1:**717–726.

80. Jones, J. T., Haegeman, A., Danchin, E. G. J., Gaur, H. S., et al. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, **14:**946–961.

81. Kanost, M. R., Jiang, H. and Yu, X. Q. (2004). Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunological Reviews*, **198:**97–105.

82. Kaya, H. K. and Burlando, T. M. (1989). Development of *Steinernema feltiae* (Rhabditida: Steinernematidae) in diseased insect hosts. *Journal of Invertebrate Pathology*, **53:**164–168.

83. Kaya, H. K. (2002). Natural enemies and other antagonists. In: Gaugler, R. (Ed.), *Entomopathogenic Nematology*. CABI Publishing, Wallingford, UK, pp. 189–204.

84. Kaya, H. K. and Gaugler, R. (1993). Entomopathogenic nematodes. *Annual Review of Entomology*, **38:**181–206.

85. Kim, Y., Ji, D., Cho, S. and Park, Y. (2005). Two groups of entomopathogenic bacteria, *Photorhabdus* and *Xenorhabdus*, share an inhibitory action against phospholipase A2 to induce host immunodepression. *Journal of Invertebrate Pathology*, **89:**258-264.

86. Kiontke, K., Barrière, A., Kolotuev, I., Podbilewicz, B., et al. (2007). Trends, stasis and drift in the evolution of nematode vulva development. *Current Biology*, **17:**1925–1937.

87. Koppenhöfer, A. M. (2000). Nematodes. In: Lacey, L.A., Kaya, H.K. (Eds.), *Field Manual of Techniques in Invertebrate Pathology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 283–301.

88. Koppenhöfer, A. M. and Kaya, H. K. (1998). Synergism of imidacloprid and an entomopathogenic nematode: a novel approach to white grub (Coleoptera: Scarabaeidae) control in turfgrass. *Journal of Economic Entomology*, **91:**618-623.

89. Koppenhöfer, A. M. and Kaya, H. K. (1997). Additive and synergistic interactions between entomopathogenic nematodes and *Bacillus thuringiensis* for scarab grub control. *Biological Control*, **8:**131–137.

90. Lacey, L. A. and Georgis, R. (2012). Entomopathogenic Nematodes for Control of Insect Pests Above and Below Ground with Comments on Commercial Production. *Journal of Nematology*, **44:**218–225.

91. Lee, J. H., Dillman, A. R. and Hallem, E. A. (2016). Temperature-dependent changes in the host-seeking behaviours of parasitic nematodes. *BMC Biology*, **14:**1-17.

92. Leulier, F., Parquet, C., Pili-Floury, S., Ryu, J. H., et al. (2003). The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. *Nature Immunology*, **4:**478–484.

93. Levy, S. E. and Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, **17:**95-115.

94. Mannion, C., Winkler, H. E., Shapiro, D. I. and Gibb, T. (2000). Interactions between halofenozide and *Heterorhabditis marelatus* for control of the Japanese beetle, *Popillia japonica. Journal of Economic Entomology*, **93:**48–53.

95. Maxmen, A. (2013). Crop pests: Under attack. *Nature*, **501:**S15-S17.

96. Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research*, **15:**1767–1776.

97. Nappi, A. J. and Ottaviani, E. (2000). Cytotoxicity and cytotoxic molecules in invertebrates. *Bioessays*, **22:**469–480.

98. Nguyen, K. B. and Smart, G. C. Jr. (1996). Identification of entomopathogenic nematodes in the Steinernematidae and Heterorhabditidae (Nemata: Rhabditidae). *Journal of Nematology*, **28:**286-300.

99. Nicol, J. M., Turner, S. J., Coyne, D. L., Nijs den, L., et al. (2011). Current nematode threats to world agriculture. In: Jones, J. T., Gheysen, G. and Fenoll, C., (Eds.). *Genomics and Molecular Genetics of Plant–Nematode Interactions*. Heidelberg: Springer, pp. 21–44.

100. Nishimatsu, T. and Jackson, J. J. (1998). Interaction of insecticides, entomopathogenic nematodes, and larvae of the western corn rootworm (Coleoptera: Chrysomelidae). *Journal of Economic Entomology*, **91:**410–418.

101. Noe, J. P. (2007). Plant-Parasitic Nematodes. In: Trigiano, R. N. (ed.). *Plant Pathology Concepts and Laboratory Exercises*, 2nd ed. Boca Raton, FL: CRC Press.

102. Nowoshilow, S., Schloissnig, S., Fei, J., Dahl, A., et al., (2018). The axolotl genome and the evolution of key tissue formation regulators. *Nature*, **554:**50–55.

103. Pareek, C. S., Smoczynski, R. and Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, **52:**413–435.

104. Pettijohn, D. E. (1996). The nucleoid. In: Neidhardt, F. C., Curtis, R., Gross, C., Ingraham, J., Lin, E. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M. and Umbarger, H. E. (Eds.). *Escherichia coli and Salmonella: Cellular and Molecular Biology, Vol. II.* ASM, Washington, DC.

105. Pigliucci, M., Murren, C. J. and Schlichting C. D. (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, **209:**2362–2367.

106. Poinar, G. O. (1990). Biology and taxonomy of Steinernematidae and Heterorhabditidae. In: R. Gaugler and H. K. Kaya, (Eds.). *Entomopathogenic nematodes in biological control.* Boca Raton, Florida: CRC Press, Inc. pp. 93-115.

107. Poinar, G. O. (1993). Origins and phylogenetic relationships of the entomophilic rhabditis, *Heterorhabditis* and *Steinernema*. *Fundamental and Applied Nematology*, **16:**333–338.

108. Poinar, G. O. (2011). The evolutionary history of nematodes: As revealed in stone, amber and mummies. Brill, Leiden.

109. Poinar, G. O. Jr, Thomas, G. M. and Lighthart, B. (1990). Bioassay to determine the effect of commercial preparations of *Bacillus thuringiensis* on entomogenous rhabditoid nematodes. *Agriculture, Ecosystems and Environment*, **30:**195–202.

110. Poinar, G.O., (1983). The natural history of nematodes. Englewood Cliffs, NJ: Prentice-Hall, Inc. pg. 323.

111. Pretty, J. and Bharucha, Z. P. (2015). Integrated Pest Management for Sustainable Intensification of Agriculture in Asia and Africa. *Insects*, **6:**152–182.

112. Renn, N. (1998). Routes of penetration of the entomopathogenic nematode *Steinernema feltiae* attacking larval and adult houseflies (Musca domestica). *Journal of Invertebrate Pathology*, **72:**281–287.

113. Robinson, M. W. and Dalton, J. P. (2009). Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiases. *Philosophical*

transactions of the Royal Society of London. Series B, Biological sciences, **364:**2763–2776.

114. Rovesti, L. and Deseö, K. V. (1991). Compatibility of pesticides with the entomopathogenic nematode, *Heterorhabditis heliothidis*. *Nematologica*, **37:**113–122.

115. Sánchez-Moreno, S., Minoshima, H., Ferris, H. and Jackson, L. E. (2006). Linking soil properties and nematode community composition: effects of soil management on soil food. *Nematology*, **8:**703–715.

116. Sanger, F., Air, G. M., Barrell, B. G., Brown, A. R., et al. (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, **265:**687–695.

117. Saxena, D., Flores, S. and Stotzky, G. (1999). Transgenic plants: Insecticidal toxin in root exudates from *Bt* corn. *Nature*, **402:**480.

118. Schierenberg, E. and Sommer, R. J. (2014). Reproduction and development. In: Schmidt-Rhaesa, A. (ed.). Handbook of zoology: *Gastrotricha*, *Cycloneuralia* and *Gnathifera*. *Nematoda* vol. 2. Berlin: De Gruyter.

119. Selvan, S., Gaugler, R. and Lewis. E. E. (1993). Biochemical energy reserves of entomopathogenic nematodes. *Journal of Parasitology*, **79:**167–172.

120. Shapiro-Ilan, D. I., Gouge, D. H., Piggott, S. J. and Fife, J. P. (2006). Application technology and environmental considerations for use of entomopathogenic nematodes in biological control. *Biological Control*, **38:**124–133.

121. Sheiner, E. K., Sheiner, E., Hammel, R. D, Potashnik, G., et al. (2003). Effect of occupational exposures on male fertility: literature review. *Industrial Health*, **41:**55–62.

122. Simón, F., Siles-Lucas, M., Morchón, R., González-Miguel, J., et al. (2012). Human and animal dirofilariasis: the emergence of a zoonotic mosaic. *Clinical Microbiology Reviews*, **25:**507–544.

123. Steiner, G. (1923). *Aplectana kraussei* n. sp., eine in der Blattwespe *Lyda* sp. parasitierende Nematodenform, nebst Bemerkungenüber das Seitenorgan der parasitischen Nematoden. *Zentralblatt fur Bakteriologie Parasitenkunde Infektionskranheiten und Hygiene Abteilung*, **59:**14–18.

124. Steiner, G. (1929). *Neoaplectana glaseri* n. g., n. sp. (Oxyuridae) a new nemic parasite of the Japanese beetle (*Popillia japonica* Newm.). *Journal of the Washington Academy of Science*, **19:**436–440.

125. Stock, S. P. and Hunt, D. J. (2005). Morphology and systematics of nematodes used in biocontrol. In P. S. Grewal, R. U. Ehlers and D. I. Shapiro-Ilan (Eds.), Nematodes as biocontrol agents (pp. 3-43). CABI Publishing, Wallingford, UK.

126. Stock, S. P., Bird, D. M., Ghedin, E. and Godrich-Blair, H. (2011). Abstracts of the second nematode-bacteria symbioses meeting. *Symbiosis*, **55:**97–109.

127. Taylor, M. J., Hoerauf, A. and Bockarie, M. (2010). Lymphatic filariasis and onchocerciasis. *Lancet*, **376:**1175–1185.

128. Torres-Barragan, A., Suazo, A., Buhler, W. G. and Cardoza, Y. J. (2011). Studies on the entomopathogenicity and bacterial associates of the nematode *Oscheius carolinensis*. *Biological Control*, **59:**123–129.

129. The *C. elegans* Genome Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, **289:**2012–2018.

130. Thomas, G. M. and Poinar, G. O. Jr. (1979). *Xenorhabdus* gen. nov., a genus of entomopathogenic, nematophilic bacteria of the family *Enterobacteriaceae*. *International Journal of Systematic Bacteriology*, **29:**352–360.

131. Torr, P., Heritage, S. and Wilson, M. J. (2004). Vibrations as a novel signal for host location by parasitic nematodes. *International Journal for Parasitology*, **34:**997–999.

132. Van der Woude, M. W. and Baumler, A. J. (2004). Phase and Antigenic Variation in Bacteria. *Clinical Microbiology Reviews*, **17:**581–611.

133. Vanninen, I. (1990). Depletion of endogenous lipid reserves in *Steinernema feltiae* and *Heterorhabditis bacteriophora* and effect on infectivity. In Proceedings and abstracts, International Colloquium on Invertebrate Pathology and Microbial Control, 5th, August 20-24. Adelaide. Australia. p. 232.

134. Viney, M. and Diaz, A. (2012). Phenotypic plasticity in nematodes. *Worm*, **1:**98-106.

135. Volgyi, A., Fodor, A., Szentirmai, A. and Forst, S. (1998). Phase Variation in *Xenorhabdus nematophilus*. *Applied and Environmental Microbiology*, **64:**1188–1193.

136. Wang, J. X. and Liu, Z. M. (1983). The safety of the nematode, *Neoaplectana glaseri* STEINER, to vertebrates. II. A test on rabbits. *Natural Enemies of Insects*, **5:**241–242.

137. Wang, J. X., Huang, J.T. and Chen, G. S. (1984). The safety of the nematode, *Neoaplectana glaseri* STE INER, to vertebrates. III. A test on monkeys, *Macaca mulatta. Natural Enemies of Insects*, **6:**41–42.

138. Wang, J. X., Qui, L. H. and Liu, Z. M. (1983). The safety of the nematode, *Neoaplectana* glaseri STEINER, to vertebrates. III. A test on rats, *Macaca mulatta. Natural Enemies of Insects*, **5:**240–241.

139. Webster, J. M., Chen, G., Hu, K. and Li, J. (2002). Bacterial metabolites. In: Gaugler, R., (Ed.), Entomopathogenic nematology. New York, NY: CABI Publishing; pp. 99–114.

140. Wharton, D. A. (1986). A Functional Biology of Nematodes. Croom Helm, London, 192 pp.

141. Wolf, L. L. and Hainsworth, F. R. (1977). Temporal pattern of feeding by hummingbirds. *Animal Behaviour*. **25:**976–988.

142. Womersley, C. (1988). Morphological and biochemical adaptations to anhydrobiosis in artificially and naturally dehydrated populations of *Ditylenchus myceliophagus* (Nematoda). *American Zoologist*, **28:**76–82.

143. Wouts, W. M., Mráček, Z., Gerdin, S. and Bedding, R. A. (1982). *Neoaplectana* Steiner, 1929 a junior synonym of *Steinernema* Travassos, 1927 (Nematoda: Rhabditida). *Systematic Parasitology*, **4:**147–154.

144. Wray, G. A, Levintin, J. S and Shapiro, L. H. (1996). Molecular evidence for deep Precambrian divergences among metazoan phyla, *Science*, **274:**568–573.

145. Zhang, C., Liu, J., Sun, J., Yang, S., et al. (2008) *Heterorhabditidoides chongmingensis* gen. nov., sp. nov. (Rhabditida: Rhabditidae), a novel member of the entomopathogenic nematodes. *Journal of Invertebrate Pathology*, **98:**153–168.

146. Zhang, M. Q. (2002). Computational prediction of eukaryotic protein-coding genes. *Genetics*, **3:**698–709.

147. Zullini, A. (1976). Nematodes as indicators of river pollutions. *Nematologia Meditterranea*, **4:**13–22.

Chapter 2

The isolation and molecular identification of indigenous entomopathogenic nematodes

2.1 Abstract

South Africa offers an assortment of undisturbed habitats with a high diversity of plant and insect species. As such, an opportunity exists for the recovery of novel entomopathogenic nematode species with a greater tolerance to local conditions and pests, which could potentially be exploited as biological control agents. A small-scale survey was conducted in the North West Province and Gauteng of South Africa, with the main aim of uncovering a potentially novel nematode species as a candidate for subsequent genomic studies using Illumina sequencing technology. From a total of 100 soil samples, only 8 samples (8%) tested positive for the presence of entomopathogenic nematodes. The nematode recovery varied significantly between both provinces, with 63% and 37% of the positive samples originating from Gauteng and the North West Province, respectively. All nematode isolates were identified to genus-level based on the molecular characterization of the internal transcribed spacer (ITS) region of partial 18S rDNA sequences. The isolates (N1 and N5) recovered from the North West Province were identified as *Heterorhabditis bacteriophora* and *Heterorhabditis zealandica*, respectively, while the isolates recovered from Gauteng was identified as a new steinernematid that clustered to a new monophyletic group called the 'Cameroonian' clade, consisting of *S. sacchari*, *S.*

***cameroonense*, *S. nyetense* and *S. fabii*. The isolation of *Steinernema* sp. HBG28 is the first report of a novel steinernematid recovered from the grassland biome of the Gauteng province. Details regarding the distribution, habitat and host preferences, as well as the virulence of this new species would need to be further investigated.**

2.2 Introduction

Entomopathogenic nematodes of the genera *Heterorhabditis* Poinar, 1976 and *Steinernema* Travassos, 1927 are obligate insect parasites that have forged unique relationships with members of the *Enterobacteriacea* family, *Photorhabdus* Boemare et al., 1993 and *Xenorhabdus* Thomas and Poinar, 1979, respectively. This symbiotic association is extremely lethal towards many different types of insects, with both nematode and its bacterial counterpart working synergistically to infect and kill susceptible hosts, particularly insects in their larval stage.

The only free-living stage of these nematodes, referred to as infective juveniles, are non-feeding and exist in an arrested state of development while actively searching for soil-borne insects to infect. The infective juvenile stage plays a crucial role in the dispersal of these nematodes in the soil, as well as vectoring entomopathogenic bacteria from one insect target to the next. Upon locating and successfully invading a host, infective juveniles release their associated bacteria into the haemocoel of an insect, where rapid bacterial proliferation commences. This coincides with the release of various virulence factors such as bacterial toxins, hydrolytic enzymes, haemolysins and antimicrobial metabolites (Eleftherianos et al., 2010; ffrench-Constant and

Bowen, 2000; ffrench-Constant et al., 2007), which inundates the immune system of the host, resulting in insect death within 48 hours after infection.

Entomopathogenic nematodes possess several attributes that make them excellent candidates as a biological control agent, i.e. they are highly virulent; they have a high reproductive rate; they have a seemingly broad host range; they actively seek out and kill cryptic insect hosts (Lewis et al., 2006); they are environmentally-friendly; they are non-toxic to humans and other vertebrate organisms and they are compatible with various chemical pesticides. For these reasons, it is important to identify indigenous species of entomopathogenic nematodes to provide the insights necessary in understanding their habitat preferences and the environmental conditions to which they are adapted and to achieve a higher level of field efficacy when exploiting these nematodes for the purpose of biological pest control.

Entomopathogenic nematodes are widely distributed in soil around the world (Adams et al., 2006), except for Antarctica (Hominick et al., 1996; Hominick, 2002; Griffin et al., 1990). More than 70 recognized species of *Steinernema* and 25 species of *Heterorhabditis* have been described worldwide, with new species found each year (Stock and Goodrich-Blair, 2012). Over the past decades, realization of the practical application of these nematodes have garnered a growing interest. Several studies have demonstrated both the successes and failures of entomopathogenic nematodes in controlling various insect pests, and therefore, there have been several soil surveys carried out for these nematodes worldwide. These surveys serve a fundamental purpose, not only to accurately identify entomopathogenic nematode species, but to gain further insight into their geographical distribution, habitat preferences and conspecific interactions (Hominick et al., 1996).

For any significant sampling effort, there exists two isolation methods, which are commonly used for the retrieval of native entomopathogenic nematodes, i.e. (i.) the recovery of already-infected insect cadavers in the field, and (ii.) insect baiting of soil samples usually through "*Galleria* traps" (Piedra-Buena et al., 2015). Once recovered, entomopathogenic nematodes may be identified through various methods. Identification based solely on morphology and morphometric analyses has been successful in distinguishing various *Steinernema* species, but this approach has failed in identifying *Heterorhabditis* isolates to species level, due to a lack of discerning morphological features (Hominick, 2002; Hunt, 2007). Additionally, this identification method proves tedious when sorting through several field-collected samples, making the use of molecular techniques a more viable option. The most popular molecular marker used in nematode taxonomy and phylogenetics is the internal transcribed spacer (ITS) region of the ribosomal DNA (ITS1, 5.8S rRNA gene and ITS2) (Freire et al., 2010; Hominick et al., 1996), which reveals a significant degree of variation among species of *Steinernema* and *Heterorhabditis* (Adams et al., 1998).

It is important to explore native communities of entomopathogenic nematodes, particularly those adapted to local, non-disturbed environments. While South Africa is unique in biodiversity, it may potentially support an abundance of native entomopathogenic nematode species due to a combination of tropical and temperate climatic conditions, as well as a spectrum of habitat types. These habitat types range from subtropical thickets and woodlands in the east, semiarid Karoo in the west, savanna and bushveld in the north, with grasslands localized in central parts of the country. This diversity both in climate and habitat may lend itself to an equally diverse population of entomopathogenic nematode species, yet to be uncovered.

The first notable report of entomopathogenic nematodes in South Africa was documented in 1953, when a *Steinernema* nematode was recovered from a maize beetle, *Heteronychus arator* in a maize field in the Eastern Cape province (Harington, 1953). A survey conducted in KwaZulu-Natal several years later led to the retrieval of one *Heterorhabditis* and three *Steinernema* species, which were used in subsequent field trials to control the sugarcane stalk-borer, *Eldana saccharina* Walker (Spaull, 1988; 1990; 1991). In the years that followed, there have been many surveys performed in the Western Cape and Eastern Cape of South Africa (Malan et al., 2006; 2008), with the prime focus of uncovering new species and strains adapted to local environments, climates and insect pest populations. Several new species have been described to date, including *Steinernema khoisanae* (Nguyen et al., 2006), *Heterorhabditis safricana* (Malan et al., 2008), *S. citrae* (Stokwe et al., 2011) and *S. sacchari* (Nthenga et al., 2014), *S. jeffreyense* (Malan et al., 2016), *S. fabii* (Abate et al., 2016).

While there are species of entomopathogenic nematodes that have been commercialized (Lacey and Georgis, 2012) and have been used with varying success to control certain problematic insect pests in other countries, these exotic species may not perform as effectively under local conditions. Importantly, the application of exotic species of entomopathogenic nematodes in South Africa poses various concerns, such as potential ecological displacement of indigenous nematode species; out-competition and replacement of naturally-occurring nematodes, which could lead to local extinction (Choo et al., 1995); effects on non-target populations (Ehlers, 2005), poor adaptation against target pests, as well as strict policies governing the importation of exotic organisms in South Africa (amendment of Act 18 of 1989 under the Agricultural Pest Act No. 36 of 1947) (Malan et al., 2011).

Although most research has focused on identifying entomopathogenic nematodes species and strains that are adapted to indigenous conditions and target pests, more to promote their use as biological control agents in native environments, these nematodes may also serve as models for understanding various processes involving biological, ecological and evolutionary mechanisms with other soil-inhabiting organisms (Burnell and Stock, 2000; Goodrich-Blair and Clarke, 2007; Stock and Goodrich-Blair, 2008).

2.2.1 Aims and objectives

The aim of this study was to not only isolate and identify indigenous entomopathogenic nematode species by conducting a small-scale survey in the North West Province and Gauteng, whilst investigating the natural distribution of entomopathogenic nematodes in these parts of South Africa, but to also uncover potentially novel nematode species that could be used for subsequent genomic studies using Illumina Technology.

2.3 Materials and methods

2.3.1 Insect source

Larvae of *Galleria mellonella* (L.) (Lepidoptera: Pyralidae), also known as the greater wax moth, was selected as the model host organism for this study as they can be easily bred under laboratory conditions and are extremely susceptible to entomopathogenic nematode infection (Kaya and Stock, 1997). The *G. mellonella* larvae were reared on an artificial diet adapted from Woodring and Kaya (1988), consisting of Pronutro wheat-free cereal, glycerol, honey, boiled water, yeast extract and sodium benzoate. All ingredients were thoroughly mixed, and the diet media was autoclaved at 121°C

for 15 min. Adult wax moths were transferred into glass containers containing fresh culture medium and wax paper. Wax moth cultures were maintained at 25°C in a growth chamber.

2.3.2 Nematode source

2.3.2.1 Regions sampled

Brits is a farming town of citrus, vegetable and grain crops in the North West Province of South Africa. The natural vegetation type in Brits is bushveld, though the savanna biome occupies most parts of the North West Province. Suikerbosrand Nature Reserve is found in Heidelberg in the Gauteng province and was the second sampling area. This region mostly consists of grassland with the climate largely influenced by altitude.

2.3.2.2 Soil samples

A total of 50 soil samples were collected from each province between April 2014 and May 2014. Samples were taken from random areas in Brits (25.6100° S, 27.7960° E), North West Province (Figure 2.1A-C). and in the Suikerbosrand Nature Reserve (26.4838° S, 28.2293° E), Gauteng (Figure 2.1D-E). Of the non-disturbed locations sampled, fifty of the sampling sites were areas consisting of various vegetation types (bushveld, thicket, shrubs, grasses) in North West Province and fifty sites comprised mainly of grassland in Gauteng. Each soil sample (approximately 1 kg) was taken from a soil depth ranging from 15-30 cm by using a spade (Kaya and Stock, 1997); and the samples were placed in sealable plastic containers, which were transported to the laboratory at the University of Witwatersrand. Soil samples were maintained at room temperature.

2.3.2.3 Insect baiting method with *Galleria mellonella* larvae

Prior to testing for the presence of entomopathogenic nematodes, all soil samples were thoroughly mixed, and the moisture content of each sample was adjusted to 8% (w/w), using sterile distilled water. Each soil sample was tested separately for the detection of entomopathogenic nematodes using the insect baiting method (Bedding and Akhurst, 1975). Fifteen to twenty-five last instar *G. mellonella* larvae were placed on the soil surface of each container. The containers were covered with a lid and maintained at 25°C for 2-5 days. Insect cadavers displaying signs of entomopathogenic nematode infection were later recovered from each container, cleaned by carefully removing remnant soil particles off the cuticles, and surface-sterilized with a 70% (v/v) ethanol solution. Cadavers were sorted according to their exhibited symptoms (i.e. colour of insect cadaver, presence and colour of spots) prior to being placed in modified White traps (Kaya and Stock, 1997).

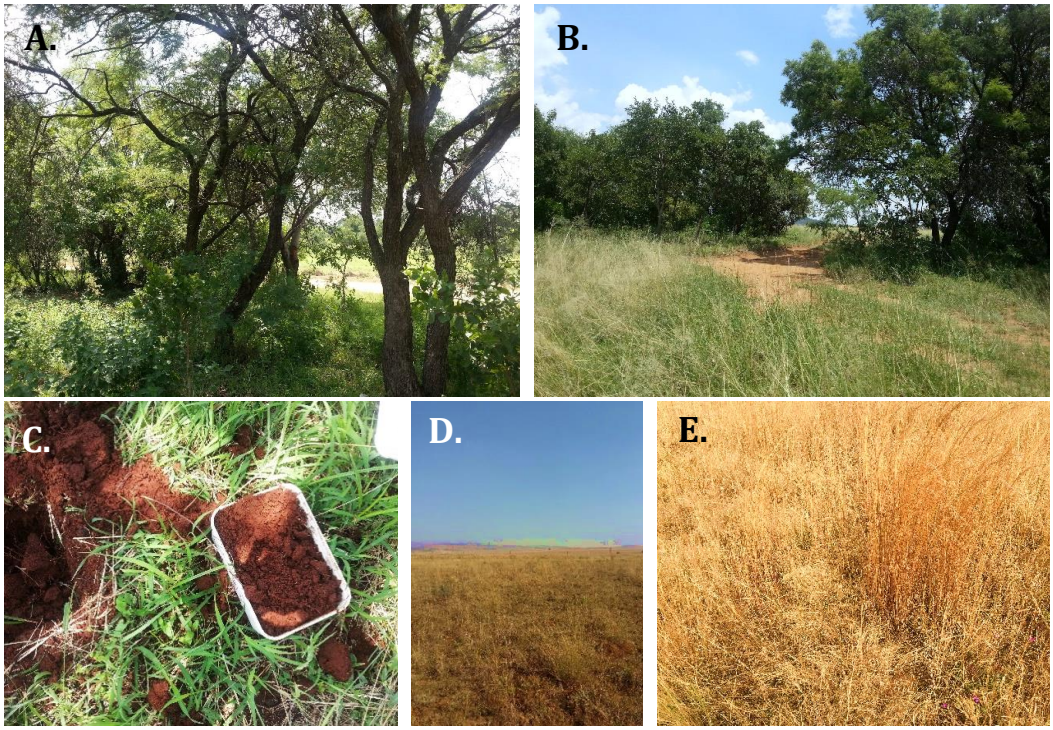


Figure 2.1. Sample areas selected for the screening of indigenous entomopathogenic nematodes. A-C: Soil samples were collected from random sites in Brits, North West Province. This part of the North West Province was dominated by shrub bushveld vegetation (Savanna) and consisted of grasses typical of the grassland biome. **D:** Soil samples were also collected from the Suikerbosrand Nature Reserve in Gauteng. Image taken with permission from Carol Motlhala (University of Witwatersrand). **E:** The sampling site in the Suikerbosrand Nature Reserve comprised open grasslands.

2.3.2.4 Nematode recovery using White traps

The modified White trap method was used to harvest putative infective juveniles, as described by Kaya and Stock (1997), making use of the natural migration of infective juveniles away from an insect cadaver. A single White trap consisted of an inverted Petri dish lid (30 mm x 10 mm), placed inside a larger Petri dish (90 mm x 15 mm) and covered with circular filter paper. Sterile tap water was added to the large Petri dish and insect cadavers were then placed in the centre of the moistened filter paper (Figure 2.4). Modified White traps were maintained at room temperature to allow for

the emergence of infective juvenile nematodes. If the presence of infective juveniles was not observed in a White trap within 14 days, it was discarded.

2.3.2.5 Koch's postulates

Koch's postulates were used to establish pathogenicity of the nematode-bacterium complex (Lacey, 1997). Koch's postulates refer to a list of criteria established to confirm whether a specific microorganism is the causative agent of disease, beginning with the isolation of the microorganism of interest, which is then used to infect a healthy host. If similar symptoms develop in the new host, the infection is considered to have originated from the microorganism in question (Poinar, 1975). To confirm the pathogenicity of putative infective juvenile nematodes, a method adapted from Kaya and Stock (1997) was used. Harvested infective juveniles recovered from White traps were used to inoculate last instar *G. mellonella* larvae in Petri dishes containing sterile river sand adjusted to a moisture content of 8% (w/w). The Petri dishes were inverted to prevent live *G. mellonella* larvae from escaping. Larval mortality was monitored after 2-3 days. Dead larvae were collected and placed on new White traps. The emergence of infective juveniles from host cadavers confirmed the pathogenicity of putative infective juveniles (Figure 2.2A and 2.2B). Host cadavers accompanied by a foul odour indicated unsuccessful culturing and were discarded.

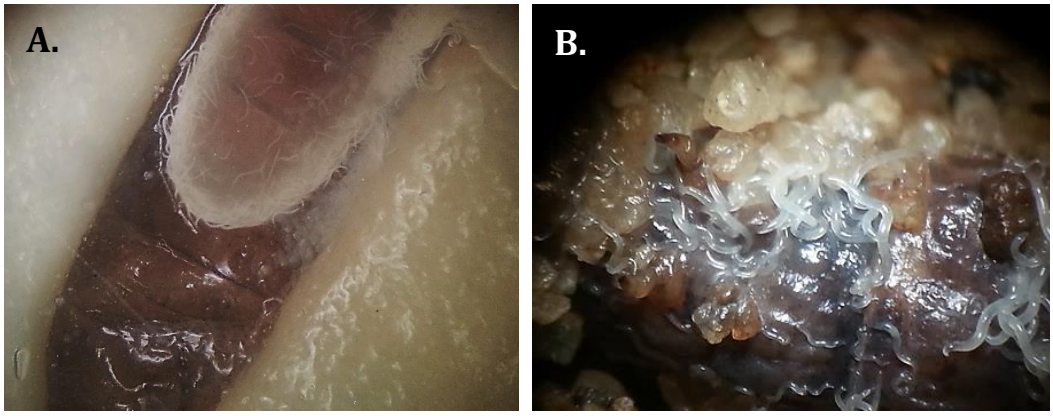


Figure 2.2. Entomopathogenic nematodes shown parasitizing *Galleria mellonella* larvae. A: Emerging nematodes pooling around a *G. mellonella* cadaver and **B:** Infective juveniles gathering on the surface of a *G. mellonella* cadaver in river sand, viewed under a dissecting microscope.

2.3.2.6 Maintenance of entomopathogenic nematodes

The infective juveniles were maintained by recycling through *G. mellonella* larvae for molecular characterization. Surplus infective juveniles were added to original soil samples intermittently and were still infective after several months (±25 months) of storage.

2.3.3 Genomic DNA extraction of infective juvenile nematodes

After several rounds of sub-culturing on White traps, newly emerged infective juveniles were harvested for genomic DNA isolation and surface sterilized in a 0.1% hypochlorite solution for 1-3 hrs. The infective juveniles were washed three times with sterile distilled water prior to DNA extraction. Total genomic DNA was extracted from the infective juvenile nematodes using the Gentra® Puregene® DNA Purification Kit (Gentra Systems 2003), according to the protocol supplied by the manufacturer. Quantification of the extracted DNA was assessed using a NanoDrop®

ND-1000 spectrophotometer (Nanodrop Technologies). DNA samples were stored at 4°C until needed.

2.3.4 Amplification by Polymerase Chain Reaction

Nuclear ribosomal regions were amplified from extracted genomic DNA by polymerase chain reaction. The TW81 forward primer (5'-GCGGATCCGTTTCCGTAGGTGAACCTGC -3') and the AB28 reverse primer (5'-GCGGATCCATATGCTTAAGTTCAGCGGGT -3') were used in the PCR reaction for amplification of the internal transcribed spacer (ITS) region of ribosomal DNA (rDNA) (Hominick et al., 1997). For amplification of the 28S ribosomal DNA (D2D3) region, the primer 502 (5'-CAAGTACCGTGAGGGAAAGTTGC-3') and 503 (5'-CCTTGGTCCGTGTTTCAAGACG-3') were used, as previously reported by Stock et al. (2001).

Each PCR reaction consisted of 3 µL of each primer, 3 µL of the DNA suspension, which served as template DNA, 14 µL nuclease-free water, together with 25 µL of DreamTaq Green PCR Master Mix (2X), which contained DNA polymerase, PCR buffer, dATP, dCTP, dGTP and dTTP (0.4 mM each), and 4 mM $MgCl_2$ (Thermo Scientific). The final reaction volume was 50 µl. The amplification reaction was performed in a thermocycler (GeneAmp PCR system 2700®), using the following cycling conditions: an initial denaturation cycle of 95°C for 3 min, followed by 35 cycles of denaturation at 94°C for 30 sec, amplification at an annealing temperature of 60°C for 45 sec and an extension period at 72°C for 90 sec. A final extension cycle at 72°C for 7 min followed the last cycle.

2.3.5 Sequencing of the 18S and 28S rDNA

The PCR products were sequenced on an automated sequencer (ABI 3130XL) at Inqaba Biotechnical Industries, Pretoria, South Africa. Sequence editing and verification of base calls were performed using the software, FinchTV version 1.4 (Geospiza, Seattle, WA, USA).

2.3.6 Multiple alignment of sequence data

The final gene sequences were queried against the National Centre for Biotechnology Information (NCBI) database to compare for similarity against sequences of other reference entomopathogenic nematode species, utilizing the Basic Local Alignment Search Tool (BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi). All nucleotide sequences were deposited in GenBank, and aligned, along with other partial gene sequences with the highest similarity, by MUSCLE (Edgar, 2004).

2.3.7 Phylogenetic analysis

Phylogenetic analyses were performed using sequences for two molecular markers, i.e. the internal transcribed spacer (ITS) and the D2D3 region of rDNA. The phylogenetic relationships of isolated entomopathogenic nematode species and other closely-related members of entomopathogenic nematodes were observed using the program, Molecular Evolutionary Genetics Analysis, version 7 (MEGA 7) (Tamura et al., 2011). Phylogenetic relationships were inferred using Maximum Likelihood (ML) and Maximum Parsimony (MP) methods. The best-fit substitution models were also selected using MEGA 7 and the phylogenetic trees were sampled with a bootstrap value of 1000 replicates.

For comparison of the ITS regions, the sequences obtained for the *Heterorhabditis* isolates were compared with the following *Heterorhabditis* taxa available on Genbank (NCBI): *H. bacteriophora* (MF535289); *H. georgiana* (HQ225901); *H. indica* (AY321483); *H. amazonensis* (DQ665222); *H. baujardi* (AF548768); *H. floridensis* (DQ372922); *H. mexicana* (AY321478); *H. zealandica* (AY321481); *H. downesi* (AY321482); *H. megidis* (AY321480); *H. marelatus* (AY321479); *H. atacamensis* (HM230723); *H. safricana* (EF488006), with *Caenorhabditis elegans* (FJ589008) used as an outgroup taxon.

Similarly, for *Steinernema* isolates, sequences of the following *Steinernema* taxa were used for comparison of the ITS regions, i.e.: *S. fabii* (KR527216); *S. sacchari* (KC633095); *S. cameroonense* (JX985267); *S. nyetense* (JX985266); *S. robustispiculum* (AY355442); *S. ashiunense* (DQ354694); *S. monticolum* (AF122017); *S. schliemanni* (HM778112); *S. akhursti* (DQ375757); *S. weiseri* (AY171268); *S. litorale* (JF892546); *S. ichnusae* (EU421129); *S. feltiae* (AY230185); *S. citrae* (FJ235074); *S. hebeiense* (DQ105794); *S. jollieti* (AY171265); *S. silvaticum* (AY230162); *S. kraussei* (AY230176); *S. texanum* (EF152568); *S. sangi* (AY355441); *S. oregonense* (AF122019); *S. xueshanense* (FJ666052); *S. cholashanense* (EF431959), with *C. elegans* (EU13007) used as an outgroup taxon.

For the D2D3 regions, sequences of the following *Steinernema* taxa were used for comparison, i.e.: *S. fabii* (KR527217); *S. sacchari* (KC633096); *S. cameroonense* (JX985265); *S. nyetense* (JX985264); *S. hermaphroditum* (AY598358); *S. scarabaei* (AY172023); *S. guangdongense* (AY169558); *S. longicaudatum* (AF331901); *S. ethiopiense* (JN651409); *S. khoisanae* (DQ314289); *S. glaseri* (AF331908); *S. cubanum* (AF331889); *S. boemarei* (FJ152415); *S. arenarium* (AF331892); *S. diaprepesi*

(DQ350609); *S. puertoricense* (AF331903); *S. austral (*FJ235126); *S. brazilense* (FJ410326); *S. citrae* (GU004534); *S. feltiae* (JF920966); *S. monticolum* (EF439651); *S. schliemanni* (HM778113); *S. scapterisci* (HM140695).

2.4 Results

2.4.1 Isolation of entomopathogenic nematodes

A total of 100 soil samples were collected from the North West Province and Gauteng. Of which, eight samples (8%) tested positive for the presence of entomopathogenic nematodes. Positive samples were collected from natural habitats at eight locations shown in Figure 2.3. Nematode recovery varied significantly between the two provinces (Table 2.1). The results from the different provinces demonstrated that the highest proportion (63%) of positive samples were detected from soil samples retrieved from the grasslands of the Suikerbosrand Nature Reserve in Gauteng. In contrast, the percentage of entomopathogenic nematodes recovered from Brits was 37%. Sample sites within Brits were dominated by shrub bushveld vegetation and consisted of grasses typical of the Grassland biome. Furthermore, entomopathogenic nematodes were recovered from soil samples consisting of sandy clay loam soil and sandy loam soil. In contrast, detection of entomopathogenic nematodes was not observed in clay soil.

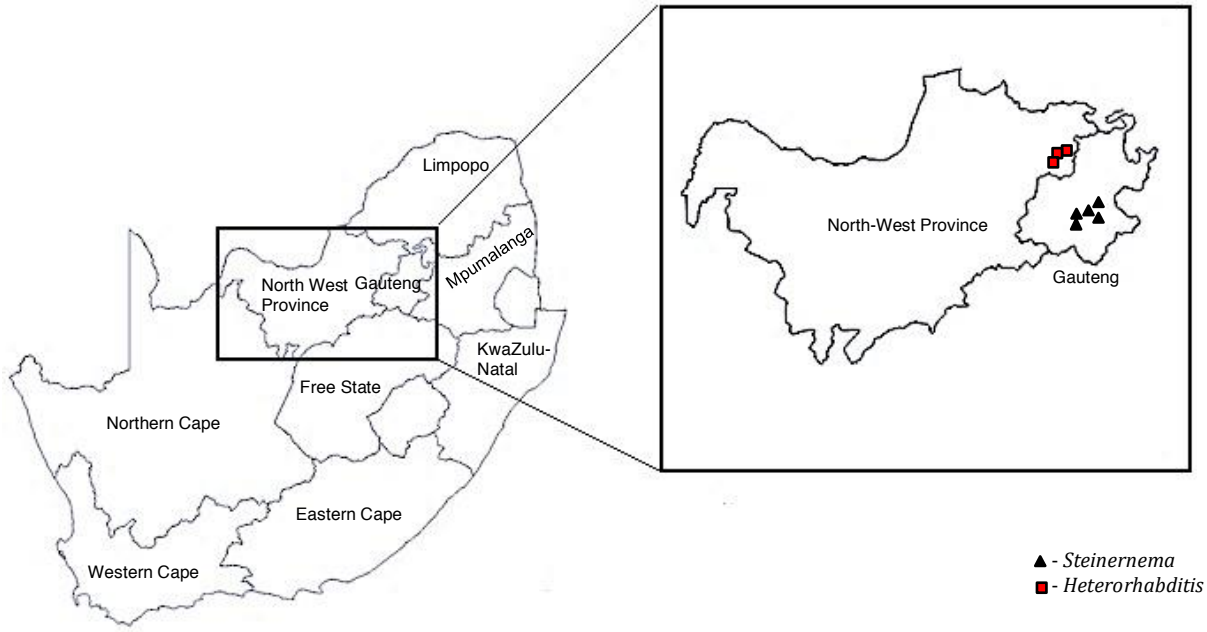


Figure 2.3. Geographical distribution of sampling sites. Soil collected from the North West Province and Gauteng tested positive for the presence of *Steinernema* and *Heterorhabditis* nematodes.

Table 2. 1. Occurrence of entomopathogenic nematodes in soil samples collected from two provinces of South Africa, including GPS coordinates and GenBank accession numbers.

Location	GPS coordinates	Isolate no.	Species	GenBank accession no.
North West Province	-25.720103, 27.764724	N1	*H. bacteriophora*	MG593947
	-25.719900, 27.764989	N2	*H. bacteriophora*	MG593948
	-25.716921, 27.746781	N5	*H. zealandica*	KT715757
Gauteng	-26.437654, 28.198494	HBG28	*Steinernema* sp.	KJ877686
	-26.437682, 28.198461	HBG29C	*Steinernema* sp.	MG272494
	-26.437627, 28.198419	HBG29F	*Steinernema* sp.	MG272495
	-26.437555, 28.198297	HBG41	*Steinernema* sp.	MG272496
	-26.437484, 28.198219	HBG44	*Steinernema* sp.	MG272497

2.4.2 Preliminary identification of nematode isolates

Infective juveniles of the heterorhabditid isolates (N1, N2, N5) were recovered from non-disturbed sites in Brits, North West Province, whilst steinernematid isolates (HBG28, HBG29C, HBG29F, HBG41, HBG44) were found in Heidelberg, Gauteng. Nematode isolates were sorted according to the symptoms exhibited by infected *G. mellonella* cadavers (Figure 2.4). Based solely on the colour of the insect cadavers, nematode isolates were categorized as possible heterorhabditids due to the dark maroon, burgundy colour observed and the absence of black melanized spots, as shown in Figure 2.4 A and C. Steinernematids were characterized by a brown, tan colour with the presence of black spots, as represented in Figure 2.4 B and D.

Furthermore, the emergence of infective juveniles from insect cadavers differed significantly between both nematode isolates. Steinernematid infective juveniles (isolate HBG28) emerged 4-5 days after insect death and was observed to adopt an intermediate foraging approach, i.e. both cruiser and ambusher foraging behaviour. Heterorhabditids (isolate N1 and N5) emerged after 7-14 days.

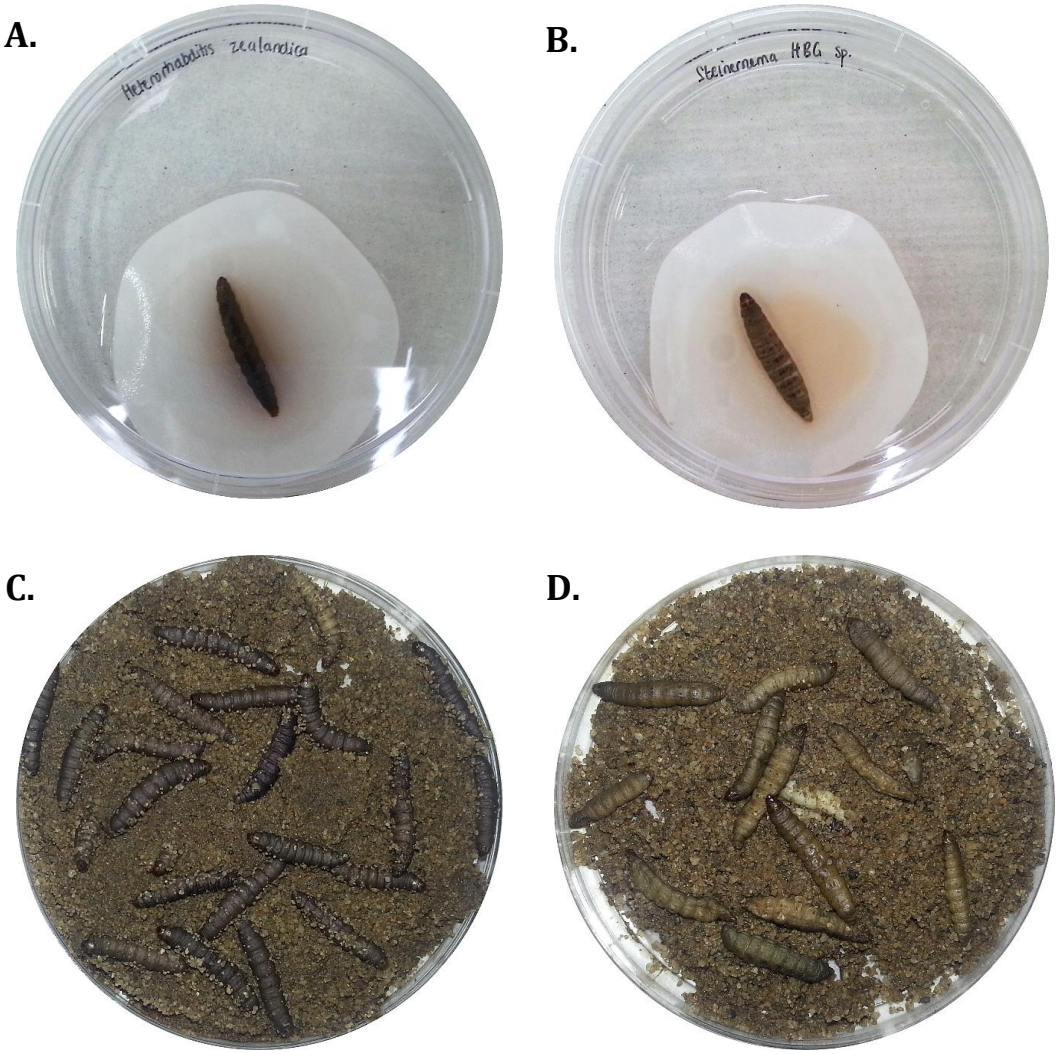


Figure 2.4. Symptom variation of *Galleria mellonella* larvae during exposure to entomopathogenic nematode isolates. Modified White trap used in the recovery of newly emerged third-stage infective juveniles from a *Galleria mellonella* insect cadaver, infected with **A:** *Heterorhabditis* (isolate N5) and **B:** *Steinernema* (isolate HBG28) entomopathogenic nematodes. **C:** Infection by possible heterorhabditid nematodes (isolate N5) and their symbiotic bacteria turned freshly killed *G. mellonella* larvae a dark maroon, burgundy colour. **D:** Infection by steinernematid nematodes (isolate HBG28) and its symbiotic bacteria turned freshly killed *G. mellonella* larvae a brown, tan colour.

2.4.3 Molecular identification of entomopathogenic nematodes

Molecular characterization of all recovered nematode isolates was performed based on the analysis of the ITS rDNA sequences. Of the eight nematode isolates, three heterorhabditids (37%) and five steinernematids (63%) were recovered in this study. Overall, three different nematode species were isolated in this survey (Table 3).

Heterorhabditis isolates N1 and N5 showed the highest homology with *Heterorhabditis bacteriophora* and *H. zealandica* (Poinar, 1990), respectively, with 100% sequence identity. Interestingly, both *Heterorhabditis* species were recovered from the North West Province, though this is the first record of *H. zealandica* occurring within this region. Both *H. bacteriophora* and *H. zealandica* appear to be well adapted to the hot semi-arid climate distinctive of the North West Province.

Recovered steinernematid isolates (HBG28, HBG29C, HBG29F, HBG41, HBG44) were molecularly characterized by analysis of rDNA sequences and were shown to represent a new *Steinernema* species at the time at which this survey was conducted. As such, the *Steinernema* species was selected for further characterization, as well as whole genome sequencing using Illumina technology. The partial ITS-rDNA nucleotide sequence of the novel *Steinernema* species was submitted to GenBank in 2014, under the name '*Steinernema* sp. HBG28', with the accession number, KJ877686. Notably, this *Steinernema* sp., appears to be well adapted to temperate, dry climates in the grassland region of South Africa.

2.4.4 Phylogenetic analyses

2.4.4.1 *Heterorhabditis* isolates

Phylogenetic analysis of all available sequences belonging to closed related *Heterorhabditis* isolates with those accumulated during the survey, based on the maximum parsimony method, are presented in Figure 2.5. Isolates N1 and N5 were shown to cluster with *H. bacteriophora* and *H. zealandica*, respectively. Estimates of the evolutionary divergence between *Heterorhabditis* isolate, N1 and other *Heterorhabditis* species, based on pairwise analyses of 14 ITS-rDNA sequences is presented in Table 2.2. *Heterorhabditis* isolate N1 was shown to be closely related to *Heterorhabditis bacteriophora*, and most divergent from *H. safricana.* Similarly, evolutionary divergence between the *Heterorhabditis* isolate, N5 and *Heterorhabditis* species, based on pairwise analyses of ITS-DNA sequences, indicated that *Heterorhabditis zealandica* was the closest species and *H. mexicana* was the most divergent species (Table 2.3).

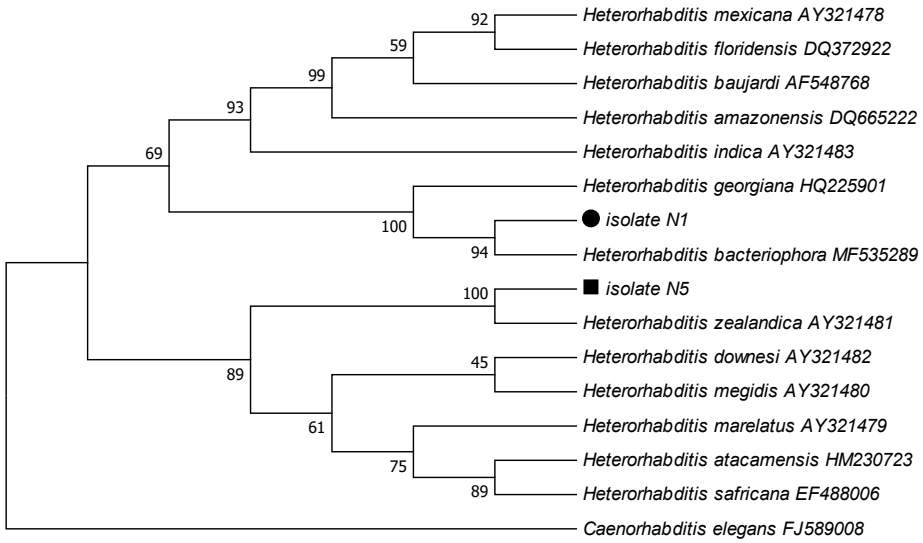


Figure 2.5. Phylogenetic analysis by Maximum Likelihood method. The evolutionary history of two *Heterorhabditis* isolates, isolate N1 (●) and isolate N5 (■) based on analysis of the ITS regions of the 18S rDNA. The phylogenetic tree was constructed by using the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura, 1980). Bootstrap values of 1000 replicates are specified below the nodes as percentages. Accession numbers have also been specified. *Caenorhabditis elegans* was selected as the outgroup taxon.

Table 2.2. Estimates of evolutionary divergence between sequences of the ITS region of 14 *Heterorhabditis* spp. The number of base substitutions per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal and were obtained by a bootstrap procedure (1000 replicates). Analyses were conducted using the Jukes-Cantor model. The analysis involved 14 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of 668 positions in the final dataset.

No.	Species	ITS Region													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	*Heterorhabditis* N1		0.000	0.005	0.019	0.019	0.019	0.020	0.021	0.019	0.016	0.018	0.016	0.016	0.016
2	*H. bacteriophora*	0.000		0.005	0.019	0.019	0.019	0.020	0.021	0.019	0.016	0.018	0.016	0.016	0.016
3	*H. georgiana*	0.018	0.018		0.019	0.019	0.020	0.020	0.021	0.020	0.017	0.018	0.016	0.016	0.017
4	*H. indica*	0.227	0.227	0.223		0.014	0.013	0.015	0.015	0.024	0.020	0.021	0.020	0.019	0.020
5	*H. amazonensis*	0.229	0.229	0.229	0.118		0.004	0.006	0.006	0.023	0.019	0.021	0.020	0.019	0.020
6	*H. baujardi*	0.229	0.229	0.231	0.116	0.012		0.005	0.006	0.024	0.020	0.021	0.020	0.019	0.020
7	*H. floridensis*	0.246	0.246	0.248	0.129	0.021	0.015		0.005	0.024	0.021	0.022	0.021	0.020	0.021
8	*H. mexicana*	0.248	0.248	0.250	0.127	0.026	0.024	0.017		0.024	0.021	0.022	0.021	0.020	0.021
9	*H. zealandica*	0.221	0.221	0.231	0.291	0.273	0.280	0.293	0.284		0.013	0.015	0.013	0.013	0.014
10	*H. downesi*	0.160	0.160	0.173	0.240	0.229	0.233	0.250	0.244	0.101		0.008	0.008	0.007	0.008
11	*H. megidis*	0.192	0.192	0.198	0.271	0.258	0.265	0.275	0.269	0.127	0.050		0.011	0.010	0.010
12	*H. marelatus*	0.158	0.158	0.171	0.242	0.227	0.231	0.248	0.242	0.101	0.046	0.077		0.006	0.006
13	*H. atacamensis*	0.152	0.152	0.161	0.225	0.221	0.227	0.242	0.235	0.096	0.035	0.067	0.024		0.005
14	*H. safricana*	0.160	0.160	0.167	0.238	0.233	0.238	0.254	0.248	0.106	0.040	0.071	0.027	0.015	

Table 2.3. Estimates of evolutionary divergence between sequences of the ITS region of 14 *Heterorhabditis* spp. The number of base substitutions per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal and were obtained by a bootstrap procedure (1000 replicates). Analyses were conducted using the Jukes-Cantor model. The analysis involved 14 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of 667 positions in the final dataset.

No.	Species	ITS Region													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	*Heterorhabditis* N5		0.002	0.014	0.013	0.013	0.013	0.020	0.021	0.023	0.024	0.025	0.013	0.025	0.025
2	*H. zealandica*	0.003		0.014	0.013	0.012	0.013	0.020	0.021	0.023	0.024	0.024	0.012	0.025	0.025
3	*H. megidis*	0.124	0.125		0.010	0.010	0.010	0.019	0.019	0.021	0.022	0.022	0.008	0.023	0.022
4	*H. marelatus*	0.101	0.101	0.081		0.006	0.007	0.016	0.016	0.020	0.021	0.021	0.008	0.022	0.022
5	*H. atacamensis*	0.098	0.098	0.069	0.026		0.005	0.016	0.016	0.019	0.021	0.021	0.007	0.022	0.021
6	*H. safricana*	0.106	0.106	0.074	0.031	0.017		0.016	0.016	0.020	0.021	0.022	0.008	0.022	0.022
7	*H. bacteriophora*	0.216	0.216	0.186	0.152	0.145	0.154		0.005	0.020	0.021	0.021	0.016	0.022	0.022
8	*H. georgiana*	0.226	0.226	0.192	0.165	0.154	0.162	0.018		0.020	0.021	0.021	0.017	0.022	0.022
9	*H. indica*	0.280	0.280	0.259	0.232	0.216	0.230	0.224	0.224		0.014	0.014	0.020	0.015	0.015
10	*H. amazonensis*	0.280	0.280	0.259	0.234	0.230	0.240	0.238	0.238	0.132		0.004	0.021	0.006	0.006
11	*H. baujardi*	0.282	0.282	0.263	0.234	0.230	0.240	0.236	0.238	0.127	0.012		0.021	0.004	0.006
12	*H. downesi*	0.099	0.099	0.051	0.048	0.035	0.042	0.152	0.165	0.230	0.234	0.234		0.022	0.021
13	*H. floridensis*	0.296	0.296	0.274	0.250	0.246	0.257	0.250	0.253	0.140	0.021	0.015	0.250		0.005
14	*H. mexicana*	0.289	0.289	0.267	0.246	0.242	0.253	0.255	0.257	0.143	0.026	0.024	0.246	0.017	

2.4.4.2 *Steinernema* isolates

Phylogenetic relationships of *Steinernema* sp. HBG28 with other related species of *Steinernema*, inferred by ITS rDNA sequences and based on maximum parsimony indicates that the unknown *Steinernema* species clusters with a new monophyletic group, referred to as the 'Cameroonian' clade (Nthenga et al., 2014; Abate et al., 2016). This new clade consists of *S. sacchari*, *S. cameroonense*, *S. nyetense* and *S. fabii* (Figure 2.6). Further analysis of the phylogenetic tree in Figure 2.6 shows that *Steinernema* sp. HBG28 shares common ancestry with *S. fabii*, with a low bootstrap support of 55%, suggesting that the *Steinernema* isolate recovered in this study may possibly be a new species that belongs to the 'Cameroonian' clade. Evolutionary divergence inferred by aligned ITS-rDNA sequences are presented in Table 2.4. *Steinernema* sp. HBG28 is shown to be closely related to *S. fabii*, followed by *S. sacchari*, *S. cameroonense* and *S. nyetense*, and most divergent to *S. cholashanense*.

Maximum parsimony analysis of the D2D3 rDNA sequences also placed *Steinernema* sp. HBG28 in the *Cameroonian*-clade and indicated shared common ancestry with *S. fabii* with bootstrap support of 97% (Figure 2.7). Also, S. *cameroonense* was grouped with *S. sacchari*, in place of *S. nyetense*, though this inferred relationship had a low bootstrap support of 47%. Evolutionary divergence, inferred by aligned D2D3-rDNA sequences, demonstrates similar results, in that *Steinernema* sp. HBG28 is shown to be closely related to *S. fabii*, *S. sacchari*, *S. cameroonse* and *S. nyetense*, however it is most divergent from *S. schliemanni* (Table 2.5).

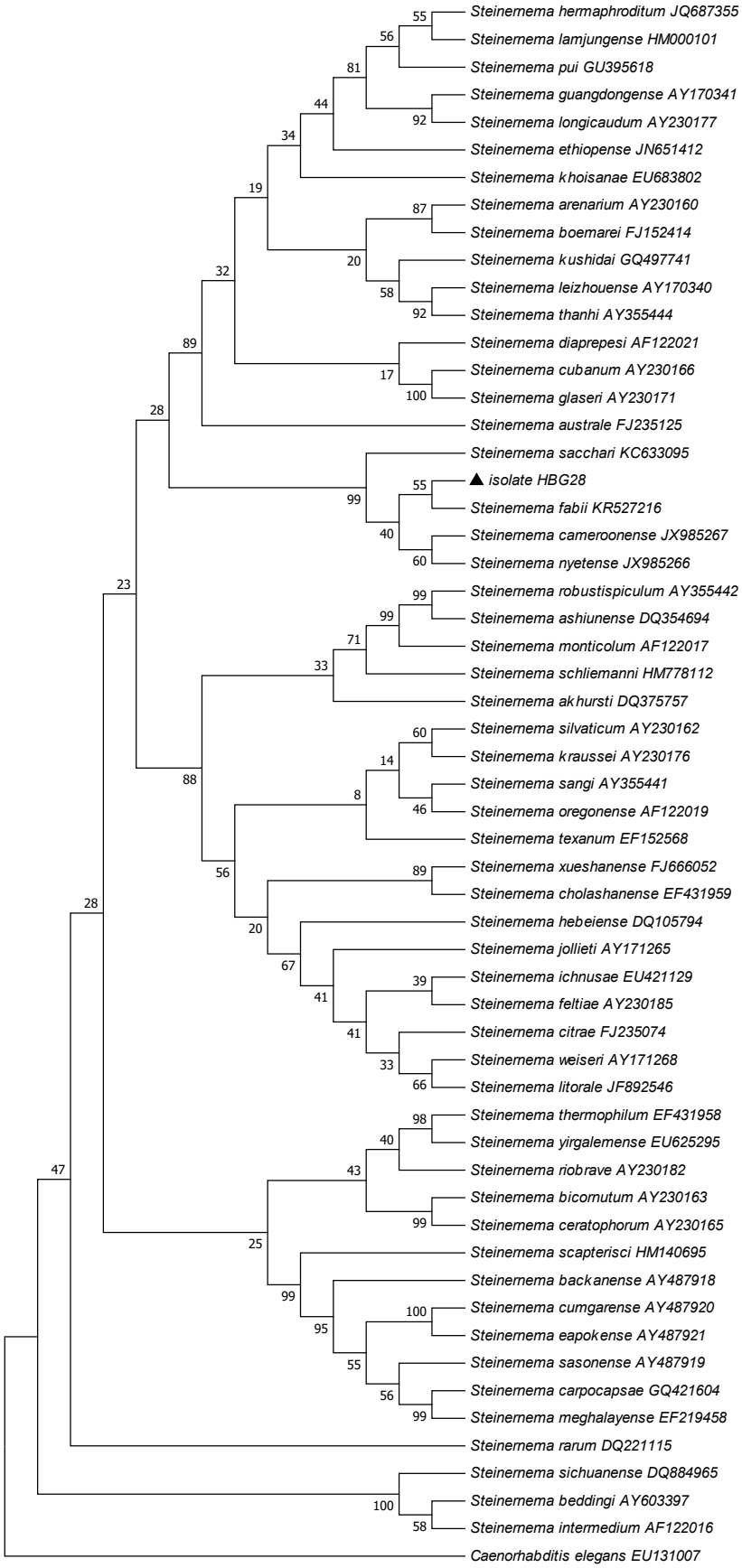


Figure 2.6. Phylogenetic relationships of *Steinernema sp.* HBG28 (▲) with 56 species of *Steinernema* based on partial ITS-rDNA nucleotide sequences from GenBank. The phylogenetic tree was constructed by using Maximum Parsimony based on the General Time Reversible (GTR) model. Bootstrap values of 1000 replicates are specified below the nodes as percentages. *Caenorhabditis elegans* was selected as the outgroup taxon.

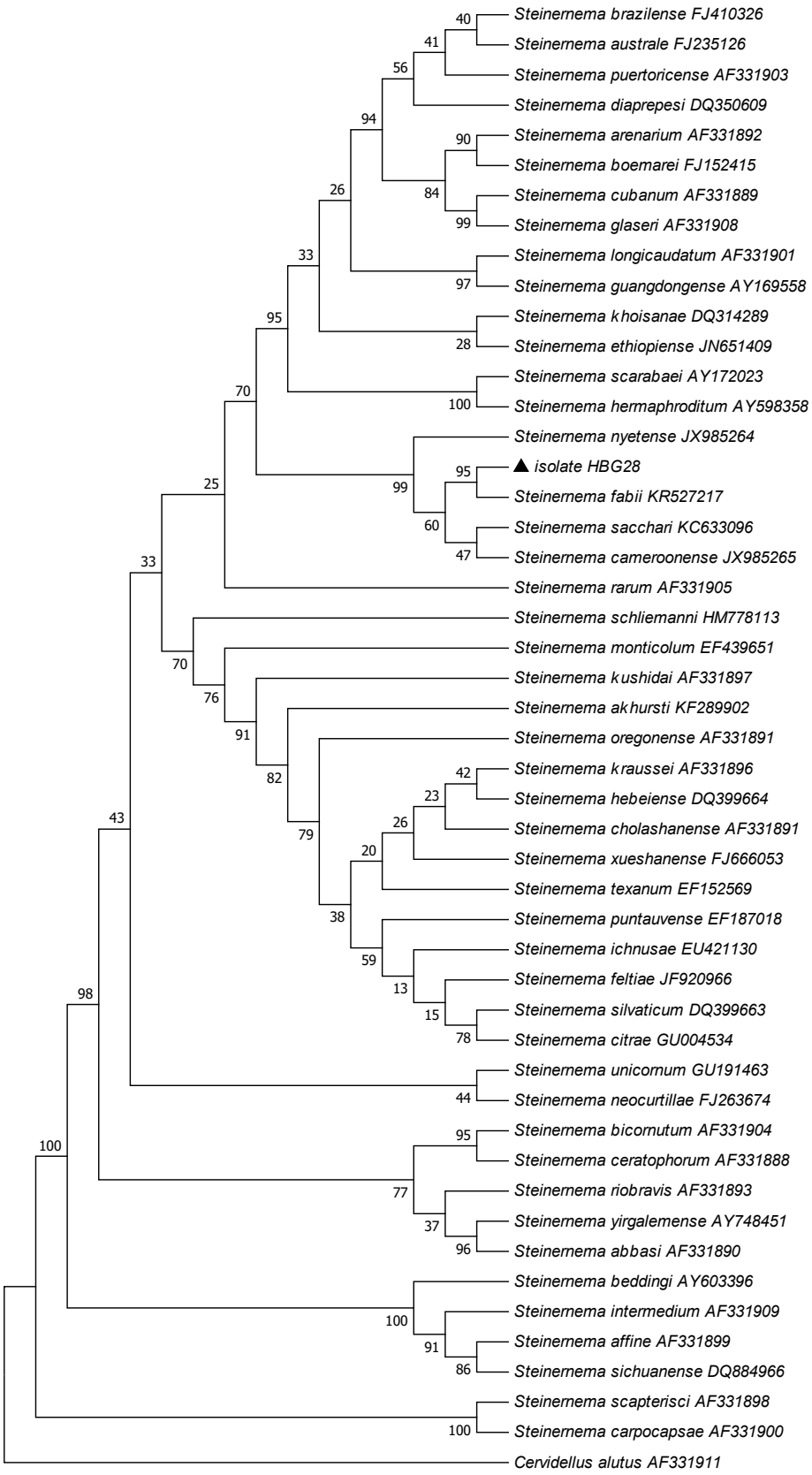


Figure 2.7. Phylogenetic relationships of *Steinernema sp.* HBG28 (▲) with 49 species of *Steinernema* based on partial D2D3-rDNA sequences from GenBank. The phylogenetic tree was constructed by using Maximum Parsimony. Bootstrap values of 1000 replicates are specified below the nodes as percentages. Accession numbers have also been specified. *Cervidellus alurus* was selected as the outgroup taxon.

Table 2.4. Estimates of evolutionary divergence between sequences. The number of base substitutions per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal and were obtained by a bootstrap procedure (1000 replicates). Analyses were conducted using the Jukes-Cantor model. The analysis involved 24 nucleotide sequences. There was a total of 511 positions in the final dataset.

No.	Species	ITS Region																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	*S.* sp. HBG28		0.009	0.013	0.012	0.013	0.020	0.020	0.019	0.023	0.019	0.021	0.021	0.020	0.021	0.021	0.022	0.020	0.021	0.021	0.021	0.020	0.021	0.020	0.020
2	*S. fabii*	0.046		0.011	0.011	0.012	0.020	0.020	0.020	0.024	0.020	0.020	0.020	0.019	0.020	0.020	0.022	0.019	0.019	0.020	0.020	0.019	0.020	0.019	0.019
3	*S. sacchari*	0.078	0.059		0.011	0.011	0.022	0.022	0.021	0.024	0.021	0.021	0.021	0.021	0.022	0.021	0.024	0.021	0.020	0.021	0.021	0.021	0.021	0.020	0.021
4	*S. cameroonense*	0.076	0.063	0.076		0.010	0.022	0.021	0.021	0.024	0.020	0.021	0.022	0.022	0.022	0.021	0.023	0.022	0.021	0.022	0.021	0.021	0.022	0.021	0.021
5	*S. nyetense*	0.091	0.076	0.065	0.049		0.023	0.023	0.022	0.024	0.021	0.022	0.022	0.022	0.022	0.022	0.024	0.022	0.021	0.022	0.022	0.022	0.022	0.021	0.021
6	*S. robustispiculum*	0.183	0.183	0.214	0.196	0.224		0.006	0.008	0.019	0.017	0.018	0.019	0.018	0.019	0.019	0.021	0.017	0.020	0.019	0.018	0.019	0.019	0.018	0.018
7	*S. ashiunense*	0.186	0.186	0.219	0.196	0.221	0.020		0.009	0.020	0.017	0.018	0.018	0.018	0.019	0.019	0.021	0.017	0.020	0.019	0.018	0.019	0.019	0.017	0.017
8	*S. monticolum*	0.181	0.181	0.201	0.191	0.214	0.034	0.046		0.017	0.017	0.017	0.017	0.016	0.017	0.017	0.019	0.016	0.018	0.017	0.017	0.018	0.017	0.016	0.016
9	*S. schliemanni*	0.235	0.240	0.243	0.237	0.240	0.166	0.181	0.146		0.020	0.020	0.019	0.019	0.020	0.019	0.021	0.019	0.020	0.020	0.021	0.021	0.020	0.019	0.019
10	*S. akhursti*	0.173	0.176	0.186	0.178	0.193	0.146	0.146	0.142	0.173		0.013	0.013	0.013	0.014	0.013	0.015	0.011	0.013	0.012	0.012	0.012	0.012	0.011	0.010
11	*S. weiseri*	0.191	0.178	0.198	0.191	0.203	0.151	0.149	0.132	0.166	0.080		0.005	0.006	0.009	0.008	0.012	0.007	0.011	0.010	0.011	0.011	0.009	0.008	0.009
12	*S. litorale*	0.191	0.178	0.198	0.193	0.206	0.154	0.151	0.135	0.163	0.083	0.012		0.007	0.008	0.008	0.012	0.008	0.011	0.011	0.011	0.011	0.009	0.008	0.009
13	*S. ichnusae*	0.186	0.176	0.201	0.196	0.203	0.154	0.151	0.135	0.166	0.083	0.020	0.024		0.008	0.009	0.012	0.007	0.011	0.010	0.011	0.011	0.009	0.008	0.008
14	*S. feltiae*	0.198	0.183	0.208	0.206	0.201	0.166	0.163	0.144	0.181	0.094	0.036	0.032	0.032		0.010	0.012	0.008	0.011	0.011	0.012	0.012	0.011	0.010	0.010
15	*S. citrae*	0.186	0.176	0.196	0.191	0.198	0.163	0.161	0.144	0.168	0.080	0.034	0.034	0.038	0.051		0.013	0.009	0.011	0.011	0.012	0.012	0.011	0.010	0.010
16	*S. hebeiense*	0.224	0.216	0.240	0.224	0.243	0.186	0.183	0.171	0.188	0.112	0.070	0.070	0.074	0.078	0.087		0.011	0.014	0.013	0.014	0.014	0.013	0.012	0.012
17	*S. jollieti*	0.173	0.171	0.193	0.191	0.193	0.139	0.137	0.123	0.156	0.065	0.030	0.030	0.028	0.040	0.046	0.070		0.011	0.010	0.010	0.010	0.009	0.008	0.008
18	*S. silvaticum*	0.186	0.166	0.183	0.198	0.201	0.173	0.173	0.149	0.181	0.087	0.061	0.061	0.059	0.065	0.072	0.098	0.061		0.008	0.011	0.011	0.010	0.009	0.009
19	*S. kraussei*	0.191	0.173	0.188	0.193	0.201	0.163	0.163	0.139	0.171	0.083	0.053	0.055	0.055	0.063	0.061	0.091	0.055	0.036		0.011	0.010	0.009	0.009	0.009
20	*S. texanum*	0.193	0.176	0.198	0.201	0.206	0.149	0.146	0.144	0.186	0.080	0.061	0.063	0.059	0.068	0.076	0.103	0.053	0.063	0.061		0.010	0.010	0.010	0.009
21	*S. sangi*	0.178	0.166	0.188	0.186	0.193	0.156	0.156	0.146	0.181	0.070	0.057	0.059	0.061	0.070	0.072	0.100	0.057	0.061	0.049	0.046		0.009	0.010	0.009
22	*S. oregonense*	0.188	0.176	0.193	0.191	0.196	0.156	0.156	0.137	0.168	0.065	0.038	0.040	0.042	0.055	0.057	0.080	0.044	0.053	0.042	0.055	0.040		0.008	0.008
23	*S. xueshanense*	0.173	0.166	0.183	0.181	0.191	0.151	0.144	0.132	0.168	0.061	0.036	0.038	0.036	0.055	0.049	0.083	0.036	0.049	0.046	0.051	0.049	0.032		0.003
24	*S. cholashanense*	0.171	0.163	0.181	0.178	0.188	0.149	0.142	0.130	0.161	0.055	0.038	0.040	0.036	0.053	0.046	0.080	0.034	0.046	0.040	0.049	0.042	0.030	0.006	

Table 2.5. Estimates of evolutionary divergence between sequences. The number of base substitutions per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal and were obtained by a bootstrap procedure (1000 replicates). Analyses were conducted using the Jukes-Cantor model. The analysis involved 23 nucleotide sequences. There was a total of 444 positions in the final dataset.

No.	Species	D2D3 region																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	*S.* sp. HBG28		0.000	0.003	0.005	0.005	0.013	0.013	0.014	0.014	0.016	0.013	0.015	0.015	0.015	0.014	0.014	0.013	0.013	0.014	0.015	0.015	0.015	0.013
2	*S. fabii*	0.000		0.003	0.005	0.005	0.013	0.013	0.014	0.014	0.016	0.013	0.015	0.015	0.015	0.014	0.014	0.013	0.013	0.014	0.015	0.015	0.015	0.013
3	*S. sacchari*	0.005	0.005		0.004	0.004	0.013	0.013	0.014	0.013	0.016	0.013	0.015	0.015	0.014	0.014	0.014	0.013	0.014	0.014	0.016	0.015	0.014	0.013
4	*S. cameroonense*	0.011	0.011	0.007		0.004	0.014	0.014	0.014	0.014	0.016	0.013	0.015	0.015	0.014	0.014	0.014	0.013	0.014	0.014	0.016	0.015	0.015	0.013
5	*S. nyetense*	0.011	0.011	0.007	0.009		0.013	0.013	0.014	0.013	0.015	0.013	0.015	0.015	0.014	0.014	0.014	0.013	0.014	0.014	0.016	0.015	0.015	0.013
6	*S. hermaphroditum*	0.071	0.071	0.071	0.078	0.068		0.000	0.011	0.010	0.012	0.010	0.013	0.013	0.012	0.011	0.011	0.011	0.012	0.012	0.016	0.015	0.017	0.014
7	*S. scarabaei*	0.071	0.071	0.071	0.078	0.068	0.000		0.011	0.010	0.012	0.010	0.013	0.013	0.012	0.011	0.011	0.011	0.012	0.012	0.016	0.015	0.017	0.014
8	*S. guangdongense*	0.078	0.078	0.078	0.083	0.078	0.054	0.054		0.006	0.011	0.010	0.012	0.012	0.011	0.011	0.010	0.010	0.012	0.011	0.016	0.015	0.016	0.014
9	*S. longicaudatum*	0.078	0.078	0.073	0.078	0.073	0.044	0.044	0.018		0.012	0.010	0.013	0.013	0.010	0.010	0.011	0.011	0.012	0.012	0.017	0.016	0.016	0.014
10	*S. ethiopiense*	0.096	0.096	0.096	0.101	0.093	0.063	0.063	0.059	0.063		0.011	0.013	0.013	0.013	0.013	0.013	0.012	0.013	0.012	0.017	0.016	0.016	0.016
11	*S. khoisanae*	0.071	0.071	0.071	0.076	0.071	0.046	0.046	0.044	0.049	0.051		0.010	0.010	0.010	0.010	0.010	0.009	0.010	0.010	0.015	0.014	0.015	0.013
12	*S. glaseri*	0.096	0.096	0.096	0.093	0.096	0.076	0.076	0.068	0.073	0.076	0.051		0.003	0.009	0.008	0.010	0.010	0.010	0.009	0.015	0.015	0.016	0.015
13	*S. cubanum*	0.096	0.096	0.096	0.093	0.096	0.076	0.076	0.063	0.073	0.076	0.051	0.005		0.010	0.009	0.009	0.009	0.010	0.010	0.015	0.015	0.016	0.015
14	*S. boemarei*	0.083	0.083	0.078	0.081	0.078	0.063	0.063	0.056	0.051	0.073	0.044	0.037	0.042		0.005	0.009	0.009	0.009	0.009	0.015	0.015	0.017	0.015
15	*S. arenarium*	0.083	0.083	0.078	0.081	0.081	0.061	0.061	0.059	0.054	0.071	0.046	0.032	0.037	0.011		0.009	0.008	0.008	0.008	0.014	0.014	0.016	0.014
16	*S. diaprepesi*	0.078	0.078	0.078	0.081	0.078	0.063	0.063	0.049	0.059	0.068	0.044	0.049	0.044	0.044	0.042		0.006	0.009	0.008	0.016	0.015	0.017	0.014
17	*S. puertoricense*	0.073	0.073	0.073	0.076	0.073	0.056	0.056	0.046	0.056	0.066	0.042	0.046	0.042	0.042	0.035	0.021		0.007	0.006	0.015	0.014	0.015	0.013
18	*S. austral*	0.076	0.076	0.081	0.083	0.081	0.068	0.068	0.061	0.063	0.076	0.051	0.046	0.051	0.035	0.032	0.035	0.023		0.006	0.015	0.015	0.017	0.015
19	*S. brazilense*	0.083	0.083	0.083	0.086	0.081	0.063	0.063	0.061	0.066	0.068	0.049	0.044	0.049	0.039	0.032	0.032	0.016	0.021		0.015	0.014	0.015	0.014
20	*S. citrae*	0.104	0.104	0.106	0.106	0.106	0.096	0.096	0.104	0.109	0.111	0.091	0.101	0.096	0.096	0.093	0.106	0.093	0.099	0.099		0.008	0.014	0.012
21	*S. feltiae*	0.096	0.096	0.096	0.101	0.096	0.086	0.086	0.093	0.099	0.096	0.081	0.099	0.093	0.093	0.091	0.093	0.086	0.099	0.091	0.028		0.012	0.011
22	*S. monticolum*	0.093	0.093	0.088	0.093	0.088	0.099	0.099	0.096	0.101	0.099	0.083	0.109	0.104	0.106	0.104	0.106	0.093	0.109	0.099	0.078	0.059		0.011
23	*S. schliemanni*	0.078	0.078	0.073	0.081	0.076	0.078	0.078	0.091	0.091	0.099	0.073	0.104	0.099	0.091	0.088	0.091	0.078	0.096	0.088	0.068	0.054	0.051	

2.5 Discussion

The aim of the present study was to isolate and identify indigenous entomopathogenic nematode species in a targeted survey carried out in the North West Province and Gauteng, whilst exploring the natural occurrence of entomopathogenic nematodes in these areas. These provinces are quite distinct in geography and habitat with a high plant diversity that could support to an equally diverse population of insects, suggesting that a significant number of different species and strains of entomopathogenic nematodes may exist within these areas.

During the present survey, two species of *Heterorhabditis* and one novel species of *Steinernema* were recovered from non-disturbed soils in the North West Province and Gauteng, respectively, having been identified using molecular techniques that involved sequences of the ITS rDNA. Despite only 8% of soil samples testing positive for the presence of entomopathogenic nematodes, the fact that three species were recovered during this small-scale survey only serves to highlight the diversity and prevalence of entomopathogenic nematode fauna within these provinces. Furthermore, given the present distribution of entomopathogenic nematodes within the sampled locations, it is likely that a more intensive survey within the North West Province and Gauteng would yield more favourable results for nematode recovery.

Typically, steinernematids are recovered more frequently over heterorhabditids during non-targeted surveys (Hominick, 2002). Previous work in South Africa have shown *Steinernema* to be the dominant genus over *Heterorhabditis* (Hatting et al., 2009; Molotsane et al., 2007) and these findings were not with those derived from the current study. In the present study, *Heterorhabditis* species were less prevalent than *Steinernema*. Of the heterorhabditids recovered in the North West Province, *H. bacteriophora* and *H. zealandica* were identified. Previously, *H. bacteriophora* has

been found in the Western Cape, Free State, KwaZulu-Natal and Mpumalanga during an extensive survey of seven geographical locations within South Africa (Hatting et al., 2009), whilst *H. zealandica* has been reported in the Eastern Cape (Malan et al., 2006), Western Cape and Mpumalanga (Malan et al., 2011), as well as more recently in KwaZulu-Natal (Steyn et al., 2017) suggesting that both *Heterorhabditis* species have a wide geographic distribution and are adapted to various habitats within South Africa. In addition, the high recovery rate of both *H. bacteriophora* and *H. zealandica* would indicate that their dispersal methods are highly optimized and is likely influenced by active (suitable hosts) or passive (wind, water and human activity) dissemination mechanisms (Hominick, 2002; Adams et al., 2006).

It is also noteworthy that while *Steinernema* nematodes were not detected in any soil samples collected from the North West Province, they were, in fact, recovered from Gauteng. At the time this survey was carried out, all *Steinernema* isolates were shown to represent a new species that belonged to a new phylogenetic group called the 'Cameroonian' clade, consisting of *S. sacchari*, *S. cameroonense* and *S. nyetense* (Nthenga et al., 2014). According to sequence data of the ITS region of rDNA, *Steinernema* sp. HBG28 was found to be phylogenetically related to *S. fabii* but divergent enough to speculate that it belonged to a new species within the aforementioned clade. Interestingly, *S. fabii* was isolated from Mpumalanga by Abate et al. (2016), and apart from differences in the locations at which these species were isolated, both appear to share similar habitat preferences, in that they are adapted to subtropical highland conditions distinctive of the grassland environment. Furthermore, the high prevalence of *Steinernema* sp. HBG28 within central parts of Gauteng may suggest that a wide range of insect hosts are susceptible towards this species.

Generally, infection by entomopathogenic nematodes is recognized by various tell-tale signs, with colour changes in healthy *Galleria mellonella* larvae being the most notable of symptoms. In the present study, *Heterorhabditis zealandica* infection resulted in *G. mellonella* larvae turning dark maroon in colour (Figure 2.4C). In the case of *Steinernema* isolates, infected *G. mellonella* cadavers turned brown and presented with black spots on the cuticle. Colour changes characteristic of entomopathogenic nematode infection are largely attributed to the associated bacteria (Koppenhöfer, 2000), especially with *Photorhabdus* that produces red pigment (Richardson et al., 1988; Inman III et al., 2012; Singh et al., 2012). Additionally, the formation of black spots on the insect cuticle is due to melanization, a cellular immune response of the host, which is often ineffective due to the sheer number of invading nematodes (Steiner, 1996; Lacey and Solter, 2012). Interestingly, the presence of black spots was not observed in *G. mellonella* larvae infected with any of the *Heterorhabditis* isolates. It has been shown that infection by *Photorhabdus*, the bacterial genus associated with *Heterorhabditis* nematodes is often accompanied by a lack of melanization (ffrench-Constant et al., 2003). This is due to inhibition of phospholipase A_2 and phenoloxidase that regulates haemocyte aggregation and nodulation by *Photorhabdus* (Kaito et al., 2002; Gillespie et al., 1997; Schmidt et al., 2001; Lavine and Strand, 2001).

During the present survey, soil type was shown to have an influence on the recovery of entomopathogenic nematodes from collected soil samples, as demonstrated in other studies (Hara et al., 1991; Griffin et al., 1994). Soil type plays a critical role in the distribution of entomopathogenic nematodes, especially where nematode movement and energy expenditure is concerned (Wallace, 1971). This, in turn, has a direct impact on the survival and infectivity of these nematodes (Kung et al., 1990), often affecting their behaviour as well. In the current study,

entomopathogenic nematodes appeared to have a greater affinity towards sandy loam and sandy clay loam soil. These findings are consistent with other studies where entomopathogenic nematodes were shown to prefer soils with high sand and low clay content (Molyneux and Bedding, 1984; Kung et al., 1990). Additionally, entomopathogenic nematodes were not detected in any collected samples containing clay soil, and this may be due to the small pore sizes and high-water retention often associated with clay soils, which can impede nematode movement, and thereby affect energy expenditure (Wallace, 1958; 1971). Typically, survival and infectivity of infective juvenile nematodes are significantly reduced in soils with a high clay content, as a result of poor aeration (Kung et al., 1990).

2.6 Conclusion

South Africa offers a range of undisturbed habitats, including the Grassland, Savanna, Albany Thicket, Forests, Desert, the Karoo (Succulent and Nama) and the Fynbos (Rutherford et al., 2006). Due to the high diversity of plant and insect species, an opportunity exists for the discovery novel entomopathogenic nematode species in a variety of ecological niches. Surveys for the retrieval of entomopathogenic nematodes in South Africa serve an important purpose for several reasons. They monitor the occurrence and natural distribution of native nematodes in various localities and habitats; allow for the discovery of new species and isolates with greater tolerance to local conditions and expands the potential of using these new species as biological control agents for local pests. Furthermore, all recovered isolates in the present study were shown to be highly infective towards *G. mellonella* larvae and could potentially be used in development of commercial strains adapted specifically to local environments and insect populations. The isolation of *Steinernema* sp. HBG28 is the first report of a novel steinernematid prevalent to the grassland biome in the Gauteng

province. Details regarding the nematode distribution, habitat and host preferences, as well as the virulence of *Steinernema* sp. HBG28 would need to be further investigated.

2.7 References

1. Abate, B. A, Malan, A. P., Tiedt, L. R., Wingfield, M. J, et al., (2016). *Steinernema fabii* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology,* **18:**235–255.
2. Adams, B. J., Burnell, A. M. and Powers, T. O. (1998). A phylogentic analysis of Heterorhabditis (Nematoda: Rhabditidae) based on internal trascribed spacer 1 DNA sequence data. *Journal of Nematology,* **30:**22–39.
3. Adams, B. J., Fodor, A., Koppenhöfer, H. S., Stackebrandt, S., et al. (2006). Biodiversity and systematic of nematode–bacterium entomopathogens. *Biological Control,* **38:**4–21.
4. Bedding, R. A. and Akhurst, R. J. (1975). A simple technique for the detection of insect parasitic rhabditid nematodes in soil. *Nematologica,* **21:**109–110.
5. Boemare N.E., Akhurst R.J. and Mourant R.G. (1993). DNA relatedness between *Xenorhabdus* spp, (*Enterobacteriaceae*), symbiotic bacteria of entomopathogenic nematodes, and a proposal to transfer *Xenorhabdus luminescens* to a new genus, *Photorhabdus* gen. nov. *International Journal of Systematic Bacteriology*, **43:**249–255.
6. Burnell, A. M. and Stock, P. S. (2000). *Heterorhabditis*, *Steinernema* and their bacterial symbionts — lethal pathogens of insects. *Nematology,* **2:**31–42.
7. Choo, H. Y., Kaya, H. K. and Stock, S. P. (1995). Isolation of entomopathogenic nematodes (Steinernematidae and Heterorhabditidae) from Korea. *Japanese Journal of Nematology,* **25:**44–51.
8. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* **32:**1792–1797.
9. Ehlers, R. U. (2005). Forum on safety and regulation. In: P. S. Grewal, R. U. Ehlers and D. I. Shapiro-Ilan, (Eds.). *Nematodes as Biocontrol Agents.* Wallingford, U.K: CABI Publishing, pp. 107–114.

10. Eleftherianos, S., Joyce, R. H. and ffrench-Constant, D. (2010). Probing the tri-trophic interaction between insects, nematodes and *Photorhabdus*. *Parasitology,* **137:**1695–1706.
11. ffrench-Constant, R. H. and Bowen, D. J. (2000). Novel insecticidal toxins from nematode-symbiotic bacteria. *Cellular and Molecular Life Sciences,* **57:**823–833.
12. ffrench-Constant, R. H., Dowling, A. and Waterfield, N. R. (2007). Insecticidal toxins from *Photorhabdus* bacteria and their potential use in agriculture. *Toxicon,* **49:**436–451.
13. ffrench-Constant, R., Waterfield, N., Daborn, P., Joyce, S., et al. (2003). *Photorhabdus*: towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiology Reviews,* **26:**433–456.
14. Freire, R., Arias, A., Méndez, J. and Insua, A. (2010). Sequence variation of the internal transcribed spacer (ITS) region of ribosomal DNA in *Cerastoderma* species (Bivalvia: Cardiidae). *Journal of Molluscan Studies,* **76:**77–86.
15. Gillespie J.P. Kanost M.R. and Trenczek T. (1997). Biological mediators of insect immunity. *Annual Review of Entomology*, **42:**611–643.
16. Goodrich-Blair, H. and Clarke, D. J. (2007). Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination. *Journal of Bacteriology,* **64:**260–268.
17. Griffin, C. T., Downes, M. J. and Block, W. (1990). Tests of Antarctic soils for insect parasitic nematodes. *Antarctic Science,* **2:**221–222.
18. Griffin, C. T., Joyce, S., Dix, I., Burnell, A. M., et al. (1994). Characterization of Heterorhabditis (Nematoda: Heterorhabditidae) from Ireland and Britain by molecular and cross-breeding techniques, and the occurrence of the genus in these islands. *Fundamental and Applied Nematology,* **17:**245–253.
19. Hara, A. H., Gaugler, R., Kaya, H. K. and Lebeck, L. M. (1991). Natural populations of entomopathogenic nematodes (Rhabditida: Steinernematidae and

Heterorhabditidae) from the Hawaiian Islands. *Environmental Entomology,* **20:**211–216.

20. Harington, J. S., (1953). South African Journal of Science - Observations on the biology, the parasites and the taxonomic position of the black maize beetle - *Heteronychus sanctae-helenae* Blanch. *South African Journal of Science,* **50:**11.
21. Hatting, J., Stock, P. S. and Hazir, S. (2009). Diversity and distribution of entomopathogenic nematodes (Steinernematidae, Heterorhabditidae) in South Africa. *Journal Invertebrate Pathology,* **102:**120–128.
22. Hazir, S., Kaya, H. K., Stock, P. and Keskin, N. (2003). Entomopathogenic Nematodes (Steinernematidae and Heterorhabditidae) for Biological Control of Soil Pests. *Turkish Journal of Biology,* **27:**181–202.
23. Hominick, W. M. (2002). Biogeography. In: R. Gaugler, (Eds.), *Entomopathogenic Nematology.* Wallingford, UK: CABI Publishing, pp. 115-143.
24. Hominick, W. M., Briscoe, B. R. and del Pino, F. G. (1997). Biosystematics of entomopathogenic nematodes: current status, protocols and definitions. *Helminthology,* **71:**271–298.
25. Hominick, W. M., Reid, A. P., Bohan, D. A. and Briscoe, B. R. (1996). Entomopathogenic nematodes: Biodiversity, geographical distribution and de convention on Biological Diversity. *Biocontrol Science and Technology,* **6:**317–331.
26. Hunt, D. J. (2007). Overview of taxonony and systematics. In: K. B. Nguyen and D. J. Hunt, (Eds.), *Entomopathogenic Nematodes: Systematics, Phylogeny and Bacterial Symbionts.* Leiden-Boston: Brill, pp. 27-57.
27. Inman III, F. L., Singh, S. and Holmes, L. D. (2012). Mass Production of the Beneficial Nematode *Heterorhabditis bacteriophora* and Its Bacterial Symbiont *Photorhabdus luminescens*. *Indian Journal of Microbiology,* **52:**316–324.
28. Jukes, T. H. and Cantor, C. R. (1969). Evolution of protein molecules. In: H. N. Munro, (Ed.), *Mammalian Protein Metabolism.*, pp. 21-132.

29. Kaito, C., Akimitsu, N., Watanabe, H. and Sekimizu, K. (2002). Silkworm larvae as an animal model of bacterial infection pathogenic to humans. *Microbial pathogenesis*, **32:**183–190.
30. Kaya, H. K. and Stock, S. P. (1997). Techniques in insect nematology. In: L. A. Lacey, (Eds.), *Manual of Techniques in Insect Pathology.* London: Academic Press, pp. 281-324.
31. Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions though comparative studies of nucleotide sequences. *Journal of Molecular Evolution,* **16:**111–120.
32. Koppenhöfer, A. M. (2000). Nematodes. In: L. A. Lacey and H. K. Kaya, (Eds.), *Field Manual of Techniques in Invertebrate Pathology.* Dordrecht: Kluwer Academic Publishers, pp. 283-301.
33. Kung, S., Gaugler, R. and Kaya, H. K. (1990). Soil type and entomopathogenic nematode persistence. *Journal of Invertebrate Pathology,* **55:**401–406.
34. Lacey, L. A. (1997). *Manual of techniques in insect pathology.* San Diego: Academic Press.
35. Lacey, L. A. and Georgis, R. (2012). Entomopathogenic Nematodes for Control of Insect Pests Above and Below Ground with Comments on Commercial Production. *Journal of Nematology,* **44:**218–225.
36. Lacey, L. A. and Solter, L. (2012). Initial handling and diagnosis of diseased invertebrates. In: L. A. Lacey, (Ed.), *Manual of Techniques in Invertebrate Pathology.* San Diego: Academic Press, pp. 1-14.
37. Lavine M.D. Strand M.R. (2001). Surface characteristics of foreign targets that elicit an encapsulation response by the moth Pseudoplusia includens. *Journal of Insect Physiology*, **47:**965–974.
38. Lewis, E. E., Campbell, J., Griffin, C., Kaya, H., et al. (2006). Behavioral ecology of entomopathogenic nematodes. *Biological Control*, **38:**66-79.

39. Malan, A. P., Knoetze, R. and Moore, S. D. (2011). Isolation and identification of entomopathogenic nematodes from citrus orchards in South Africa and their biocontrol potential against false codling moth. *Journal of Invertebrate Pathology,* **108:**115–125.
40. Malan, A. P., Knoetze, R. and Tiedt, L. R. (2016). *Steinernema jeffreyense* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Journal of helminthology*, **90:**262–278.
41. Malan, A. P., Nguyen, K. B. and Addison, M. F. (2006). Entomopathogenic nematodes (Steinernematidae and Heterorhabditidae) from the southwestern parts of South Africa. *African Plant Protection,* **12:**65–69.
42. Malan, A. P., Nguyen, K. B., De Waal, J. Y. and Tiedt, L. R. (2008). *Heterorhabditis safricana* n. sp. (Rhabditida: Heterorhabditidae), a new entomopathogenic nematode from South Africa. *Nematology,* **10:**381–396.
43. Molotsane, R., Ngoma, L. and Gray, V. (2007). A biogeographic survey of entomopathogenic nematodes of Gauteng and the North West Province of South Africa. *South African Journal of Plant and Soil,* **24:**256.
44. Molyneux, A. S. and Bedding, R. A. (1984). Influence of soil texture and moisture on the infectivity of Heterorhabditis sp. D1 and Steinernema glaseri for larvae of the sheep blowfly, *Lucilia cuprina*. *Nematologica,* **30:**358–365.
45. Nguyen, K. B., Malan, A. P. and Gozel, U. (2006). *Steinernema khoisanae* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology,* **8:**157–175.
46. Nthenga, I., Knoetze, R., Berry, S., Tiedt, L. R., et al. (2014). *Steinernema sacchari* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology*, **16:**475–494.
47. Piedra-Buena, A., López-Cepero, J. and Campos-Herrera, R., (2015). Entomopathogenic Nematode Production and Application: Regulation, Ecological

Impact and Non-target Effects. In: R. Campos-Herrera, (Eds.), *Nematode Pathogenesis of Insects and Other Pests: Ecology and Applied Technologies for Sustainable Plant and Crop Protection.* Zurich, Switzerland: Springer International Publishing, pp. 255-284.

48. Poinar, G. O. (1975). Entomogenous nematodes. A manual and host list of insect-nematode associations. In: Leiden, The Netharlands: E.J. Brill, 317.
49. Poinar, G. O. Jr. (1990). Taxonomy and biology of steinernematidae and heterorhabditidae. In: R. Gaugler and H. K. Kaya, (Eds.), *Entomopathogenic Nematodes in biological control.* Boca Raton: FL: CRC Press, pp. 23-60.
50. Poinar, G. O., Jr. (1976). Description and biology of a new insect parasitic rhabditoid, *Heterorhabditis bacteriophora* n. gen. n. sp. (Rhabditida; Heterorhabditidae n. fam.). *Nematologica*, **21:**463–470.
51. Richardson, W. H., Schmidt, T. M. and Nealson, K. H. (1988). Identification of an anthraquinone pigment and a hydroxystilbene antibiotic from *Xenorhabdus luminescens*. *Applied Environmental Microbiology,* **54:**1602–1605.
52. Rutherford, M. C., Mucina, L. and Powrie, L. W. (2006). Biomes and bioregions of southern Africa. In: L. Mucina and M. C. Rutherford, (Eds.), *The vegetation of South Africa, Lesotho and Swaziland.* South African National Bioversity Institute, Pretoria: Strelitzia, pp. 30-51.
53. Schmidt, O., Theopold, U. and Strand, M. (2001). Innate immunity and its evasion and suppression by hymenopteran endoparasitoids. *BioEssays*, **23:**344–351.
54. Singh, S., Eric, M., Floyd, I. and Leonard, H. D. (2012). Characterization of *Photorhabdus luminescens* Growth for the Rearing of the Beneficial Nematode *Heterorhabditis bacteriophora*. *Indian Journal of Microbiology,* **52:**325–331.
55. Spaull, V. W. (1988). A preliminary evaluation of the entomogenous nematodes to control the African sugarcane stalk borer *Eldana saccharina* (Lepidoptera:

Pyralidae). *Proceedings of the South African Sugar Technologists' Association,* **62:**120–123.

56. Spaull, V. W. (1990). Field tests to control the phylarid, Eldana saccharina, with an entomogenous nematode, *Heterorhabditis* sp. *Proceedings of the South African Sugar Technologists' Association,* **64:**103–106.
57. Spaull, V. W. (1991). *Heterorhabditis* and *Steinernema* species (Nematode: Rhabditida) for the control of a sugar cane stalk borer in South Africa. *Phytophylactica,* **23:**213–215.
58. Steiner, W. A. (1996). Melanization of *Steinernema feltiae* Filipjev and *S. kraussei* Steiner in larvae of *Otiorhynchus sulcatus* (F.). *Fundamental and Applied Nematology,* **19:**67–70.
59. Steyn, W. P., Malan, A. P., Daneel, M. S. and Slabbert, R. M. (2017). Entomopathogenic nematodes from north-eastern South Africa and their virulence against false codling moth, *Thaumatotibia leucotreta* (Lepidoptera: Tortricidae). *Biocontrol Science and Technology,* **27:**1265–1278.
60. Stock, P. S., Campbell, J. F. and Nadler, S. A. (2001). Phylogeny of *Steinemema* Travassos, 1927 (Cephalobina: Steinernematidae) inferred from ribosomal DNA sequences and morphological characters. *Journal of Parasitology*, **87:**877–889.
61. Stock, P. S. and Goodrich-Blair, H. (2008). Entomopathogenic nematodes and their bacterial symbionts: the inside out of a mutualistic associations. *Symbiosis,* **46:**65–75.
62. Stock, S. P. and Goodrich-Blair, H. (2012). Nematode parasites, pathogens and associates of insects and invertebrates of economic importance. In: L. A. Lacey and H. K. Kaya, (Eds.), *Field Manual of Techniques in Invertebrate Pathology.* San Diego: Academic Press, pp. 373–426.

63. Stokwe, N. F., Malan, A. P., Nguyen K. B. and Knoetze, R. (2011). *Steinernema citrae* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology,* **13:**569–587.

64. Tamura, K., Peterson, D. and Peterson, N. (2011). mega7: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution,* **28:**2731–2739.

65. Thomas, G. M., and Poinar, G. O. Jr. (1979). *Xenorhabdus* gen. nov., a genus of enteropathogenic, nematophilic bacteria of the family *Enterobacteriaceae*. *International Journal of Systematic Bacteriology*, **29:**352–360.

66. Wallace, H. R. (1958). Movement of Eelworms. *Annals of Applied Biology,* **46:**662–668.

67. Wallace, H. R. (1971). Abiotic influences in the soil environment. In: B. M. Zuckerman, W. F. Mai and R. A. Rohde, eds. *Plant Parasitic Nematodes, Vol. I.* New York: Academic Press, pp. 257-280.

68. Woodring, J. L. and Kaya, H. K. (1988). Steinernematid and Heterorhabditid nematodes: handbook of biology and techniques. *South cooperative Series Bulletin Arkansas Agricultural Experimental Station Fayetteville,* **331:**1–30.

Chapter 3

Morphological characterization of *Steinernema* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from the grasslands of South Africa

---

3.1 Abstract

A new species of entomopathogenic nematode, *Steinernema* n. sp. (isolate HBG28), was isolated from soil in the Suikerbosrand Nature Reserve in the Gauteng province of South Africa. The new species was morphologically characterized by the length of the infective juvenile (IJ) of 655 (518-720) µm, a wide body diameter of 40 (28-51) µm, a tail length of 58 (51-67) µm, ratio a = 17 (14-23), H% = 51 (48-55) and E% = 79 (71-94). The male of the first generation was recognized by a long, orange spicule with a length of 85 (82-88) µm and a gubernaculum of 70 (68-73) µm in length; D% = 58 (54-63) and GS% = 82 (76-89). The first generation female can be distinguished by slightly protruding vulvae and double-flapped epiptygmata, as well as the presence of a slight postanal swelling. The second generation varies by having a prominent postanal swelling and a conoid tail tapering to a sharp tip. Previous molecular characterization based on the ITS and D2D3 gene regions revealed that *Steinernema* n. sp. (isolate HBG28) was distinct from other *Steinernema* species, clustering to a new monophyletic group, referred to as the 'Cameroonian' clade, comprising *S. fabii*, *S. sacchari*, *S. cameroonense* and *S. nyetense.* Morphological

and morphometric analyses indicated certain similarities and differences between *Steinernema* n. sp. (isolate HBG28) and these closely-related *Steinernema* species.

3.2 Introduction

Entomopathogenic nematodes of the genus *Steinernema* (Travassos, 1927) are obligate and lethal insect parasites that share an exclusively symbiotic relationship with enteric Gram-negative bacteria of the genus *Xenorhabdus* (Thomas and Poinar, 1979). The only free-living and non-feeding stage of entomopathogenic nematodes, known as the third-stage infective juvenile seeks out an insect host to infect by gaining entry into their haemocoel. Once inside, the infective juveniles expel their symbiotic bacteria, which rapidly multiplies and metabolizes the insect tissues, providing nutrient conditions conducive for nematode growth and development. The insect host succumbs to death, usually within 24-48 hours after initial penetration of the infective juveniles and the nematodes resume development, moulting to the fourth stage (J4). Reproductive males and females appear within two days in *S. carpocapsae* (Weiser, 1955) when cultured *in vivo* in *Galleria mellonella* larvae, where amphimictic reproduction can occur for two generations (Wang and Bedding, 1996). By the third generation, egg-laying ceases in female nematodes of *Steinernema* spp. and their offspring develop through *endotokia matricida*, usually in response to nutrient depletion (Ayako et al., 2010). Unlike in *Heterorhabditis* spp., juvenile nematodes arising from *endotokia matricida* only develop into infective juveniles upon exiting the maternal nematode. Once developed, the infective juveniles emerge from the insect cadaver into the soil where they may survive up to several months before a suitable insect host is found.

Entomopathogenic nematodes have demonstrated the ability to serve as biological control agents against several insect pests of agricultural importance (Shapiro-Ilan et al., 2002; Shapiro-Ilan and Gaugler, 2002). Due to the biological control relevance of these nematodes, several surveys have been conducted, indicating that entomopathogenic nematodes have a wide distribution worldwide, except for Antarctica (Hominick, 2002). The discovery of novel entomopathogenic nematode species can provide an important opportunity to develop commercial strains that are well-adapted to local pests and environmental conditions for the purpose of biological control (Kaya and Gaugler, 1993).

There are more than 70 recognized species of *Steinernema* that have been described thus far (Stock and Goodrich-Blair, 2012). In recent years, several new species have been uncovered in South Africa, including *S. citrae* Stokwe, Malan, Nguyen, Knoetze and Tiedt, 2011 (Stokwe et al., 2011); *S. khoisanae* Nguyen, Malan and Gozel, 2006 (Nguyen et al., 2006); *S. innovationi* Çimen, Lee, Hatting, Hazir and Stock, 2014 (Çimen et al., 2014); *S. sacchari* Nthenga, Knoetze, Berry, Tiedt and Malan, 2014 (Nthenga et al., 2014*); S. tophus* Çimen, Lee, Hatting, Hazir and Stock, 2014 (Çimen et al., 2014); and *S. jeffreyense* Malan, Knoetze and Tiedt, 2015 (Malan et al., 2015), *S. fabii* (Abate et al., 2016) and *S. litchii* Steyn, Knoetze, Tiedt and Malan (Steyn et al., 2017).

3.2.1 Aims and objectives

In 2014, a survey for entomopathogenic nematodes was conducted, which resulted in the isolation of a novel species belonging to the genus *Steinernema* Travassos, 1927 from a soil sample collected in the Suikerbosrand Nature Reserve in Heidelberg, Gauteng province of South Africa. The objective of this study was to perform

morphological and morphometric characterization to differentiate this species from other closely related *Steinernema* species that have been described.

3.3 Materials and methods

3.3.1 Nematode source

Soil samples were collected from the Suikerbosrand Nature Reserve (26.4838° S, 28.2293° E) in Heidelberg, Gauteng in June of 2014. Infective juvenile nematodes were recovered by baiting the soil samples with *Galleria mellonella* (L.) (Lepidoptera: Pyralidae), according to the method described by Stock and Goodrich-Blair (2012). *G. mellonella* larvae were exposed to infective juvenile nematodes for the purpose acquiring nematodes at different life stages through dissection. Adult nematodes of the first and second generations were obtained 3-5 days and 6-7 days post infection, respectively (Nthenga et al., 2014). Third-stage infective juveniles were recovered after 14 days post infection and were harvested during the first week of emergence. Dissections were carried out, with the aid of a dissecting needle, in Ringer's solution to prevent osmotic shock of the nematode specimens (Kaya and Stock, 1997).

3.3.2 Light microscopy

3.3.2.1 Heat killing and fixing

Nematodes were killed by heat exposure and were subsequently fixed in a solution of TAF (2% triethanolamine, 8% formalin in distilled water) and Ringer's solution at 85°C for 24 hrs (Steinhorst, 1966). The TAF solution was replaced with double-strength TAF and the specimens were stored at 4°C to induce relaxing of the nematodes (Kaya and Stock, 1997).

3.3.2.2 Processing to glycerine

Fixed nematodes were processed to glycerine, using the slow evaporation method (Seinhorst, 1959). Specimens were soaked in Seinhorst solution I (95% ethanol, glycerine and distilled water in the ratio: 20:1:79), which was placed in a small desiccator. To allow slow evaporation of the ethanol, the desiccator was maintained overnight at 35°C. The specimens were submerged in Seinhorst solution II (95% ethanol and glycerine in the ratio: 95:5) and maintained at 40°C for 3 hrs.

3.3.2.2 Morphological observation

For light microscopy, permanent slides were prepared by mounting nematode specimens in pure glycerine with the aid of a dissecting needle. Coverslips were positioned over the specimens and sealed with clear nail polish. For morphometric measurements, the permanent slides were observed using an Olympus BX63 OM/FM fluorescence microscope with 20 nematodes of each life stage examined. For direct observations of live specimens at different stages and drawings, a Zeiss microscope with a Imager MZ and an Axiocam 506 colour camera were used.

3.4 Results

Steinernema n. sp. (isolate HBG28)

3.4.1 Description

3.4.1.1 Infective juvenile

Body straight and slightly curved at the ends when in a heat-relaxed state, tapering towards the posterior and anterior regions (Figure 3.2A). Head region slightly rounded and continuous with body shape. Pharynx long and thin, comprising of a uniform procorpus, marginally swollen metacorpus, narrow isthmus and a distinct

pyriform basal (terminal) bulb (Figure 3.2B). Hemizonion found near the mid-level of the basal bulb and excretory pore is obscure. Conoid tail, slightly curved towards the ventral side, tapering to a pointed terminus. Sheath (second-stage cuticle) observed with live specimens (Figure 3.2C).

3.4.1.2 First generation male

Body slim, ventrally curved with a habitus J-shaped posterior when heat-killed and fixed with TAF (Figure 3.2D). Male specimens smaller than the females. Cuticle appears smooth under light microscope. Head rounded and slightly tapered at the anterior end (Figure 3.2E). Stoma shallow and cup-shaped. Pharynx muscular, with cylindrical procorpus, slightly widened metacorpus and narrow isthmus. Nerve ring surrounds anterior part of isthmus and located between the metacorpus and pyriform basal bulb. Excretory duct distinct with excretory pore adjacent to the metacorpus and anterior to the nerve ring. Spicules paired, soft-orange colour (according to a HEX triplet colour chart, i.e. #DDAA4F) and ventrally curved (Figure 3.2F). Spicule head (manubrium) longer than width, shaft (calamus) short and dorsal lobe distinctly curved. Vellum thin, blade arcuate, tapering slightly with blunt spicule terminus. Gubernaculum boat-shaped and swollen in the middle. Tail conoid, terminus bluntly rounded with mucron present in some specimens and absent in others.

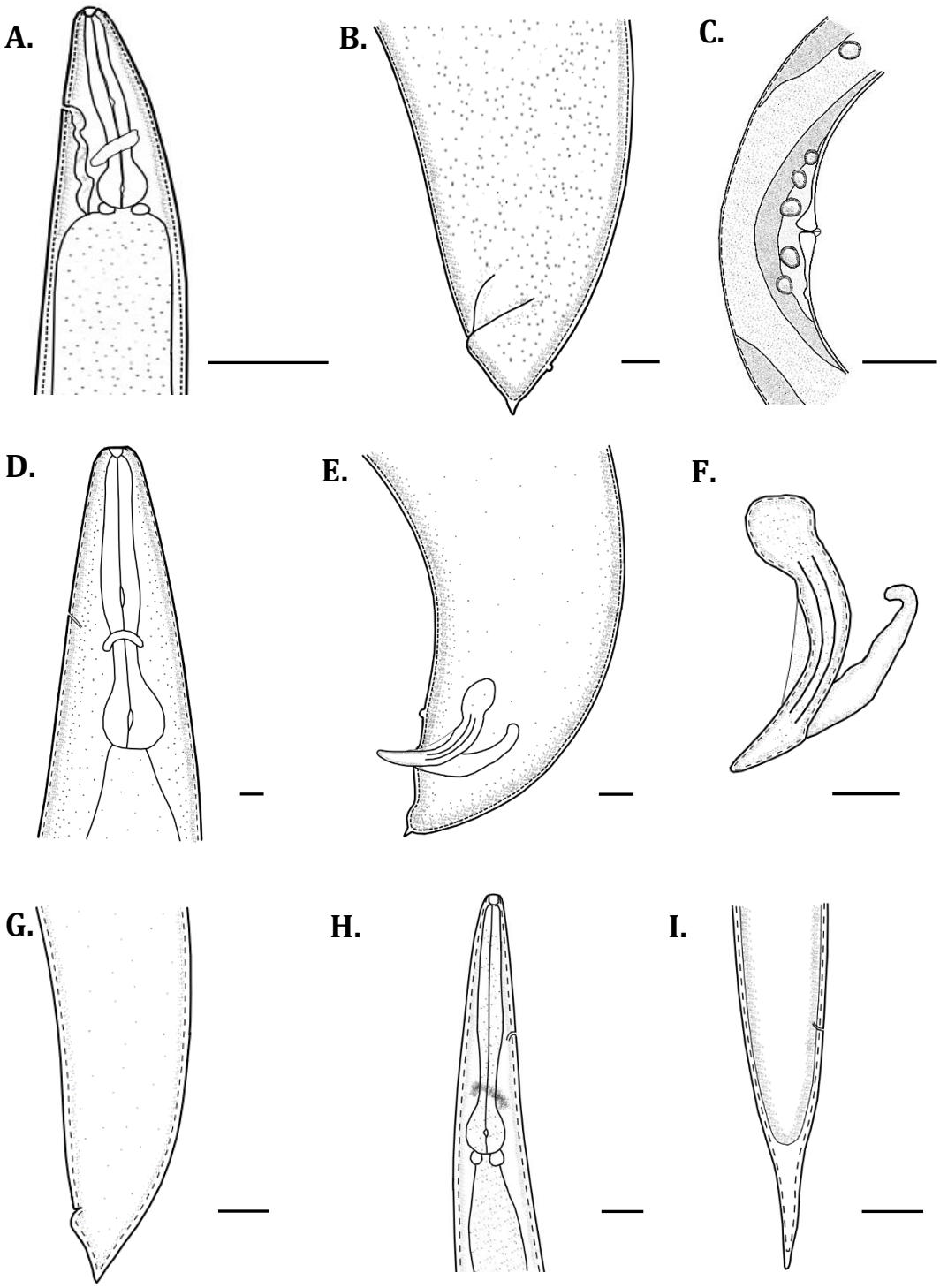


Figure 3.1. *Steinernema* n. sp. (isolate HBG28). First generation female (A-C), **A:** Anterior region; **B:** Tail region; **C:** Vulva with double-flapped epiptygmata. First generation male (D-F), **D:** Anterior region; **E:** Lateral view of tail region; **F:** Spicule with thin vellum and gubernaculum. **G:** Tail of second generation female. Infective juvenile (H-I), **H:** Anterior region; I: Tail region. (Scale bars: A: 100 µm; B, D, E, F, H: 20 µm; C, G: 200 µm; I: 50 µm).

3.4.1.3 Second generation male

Although smaller in size, second generation males have similar morphology to first generation males. Tail conoid, with mucron present.

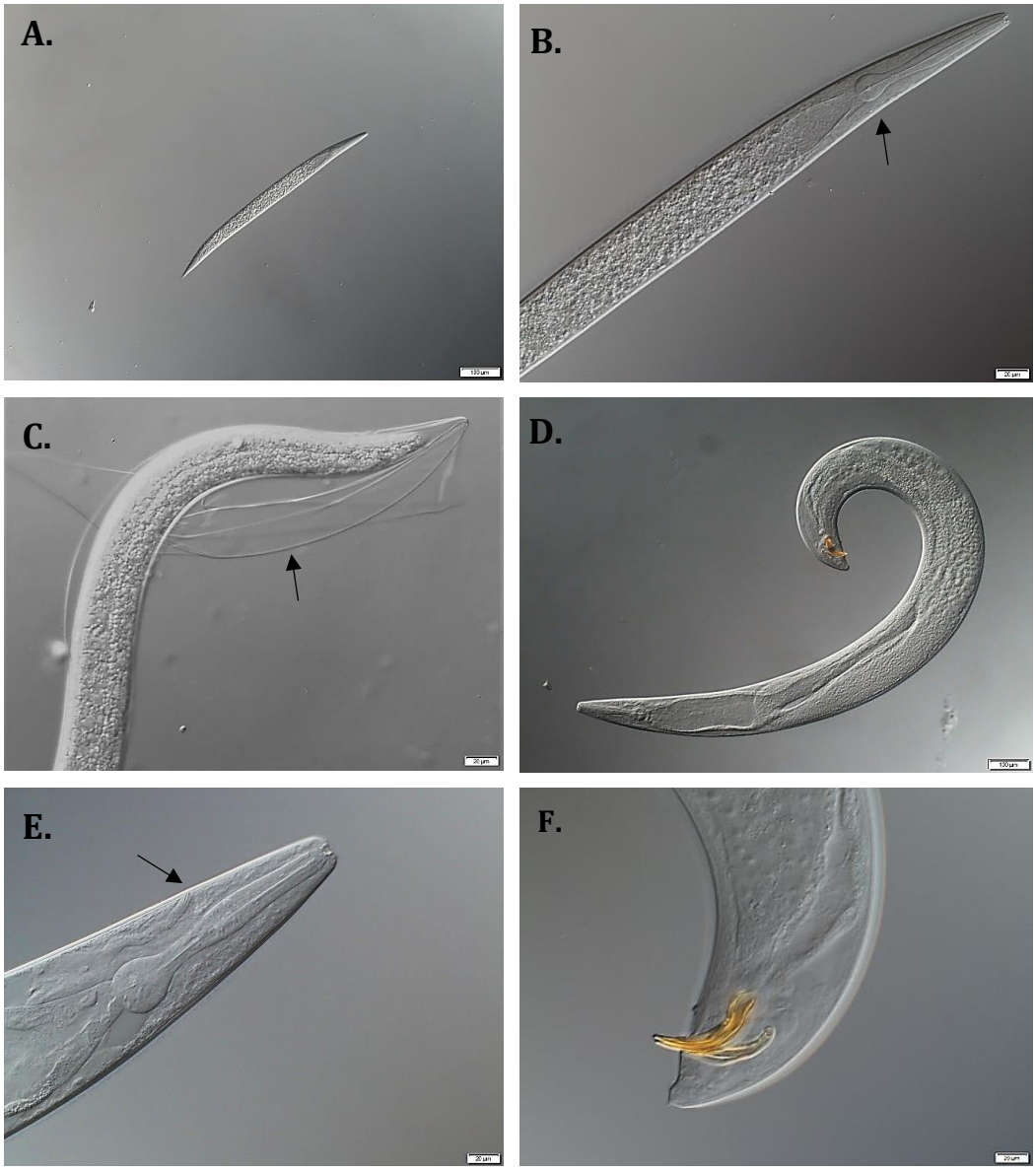


Figure 3.2. *Steinernema* sp. HBG28. Infective juvenile (A-C), **A:** Entire length; **B:** Anterior region displaying pharynx (arrow) and intestine. **C:** Tail region with sheath (arrow) retained from the second stage (J2) cuticle. First generation male (D-F), **D:** Entire length; **E:** Anterior region displaying excretory pore (arrow); **F:** Lateral view of tail region showing spicule and gubernaculum. Scale bars: A, D: 100 µm; B, C, E, F: 20 µm.

3.4.1.4 First generation female

Body robust and habitus C-shaped or coiled to form a closed 6-shape when heat-killed and fixed (Figure 3.3A). Cuticle appears smooth under a light microscope. Head slightly rounded, tapered smoothly at anterior end and continuous with body shape (Figure 3.3B). Stoma shallow and funnel-shaped. Pharynx with cylindrical procorpus, slightly swollen metacorpus, distinct isthmus and pyriform, valvate basal bulb. Nerve ring located anterior to basal bulb, surrounding isthmus. Excretory pore located mid-pharynx level, posterior to the metacorpus with cuticularized excretory duct (Figure 3.3C). Vulva a median transverse slit, situated 53% from head region with slightly protruding vulva lips and double-flapped epiptygmata present (Figure 3.3D-E). Tail short, bluntly conoid and appearing dome-shaped with a terminal peg (Figure 3.3F).

3.4.1.5 Second generation female

Similar morphology to first generation females but smaller in size. Body slender and slightly curved into a C-shape when heat-relaxed. Tail conical, tapering to sharp tip, mucron absent and slight postanal swelling present.

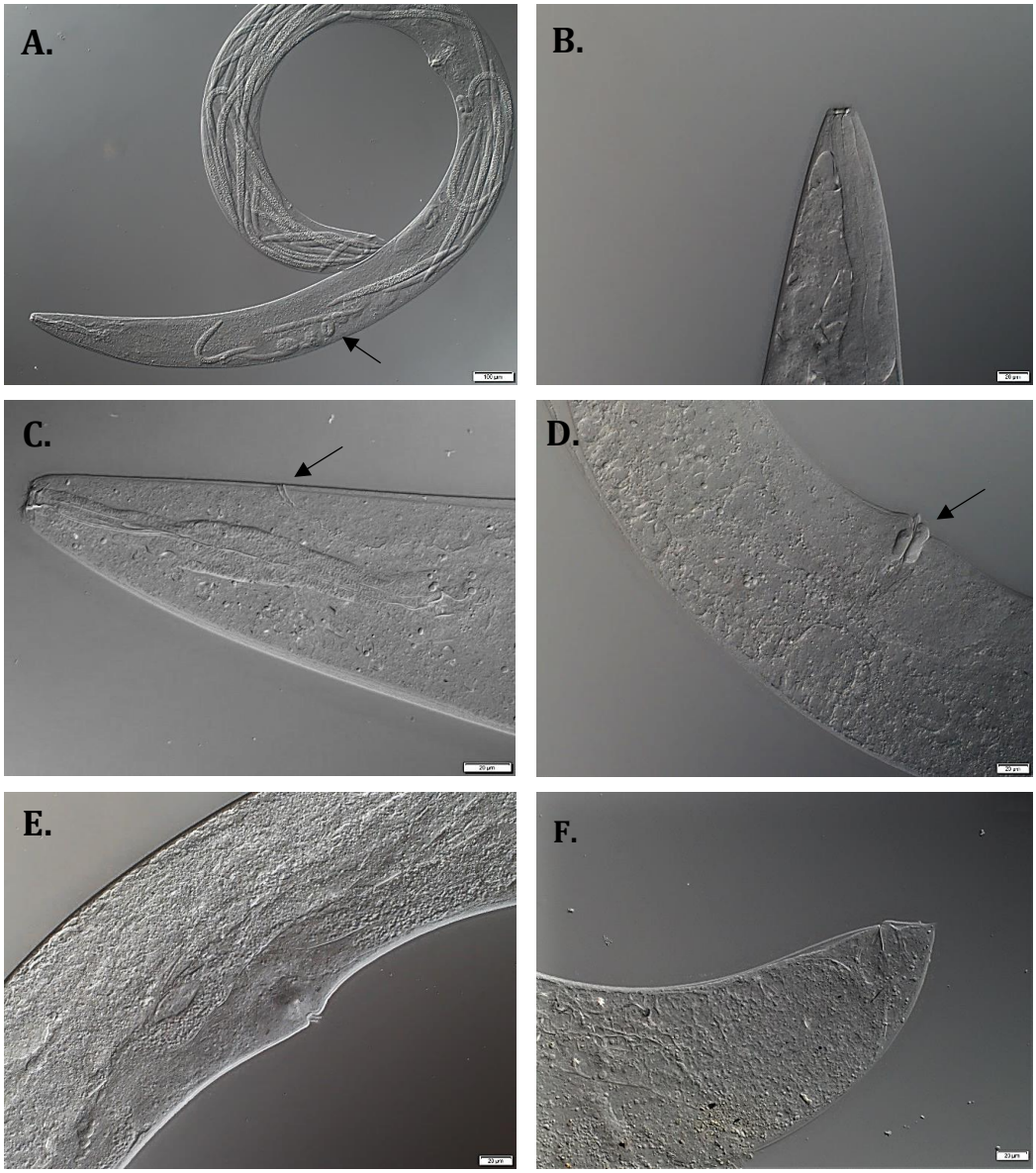


Figure 3.3. *Steinernema* sp. HBG28. First generation female (A-F), **A:** Entire length of female exhibiting *endotokia matricida* with first stage juveniles (J1) developing into J2; **B:** Anterior region. **C:** Anterior region showing excretory pore and excretory duct (arrow); **D**, **E:** Vulva with slightly protruding lips and double-flapped epiptygmata (arrow); **F:** Tail region showing a terminal peg at the tip. Scale bars: A: 100 µm; B, C, D, E, F: 20 µm.

3.4.2 Measurements

See Table 3.1.

3.4.3 Type host and locality

Natural hosts are unknown. *Steinernema* n. sp. (isolate HBG28) was isolated from a soil sample collected from the Suikerbosrand Nature Reserve in Gauteng, South Africa by baiting with *G. mellonella* larvae.

3.4.4 Diagnosis and relationships

Steinernema n. sp. (isolate HBG28) can be characterized from other related species in the genus by differences in morphological and morphometric features of various nematode life (Table 3.1). The infective juvenile of the new species is distinguished from all recognized *Steinernema* species by a body length of 655 (518-720) µm, body diameter of 40 (28-51) µm, distance from anterior end to excretory pore of 46 (40-49) µm, distance from anterior end to nerve ring of 59 (53-65) µm, distance from anterior end to base of pharynx of 96 (79-112) µm, tail length of 58 (51-67) µm, anal body diam. of 25 (17-33) µm, D% = 49 (39-60), E% = 79 (71-94) and hyaline tail portion almost equal to half the tail length, i.e. H% = 51 (48-55) (Table 3.1).

The first generation male of the new species is characterized by a long, soft-orange spicule with a length of 85 (82-88) µm long and a developed velum, along with a gubernaculum of 70 (68-73) µm in length. Other diagnostic characters include D% = 58 (54-63), E% = 287 (278-303), SW% = 175 (158-183) and GS% = 82 (76-89). The second-generation males are comparable to the first generation in terms of morphological characters but are smaller in size and have a wider body diameter. Both first and second generation males possess mucrons. The first generation females

have slightly protruding vulvae with double-flapped epiptygmata. The first generation female has a dome-shaped tail with a terminal peg, as well as a slightly developed postanal swelling, whilst the second-generation female has a conical and sharply-pointed tail, with a well-developed postanal swelling. *Endotokia matricida* was also observed amongst some first generation females.

Phylogenetic analyses of sequences of ITS and 28S rDNA-D2D3 regions used in Chapter 2 indicate that *Steinernema* n. sp. (isolate HBG28) shares common ancestry with S*. fabii*, clustering to a new monophyletic group, referred to as the 'Cameroonian' clade, which consists of *S. sacchari*, S*. cameroonense* and *S. nyetense* (Nthenga et al., 2014; Abate et al., 2016). The new species can be distinguished from all closely-related *Steinernema* species within the aforementioned clade by various diagnostic features (Tables 3.1, 3.2), including a wider body diameter of the infective juvenile 40 (28-51) µm and a longer length of the first-generation male gubernaculum 70 (68-73) µm. *Steinernema* n. sp. (isolate HBG28) can be differentiated from *S. fabii* by a wider body diameter and longer pharynx length of the first generation males (Table 3.3). The body diameter of the first generation male of *Steinernema* n. sp. (isolate HBG28) at 167 (156-178) µm is wider than *S. fabii*, *S. sacchari*, *S. cameroonense* and *S. nyetense*.

The first generation females of *Steinernema* n. sp. (isolate HBG28) differ from those of *S. cameroonense* and *S. nyetense* by not possessing a mucron. Furthermore, females of the new species have slightly protruding vulvae and double-flapped epiptygmata, whereas females of *S. sacchari* have non-protruding vulvae, in spite of having double-flapped epiptygmata as well.

When comparing the infective juveniles, the body length of *Steinernema* n. sp. (isolate HBG28) is longer than *S. fabii*, *S. nyetense* and *S. cameroonense*, but is shorter

than *S. sacchari* (Table 3.2). Additionally, the first generation males of *Steinernema* n. sp. (isolate HBG28) has a mucronate tail whilst *S. sacchari* lacks this feature.

Table 3.1 Morphometrics of *Steinernema* n. sp. (isolate HBG28). All measurements are in µm and in the form: mean ± standard deviation (range).

Character	First generation		Second generation		Infective juvenile
	Male	Female	Male	Female	
n	20	20	20	20	20
L	1880 ± 100.6 (1765-1950)	4171 ± 177 (4025-4368)	1148 ± 80.3 (1101-1241)	1506 ± 145 (1360-1651)	655 ± 80.3 (518-720)
a	11 ± 1.1 (10-13)	19 ± 1.2 (18-21)	16 ± 0.65 (15-17)	14 ± 1.7 (12-16)	17 ± 3.7 (14-23)
b	10.8 ± 0.9 (9.8-11.5)	19.9 ± 1.1 (18-21)	8 ± 0.9 (7-8.8)	9 ± 0.7 (8-10)	7 ± 1 (5.8-8.5)
c	56 ± 5.7 (50-60)	104 ± 12.6 (91-117)	41 ± 2.4 (40-44)	27 ± 2.5 (24-30)	11 ± 1.7 (10-14)
ć	0.7 ± 0.1 (0.7-0.8)	0.6 ± 0.1 (0.5-0.8)	0.8 ± 0.1 (0.7-0.9)	1.4 ± 0.5 (0.9-1.9)	2.4 ± 0.6 (1.6-3)
V%	– –	53 ± 1.6 (52-55)	– –	54 ± 1.4 (52-56)	– –
Body diam. (BD)	167 ± 11 (156-178)	217 ± 20.8 (200-240)	74 ± 2 (72-76)	106 ± 3.3 (102-110)	40 ± 9.7 (28-51)
Excretory pore (EP)	101 ± 4.2 (98-106)	106 ± 9.6 (95-113)	86 ± 7.9 (77-92)	82 ± 11.5 (70-94)	46 ± 3.6 (40-49)
Nerve ring (NR)	116 ± 6.4 (111-123)	137 ± 7.5 (130-145)	110 ± 14 (96-124)	105 ± 3.4 (101-109)	59 ± 5.1 (53-65)
Pharynx length (ES)	175 ± 6.11 (168-180)	210 ± 7.2 (204-218)	143 ± 12.3 (133-157)	167 ± 4 (161-172)	96 ± 14.9 (79-112)
Tail length (T)	35 ± 3.5 (33-38)	41 ± 6.6 (35-38)	28 ± 3 (25-31)	56 ± 5.6 (50-62)	58 ± 7.5 (51-67)
Anal body diam. (ABD)	49 ± 3.9 (45-53)	66 ± 3 (64-70)	36 ± 1.9 (33-38)	41 ± 4.7 (36-46)	25 ± 6.9 (17-33)
Hyaline region (H)	– –	– –	– –	– –	30 ± 2.8 (27-34)
Spicule length (SL)	85 ± 3.3 (82-88)	– –	70 ± 0.6 (69-71)	– –	– –
Spicule width (SW)	14 ± 2.1 (11-16)	– –	12 ± 1.4 (11-14)	– –	– –
Gubernaculum length (GL)	70 ± 3.2 (68-73)	– –	48 ± 3.2 (46-52)	– –	– –
Gubernaculum width (GW)	9 ± 1 (7-10)	– –	7 ± 0.4 (7-8)	– –	– –
D% = EP/ES x 100	58 ± 4.5 (54-63)	50 ± 3.9 (46-55)	61 ± 10.4 (49-70)	49 ± 6.4 (42-56)	49 ± 10.7 (39-60)
E% = EP/T x 100	287 ± 13.6 (278-303)	265 ± 39 (235-310)	311 ± 49.5 (277-368)	146 ± 19 (127-165)	79 ± 9.3 (71-94)
SW% = SL/ABD x 100	175 ± 14 (158-183)	– –	195 ± 5.9 (188-200)	– –	– –
GS% = GL/SL x 100	82 ± 6.5 (76-89)	– –	69 ± 4.9 (65-75)	– –	– –
H% = H/T x 100	– –	– –	– –	– –	51 ± 2.4 (48-55)

n=number of specimens
L=body length
a=(L/BD)
b=(L/ES)
c=(L/T)

Table 3.2 Comparative morphometrics of infective juveniles of *Steinernema* n. sp. (isolate HBG28) and related *Steinernema* spp. belonging to the Cameroonian group (in descending order of body length). All measurements are in µm and in the form: mean (range).

Species	Morphometric character											References
	L	BD	EP	NR	ES	T	a	b	c	D%	E%	
S. sacchari	680	37	53	84	113	64	19	6	11	47	82	Nthenga et al., 2014
	(630-722)	(30-47)	(49-58)	(78-97)	(104-127)	(51-74)	(14-23)	(6-7)	(10-12)	(41-54)	(70-109)	
S. n. sp. (HBG28)	655	40	46	59	96	58	17	7	11	49	79	–
	(518-720)	(28-51)	(40-49)	(53-65)	(79-112)	(51-67)	(14-23)	(5.8-8.5)	(10-14)	(39-60)	(71-94)	
S. nyetense	648	32	52	85	114	82	21	6	8	46	66	Kanga et al., 2012
	(565-708)	(25-37)	(46-57)	(72-102)	(104-128)	(54-113)	(19-26)	(5-6)	(6-11)	(37-50)	(44-89)	
S. fabii	641	28	53	65	132	58	24	4.8	11	41	93	Abate et al., 2016
	(590-697)	(26-31)	(49-57)	(55-84)	(120-146)	(52-64)	(21-41)	(4.4-5.3)	(10-12)	(35-60)	(83-105)	
S. cameroonense	622	30	54	85	113	76	21	6	9	48	75	Kanga et al., 2012
	(490-694)	(24-35)	(45-64)	(69-100)	(105-125)	(52-107)	(17-25)	(5-6)	(6-12)	(42-56)	(48-116)	

Table 3.3 Comparative morphometrics of first generation males of *Steinernema* n. sp. (isolate HBG28) and related *Steinernema* spp. belonging to the Cameroonian group (in descending order of spicule length (L)). All measurements are in µm and in the form: mean (range).

Species	Morphometric character					
	Spicule	Gubernaculum	BD	D%	SW%	GS%
S. fabii	90	66	126	64	177	73
	(79-106)	(56-77)	(97-153)	(52-75)	(126-224)	(63-86)
S. n. sp. (HBG28)	85	70	167	58	175	82
	(82-88)	(68-73)	(156-178)	(54-63)	(158-183)	(76-89)
S. sacchari	83	61	145	67	171	73
	(73-89)	(50-68)	(86-205)	(54-88)	(146-210)	(66-81)
S. nyetense	80	53	106	55	199	66
	(67-98)	(40-62)	(62-159)	(40-70)	(125-283)	(51-77)
S. cameroonense	69	45	90	64	170	64
	(51-85)	(37-57)	(65-124)	(48-76)	(131-201)	(47-76)

3.5 Discussion

Approaches utilizing classical taxonomy have been on a continual decline over the years (Andre et al., 2001; Coomans, 2002), with only a few qualified nematode taxonomists still active. The advent of new technologies has resulted in a reliance on molecular-based diagnostic protocols for nematode identification; and this is especially true for the identification of entomopathogenic nematodes. The sole application of morphological and morphometric characterization of nematodes can be challenging on account of overlapping morphometric measurements, morphological similarity, inadequate number of unique taxonomic characters (Oliveira et al., 2011) and the use of hazardous fixing agents. Nevertheless, to improve the reliability of nematode systematics, studies should preferably incorporate the use of both morphological and molecular techniques (De Ley and Blaxter, 2002).

In the present study, morphological and morphometric characterization was performed on the third-stage infective juvenile and adult stages of a *Steinernema* isolate recovered from the grasslands of Gauteng, South Africa. The current diagnostic morphological and morphometric findings, together with molecular data based on the ITS-rDNA region (in Chapter 2), indicate that *Steinernema* n. sp. (isolate HBG28) was a unique species that belongs to the Cameroonian clade.

3.6 Conclusion

The recovered species is adapted to a grassland environment and could potentially be used to control local insect pests acclimated to the same environmental conditions. Further morphological studies would need to be performed, including the application of scanning electron microscopy (SEM) and cross hybridization studies involving nematodes within the Cameroonian clade to validate the current results.

3.7 References

1. Andre, H. M., Ducarme, X., Anderson, J. M., Crossley, Jr., et al. (2001). Skilled eyes are needed to go on studying the richness of the soil. *Nature*, **409:**761.
2. Ayako, H., Ehlers, R. U. and Strauch, O. (2010). Life cycle and population development of the entomopathogenic nematodes *Steinernema carpocapsae* and *S. feltiae* (Nematoda, Rhabditida) in monoxenic liquid culture. *Nematology*, **12:**201–210.
3. Çimen, H., Lee, M. M., Hatting, J., Hazir, S., et al. (2014). *Steinernema tophus* sp. n. (Nematoda: Steinernematidae), a new entomopathogenic nematode from South Africa. *Zootaxa*, **3821:**337–353.
4. Çimen, H., Lee, M. M., Hatting, J., Hazir, S., et al. (2015). *Steinernema innovationi* n. sp. (Panagrolaimomorpha: Steinernematidae), a new entomopathogenic nematode species from South Africa. *Journal of Helminthology*, **89:**415–427.
5. Coomans, A. (2002). Present status and future of nematode systematics. *Nematology*, **4:**573–582.
6. De Ley, P. and Blaxter, M. L. (2002). Systematic position and phylogeny. In: The Biology of Nematodes, D. L. Lee, (Eds.), London: Taylor and Francis, pp. 1–30.
7. Kanga, F. N., Trinh Quang, P., Waeyenberge, L., Spiridonov, S. E., et al. (2012). Two new species of *Steinernema Travassos*, 1927 from the humid forest of southern Cameroon. *Russian Journal of Nematology*, **20:**15–36.
8. Kaya, H. K. and S. P. Stock. (1997). Techniques in insect nematology. In Manual of techniques in insect pathology, L. A. Lacey (ed.). Biological Techniques Series, Academic Press, San Diego, California, p. 281–324.
9. Malan, A. P., Knoetze, R. and Tiedt, L. R. (2015). *Steinernema jeffreyense* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Journal of Helminthology*, **90:**262–278.

10. Nguyen, K. B., Malan, A. P. and Gozel, U. (2006). *Steinernema khoisanae* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology*, **8:**157–175.
11. Nthenga, I., Knoetze, R., Berry, S., Tiedt, L. R., et al. (2014). *Steinernema sacchari* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology*, **16:**475–494.
12. Oliveira, C. M. G., Monteiro, A. R. and Blok, V. C. (2011). Morphological and molecular diagnostics for plant-parasitic nematodes: Working together to get identification done. *Tropical Plant Pathology*, **36:** 65–73.
13. Seinhorst, J. W. (1959). A rapid method for the transfer of nematodes from fixative to anhydrous glycerin. *Nematologica*, **4:**67–69.
14. Shapiro-Ilan, D. I. and Gaugler, R. (2002). Production technology for entomopathogenic nematodes and their bacterial symbionts. *Journal of Industrial Microbiology and Biotechnology*, **28:**137–146.
15. Shapiro-Ilan, D. I., Gouge, D. H. and Koppenhöfer, A. M. (2002). Factors affecting commercial success: Case studies in cotton, turf and citrus. pp. 333–356 in R. Gaugler, (Eds.), *Entomopathogenic Nematology*. Wallingford, UK: CABI Publishing.
16. Steyn, W. P., Knoetze, R., Tiedt, L. R., and Malan, A. P. (2017). *Steinernema litchii* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology*, **19:**1157–1177.
17. Stock, S. P. and Goodrich-Blair, H. (2012). Nematode parasites, pathogens and associates of insects and invertebrates of economic importance. In: Lacey, L.A. (Ed.). Manual of techniques in invertebrate pathology, 2nd edition. Amsterdam, The Netherlands, Elsevier, pp. 373-426.
18. Stokwe, N. F., Malan, A. P., Nguyen, K. B., Knoetze, R., et al. (2011). *Steinernema citrae* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology*, **13:**569–587.

19. Thomas, G. M. and Poinar, G. O. Jr. (1979). *Xenorhabdus* gen. nov., a genus of entomopathogenic, nematophilic bacteria of the family *Enterobacteriaceae*. *International Journal of Systematic Bacteriology*, **29:**352–360.
20. Wang, J. and Bedding, R. A. (1996). Population development of *Heterorhabditis bacteriophora* and *Steinernema carpocapsae* in larvae of *Galleria mellonella*. *Fundamental and Applied Nematology*, **19:**363–367.
21. Weiser, J. (1995). *Neoaplectana carpocapsae* n. sp. (Anguillulata, Steinernematidae) novy cizopasnic housenik obalece jablecneho, *Carpocapsa* pomonella L. *Vestnik Ceskoslovenske Zoologicke Spolecnosti*, **19:**44–52.

Chapter 4

Isolation and molecular identification of bacterial endosymbionts, *Xenorhabdus khoisanae* and *Photorhabdus heterorhabditis*

4.1 Abstract

During a small-scale survey conducted in the North West Province and Gauteng of South Africa, two entomopathogenic nematode spp. were recovered, i.e. *Heterorhabditis zealandica* and an undescribed steinernematid. The aim of this study was to molecularly identify the bacterial symbionts associated with the recovered entomopathogenic nematodes by sequencing the 16*S* rDNA and performing phylogenetic analyses on the partial sequences. Bacterial symbionts (isolates VMG and NB28) recovered from *H. zealandica* and *Steinernema* sp. HBG28, respectively, were identified as *Photorhabdus heterorhabditis* and *Xenorhabdus* khoisanae, respectively. This study is also the first report of swarming behaviour exhibited by phase II variants of *Xenorhabdus*, as was observed in *X. khoisanae* strain MCB on solid agar.

4.2 Introduction

Members of the genera *Photorhabdus* (Boemare, Akhurst and Mourant, 1993) and *Xenorhabdus* (Thomas and Poinar, 1979) are Gram-negative, pathogenic γ-Proteobacteria, belonging to the family *Enterobacteriaceae*. Unlike other genera within the family, *Photorhabdus* and *Xenorhabdus* represent a unique group of bacteria in terms of phenotypic (Holt et al., 1994) and genotypic (Brenner and Farmer, 2005) characteristics. A key trait shared by all species of *Photorhabdus* and *Xenorhabdus* is their ability to form symbiotic relationships with entomopathogenic nematodes of the families Heterorhabditidae and Steinernematidae, respectively (Adams et al., 2006). Among the species of *Photorhabdus*, one exception can be found in *P. asymbiotica*, which is capable of infecting both insect and human hosts (Farmer et al., 1989; Gerrard et al., 2004).

The bacterial symbionts undergo intricate lifecycles that revolve around two eukaryotic hosts: a free-living, infective juvenile form of the nematodes they associate with, called infective juveniles and a susceptible larval stage of insects that they parasitize. The bacterium-nematode association is high dynamic and mutually beneficial, whilst the relationship that exists between the insect host and bacterium is strictly parasitic in nature. The bacteria are harboured as symbionts in the intestinal tract of the infective juvenile. Free-living forms of the symbiotic bacteria are rarely found in nature in the absence of a nematode or insect host (Akhurst et al., 2004; Tailiez et al., 2006). This suggests that the bacterial symbionts require their nematode hosts for survival, reproduction and dissemination (Forst et al., 1997). The infective juveniles are highly resilient in a soil environment, persisting for long periods of time while searching for suitable insect hosts to parasitize. When a suitable host is encountered, the infective juvenile nematodes penetrate the insect either

through natural openings or through the cuticle, migrating to the haemocoel. The bacterial symbionts are released into the haemolymph where they begin to rapidly proliferate to high numbers. During this time, the bacteria release anti-immune and antimicrobial proteins, which enables the bacteria to overcome the host immune system and inhibit the growth of competitor microorganisms (Forst and Clarke, 2002). This process also coincides with the release of bacterial toxins, which results in rapid insect death through septicaemia and toxaemia. The host tissues are metabolized by the symbiotic bacteria. The bacterial biomass serves as a nutritional resource for the developing nematodes, which reproduce, resulting in thousands of offspring. Following the depletion of all available nutrients, the bacterial symbionts re-associate with the new nematode progeny before exiting the insect cadaver. The bacterial symbionts are retained by the infective juveniles until a new insect host can be infected.

Over the years, the bacterial symbionts and their nematode partners have received considerable attention due to their high insecticidal activity against various insect pests; and have been successfully implemented in biological control and integrated pest management programs (Grewal et al., 2005). Previous research has also focused on the symbiotic bacteria and the secondary metabolites they produce, in the attempt of identifying biologically-active compounds that are antimicrobial and insecticidal in nature.

While *Photorhabdus* and *Xenorhabdus* spp. can be propagated under standard laboratory conditions, both *Photorhabdus* and *Xenorhabdus* spp. are known to spontaneously switch between two cell types or forms due to phase variation (Volgyi et al., 1998). Typically, this phenotypic switching in *Xenorhabdus* and *Photorhabdus* bacteria is highly variable, occurring in response to changes in the environment or

growth conditions, though it is reversible. The two forms of bacteria are distinct in morphology and physiology. The primary form or phase I variants refers to the bacteria naturally retained by the infective juvenile nematodes and contribute to nematode growth, development and pathogenicity (Boemare and Akhurst, 1988). In contrast, the secondary form or phase II variants are not associated with insect virulence, nor provides conditions suitable for nematode growth and development. Though the underlying mechanism of phase variation remains to be fully understood, it has been indicated that DNA rearrangements (Akhurst et al., 1992), changes in plasmid content (Leclerc and Boemare, 1991) and recombination-mediated processes (Pinyon et al., 2000) are not the driving force behind this process (Sugar et al., 2012).

To date, four highly diverse species of *Photorhabdus* have been described: *P. luminescens*, *P. temperate* and *P. asymbiotica*. A new species, *P. heterorhabditis* has been added to the *Photorhabdus* genus, having been isolated in South Africa (Ferreira et al., 2014). Of the *Photorhabdus* spp., three have been further split into subspecies based on 16S rDNA sequence data, i.e.: *P. luminescens* subsp. *luminescens* (Thomas and Poinar 1979), *P. asymbiotica* subsp. *asymbiotica*, *P. luminescens* subsp. *laumondii*, *P. luminescens* subsp. *akhurstii*, *P. temperate*, *P. temperata* subsp. *temperata* (Fischer-Le Saux et al., 1999), *P. asymbiotica* subsp. *australis* (Akhurst et al., 2004), *P. luminescens* subsp. kayaii, *P. luminescens* subsp. *thracensis*, *P. temperata* subsp. thracensis (Hazir et al., 2004), *P. temperata* subsp. *cinerea* (Tóth and Lakatos, 2008), *P. luminescens* subsp. *caribbeanensis*, *P. temperata* subsp. *khaini*, *P. temperata* subsp. *tasmaniensis*, *P. luminescens* subsp. *hainanensis* (Tailliez et al. 2010), *P. luminescens* subsp. *kleinii*, *P. temperata* subsp. *stackebrandtii* (An and Grewal, 2011), *P. luminescens* subsp. *namnaonensis* (Glaeser et al., 2017), *P. luminescens* subsp. *noenieputensis* (Ferreira et al., 2013).

Within the genus of *Xenorhabdus,* there are twenty-six recognized species, which includes: *X. nematophila* (Poinar and Thomas, 1979), *X. bovienii* (Akhurst, 1988), *X. poinarii*, *X. beddingii*, *X. japonica* (Nishimura et al., 1994), *X. budapestensis* (Lengyel et al., 2005), *X. ehlersii*, *X. innexi*, *X. szentirmaii*, *X. cabanillasii* (Tailliez et al., 2006), *X. doucetiae*, *X. griffiniae*, *X. hominickii*, *X. koppenhoeferi*, *X. kozodoii*, *X. mauleonii*, *X. miraniensis*, *X. romanii*, *X. stockiae*, *X. indica* (Somvanshi et al., 2009), *X. vietnamensis* (Tailliez et al., 2010), *X. magdalenensis* (Tailliez et al., 2012), *X. ishibashii* (Kuwata et al., 2013), *X. khoisanae* (Ferreira et al., 2013), *X. eapokensis* (Kämpfer et al., 2017), *X. thuongxuanensis*.

Comparison of 16S rRNA gene sequences is useful for the accurate identification of bacterial symbionts associated with entomopathogenic nematodes (Szállás et al., 1997). The relationship that exists between the symbiotic bacterium and entomopathogenic nematode is one that could provide insight into the mechanisms that are involved in symbiont specificity, host selection and bacterial function.

4.2.1 Aims and objectives

During a survey conducted in the North West Province and Gauteng of South Africa, two species of entomopathogenic nematodes were collected, i.e. *Heterorhabditis zealandica* and an unknown species of *Steinernema*. The aim of this study was to isolate and molecularly identify the bacterial symbionts associated with these recovered species of entomopathogenic nematodes by sequencing the internal transcribed spacer (ITS) regions of the 16S rDNA and perform phylogenetic analyses on the partial sequences.

4.3 Materials and methods

4.3.1 Source of insects and nematodes

Larvae of *Galleria mellonella* (L.) (Lepidoptera: Pyralidae) were reared on an artificial diet adapted from Woodring and Kaya (1988), consisting of Pronutro wheat-free cereal, glycerol, honey, boiled water, yeast extract and sodium benzoate. All experiments were conducted using entomopathogenic nematodes that were recovered from soil samples collected in Brits, North-West Province and the Suikerbosrand Nature Reserve, Gauteng, which included *Heterorhabditis zealandica* (GenBank accession number: KT715757) and a novel *Steinernema* sp. HBG28 (GenBank accession number: KJ877686), respectively. *In-vivo* rearing of infective juveniles was performed using *G. mellonella* larvae. The infective juveniles were recovered using modified White traps (Kaya and Stock, 1997), and were maintained at room temperature, according to the methodology used in Chapter 2 (section 2.3.2.4). Additionally, *G. mellonella* larvae exposed to infective juveniles were collected and used for subsequent experiments involving bacterial isolation.

4.3.2 Bacterial isolation and growth conditions

Bacterial isolates VMG and NB28 of *Heterorhabditis zealandica* and *Steinernema* sp. HBG28, respectively were isolated indirectly from the haemolymph of *G. mellonella* larvae within 48 hours of infection by infective juvenile nematodes, according to the procedure described by Akhurst (1980). The insect cadavers were surface-sterilized in 70% (v/v) ethanol, briefly ignited and quenched by submerging into sterile distilled water. A drop of haemolymph was extracted by piercing the prothoracic region of the insect cadaver using a sterilized 28-gauge needle and syringe and plated onto nutrient agar supplemented with 0.025% bromothymol blue and 0.004% triphenyltetrazolium chloride (NBTA) to determine dye adsorption. All plates were

incubated in the dark at 25°C for 48-72 hrs. Preliminary identification of the bacterial isolates was performed by observing colony morphology (i.e. colony shape, colour and shape). Phase I variants of the symbiotic bacteria were obtained by selecting blue-green or green colonies on NBTA medium and were routinely sub-cultured on NBTA plates and in 10 mL of nutrient broth at 25°C for 48 hrs. Bacterial cultures were maintained in 40% glycerol at -80°C.

4.3.3 *In vitro* culturing to confirm nematode-bacterium association

To confirm the obligate association between the isolated bacteria and their respective nematode hosts, bacterial lawns were prepared on lipid agar (LA; 2.5% nutrient agar, 1% honey, 0.5% yeast extract, 0.25% cod liver oil and 0.2% $MgCl_2.6H2O$), adapted from Kaya and Stock (1997). Blue and blue-green colonies were selected from NBTA plates and transferred to nutrient broth (NB) (Biolab). Nutrient broth cultures were incubated overnight at 25°C. Spread plates of the nutrient broth culture were prepared on LA plates and incubated for 24-48 hours at 25°C. Newly emerged infective juveniles were harvested and surface-sterilized in a 0.1% (v/v) sodium hypochlorite for 1 hour and was washed three times in Ringer solution. Bacterial lawns were inoculated with infective juveniles (500 IJs/ml) corresponding to their respective nematode hosts and the plates were incubated at 25°C to assess if the bacterial isolates supported nematode growth and development. LA plates were monitored over a period of 10-14 days, and detection of developing nematodes confirmed the symbiotic nature of the bacterial isolate in question, prior to molecular characterization.

4.3.4 Genotypic characterization

Genomic DNA was isolated from bacterial colonies prepared on NBTA plates, using the ZR fungal/bacterial DNA MiniPrep kit (Zymo Research), according to the protocol supplied by the manufacturer. Quantification of the bacterial DNA extractions was assessed using a NanoDrop® ND-1000 spectrophotometer (Nanodrop Technologies). The 16S rDNA region was amplified using the forward primer EUB968 (5'-AACGCGAAGAACCTTAC-3') (Zheng et al., 2012) and the reverse primer 1392R (5'-ACGGGCGGTGTGTRC -3') (Lane et al., 1991).

Each PCR reaction consisted of 3 µL of each primer, 3 µL of the DNA suspension, which served as template DNA, 14 µL nuclease-free water, together with 25 µL of DreamTaq Green PCR Master Mix (2X), which contained DNA polymerase, PCR buffer, dATP, dCTP, dGTP and dTTP (0.4 mM each), and 4 mM $MgCl_2$ (Thermo Scientific). The final reaction volume was 50 µl. All amplification reactions were performed using a thermocycler (GeneAmp PCR system 2700®), with an initial denaturation cycle of 95°C for 3 min, 35 cycles of denaturation at 94°C for 30 sec, amplification at an annealing temperature of 60°C for 45 sec and an extension period at 72°C for 90 sec. A final extension cycle at 72°C for 7 min was performed after the last cycle. All PCR products were sequenced on an ABI 3130XL sequencer ABI 3130xl (Applied Biosystems, Foster City, California, USA) at Inqaba Biotechnical Industries, Pretoria, South Africa. Sequence editing and verification of base calls were performed using the software, FinchTV version 1.4 (Geospiza, Seattle, WA, USA).

4.3.5 Phylogenetic analyses

The 16S rRNA gene sequences were queried against the National Centre for Biotechnology Information (NCBI) database to compare for nucleotide similarity

against sequences of the closest relatives, utilizing the Basic Local Alignment Search Tool (BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi). Five housekeeping (*recA*, *gyrB*, *dnaN*, *gltX* and *infB*) gene sequences were obtained directly from the genomes of *Photorhabdus heterorhabditis* strain VMG (accession no. LJCS00000000) and *Xenorhabdus khoisanae* strain MCB (accession no. LFCV01000000, where homologous housekeeping gene sequences of *Photorhabdus* and *Xenorhabdus* spp. were identified and queried against the genomes using BLAST.

The 16S rRNA and housekeeping gene sequences were aligned using MUSCLE (Edgar, 2004) and the sequence alignments were concatenated. The phylogenetic relationships of the bacterial isolates with other closely related species of bacteria were observed using the program, Molecular Evolutionary Genetics Analysis, version 7 (MEGA 7) (Tamura et al., 2011). The best nucleotide substitution model using the Bayesian Information Criterion (BIC) was determined for all gene sequences with MEGA7. All phylogenetic trees were constructed with outgroups.

4.4 Results

4.4.1 Isolation of the symbiotic bacteria

Symbiotic bacteria were successfully isolated indirectly from the haemolymph of *G. mellonella* larvae infected with *H. zealandica* and *Steinernema* sp. HBG28 nematodes. The putative *Photorhabdus* sp. VMG (isolate VMG) and *Xenorhabdus* sp. NB28 (isolate NB28) bacteria were sub-cultured on NBTA media to acquire pure cultures prior to 16S rDNA sequencing. All isolates tested were found to absorb bromothymol blue dye, a characteristic indicative of phase I variants.

The symbiotic bacterium of *H. zealandica*, isolate VMG produced glossy, dark green colonies that were round, convex and viscid in nature (Figure 4.1A). By comparison, the bacterial symbiont of *Steinernema* sp. HBG28, isolate NB28 produced small to medium-sized, round colonies that were blue-green in colour, smooth in texture and glossy in appearance (Figure 4.1B). Phenotypic variants were observed for both *Photorhabdus* and *Xenorhabdus* isolates, albeit the bacterial colonies of only *Xenorhabdus* variants are shown in Figure 4.1C Phase II variants were characterized by a loss in dye-binding ability and appeared red-brown in colour.

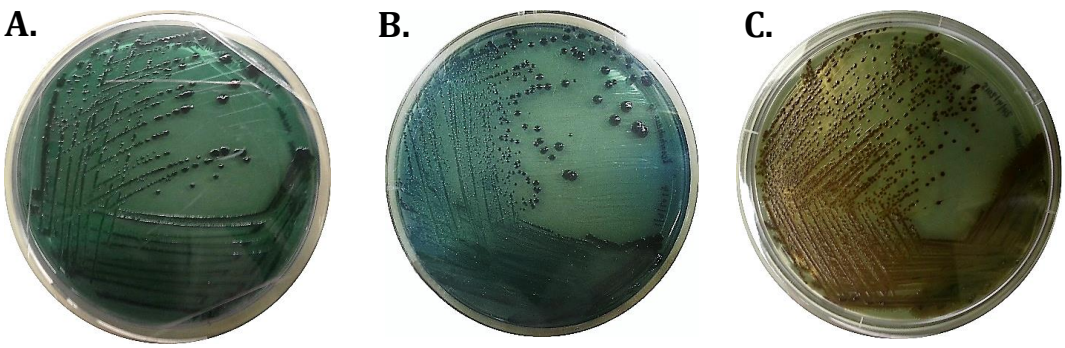


Figure 4.1. Bacterial colonies of the symbiotic bacteria. A: Dark-green colonies of a putative *Photorhabdus* sp. (isolate VMG) isolated from *Heterorhabditis zealandica* were observed, later identified as *P. heterorhabditis*. **B:** Blue-green bacterial colonies of a putative *Xenorhabdus* sp. (isolate NB28) isolated from an undescribed *Steinernema* sp. were observed, later identified as *X. khoisanae*. **C:** Phase II bacterial colonies of the *Xenorhabdus* sp. NB28 were observed, characterized by red-brown colonies, due to a lack of dye adsorption.

4.4.2 Swarming behaviour of isolate NB28

Some bacteria exhibit a distinct behaviour on solid agar referred to as 'swarming', which involves a special form of bacterial surface translocation (Sharma and Anand, 2002). Phase I and II variants of symbiotic bacterium associated with *Steinernema* sp. HBG28 demonstrated swarming behaviour, as shown in Figure 4.2A and 4.2B, respectively. Swarming colonies formed dendritic (branch-like) patterns for both phenotypic variants. Swarming was not observed in the symbiotic bacterium (isolate VMG) associated with *H. zealandica*.

4.4.3 Nematode growth and development on symbiont lawns

Lawns of phase I bacteria were prepared on LA plates (Figure 4.3A) to establish if the bacterial isolates were able to support nematode growth and development. Both bacterial isolates, VMG and NB28 were able to facilitate nematode propagation (Figure 4.3B), thus confirming the symbiotic nature of the bacterial isolates in question. The entomopathogenic nematodes were able to undergo 2-3 rounds of reproduction on the bacterial lawns within a period of 7-14 days. *Endotokia matricida* was observed in female nematodes after two rounds of reproduction (Figure 4.3C).

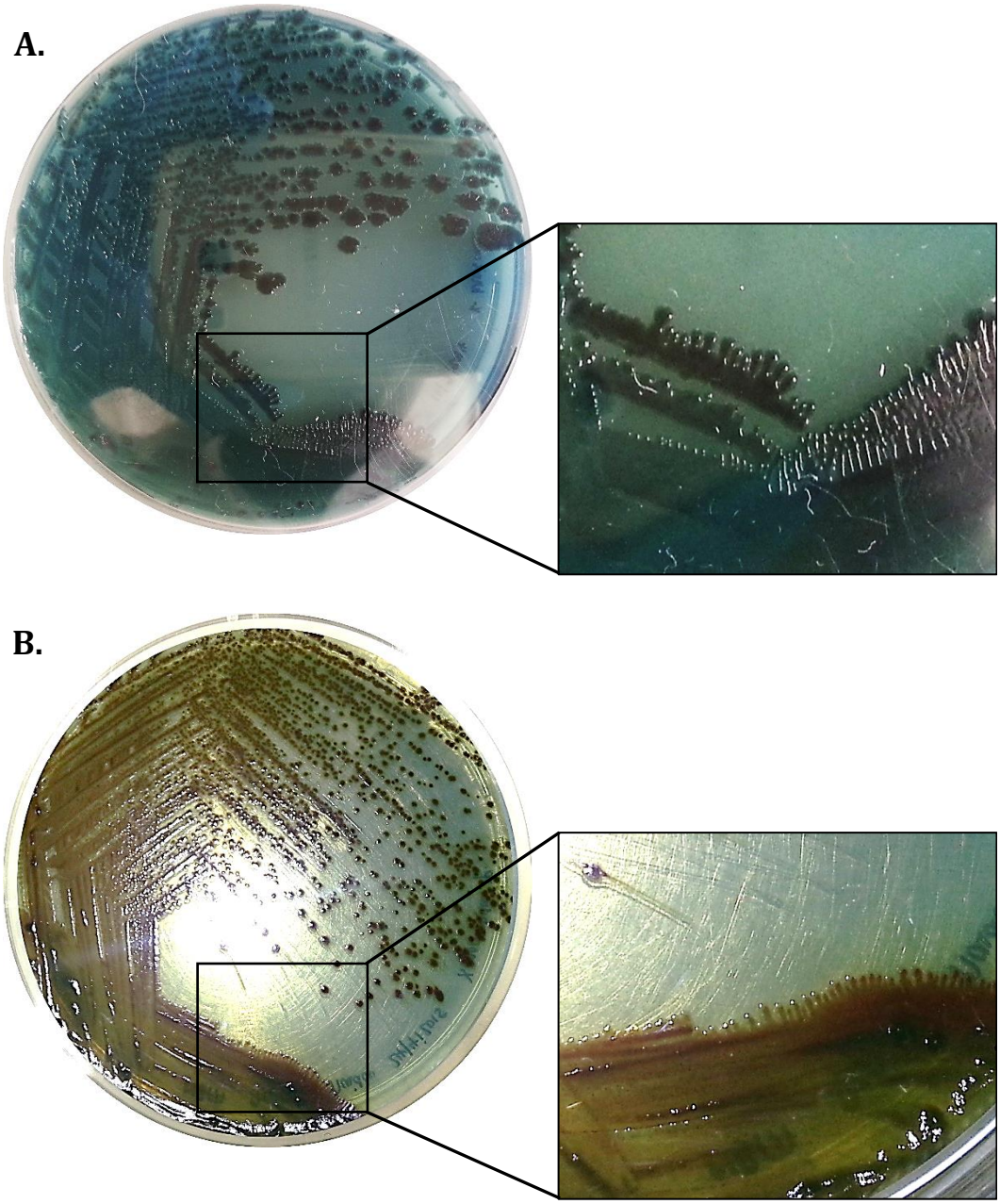


Figure 4.2. Swarming behaviour of isolate NB28, later identified as *Xenorhabdus khoisanae*. Swarming presented as tendril-like colonies on NBTA agar. **A:** Swarming exhibited in phase I variants; **B:** Swarming exhibited in phase II variants of *X. khoisanae,* with zoomed portions showing colour and colony morphology.

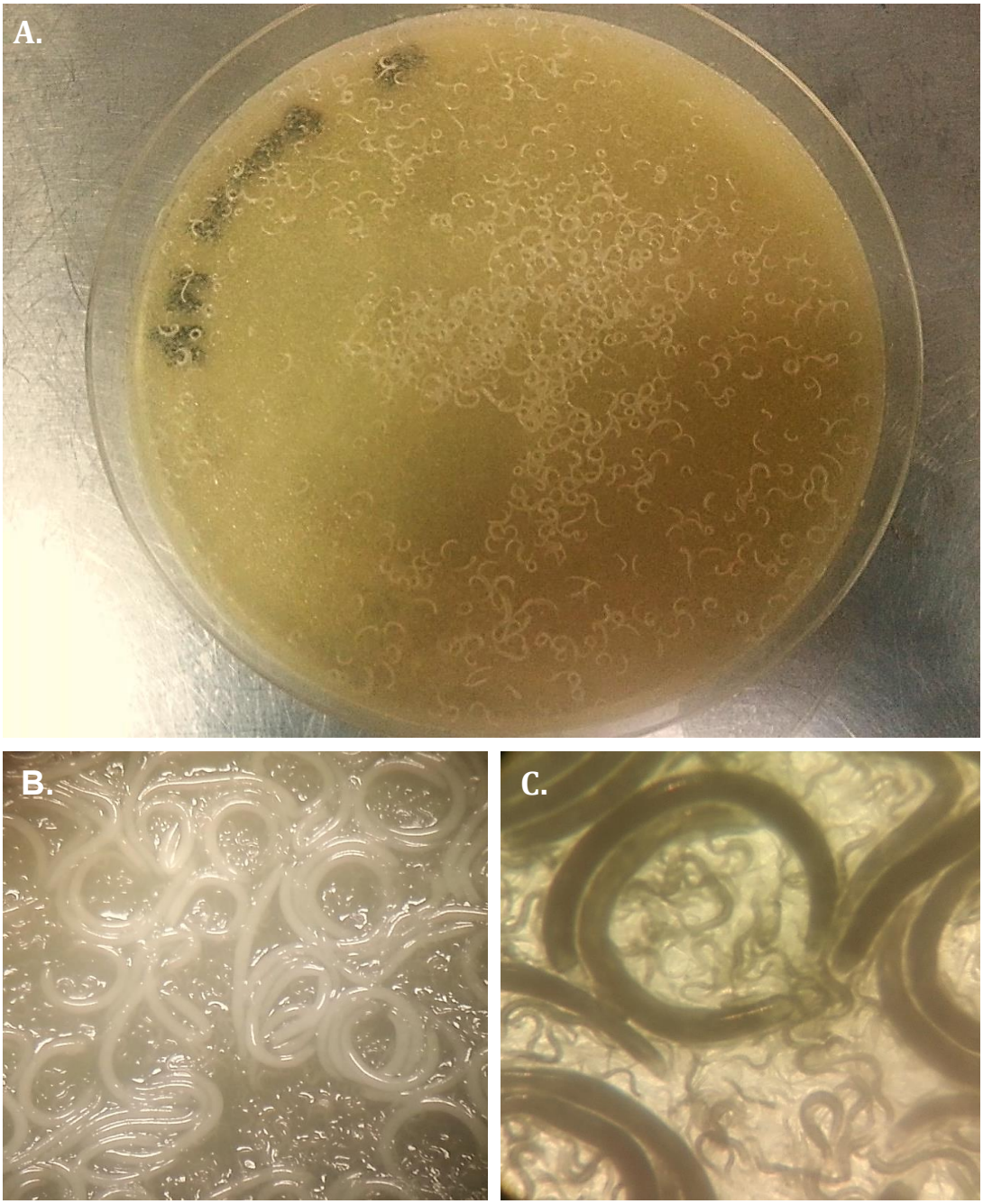


Figure 4.3. *In-vitro* culturing of entomopathogenic nematodes with their symbiotic bacteria on lipid agar to confirm obligate symbiosis. Lawns of the symbiotic bacteria were prepared on lipid agar, inoculated with infective juveniles (500 IJs/ml) and incubated for 3-5 days at 25°C. The observation of nematode growth and development on bacterial lawns confirmed the dependency of the nematodes on the bacterial isolates in question. **A:** Entomopathogenic nematodes, *Heterorhabditis zealandica* undergoing their life cycle on a lawn of symbiotic bacteria (later identified as *Photorhabdus heterorhabditis*); **B:** First- and second-generation adults of entomopathogenic nematodes were visible to the naked eye; **C:** Second-generation female nematodes undergoing *endotokia matricida*, observed under a dissecting microscope.

4.4.4 Molecular characterization of the symbiotic bacteria

The 16S rRNA gene sequences of isolate VMG and NB28 were sequenced in the lengths of 509 and 374 nucleotides, respectively. All sequences were submitted to BLAST and queried against the NCBI nucleotide database for a similarity search. The symbiotic bacteria of *H. zealandica* and *Steinernema* sp. HBG28 were confirmed to be *Photorhabdus* and *Xenorhabdus* species, respectively, based on the analysis of their 16S rRNA gene sequences. Sequence alignment with those of other *Photorhabdus* spp. indicated that *Photorhabdus* sp. VMG had the highest similarity (97%) with *P. heterorhabditis*. Similarly, when compared with gene sequences of other representative *Xenorhabdus* spp., *Xenorhabdus* sp. NB28 was shown to have the highest similarity (99%) with *X. khoisanae.* The partial 16S rDNA nucleotide sequence of isolate VMG and NB28 were submitted to GenBank, with the accession numbers, KT692968 and KT122844, respectively.

4.4.5 Phylogenetic analyses

The 16S rRNA gene sequence of isolate VMG were compared to those of closely related *Photorhabdus* strains, species and subspecies. The phylogenetic analysis indicated that the bacterial symbiont that was associated with *H. zealandica* clustered with *P. heterorhabditis* strain SF783 with 90% bootstrap support (Figure 4.4). From the phylogenetic analysis based on the five concatenated housekeeping gene sequences, it was revealed that isolate VMG clustered with *P. heterorhabditis* strain SF41 (Figure 4.5).

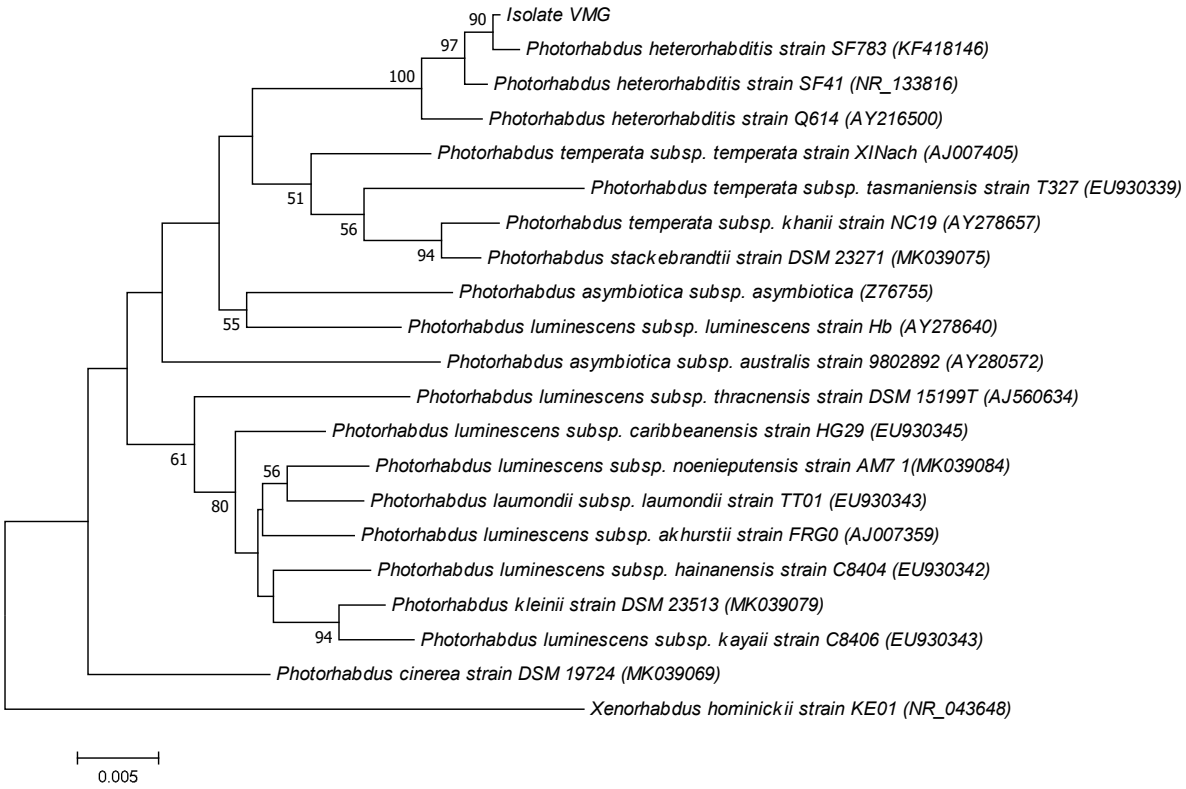


Figure 4.4. Phylogenetic relationship of isolate VMG with 19 *Photorhabdus* strains, species and subspecies based on the sequences of the 16S rRNA gene. The tree was constructed using the Maximum Likelihood method based on the General Time Reversible model of substitution with gamma-distributed with invariant sites (G+I) as determined by MEGA to best fit with the data using the BIC criterion. The sequence of *X. hominickii* strain KE01 as an outgroup. Bootstrap values of 1000 replicates are indicated below the nodes as percentages with values above 50% displayed at the branch point.

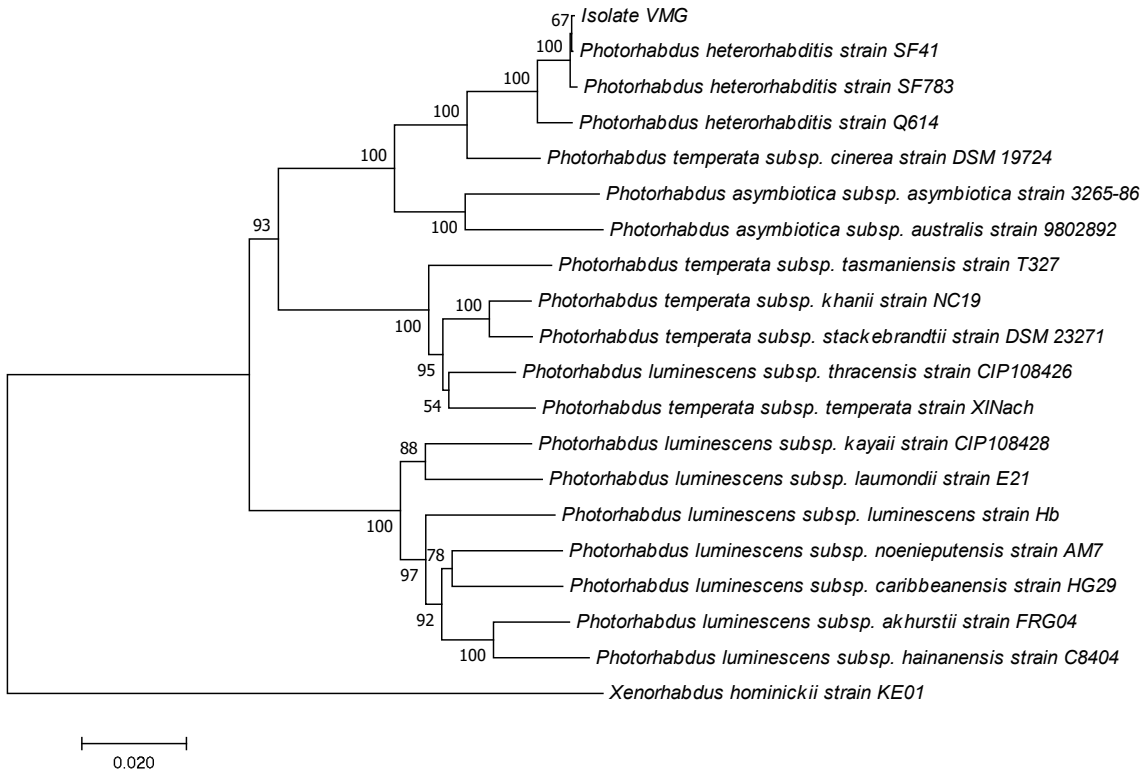


Figure 4.5. Phylogenetic relationship of isolate VMG with 18 *Photorhabdus* strains, species and subspecies based on five concatenated housekeeping gene sequences (*recA*, *gyrB*, *dnaN, gltX* and *infB*). The tree was constructed using the Maximum Likelihood method based on the Kimura-2 parameter model of substitution with gamma-distributed with invariant sites (G+I) as determined by MEGA to best fit with the data using the BIC criterion. The concatenated sequences of *X. homonickii* strain KE01 as an outgroup. Bootstrap values of 1000 replicates are indicated below the nodes as percentages with values above 50% displayed at the branch point.

The concatenated (16S rRNA, recA, gyrB, dnaN, gltX and infB) gene sequences of isolate NB28 were compared to sequences of closely related *Xenorhabdus* strains and species. Results obtained for the phylogenetic analysis indicated that isolate NB28 was closely related to *X. khoisanae*, with a high bootstrap support (Figure 4.6).

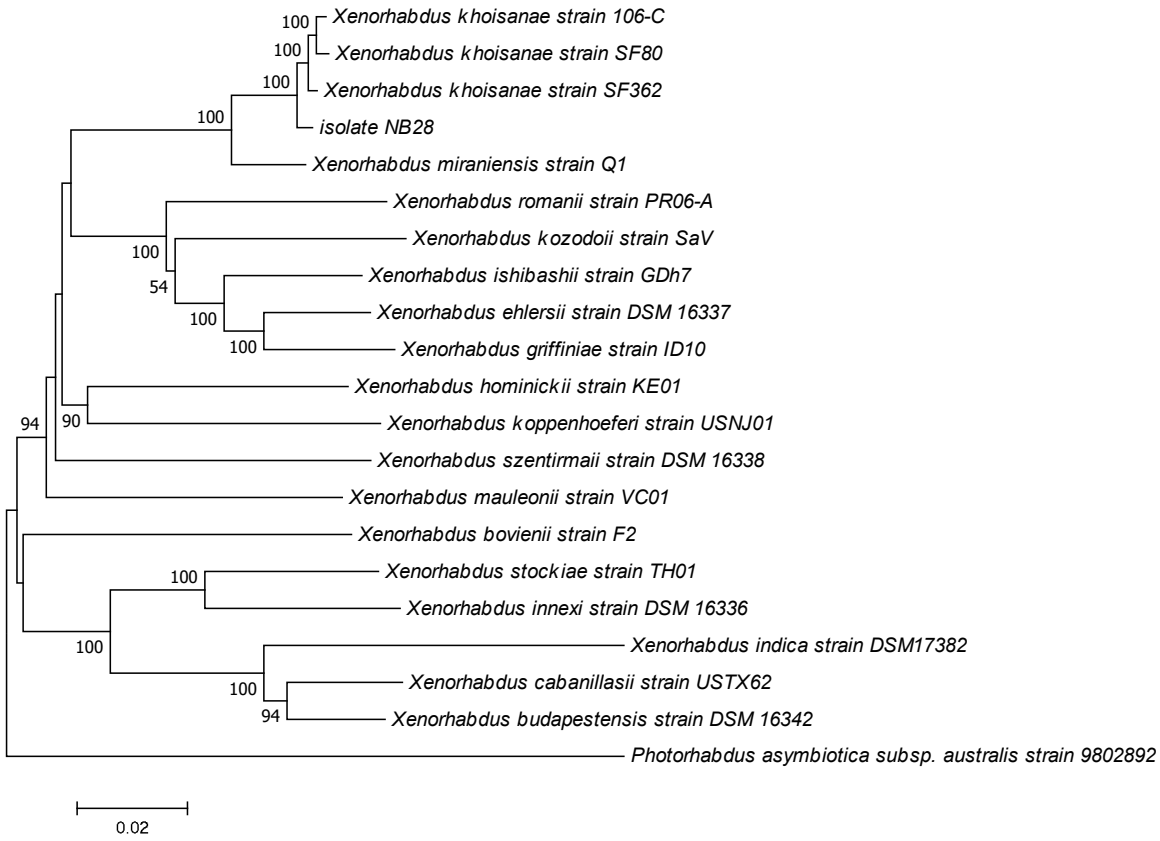


Figure 4.6. Phylogenetic relationship of isolate NB28 with 19 *Xenorhabdus* strains and species based on concatenated (16S rRNA, *recA*, *gyrB*, *dnaN*, *gltX* and *infB*) gene sequences. The tree was constructed using the Maximum Likelihood method based on the General Time Reversible model of substitution with gamma-distributed with invariant sites (G+I) as determined by MEGA to best fit with the data using the BIC criterion. The concatenated sequences of *X. homonickii* strain KE01 as an outgroup. Bootstrap values of 1000 replicates are indicated below the nodes as percentages with values above 50% displayed at the branch point.

4.5 Discussion

The aim of the present study was to isolate and identify the symbiotic bacteria associated with indigenous entomopathogenic nematode species recovered during a survey of the North West Province and Gauteng of South Africa. The symbiotic bacteria of *H. zealandica* and a new *Steinernema* sp. HBG28 were successfully isolated indirectly from the haemolymph of *G. mellonella* larvae infected with infective juvenile nematodes.

Phase I and II variants were observed for both bacterial isolates, exhibiting differences in colony morphology and dye adsorption. Phase I bacterial symbionts were characterized by the ability to bind bromothymol blue dye, producing blue-green colonies on NBTA medium. Phase II variants presented as small-sized, red colonies on NBTA medium, and the loss in dye-binding ability. These observations are consistent with the findings of other related studies. Phenotypic variants are commonly observed among *Xenorhabdus* and *Photorhabdus* spp. (Akhurst, 1980). Typically, phase I cells are distinctively larger (Givaudan et al., 1995) and are known to adsorb specific dyes on agar plates (Akhurst and Boemare, 1988). In contrast, these characteristics are either absent or significantly reduced in phase II variants (Akhurst, 1980; Boemare and Akhurst , 1988; Forst et al., 1997; Givaudan and Lanois, 2000). Phase II variants are known to occur with prolonged incubation *in vitro* (Akhurst, 1982; Bleakley and Nealson et al., 1988), as was observed in the current study.

The present study also included the occurrence of swarming behaviour for at least one bacterial isolate on solid agar, i.e. the putative *Xenorhabdus* sp. NB28 (isolate NB28). Both phase I and II variants of *Xenorhabdus* sp. NB28 exhibited swarming motility, presenting as highly-branched patterns away from the central colonies.

Swarming is a specialized form of bacterial surface movement on solid media, facilitated by rotating flagella (Henrichsen, 1972). It is considered as a behavioural response to a surface, allowing flagellated bacteria to mimic a multicellular population in the colonization of nutrient-rich solid substrates (Shapiro, 1998). Notably, swarming has also been observed in several bacterial genera, including *Bacillus*, *Clostridium*, *Escherichia*, *Proteus*, *Vibrio*, *Pseudomonas*, *Salmonella*, *Serratia*, (Kearns, 2010), in addition to *Xenorhabdus* (Givaudan et al., 1995) and *Photorhabdus* (Hurst et al., 2015). In a previous study by Givaudan et al. (1995), it was demonstrated that swarming behaviour only occurred in phase I variants of *X. nematophila*, *X. bovienii*, *X. poinarii* and *X. beddingii*., but not in phase II cells. It was further implied that the lack of motility in phase II variants was as a result of a defect that occurred during flagella synthesis. It is interesting that the phenotypic variants of *Xenorhabdus* sp. NB28 presented in this study were still able to retain the ability to swarm despite undergoing phase variation, while this form of motility is entirely absent in phase II variants of other *Xenorhabdus* species such as *X. nematophila* and *X. bovienii*. Under natural conditions, the ability of swarming in symbiotic bacteria could play an essential role in mutualism and parasitism, allowing for rapid and coordinated migration during the colonization of the nematode host and the insect body cavity (Givaudan et al., 1995).

One key feature of the symbiosis between entomopathogenic nematodes and their associated bacterial symbionts is the nutritional dependency of these nematodes on their bacteria (Han and Ehlers, 2001). Isolates VMG and NB28 both demonstrated the ability to support the nematode growth and development on lipid agar, resulting in high progeny production. Adult development was observed for *H. zealandica* and *Steinernema* sp. HBG28 nematodes in the presence of their bacterial

symbionts. This, in turn, confirmed the symbiotic nature of the isolates NB28 and VMG.

The bacterial isolates were molecularly identified by performing phylogenetic analyses based on the 16S rRNA and five housekeeping (*recA*, *gyrB*, *dnaN*, *gltX* and *infB*) gene sequences. For isolate VMG, gene sequence analysis indicated that the bacterium belonged to the genus *Photorhabdus*. When compared with sequences of other closely related *Photorhabdus* strains, species and subspecies, it was shown that isolate VMG clustered with *P. heterorhabditis* strains. The species, *Photorhabdus heterorhabditis* is a new addition to the *Photorhabdus* genus and has similarly been shown to associate with other *Heterorhabditis zealandica* nematodes in South Africa as previously reported by Ferreira et al., 2014. By comparison, a nucleotide similarity search indicated that the symbiotic bacterium of *Steinernema* sp. HBG28 belonged to the genus *Xenorhabdus*. The 16S rRNA and five housekeeping gene sequences of isolate NB28 were compared with sequences of closely related species of *Xenorhabdus* and the results indicated that it was closely related to *X. khoisanae*.

The symbiotic bacterium, *X. khoisanae* has been found to associate with more than one *Steinernema* hosts, including *S. sacchari* (Nthenga et al., 2014) and *S. jeffreyense* (Malan et al., 2016). *Steinernema sacchari* and *Steinernema* sp. HBG28 belong to a new group, the 'Cameroonian' clade, while *S. jeffreyense* belongs to *glaseri*-clade; and thus, reflecting the ability of *X. khoisanae* to associate with several nematodes of the same clade, as well as distantly related clades (Dreyer et al., 2017). While *Steinernema* spp. will only associate with a single species of *Xenorhabdus* bacteria, the *Xenorhabdus* sp. does not demonstrate the same level of fidelity, forging several associations with different *Steinernema* hosts at a time (McMullen et al., 2017). Such host promiscuity in *Xenorhabdus* bacteria is not uncommon among

Steinernema-Xenorhabdus symbioses (Boemare, 2002; Stock and Goodrich-Blair, 2012), and has been described extensively in *X. bovienii*, which associates with more than one *Steinernema* species at a time (Lee and Stock, 2010a; Lee and Stock, 2010b; Tailliez et al., 2006). As such, it is not surprising that *X. khoisanae* would exhibit similar tendencies in having symbiotic associations were multiple *Steinernema* spp. found in South Africa. A possible explanation for the high prevalence of *X. khoisanae* among several indigenous *Steinernema* spp. may be attributed to the local distribution of these nematodes, which has limited gene duplication opportunities for *X. khoisanae.* It can further be argued that the preservation of *X. khoisanae* among local *Steinernema* spp. may allude to how ancient this bacterial symbiont is and that it has been in existence prior to the speciation events of local *Steinernema* spp. with which it now associates. As a consequence, the symbiont has been retained over time, having been carried over from its ancestral species of *Steinernema* to sister host species.

4.6 Conclusion

The isolation of bacterial symbionts from entomopathogenic nematodes retrieved during survey in South Africa is important for several reasons. It provides an indication of the natural occurrence and distribution of symbiotic bacteria amongst native nematodes in various localities and habitats. More importantly, it can allow for the discovery of new species and isolates with greater infectivity towards local insect pests. This study, for the first time, demonstrates swarming behaviour in phase II variants of *Xenorhabdus* bacteria. It is clear that flagellar motility in *Xenorhabdus* bacteria plays an important role in the course of insect infection and the recolonization of the nematode host. To some extent, this swarming phenomenon may even contribute to the high prevalence of *X. khoisanae* observed among several *Steinernema* spp. in South Africa; and, as such, would require investigation.

4.7 References

1. Adams, B. J., Fodor, A., Koppenhöfer, H. S., Stackebrandt, E., et al. (2006). Biodiversity and systematics of nematode-bacterium entomopathogens. *Biological Control*, **37:**32–49.
2. Akhurst, R. J. (1980). Morphological and functional dimorphism in *Xenorhabdus* spp., bacteria symbiotically associated with the insect pathogenic nematodes *Neoplectana* and *Heterorhabditis*. *Journal of General Microbiology*, **121:**303–309.
3. Akhurst, R. J. (1980). Morphological and functional dimorphism in *Xenorhabdus* spp., bacteria symbiotically associated with the insect pathogenic nematodes *Neoplectana* and *Heterorhabditis*. *Journal of General Microbiology*, **121:**303–309.
4. Akhurst, R. J. (1982). Antibiotic activity of *Xenorhabdus* spp., bacteria symbiotically associated with insect pathogenic nematodes of the families Heterorhabditidae and Steinernematidae. *Journal of General Microbiology*, **128:**3061–3065.
5. Akhurst, R. J. and Boemare, N. E. (1988). A numerical taxonomic study of the genus *Xenorhabdus* (*Enterobacteriaceae*) and proposed elevation of the subspecies of *X. nematophilus* to species. *Journal of General Microbiology*, **134:**1835–1845.
6. Akhurst, R. J., Boemare, N. E., Janssen, P. H., Peel, M. M., et al. (2004). Taxonomy of Australian clinical isolates of the genus *Photorhabdus* and proposal of *Photorhabdus asymbiotica* subsp. *asymbiotica* subsp. nov. and *P. asymbiotica* subsp. *australis* subsp. nov. *International Journal of Systematic and Evolutionary Microbiology*, **54:**1301–1310.
7. Akhurst, R. J., Smigielski, A. J., Mari, J., Boemare, N. et al. (1992). Restriction analysis of phase variation in *Xenorhabdus* spp. (*Enterobacteriaceae*), entomopathogenic bacteria associated with nematodes. *Systematic and Applied Microbiology*, **5:**469–473.

8. An, R. and Grewal, P. (2010). *Photorhabdus temperata* subsp. *stackebrandtii* subsp. nov. (Enterobacteriales: *Enterobacteriaceae*). *Current Microbiology*, **61:**291–297.
9. Bleakley, B. and Nealson, K. H. (1988). Characterization of primary and secondary forms of *Xenorhabdus luminescens* strain Hm. *FEMS Microbiology Ecology*, **53:**241–250.
10. Boemare, N. E. (2002). Biology, taxonomy, and systematics of *Photorhabdus* and *Xenorhabdus*. In: Gaugler R, editor. Entomopathogenic Nematology; p. 35–56.
11. Boemare, N. E. and Akhurst, R. J. (1988). Biochemical and physiological characterization of colony form variants in *Xenorhabdus* spp. (*Enterobacteriaceae*). *Journal of General Microbiology*, **134:**751–761.
12. Boemare, N. E., Akhurst, R. J. and Mourant, R. G. (1993). DNA relatedness between *Xenorhabdus* spp. (*Enterobacteriaceae*), symbiotic bacteria of entomopathogenic nematodes, and a proposal to transfer *Xenorhabdus luminescens* to a new genus, *Photorhabdus* gen. nov. *International Journal of Systematic Bacteriology*, **43:**249–255.
13. Brenner, D. J. and Farmer, J. L. (2005). Family 1. *Enterobacteriaceae*. In: Garrity, G. M., Brenner, D. J., Krieg, N. R. and Staley, J. T. (eds), Bergey's Manual of Systematic Bacteriology, Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria, 2nd ed., Springer, New York, USA, pp. 587–607.
14. Dreyer, J., Malan, A. and Dicks, L. M. T., (2017). Three Novel *Xenorhabdus-Steinernema* Associations and Evidence of Strains of Switching Between Different Clades. *Current Microbiology*, **74:**938–942.
15. Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5:**113.
16. Farmer, J. J., Jorgensen, J. H., Grimont, P. A., Akhurst, R. J., et al. (1989). *Xenorhabdus luminescens* (DNA hybridization group 5) from human clinical specimens. *Journal of Clinical Microbiology*, **27:**1594-1600.

17. Ferreira, T., van Reenen, C. A., Endo, A., Spröer, C., et al. (2013). Description of *Xenorhabdus khoisanae* sp. nov., the symbiont of the entomopathogenic nematode *Steinernema khoisanae*. *International Journal of Systematic and Evolutionary Microbiology*, **63:**3220–3224.
18. Ferreira, T., van Reenen, C. A., Endo, A., Tailliez, P., et al. (2014). *Photorhabdus heterorhabditis* sp. nov., a symbiont of the entomopathogenic nematode *Heterorhabditis zealandica*. *International Journal of Systematic and Evolutionary Microbiology*, **64:**1540–1545.
19. Ferreira, T., van Reenen, C. A., Pagès, S., Tailliez, P., et al. (2013). *Photorhabdus luminescens* subsp. *noenieputensis* subsp. nov., a symbiotic bacterium associated with a novel *Heterorhabditis* species related to *Heterorhabditis indica*. *International Journal of Systematic and Evolutionary Microbiology*, **63:**1853–1858.
20. Fischer-Le Saux, M., Viallard, V., Brunel, B., Normand, P., et al. (1999). Polyphasic classification of the genus *Photorhabdus* and proposal of new taxa: *P. luminescens* subsp. *luminescens* subsp. nov., *P. luminescens* subsp. *akhurstii* subsp. nov., *P. luminescens* subsp. *laumondii* subsp. nov., *P. temperata* sp. nov., *P. temperata* subsp. *temperata* subsp. nov. and *P. asymbioyica* sp. nov. *International Journal of Systematic Bacteriology*, **49:**1645–1656.
21. Forst, S. and Clarke, D. (2002). Bacteria-nematode symbiosis. In: Gaugler, R. (eds), *Entomopathogenic Nematology*, CABI Publishing, Wallingford, UK, pp. 57–77.
22. Forst, S., Dowds, B., Boemare, N. and Stackebrandt, E. (1997). Bugs that kill bugs. *Annual Review of Microbiology*, **51:**47–72.
23. Gerrard, J., Waterfield, N., Vohra, R. and ffrench-Constant, R. (2004). Human infection with *Photorhabdus asymbiotica*: an emerging bacterial pathogen. *Microbes and Infection*, **6:**229-237.

24. Givaudan, A. and Lanois, A. (2000). flhDC, the flagellar master operon of *Xenorhabdus nematophilus*: Requirement for motility, lipolysis, extracellular hemolysis, and full virulence in insects. *Journal of Bacteriology*, **182:**107–115.
25. Givaudan, A., Baghdiguian, S., Lanois, A. and Boemare, N. (1995). Swarming and swimming changes concomitant with phase variation in *Xenorhabdus nematophilus*, *Applied and Environmental Microbiology*, **61:**1408-1413.
26. Glaeser, S. P., Tobias, N. J., Thanwisai, A., Chantratita, N., et al. (2017). *Photorhabdus luminescens* subsp. *namnaonensis* subsp. nov., isolated from *Heterorhabditis baujardi* nematodes. *International Journal of Systematic and Evolutionary Microbiology*, **67:**1046–1051.
27. Grewal, P. S., Ehlers, R. U. and Shapiro-Ilan, D. I., editors. (2005). Nematodes as Biocontrol Agents. New York: CABI.
28. Han, R. C. and Ehlers, R. U. (2001). Effect of *Photorhabdus luminescens* phase variants on the *in vivo* and *in vitro* development and reproduction of the entomopathogenic nematodes *Heterorhabditis bacteriophora* and *Steinernema carpocapsae*. *FEMS Microbiology Ecology*, **35:**239–247.
29. Hazir, S., Stackebrandt, E., Lang, E., Schumann, P., et al. (2004). Two new subspecies of *Photorhabdus luminescens*, isolated from *Heterorhabditis bacteriophora* (Nematoda: Heterorhabditidae): *Photorhabdus luminescens* subsp. *kayaii* subsp. nov. and *Photorhabdus luminescens* subsp. *thracensis* subsp. nov. *Systematic and Applied Microbiology*, **27:**36–42.
30. Henrichsen, J. (1972). Bacterial surface translocation: a survey and a classification. *Bacteriological Reviews*, **36:**478–503.
31. Holt, J. G., Krieg, N. R., Sneath, P. H., Staley, J. T., et al. (1994). Bergey's manual of determinative bacteriology, 9th ed. 188–189.
32. Hurst, S., Rowedder, H., Michaels, B., Bullock, H., et al. (2015). Elucidation of the *Photorhabdus temperata* Genome and generation of a transposon mutant library

to identify motility mutants altered in Pathogenesis. *Journal of Bacteriology*, **197:**2201–2216.

33. Inman, F. L. III and Holmes, L. D. (2012a). Mass production of the beneficial nematode *Heterorhabditis bacteriophora* and its bacterial symbiont *Photorhabdus luminescens*. *Indian Journal of Microbiology*, **52:**316–324.
34. Inman, F. L. III and Holmes, L. D. (2012b). The effects of trehalose on the bioluminescence and pigmentation of the phase I variant of *Photorhabdus luminescens*. *Journal of Life Sciences*, **6:**119–129.
35. Kämpfer, P., Tobias, N. J., Ke, L. P., Bode, H. B., et al. (2017). *Xenorhabdus thuongxuanensis* sp. nov. and *Xenorhabdus eapokensis* sp. nov., isolated from *Steinernema* species. *International Journal of Systematic and Evolutionary Microbiology*, **67:**1107–1114.
36. Kaya, H. K. and Stock, S. P. (1997). Techniques in insect nematology. In: Lacey, L.A. (ed.), Manual of techniques in insect pathology, Academic Press, London, UK, pp. 281-324.
37. Kearns, D. B. (2010). A field guide to bacterial swarming motility. *Nature Reviews Microbiology*, **8:**634–644.
38. Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions though comparative studies of nucleotide sequences. *Journal of Molecular Evolution,* **16:**111-120.
39. Kuwata, R., Qiu, L. H., Wang, W., Harada, Y., et al. (2013). *Xenorhabdus ishibashii* sp. nov., isolated from the entomopathogenic nematode *Steinernema aciari*. *International Journal of Systematic and Evolutionary Microbiology*, **63:**1690–1695.
40. Lane, D. J. (1991). 16S/23S rRNA sequencing. In: Stackebrandt, E. and Goodfellow, M. (eds.), Nucleic Acid Techniques in Bacterial Systematics, John Wiley & Sons, New York, pp. 115-175.

41. Leclerc, M. C. and Boemare, N. E. (1991). Plasmids and phase variation in *Xenorhabdus* spp. *Applied and Environmental Microbiology*, **57:**2597–2601.
42. Lee, M. M. and Stock, S. P. (2010a). A multigene approach for assessing evolutionary relationships of *Xenorhabdus* spp. (λ-Proteobacteria), the bacterial symbionts of entomopathogenic *Steinernema* nematodes. *Journal of Invertebrate Pathology*, **104:**67–74.
43. Lee, M. M. and Stock, S. P. (2010b). A multilocus approach to assessing co-evolutionary relationships between *Steinernema* spp. (Nematoda: Steinernematidae) and their bacterial symbionts *Xenorhabdus* spp. (γ-.Proteobacteria: Enterobacteriaceae). *Systematic Parasitology*, **77:**1–12.
44. Lengyel, K., Lang, E., Fodor, A., Szállás, E., et al. (2005). Description of four novel species of *Xenorhabdus*, family *Enterobacteriaceae*: *Xenorhabdus budapestensis* sp. nov., *Xenorhabdus ehlersii* sp. nov., *Xenorhabdus innexi* sp. nov., and *Xenorhabdus szentirmaii* sp. nov. *Systematic and Applied Microbiology*, **28:**115–122.
45. Malan, A. P., Knoetze, R. and Tiedt, L. R. (2015). *Steinernema jeffreyense* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa*. Journal of Helminthology*, **90:**262–278.
46. McMullen, J. G. II, Peterson, B. F., Forst, S., Blair, H. G., et al. (2017). Fitness costs of symbiont switching using entomopathogenic nematodes as a model. *BMC Evolutionary Biology*, **17:**100.
47. Nishimura, Y., Hagiwara, A., Suzaki, T. and Yamanaka, S. (1994). *Xenorhabdus japonicus* sp. nov. associated with the nematode *Steinernema kushidai*. *World Journal of Microbiology and Biotechnology*, **10:**207–210.
48. Nthenga, I., Knoetze, R., Berry, S., Tiedt, L. R., et al. (2014). *Steinernema sacchari* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology*, **16:**475–494.

49. Pinyon, R. A., Hew, F. H. and Thomas, C. J. (2000). *Xenorhabdus bovienii* T228 phase variation and virulence are independent of RecA function. *Microbiology*, **146:**2815–2824.

50. Shapiro, J. A. (1998). Thinking about bacterial populations as multicellular organisms. *Annual Review of Microbiology*, **52:**81–104.

51. Sharma, M. and Anand, S. K. (2002). Swarming: A coordinated bacterial activity. *Current Science*, **83:**707-715.

52. Stock, S. P. and Goodrich-Blair, H. (2012). Nematode parasite, pathogens, and associates of insects and invertebrates of economic importance. In: Lacey, L. A., (ed.). Manual of Techniques in Invertebrate Pathology; pp. 373–426.

53. Sugar, D. R., Murfin, K. E., Chaston, J. M., Ansersen, A. W., et al. (2012). Phenotypic variation and host interactions of *Xenorhabdus bovienii* SS-2004, the entomopathogenic symbiont of *Steinernema jollieti* nematodes, *Environmental Microbiology*, **14:**924–939.

54. Szállás, E., Koch, C., Fodor, A., Burghardt, J., et al. (1997). Phylogenetic evidence for the taxonomic heterogeneity of *Photorhabdus luminescens*. *International Journal of Systematic Bacteriology*, **47:**402–407.

55. Tailliez, P., Laroui, C., Ginibre, N., Paule, A., et al. (2010). Phylogeny of *Photorhabdus* and *Xenorhabdus* based on universally conserved protein-coding sequences and implications for the taxonomy of these two genera. Proposal of new taxa: *X. vietnamensis* sp. nov., *P. luminescens* subsp. *hainanensis* subsp. nov., *P. temperata* subsp. *tasmaniensis* subsp. nov., and the reclassification of *P. luminescens* subsp. *thracensis* as *P. temperata* subsp. *thracensis* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, **60:**1921–1937.

56. Tailliez, P., Pagès, S., Edington, S. and Tymo, L. M. (2012). Description of *Xenorhabdus magdalenensis* sp. nov., the symbiotic bacterium associated with

Steinernema australe. *International Journal of Systematic and Evolutionary Microbiology*, **62:**1761–1765.

57. Tailliez, P., Pagès, S., Ginibre, N. and Boemare, N. (2006). New insight into diversity of the genus *Xenorhabdus*, including the description of ten novel species. *International Journal of Systematic and Evolutionary Microbiology*, **56:**2805–2818.
58. Tailliez, P., Pagès, S., Ginibre, N. and Boemare, N. E. (2006). New insight into the diversity of the genus *Xenorhabdus*, including the description of ten novel species. *International Journal of Systematic and Evolutionary Microbiology*, **56:**2805–2818.
59. Tamura, K., Peterson, D., Peterson, N., Stecher, G., et al. (2011). mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution,* **28:**2731–2739.
60. Thomas, G. M. and Poinar, G. O. Jr. (1979). *Xenorhabdus* gen. nov., a genus of entomopathogenic nematophilic bacteria of the family *Enterobacteriaceae*. *International Journal of Systematic Bacteriology*, **29:**352-360.
61. Thomas, G. M., Poinar, G. O. Jr. (1979). *Xenorhabdus* gen. nov., a genus of entomopathogenic, nematophilic bacteria of the family *Enterobacteriaceae*. *International Journal of Systematic Bacteriology*, **29:**352–360.
62. Tóth, T. and Lakatos, T. (2008). *Photorhabdus temperata* subsp. *cinerea* subsp. nov., isolated from *Heterorhabditis* nematodes. *International Journal of Systematic and Evolutionary Microbiology*, **58:**2579–2581.
63. Volgyi, A., Fodor, A., Szentirmai, A. and Forst, S. (1998). Phase Variation in *Xenorhabdus nematophilus. Applied and Environmental Microbiology*, **64:**1188–1193.
64. Woodring, J. L. and Kaya, H. K. (1988). Steinernematid and Heterorhabditid nematodes: handbook of biology and techniques. *South cooperative Series Bulletin Arkansas Agricultural Experimental Station Fayetteville,* **331:**1–30.

65. Zheng, X. W., Yan, Z., Han, B. Z., Zwietering, M. H., et al., (2012). Complex microbiota of a Chinese “Fen” liquor fermentation starter (*Fen-Daqu*), revealed by culture-dependent and culture-independent methods. *Food Microbiology*, **31:**293–300.

Chapter 5

Genomic sequence and comparative analysis of the entomopathogenic bacterial symbiont, *Xenorhabdus khoisanae* strain MCB

5.1 Abstract

This chapter describes, for the first time, the genome assembly and annotation of *Xenorhabdus khoisanae* strain MCB, a bacterial symbiont associated with a *Steinernema* sp. recovered from the grasslands of South Africa. Additionally, *X. khoisanae* has been shown to associate with several *Steinernema* nematode spp. in South Africa, indicating an underlying genetic plasticity. The draft assembly of *X. khoisanae* strain MCB consisted of 266 contigs with an N50 of 38,931 bp. The genome spanned a length of 4,653,115 bp at 90X sequence coverage with a G+C content of 43.60%. Genome annotation results predicted a total of 3,869 protein-coding genes, 83 ribosomal RNA genes and 65 tRNA genes. Comparative genomic analyses with three previously reported entomopathogenic bacteria, *X. bovienii* SS-2004, *X. nematophila* ATCC 19061 and *Photorhabdus luminescens* DSM 33684 indicated that these four genomes shared 1,786 unique genes that corresponded to less than half of the predicted protein-coding sequences in each representative genome. Further analysis led to the detection of ~47 genomic islands, enriched in genes related to putative functions, phage remnants and virulence, likely to have been acquired through horizonal gene transfer (HGT). Furthermore, *in silico* analysis identified fourteen flagellar

motility genes in *X. khoisanae* strain MCB, with two chemotaxis gene products, cheA and cheY found to be implicated in bacterial swarming behaviour during phase variation. This genome represents a valuable resource for understanding mechanisms involved in bacterial swarming during phase variation in *X. khoisanae*, which is the first report of this behaviour in phase II variants of *Xenorhabdus* bacteria.

5.2 Introduction

Members of the genus *Xenorhabdus* (Thomas and Poinar, 1979) are motile Gram-negative bacteria belonging to the family *Enterobacteriaceae*. Having adopted two distinct lifestyles, *Xenorhabdus* spp. can serve as either parasites to insects or symbionts to entomopathogenic nematodes in the genus *Steinernema* (Forst and Nealson, 1996). The only free-living and non-feeding stage of entomopathogenic nematodes, known as the third-stage infective juvenile harbours the bacterial symbiont in a specialized intestinal vesicle called a receptacle (Bird and Akhurst, 1983). The infective juvenile, in effect, serves as a vector for the transmission of *Xenorhabdus* spp. to various susceptible insects, while also offering a protective environment for the bacterial symbiont.

The infective juvenile nematodes can adopt various foraging strategies when seeking out an insect host to parasitize. Upon encountering such a host, *Steinernema* nematodes gain entry into the haemocoel through natural openings, where their associated bacterial symbiont is expelled. The symbiotic bacteria begin to rapidly multiply, proliferating to high levels in the haemolymph, all whilst releasing an array of secondary metabolites that aid in disabling the insect immune system (Dunphy and Webster, 1991; Park and Forst, 2006) and, eventually, killing the insect through

secreted insecticidal toxins (Akhurst, 1982; Forst et al., 1997). Additionally, the bacterial symbiont also secretes antimicrobial compounds (Göt et al., 1981) that precludes the co-habitation of the insect cadaver by other competing organisms.

Steinernematid nematodes have evolved obligate associations with *Xenorhabdus* bacteria, which are founded on strict partner recognition and species-specific fidelity (McMullen et al., 2017). While *Steinernema* nematodes often form single-species associations with *Xenorhabdus* bacteria, *Xenorhabdus* spp. can foster multiple associations with several *Steinernema* hosts at one time. Such host promiscuity in *Xenorhabdus* spp. can perhaps be best illustrated with the bacterium described in this study, i.e. *Xenorhabdus khoisanae* strain MCB. Additionally, *X. khoisanae* has been found to associate with other *Steinernema* hosts indigenous to South Africa, including *S. khoisanae* (Ferreira et al., 2013), *S. jeffreyense* (Malan et al., 2016) and *S. sacchari* (Nthenga et al., 2014) among others, pointing to an underlying genetic plasticity in *X. khoisanae*.

5.2.1 Aims and objectives

The genome sequence of *Xenorhabdus khoisanae* strain MCB, an insect-pathogenic bacterial symbiont associated with a *Steinernema* sp. recovered from the grasslands of South Africa is reported for the first time. The aim of this chapter was to assemble and annotate the genome of *X. khoisanae* strain MCB so that its predicted functional proteome could be used to perform comparative genomic analyses with other entomopathogenic bacterial species, allowing for the detection of key genes implicated in swarming motility.

5.3 Materials and methods

5.3.1 Genome sequencing of *X. khoisanae* strain MCB

5.3.1.1 Bacterial isolation and molecular identification

The symbiotic bacteria were isolated indirectly from the haemolymph of *Galleria mellonella* larvae within 48 hours of infection by *Steinernema* sp. HBG28 infective juvenile nematodes, according to the methodology used in Chapter 4 (section 4.3.2). Briefly, haemolymph was streaked on nutrient agar supplemented with 0.025% bromothymol blue and 0.004% triphenyltetrazolium chloride (NBTA) and incubated in the dark at 25°C for 48-72 hrs. Blue-green bacterial colonies, indicative of phase I variants, were routinely sub-cultured on NBTA plates prior to DNA isolation. Following isolation, the symbiotic bacterial species was found to have 99% sequence similarity to *Xenorhabdus khoisanae* (Ferreira et al., 2013) based on a partial 16S rRNA gene sequence, which was deposited in GenBank, with the accession number, KT122844.

5.3.1.2 Extraction of bacterial genomic DNA

Total genomic DNA was extracted from solid bacterial colony cultures by using the ZR fungal/bacterial DNA MiniPrep kit (Zymo Research), according to the protocol supplied by the manufacturer. Bacterial DNA isolates were quantified using a Qubit™ 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA).

5.3.1.3 Library preparation and sequencing

The Nextera DNA sample preparation kit (Illumina) was used in the generation of genomic DNA paired-end (PE) libraries. Paired-end (2 × 300 bp) sequencing was performed on an Illumina MiSeq instrument using version 3 MiSeq chemistry at the

Agricultural Research Council (ARC) Biotechnology Platform. A total of 2,632,914 PE reads were generated from the sequencing of genomic DNA.

5.3.1.4 Raw data trimming

Adapter contamination and quality of the PE sequence reads were assessed using CLC Genomics Workbench version 6.5.1 (CLC bio). Nextera transposase adapter sequences and low-quality bases with a Phred quality score (Q) of below Q20 were removed from all the paired-end reads using CLC Genomics Workbench.

5.3.2 Genome assembly and annotation of *X. khoisanae* strain MCB

The *de novo* genome assembly of *X. khoisanae* strain MCB was generated using CLC Genomic Workbench with 1,940,231 trimmed PE reads, a *k*-mer length of 64 and a bubble size of 80. Additionally, the PE reads were mapped to the assembled contigs to establish genome ends relative to possible sequencing artefacts. The genome assembly (including genes, proteins, rRNAs and tRNAs) was annotated through the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/).

5.3.3 Bioinformatics analyses

The predicted protein-coding sequences were assigned Clusters of Orthologous Groups (COGs) classifications to deduce putative gene function of *X. khoisanae* strain MCB using the NCBI COG 2/2/2011 database (prokaryotic proteins) by the WebMGA portal (Wu et al., 2011; http://weizhong-lab.ucsd.edu/webMGA/server/kog/), using an E-value cut-off of $1x10^{-5}$. The MAUVE alignment tool version 2.4.0 (Darling et al., 2004) was used to align the genomes of *X. khoisanae* strain MCB, *X. bovienii* SS 2004

and *X. nematophila* ATCC 19061 by tracing orthologous locally colinear blocks (LCBs) through each *Xenorhabdus* genome after contig ordering. The number of genes shared among the *Xenorhabdus* genomes and *Photorhabdus luminescens* DSM 33684 were also determined using MAUVE. The visualization of genomic islands within the *X. khoisanae* strain MCB genome was carried out through the use of the IslandViewer web application (http://www.pathogenomics.sfu.ca/islandviewer/) (Dhillon et al., 2015). The draft genome of *X. khoisanae* strain MCB was circularly represented by using Basic Local Alignment Search Tool (BLAST) Ring Image Generator (BRIG) (Alikhan et al., 2011). Additionally, the genome was also submitted to the Rapid Annotation Subsystem Technology (RAST) server (http://rast.nmpdr.org) (Overbeek et al., 2014) for baseline gene calling and functional annotation of the *X. khoisanae* strain MCB genome.

5.4 Results

5.4.1 Genome assembly of *X. khoisanae* strain MCB

The trimmed PE reads of *X. khoisanae* strain MCB were assembled into 266 contigs, yielding a draft genome of 4,653,115 bp in length at 90X sequence coverage and with a G+C content of 43.60%. The N50 of the draft genome was 38,931 bp with the largest and shortest contigs being 133,131 and 439 bp in length, respectively. The genome metrics of the draft genome assembly of *X. khoisanae* strain MCB is summarized in Table 5.1.

Table 5.1 Genome assembly metrics of *Xenorhabdus khoisanae* strain MCB

Genomic features	*Xenorhabdus khoisanae* strain MCB
Genome size (bp)	4,653,115
G+C content (%)	43.60
No. of paired-end reads (trimmed)	1,940,231
Total bases in paired-end reads	419,965,962
Number of contigs	266
Maximum contig length (bp)	133,131
Minimum contig length (bp)	439
N50 (bp)	38,931
Mean contig length (bp)	16,147

5.4.2 Genome annotation overview of *X. khoisanae* strain MCB

The resultant assembly for *X. khoisanae* was compared to previously sequenced entomopathogenic bacterial genomes to confirm the validity of the annotation results, as shown in Table 5.2. The draft genome of *X. khoisanae* strain MCB comprised approximately 3,869 protein-coding genes with a total of 65 and 83 tRNA and rRNA genes detected, respectively. The genomic features of three previously sequenced *Xenorhabdus* and *Photorhabdus* bacterial species were compared to that of *X. khoisanae* strain MCB. While, the genome size, G+C content and number of putative protein-coding genes of *X. khoisanae* strain MCB appeared to be in range with those of *X. bovienii* SS 2004 and *X. nematophila* ATCC 19061, the number of tRNAs and rRNAs were significantly fewer in *X. khoisanae*. By comparison, the core genome of *P. luminescens* DSM 33684 – which is the bacterial symbiont associated with *H. bacteriophora* nematodes, contained 4,154 protein-coding genes with 85 and 122 tRNA and rRNA genes, respectively. The annotated draft genome of *X. khoisanae* strain MCB was deposited to GenBank, under the accession no. LFCV01000000.

Table 5.2 Genome features of *Xenorhabdus khoisanae* strain MCB

Genomic features	*X. khoisanae* MCB[1]	*X. bovienii* SS-2004[2]	*X. nematophila* ATCC 19061[3]	*P. luminescens* DSM 3368[4]
Genome size (bp)	4,653,115	4,225,498	4,432,590	5,464,970
G+C content (%)	43.60	44.70	44.2	42.60
Genes (total)	4,155	3,864	4,223	4,631
CDS (coding)	3,869	3,499	3,754	4,154
RNA genes	83	109	106	122
rRNAs	14	22	22	28
rRNAs (5S, 16S, 23S)	6, 3, 5	8, 7, 7	8, 7, 7	8, 10, 10
Complete rRNAs (5S)	6	8	8	8
Partial rRNAs (16S, 23S)	3, 5	7, 7	7, 7	10, 10
tRNA	65	82	79	85
ncRNAs	4	5	5	11
Pseudogenes	257	256	363	353

[1] *Xenorhabdus khoisanae* strain MCB (Naidoo et al., 2015; GenBank accession no. LFCV01000000).

[2] *Xenorhabdus bovienii* SS-2004 (Chaston et a., 2011; GenBank accession no. NC_013892).

[3] *Xenorhabdus nematophila* AN6/1 (Chaston et al., 2011; GenBank accession no. NC_014228).

[4] *Photorhabdus luminescens* strain DSM 3368 (unpublished, GenBank accession no. JXSK00000000).

5.4.3 Functional annotation of protein-coding sequences (CDS)

5.4.3.1 COG functional analysis

Analysis of the functional Clusters of Orthologous Groups (COG) categories and the subsequent COG distribution of protein-coding genes in *X. khoisanae* strain MCB indicated that the most frequent COG categories consisted of genes involved in general function prediction (R), amino acid transport and metabolism (E), and energy production and conversion (C), consisting of 702, 270 and 258 genes, respectively (Figure 5.1). It can be proposed that these COG-assigned genes in *X. khoisanae* strain MCB may be critical for a lifestyle dedicated towards symbiosis with *Steinernema* sp. HBG28 and virulence against various insect hosts. Additionally, among the assigned COGs, nuclear structure (Y) and extracellular structures (W) were the least frequent categories, consisting of 2 and 1 gene(s), respectively.

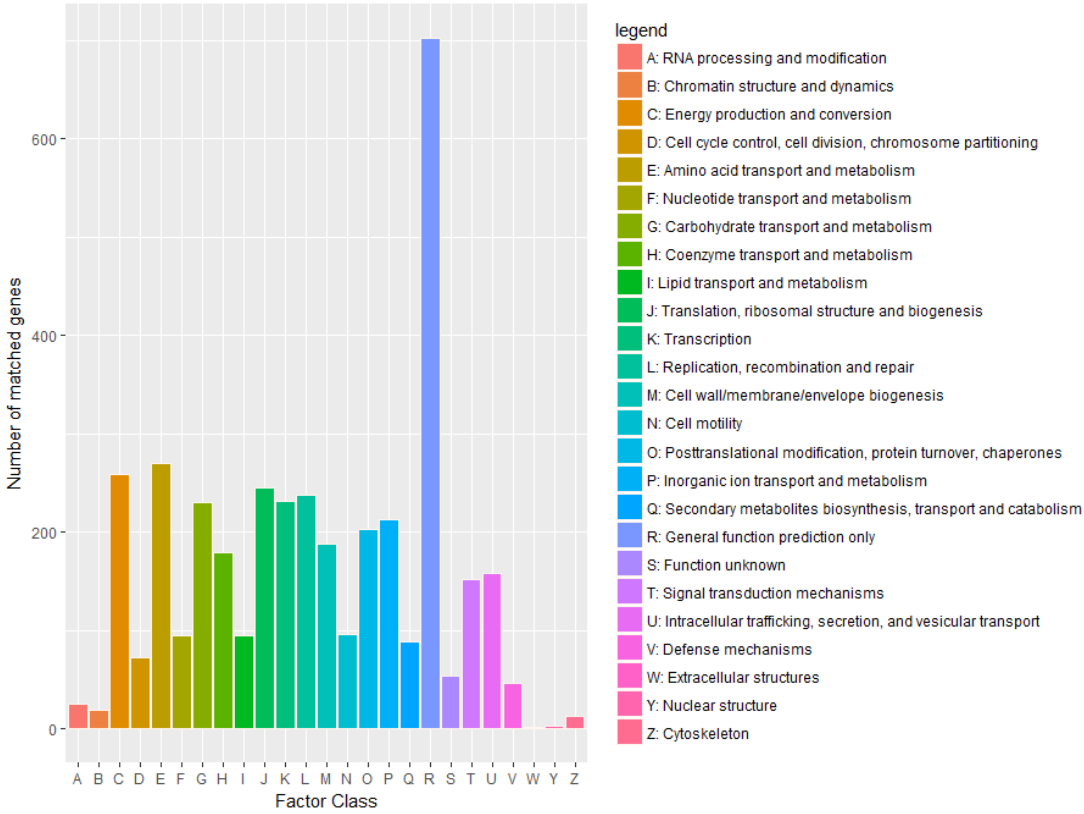


Figure 5.1. COG functional classification (A-Z) of *X. khoisanae* proteome. COG functional categories descriptions are: (A) RNA processing and modification; (B) Chromatin structure and dynamics; (C) Energy production and conversion; (D) Cell cycle control, cell division and chromosome partitioning; (E) Amino acid metabolism and transport; (F) Nucleotide transport and metabolism; (G) Carbohydrate transport and metabolism; (H) Coenzyme transport and metabolism; (I) Lipid transport and metabolism; (J) Translation, ribosomal structure and biogenesis; (K) Transcription; (L) Replication, recombination and repair; (M) Cell wall/membrane/envelope biogenesis; (N) Cell motility; (O) Post-translational modification, protein turnover, chaperones; (P) Inorganic ion transport and metabolism; (Q) Secondary metabolites biosynthesis, transport and catabolism; (R) General function prediction only; (S) Function unknown; (T) Signal transduction mechanisms; (U) Intracellular trafficking, secretion and vesicular transport; (V) Defence mechanisms; (W) Extracellular structures; (Y) Nuclear structure; (Z) Cytoskeleton.

5.4.3.2 Genomic islands

The presence of genomic islands is often observed in several bacterial genomes through the acquisition and exchange of genes through lateral gene transfer (Zhang et al., 2015). Typically, genomic islands vary according to their G+C content in relation to the core genome, encoding for various additional activities of unique functions, particularly those involved in symbiotic interactions and virulence (He et al., 2004). The *X. khoisanae* strain MCB genome was found to possess 47 genomic islands spanning 614 genes, as shown in Figure 5.2. The largest genomic island was found to be 106,380 bp in length, ranging from 4,639,942 to 4,746,322 bp. This genomic region consisted of 80 genes, encoding for several hypothetical proteins, as well as phage-related proteins explicitly involved in virulence (Table 5.3). In particular, these virulent gene products, including the type VI secretion system tube protein Hcp and Hcp1, and the type IV secretion protein Rhs thereof, were likely have been acquired through horizonal gene transfer (HGT), as evident by the presence of phage gene products.

Table 5.3 Secretion and toxin-antitoxin systems in the largest genomic island

Locus	Gene start	Gene end	Product
AB204_RS20355	4643962	4644441	type VI secretion system tube protein Hcp
AB204_RS21835	4745553	4746032	Hcp1 family type VI secretion system effector
AB204_RS17270	4711676	4713782	type IV secretion protein Rhs

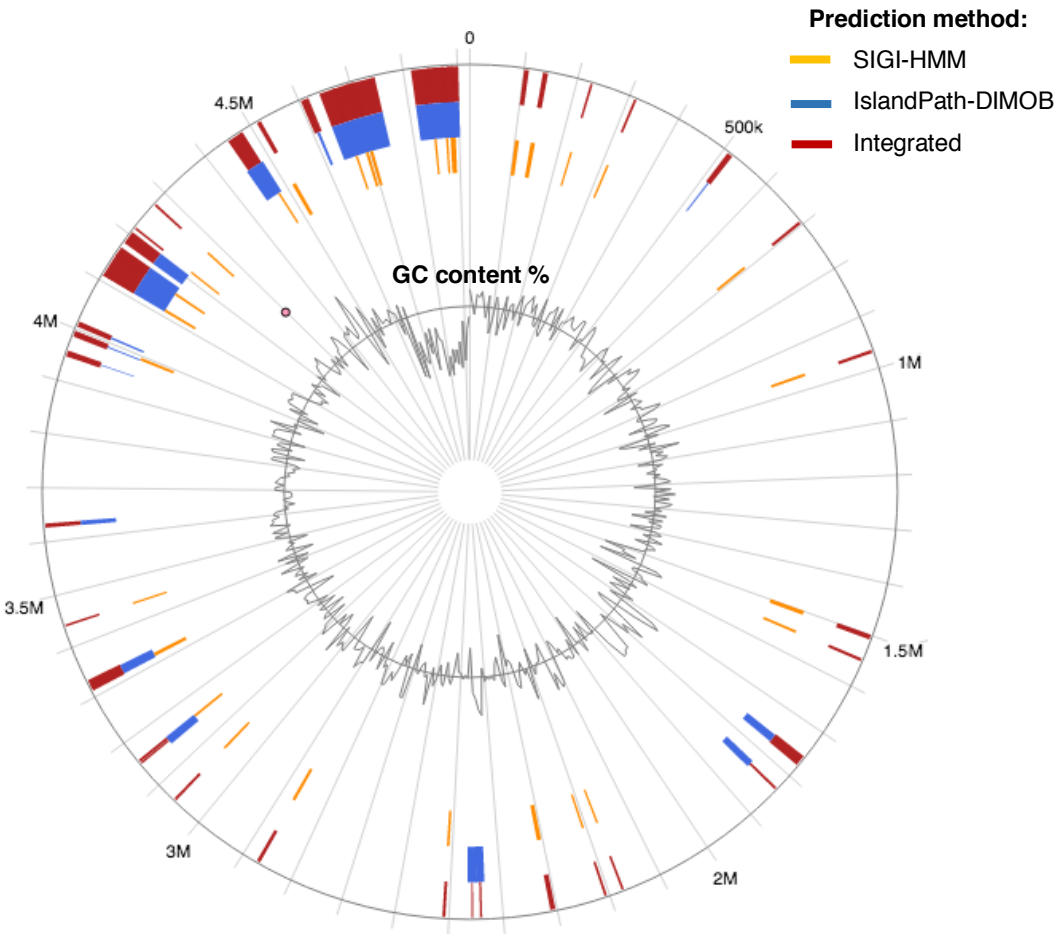


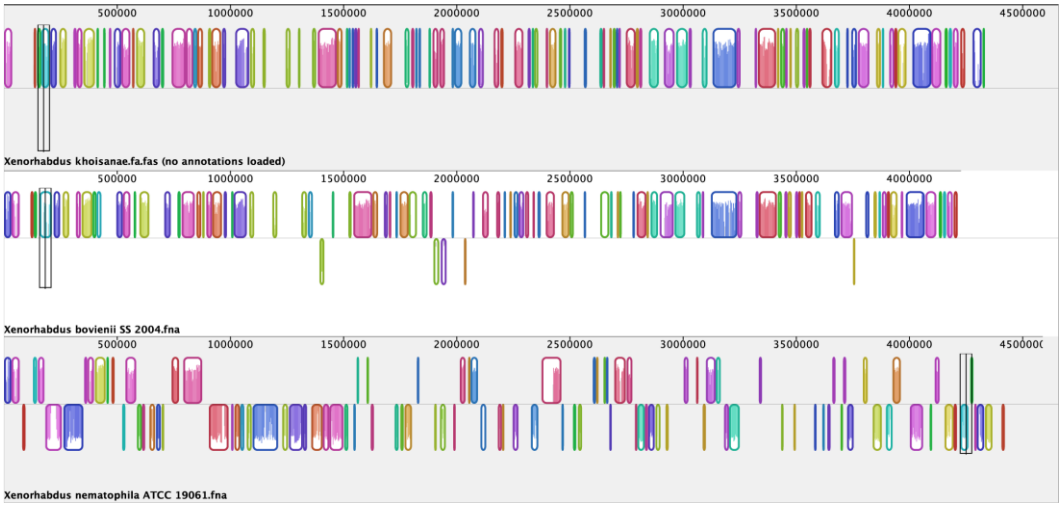
Figure 5.2. Genomic islands in the genome of *X. khoisanae* strain MCB. The genome of *X. khoisanae* strain MCB was found to possess 47 genomic islands spanning 614 genes. The largest genomic island was found to be 106,380 bp in length, ranging from 4,639,942 bp to 4,746,322 and consisting of 80 genes.

5.4.4 Comparative genomic analyses

5.4.4.1 MAUVE alignments of *Xenorhabdus* genomic sequences

Whole genome alignments of *X. khoisanae* strain MCB with *X. bovienii* SS 2004 and *X. nematophila* ATCC 19061 revealed a high level of conservation across these genomes, according to the synteny blocks (in colour) of aligned genomic sequences, which were homologous (Figure 5.3A). Blank regions outside the syntenic blocks represented unaligned sequences that were presumably non-homologous. The three *Xenorhabdus* were highly syntenic, as indicated by the lines found between each genome, tracing each locally colinear block (LCB) through every *Xenorhabdus* genome (Figure 5.3B).

A.



B.

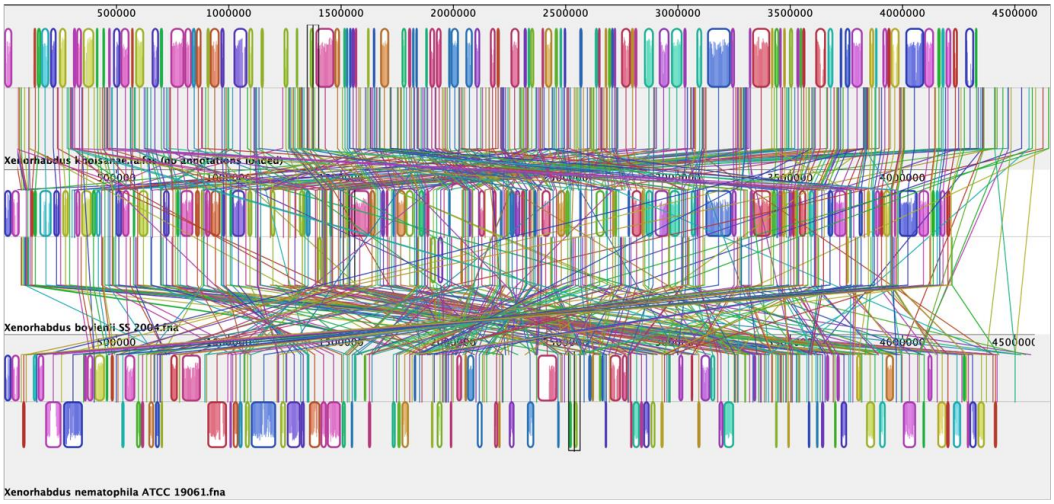


Figure 5.3. Genome pairwise alignment of *Xenorhabdus* genomes. A: Synteny blocks were shown in colour for the genomes of *Xenorhabdus khoisanae*, *X. bovienii* SS 2004 and *X. nematophila* ATCC 19061 and were generated by the MAUVE software (Darling et al., 2004). The coloured blocks represented aligned genomic sequences that were homologous. White regions outside the blocks were sequences not aligned and therefore non-homologous. Blocks below the centre line indicated regions that were aligned in the reverse complement orientation. **B:** The lines found between the genomes traced each orthologous locally collinear alignment block (LCB) through every *Xenorhabdus* genome.

5.4.4.2 Genomic comparison of entomopathogenic bacteria

A comparison of the genomes of *X. khoisanae* strain MCB, *X. bovienii* SS 2004, *X. nematophila* ATCC 19061 and *P. luminescens* DSM 33684 using a MAUVE progressive alignment at 70% coverage and 60% identity revealed a total of 1,786 core genes common to four bacterial genomes (Figure 5.4), indicative of their conservation within the *Enterobacteriaceae* family. Moreover, *X. khoisanae* strain MCB, *X. bovienii*, *X. nematophila* contained 384, 326 and 395 unique genes, respectively, while the *P. luminescens* genome had the highest number of unique genes, i.e. 530. This analysis also showed that *Xenorhabdus* genomes had more genes in common to each other (158 genes for *X. khoisanae* and *X. bovienii*; 193 genes for *X. khoisanae* and *X. nematophila*; 104 genes for *X. bovienii* and *X. nematophila*) than with *Photorhabdus-Xenorhabdus* pairs (54 genes for *P. luminescens* and *X. khoisanae*; 15 genes for *P. luminescens* and *X. bovienii*; 58 genes for *P. luminescens* and *X. nematophila*). An additional 2804 genes were shared between two or more species.

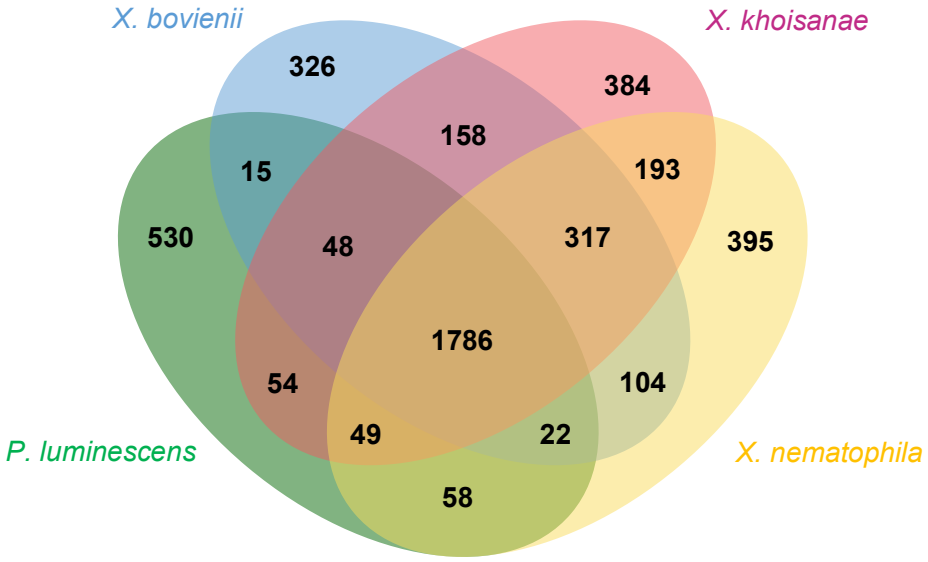


Figure 5.4. Venn diagram indicating the distribution of unique genes among the genomes, *X. khoisanae* strain MCB, *X. bovienii* SS 2004, *X. nematophila* ATCC 19061 and *P. luminescens* DSM 33684.

5.4.4.3 Whole genome comparative analysis

The whole genome comparison of the 12 *Xenorhabdus* genomes revealed that the genomic sequences of *X. khoisanae* strain MCB were significantly similar to these genomes, with a large majority of compared genomic regions indicating identity of over 70% against the *X. khoisanae* strain MCB genome (Figure 5.5). There were, however, several genomic regions that were blank, indicating an identity less than 70%, as shown numbered labels in Figure 5.5. Coincidentally, these blank regions corresponded to some locations containing genomic islands (Figure 5.2). Further investigation into the missing genomic islands would need to be performed.

Additionally, fourteen genes (*motA*, *motB*, *fliA*, *cheA*, *cheY*, *fliR*, *fliN*, *fliM*, *flhB*, *flhA*, *flil*, *rpoN*, *flgh* and FlgD) implicated in flagellar motility were identified through the RAST server and their relative positions within the *X. khoisanae* genome were observed. Perhaps the most notable observation was the absence of certain flagellar motility genes within some *Xenorhabdus* genomes being compared, such as with *X. nematophila* and *X. bovienii*, which showed little to no significant sequence identity at these genomic positions. Moreover, it was found that the entomopathogenic bacterium, *X. griffiniae*, which has also been reported in South African *Steinernema* nematodes (Mothupi et al., 2015; Dreyer et al., 2018), also possessed most of the flagellar motility genes that featured in the *X. khoisanae* genome.

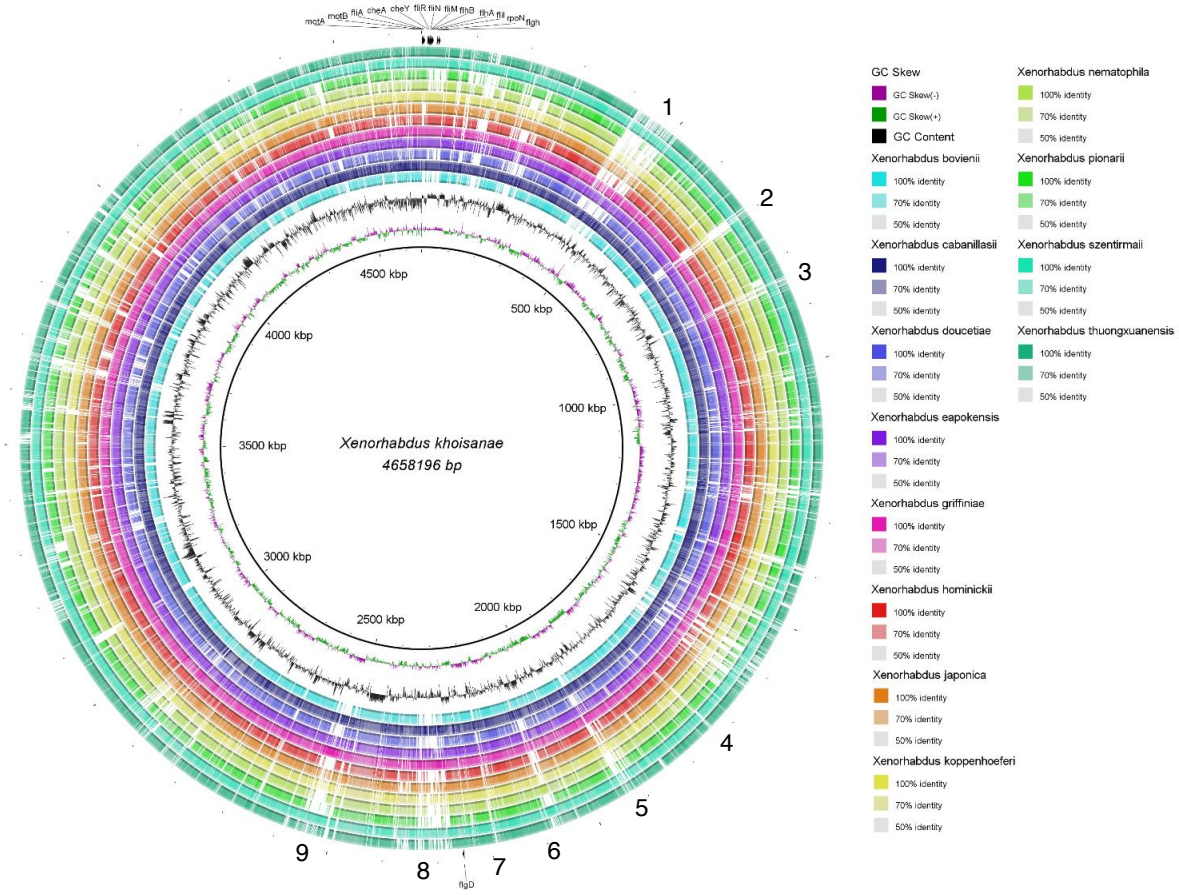


Figure 5.5. Genome comparison of the draft genome of *Xenorhabdus khoisanae* strain MCB against twelve other *Xenorhabdus* genomes. The innermost circle represented the reference genome, *X. khoisanae* strain MCB and the outermost circles represented the percentage similarity of each genome, indicated in different colours. The GC skew and content are also shown on the circular genome plot, along with identity labels of fourteen genes implicated in flagellar motility. Some genomic regions, as indicated by the numbered labels, mirrored the locations of certain genomic islands previously observed.

5.4.4.4 Identification of genes involved in flagellar motility

Using the RAST server, key genes implicated in flagellar motility of *X. khoisanae* strain MCB were considered in order to determine their involvement in bacterial swarming behaviour, especially during phase variation. This led to the identification of 14 motility-related genes in *X. khoisanae* strain MCB, many of which were also found in the genomes of *X. bovienii* SS 2004, *X. nematophila* ATCC 19061 and *P. luminescens* DSM 33684, except for two genes, *cheA* and *cheY* (Table 5.4). The presence of the *cheA* and *cheY* genes in *X. khoisanae* would suggest that they have direct roles in bacterial swarming during phase variation (discussed in Chapter 4). The notable absence of *cheA* and *cheY* in *X. bovienii* and *X. nematophila* could offer a possible explanation as to why swarming behaviour has never been observed in phase II variants of *X. bovienii* and *X. nematophila* (Givaudan et al., 1995).

Table 5.4 The distribution of key genes in entomopathogenic bacterial genomes

Function	Gene product	*X. khoisanae*	*X. bovienii*[1]	*X. nematophila*	*P. luminescens*
Flagella motility	FliA	1	1	1	1
	FliR	1	1	1	1
	FlgD	1	1	1	1
	FliN	1	1	1	1
	FliM	1	1	1	1
	FlhB	1	1	1	1
	FlhA	1	1	1	1
	FliI	1	1	1	1
	flgh	1	1	1	1
	MotA	1	1	1	1
	MotB	1	1	1	1
	CheA	1	-	-	-
	CheY	1	-	-	-

*The gene distribution in X. *bovienii*, *X. nematophila* and *P. luminescens* were obtained using the RAST server (Overbeek et al., 2014).

5.5 Discussion

This chapter introduces the genome assembly and annotation of a bacterial symbiont, *X. khoisanae* strain MCB, that is associated with a *Steinernema* sp. HBG28, recovered from the Suikerbosrand Nature Reserve in the Gauteng province of South Africa.

The aim of this chapter was to assemble and annotate the genome of the bacterial species and, in doing so, determine key genes implicated in bacterial swarming during phase variation. In addition, the genome assembly of *X. khoisanae* strain MCB would facilitate the removal of symbiont contamination during the subsequent sequencing and assembly of the nematode host, which will be discussed in Chapter 6. Briefly, this would entail using a genome mapping approach, whereby the resultant nematode sequence reads would be mapped against a reference bacterial genome to eliminate contaminating bacterial reads prior to the nematode assembly.

5.5.1 Genome structure and annotation of *X. khoisanae* strain MCB

The core genome of *X. khoisanae* strain MCB was 4.6 Mbp in length with a G+C content of 43.60% and comprised approximately 3,869 protein-coding genes. By comparison, the genome of *P. luminescens* consists of a greater number of protein-coding genes (i.e. 4,154) relative to *X. khoisanae* strain MCB. The obligate relationship that exists between *P. luminescens* and its associated nematode host, *H. bacteriophora* has been instrumental in shaping the genomic architecture and content of the nematode genome (Waterfield et al., 2009; Li et al. 2016). As such, *H. bacteriophora* nematodes are highly reliant on their associated bacteria for immune suppression/evasion, host tissue degradation and insect-killing. This is reflected by the large repertoire of genes contained within the genome of *P. luminescens*, facilitating many of these functions.

By comparison, *Steinernema* entomopathogenic nematodes have a more active role in insect pathogenicity other than simply conveying their symbiotic bacteria to various hosts (Binda-Rossetti et al., 2016; Toubarro et al., 2009). It can be surmised that *Xenorhabdus* genomes have a reduced number of protein-coding genes relative to *Photorhabdus* genomes due to the almost equal contribution of the bacterial symbiont with their nematode host in insect virulence.

5.5.2 Comparative genomics analysis of *Xenorhabdus* and *Photorhabdus*

Chaston et al. (2011) has shown that *Xenorhabdus* and *Photorhabdus* bacteria are more phylogenetically related to each other than their nematode partners. This was supported by the results obtained for comparative analysis between *Xenorhabdus* genomes and *P. luminescens*, which revealed a total of 1,786 unique genes shared among all four bacterial genomes, corresponding to less than half of the predicted coding sequences in each representative genome (Figure 5.3). Furthermore *X. khoisanae* strain MCB shared the most genes with *X. nematophila* ATCC 19061 and had the least number of genes in common with *P. luminescens* DSM 33684, i.e. 193 and 54 genes, respectively. The former observation was unexpected, considering that *X. khoisanae* strain MCB is more phylogenetically similar to *X. bovienii* than *X. nematophila*, based on the 16S rRNA gene sequences (shown in Chapter 4).

5.5.3 Genomic islands of *X. khoisanae* strain MCB

In silico analysis identified several genomic islands in *X. khoisanae* strain MCB, rich in phage remnants, evident by the large number of phage proteins contained within them. Of particular interest, was the genomic island that occupied the largest region

of 106,380 bp in length within the *X. khoisanae* genome. This region was found to contain a type VI secretion system tube haemolysin co-regulated protein (Hcp).

A common strategy adopted by bacteria for controlling neighbouring cells is their secretion of toxins or effector proteins through specialized secretion systems (Costa et al., 2015). Some secretion systems can function by secreting toxins to extracellular spaces; while others, such as the type VI secretion system (T6SS), employ a more direct approach in toxin delivery to nearby cells (Costa et al., 2015). The T6SS system mediates interbacterial interactions and is common in several Gram-negative bacteria (Pukatzki et al., 2006; Mougous et al., 2006). The system, itself, comprises various components including the Hcp (Mougous et al., 2006), which facilitates the secretion of effector proteins. These effector proteins have been implicated in bacterial toxicity and are encoded next to a gene whose product confers immunity against the secreted toxin, thereby avoiding self-intoxication (Russell et al., 2011).

Clearly, the T6SS apparatus serves an important purpose, especially when the niche requirements between two competing bacterial species coincide (Russell et al., 2014). The implications of this mechanism can perhaps be best reflected in *Xenorhabdus* bacteria where some species can be in fierce competition to occupy the same insect resource, particularly if two different *Steinernema* nematodes co-infect the same insect host. This could possibly give cause to the prevalence of *X. khoisanae* as symbionts to several *Steinernema* spp. in South Africa, including the *Steinernema* nematode isolated in the current study. It is possible that the mechanisms governing the T6SS is more optimized in *X. khoisanae*, thereby making this species more virulent. In a study by Sicard et. al. (2006), it was shown that the co-infection of the same insect host, *G. mellonella* by *S. carpocapsae* and *S. scapterisci* nematodes and their respective symbionts, *X. nematophila* and *X. innexi*, resulted in bacterial dominance in favour of

only one species of bacterial symbiont within the insect haemolymph, i.e. *X. nematophila*. In such a situation involving co-infection of the same insect host by two local *Steinernema* spp., it can be proposed that the growth of *X. khoisanae* may preclude that of another symbiont within the insect haemolymph, due to the highly virulent mechanisms involved in its T6SS. While *Steinernema-Xenorhabdus* interactions are regulated by strict partner recognition and species-specific fidelity, *Steinernema* associations with non-cognate symbionts have been demonstrated by McMullen et al. (2017), and may not be entirely uncommon in nature, despite the fitness costs and reduced virulence associated with mismatched pairings. As a consequence, it is likely that the highly dominant *X. khoisanae*, having outcompeted the less virulent *Xenorhabdus* symbiont, is then taken up by a non-cognate *Steinernema* host – thus perpetuating the prevalence of *X. khoisanae* bacteria among local *Steinernema* nematodes.

Another protein that was found in the largest genomic island of *X. khoisanae* was that of the type IV secretion Rhs protein, which is encoded by *rhs* genes, derived from the name “rearrangement hot spot" (Lin et al., 1984). The gene products of *rhs* appears to be widely featured in many Gram-negative bacteria, including members of the *Enterobacteriaceae* family (Kung et al., 2012), such as *Photorhabdus* and *Xenorhabdus*. In spite of their ubiquity among several bacterial species, their definitive function remains unknown (Hill et al., 1994). It has been proposed that some *rhs* gene products closely resembles bacteriocins or capsule transport proteins (Sisto et al., 2010; Roh et al. 2008; McNulty et al., 2006). Additionally, Rhs proteins demonstrate sequence similarity to toxins released by *Photorhabdus* bacteria, thus reflecting their possible role in insect infection. As such, the presence of the Rhs gene product in *X. khoisanae* strain MCB is understandable given that bacterial members of

Xenorhabdus and *Photorhabdus* are phylogenetically related, much unlike their nematode hosts (Chaston et al., 2011).

5.5.4 The role of *cheA* and *cheY* genes in bacterial swarming

The close relationships observed among the *Xenorhabdus* genomes provided an ideal opportunity for detecting key genes essential for promoting bacterial swarming. Bioinformatic analysis of the draft genome of *X. khoisanae* MCB revealed fourteen genes that encode for proteins involved in flagellar motility. Perhaps the most striking observation in this analysis was the absence of two flagellar genes, *cheA* and *chey* genes in both the *X. bovienii* and *X. nematophila* genomes. It has been previously noted that phase II variants of *X. bovienii* and *X. nematophila* exhibit a lack of motility, reflected by their inability to swarm during phenotypic switching (Givaudan et al., 1995). By comparison, *X. khoisanae* strain MCB demonstrates swarming behaviour in both phase I and II variants (see Chapter 4), as well as containing the *cheA* and *chey* genes.

Motility allows bacteria to migrate from one location to another, serving as a necessary step toward biofilm development. Most bacteria feature the ability to swarm by virtue of rotating flagella. In *Xenorhabdus* spp. for example, swarming is thought to have implications in mutualism and parasitism by allowing bacteria to rapidly colonize its associated nematode host, as well as its insect host (Givaudan et al., 1995). The CheA and CheY proteins play important roles in swarming as they directly influence the direction at which flagella rotate (Bourret et al., 1989). Briefly, the recognition of a signal by CheA is initiated through ligand-binding brought about by conformational changes of the cytoplasmic receptor. This results in the donation of a phosphate by the CheA protein to the response regulator, CheY. The

phosphorylated CheY then interacts with the FliM target protein (Lee et al., 2001), altering flagellar rotation from the default counter-clockwise state to a clockwise rotation. This rotation switch triggers the cell to tumble forward, thereby changing the direction of motility (Barak and Eisenbach, 1992, Sagi et al., 2003). Essentially, external signals mediate changes in swarming patterns of bacteria through the alteration of phosphorylated CheY levels.

The ability to swarm, especially during phase variation may contribute to the success of *X. khoisanae* as a prevalent bacteria symbiont among local *Steinernema* nematodes. Moreover, being associated to a non-cognate nematode host may create conditions that give rise to phenotypic switching of *X. khoisanae.* As phase II variants of *X. khoisanae* can undergo swarming, the bacterial cells are likely to adhere within the specialized intestinal vesicle of the nematode and are not expelled prematurely. Over time, this mismatched association may have evolved, such that *X. khoisanae* has inherently become the bacterial symbionts of several *Steinernema* nematode hosts indigenous to South Africa. Another interesting observation was that the entomopathogenic bacterium, *X. griffiniae*, which has also been found to associate with local *Steinernema* nematode species (Mothupi et al., 2015; Dreyer et al., 2018), also possesses all flagellar motility genes featured in the *X. khoisanae* genome with the exception of the *cheA* and *cheY* genes. This finding only serves to highlight the importance of bacterial swarming in the colonization of *Steinernema* nematode species, especially during phenotypic bacterial switching.

5.6 Conclusion

Overall, the genomic composition of insect-pathogenic bacteria can be viewed as highly dynamic, prone to changes that occur due to genome reduction (Moran, 2002), gene duplication, divergence (Ohno, 1970) and horizontal gene transfer, all of which have been largely influenced by fierce competition within a soil environment (Chaston et al., 2011). While *Xenorhabdus* species may not have entirely lost its genetic machinery as a former free-living organism, in light of the fact that it can be isolated and cultured under standard laboratory conditions, it has over time acquired genes that facilitate insect pathogenicity and the exclusion of other bacterial competitors at an intra- and inter-species level. This has allowed Xenorhabdus spp. to occupy a unique niche.

This study presented a genomic analysis of an entomopathogenic bacterial species, *X. khoisanae* that is widely associated to several *Steinernema* nematode spp. in South Africa. The identification of two chemotaxis-related genes is thought to be crucial for bacterial swarming in *X. khoisanae*, especially during phase variation. This trait, along with other key features noted in this study, could suggest that *X. khoisanae* evolved under high competition pressure to be able to associate with many *Steinernema* nematode species. This genome represents a valuable resource for the engineering of novel antibacterial and insecticidal products that can be exploited for medical and agricultural application.

5.7 References

1. Abebe-Akele, F., Tisa, L. S., Cooper, V. S., Hatcher, P. J., et al. (2015). Genome sequence and comparative analysis of a putative entomopathogenic *Serratia* isolated from *Caenorhabditis briggsae*. *BMC Genomics*, **16:**531.
2. Akhurst, R. J. (1982). Antibiotic activity of *Xenorhabdus* spp., bacteria symbiotically associated with insect pathogenic nematodes of the families Heterorhabditidae and Steinernematidae. *Journal of General Microbiology*, **128:**3061–3065.
3. Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. and Beatson S.A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, **12**:402.
4. Barak, R. and Eisenbach, M. (1992). Correlation between phosphorylation of the chemotaxis protein CheY and its activity at the flagellar motor. *Biochemistry*, **31:**1821–1826.
5. Binda-Rossetti, S., Mastore, M., Protasoni, M. and Brivio, M. F. (2016). Effects of an entomopathogen nematode on the immune response of the insect pest red palm weevil: Focus on the host antimicrobial response. *Journal of Invertebrate Pathology*, **133:**110–119.
6. Bird, A. F. and Akhurst, R. J. (1983). The nature of the intestinal vesicle in nematodes of the family Steinernematidae. *International Journal for Parasitology*, **13:**599–606.
7. Chaston, J. M., Suen, G., Tucker, S. L., Andersen, A. W., et al. (2011). The entomopathogenic bacterial endosymbionts *Xenorhabdus* and *Photorhabdus*: convergent lifestyles from divergent genomes. *PloS One*, **6:**e27909.

8. Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., et al. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews. Microbiology*, **13:**343–359.
9. Darling, A. C. E., Mau, B., Blattner, F. R. and Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, **14:**1394–403.
10. Dhillon, B. K., Laird, M. R., Shay, J. A., Winsor, G. L., et al. (2015). IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Research*, **43:**W104–W108.
11. Dreyer, J., Malan, A. and Dicks, L. M. T. (2017). Three Novel *Xenorhabdus-Steinernema* Associations and Evidence of Strains of Switching Between Different Clades. *Current Microbiology*, **74:**938–942.
12. Dreyer, J., Malan, A. P. and Dicks, L. M. T. (2018). First report of a symbiotic relationship between *Xenorhabdus griffiniae* and an unknown *Steinernema* from South Africa. *Archives of Microbiology*, **200:**349–353.
13. Dunphy, G. B. and Webster, J. M. (1991). Antihemocytic surface components of *Xenorhabdus nematophilus* var. *dutki* and their modification by serum of nonimmune larvae of *Galleria mellonella*. *Journal of Invertebrate Pathology*, **58:**40–51.
14. Ehlers, R. (2001). Mass production of entomopathogenic nematodes for plant protection. *Applied Microbiology and Biotechnology*, **56:**623–633.
15. Ferreira, T., van Reenen, C. A., Endo, A., Spröer, C., et al. (2013). Description of *Xenorhabdus khoisanae* sp. nov., the symbiont of the entomopathogenic nematode *Steinernema khoisanae*. *International Journal of Systematic and Evolutionary Microbiology*, **63:**3220–3224.

16. Forst S. and Nealson K. (1996). Molecular biology of the symbiotic-pathogenic bacteria *Xenorhabdus* spp. and *Photorhabdus* spp.. *Microbiology Reviews*, **60:**21–43.
17. Forst, S., Dowds, B., Boemare, N. and Stackebrandt E. (1997). *Xenorhabdus* and *Photorhabdus* spp.: bugs that kill bugs. *Annual Review of Microbiology*, **51:**47–72.
18. Givaudan, A., Baghdiguian, S., Lanois, A. and Boemare, N. (1995). Swarming and Swimming Changes Concomitant with Phase Variation in *Xenorhabdus nematophilus*. *Applied and Enviromental Microbiology*, **61:**1408–1413.
19. Götz, P., Boman, A. and Boman, H. G. (1981). Interactions between insect immunity and an insect-pathogenic nematode with symbiotic bacteria. *Proceedings of the Royal Society B: Biological Sciences*, **212:**333–350.
20. He, J., Baldini, R. L., Deziel, E., Saucier, M., et al. (2004). The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proceedings of the National Academy of Sciences of the United States of America*, **101:**2530–2535.
21. Hill, C. W., Sandt, C. H. and Vlazny, D. A. (1994). Rhs elements of *Escherichia coli*: A family of genetic composites each encoding a large mosaic protein. *Molecular Microbiology*, **12:**865–871.
22. Kung, V. L., Khare, S., Stehlik, C., Bacon, E. M., et al. (2012). An rhs gene of *Pseudomonas aeruginosa* encodes a virulence protein that activates the inflammasome. *Proceedings of the National Academy of Science*, **109:**1275–1280.
23. Lee, S. Y., Cho, H. S., Pelton, J. G., Yan, D., et al. (2001). Crystal structure of an activated response regulator bound to its target. *Nature Structural Biology*, **8:**52–56.
24. Lin, R. J., Capage, M. and Hill, C. W. (1984). A repetitive DNA sequence, rhs, responsible for duplications within the *Escherichia coli* K-12 chromosome. *Journal of Molecular Biology*, **177:**1–18.

25. Lu, D., Baiocchi, T. and Dillman., A. R. (2017). Genomics of Entomopathogenic Nematodes and Implications for Pest Control. *Trends in Parasitology*, **32:**588–598.
26. Malan, A. P., Knoetze, R. and Tiedt, L. R. (2015). *Steinernema jeffreyense* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa*. Journal of Helminthology*, **90:**262–278.
27. Mariconda, S., Wang, Q. and Harshey, M. R. (2006). A mechanical role for the chemotaxis system in swarming motility. *Molecular Microbiology*, **60:**1590–1602.
28. McMullen, J. G. II, Peterson, B. F., Forst, S., Blair, H. G., et al. (2017). Fitness costs of symbiont switching using entomopathogenic nematodes as a model. *BMC Evolutionary Biology*, **17:**100.
29. McNulty, C., Thompson, J., Barrett, B., Lord, L., et al. (2006). The cell surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. *Molecular Oral Microbiology*, **59:**907–922.
30. Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell*, **108:**583–586.
31. Mougous, J. D., Cuff, M. E., Raunser, S., Shen, A., et al. (2006). A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science*, **312:**1526–1530.
32. Mothupi, B., Featherston, J. and Gray, V. (2015). Draft whole-genome sequence and annotation of *Xenorhabdus griffiniae* strain BMMCB associated with the South African Entomopathogenic nematode *Steinernema khoisanae* strain BMMCB. *Genome Announcements*, **3:**e00785-15.
33. Naidoo, S., Featherston. J. and Gray, V. M. (2015). Draft Whole-Genome Sequence and Annotation of the Entomopathogenic Bacterium *Xenorhabdus khoisanae* Strain MCB. *Genome Announcements*, **3:**e01279–15.

34. Nthenga, I., Knoetze, R., Berry, S., Tiedt, L. R. et al. (2014). *Steinernema sacchari* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology*, **16:**475–494.

35. Ohno, S. (1970). Evolution by Gene Duplication. New York: Springer. 160 p.

36. Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, **42:**206–214.

37. Park, D. and Forst, S. (2006). Co-regulation of motility, exoenzymes, and antibiotic production by the EnvZ-OmpR-FlhDC-FliA pathway in *Xenorhabdus nematophila*. *Molecular Microbiology*, **6:**1397–1412.

38. Pukatzki, S., Ma, A. T., Sturtevant, D., Krastins, B., et al. (2006). Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proceedings of the National Academy of Sciences of the United States of America*, **103:**1528–1533.

39. Roh, E., Heu, S. and Moon, E. (2008). Genus-specific distribution and pathovar-specific variation of the glycinecin R gene homologs in *Xanthomonas* genomes. *The Journal of Microbiology*, **46:**681–686.

40. Russell, A. B., Hood, R. D., Bui, N. K., LeRoux, M., et al. (2011). Type VI secretion delivers bacteriolytic effectors to target cells. *Nature*, **475:**343–347.

41. Russell, A. B., Peterson, S. B., and Mougous, J. D. (2014). Type VI secretion system effectors: poisons with a purpose. *Nature Reviews. Microbiology*, **12:**137–48.

42. Sagi, Y., Khan, S. and Eisenbach, M. (2003). Binding of the chemotaxis response regulator CheY to the isolated, intact switch complex of the bacterial flagellar motor: lack of cooperativity. *The Journal of Biological Chemistry*, **278:**25867–25871.

43. Sisto, A., Cipriani, M. G., Lonigro, S. L., Valerio, F., et al. (2010). An Rhs-like genetic element is involved in bacteriocin production by *Pseudomonas savastanoi* pv. *savastanoi*. *Antonie Van Leeuwenhoek*, **98:**505–517.
44. Thomas, G. M. and Poinar, G. O. Jr. (1979). *Xenorhabdus* gen. nov., a genus of entomopathogenic, nematophilic bacteria of the family *Enterobacteriaceae*. *International Journal of Systematic Bacteriology*, **29:**352–360.
45. Toubarro, D., Lucena-Robles, M., Nascimento, G., Costa, G., et al. (2009). An apoptosis-inducing serine protease secreted by the entomopathogenic nematode *Steinernema carpocapsae*. *International Journal for Parasitology*, **39:**1319–1330.
46. Waterfield, N. R., Ciche, T. and Clarke, D. (2009). *Photorhabdus* and a host of hosts. *Annual Review of Microbiology*, **63:**557–574.
47. Waterfield. N.R., Bowen, D. J., Fetherston, J. D., Perry, R. D., et al. (2001). The tc genes of *Photorhabdus*: A growing family. *Trends in Microbiology*, **9:**185–191.
48. Wu, S., Zhu, Z., Fu, L., Niu, B., et al. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, **12:**444.
49. Zhang, X., Peng, C., Zhang, G. and Gao, F. (2015). Comparative analysis of essential genes in prokaryotic genomic islands. *Scientific Reports*, **5:**12561.

Chapter 6

Genome assembly of a novel *Steinernema* entomopathogenic nematode species

6.1 Abstract

This chapter describes the generation of a draft genome for the entomopathogenic nematode, *Steinernema* sp. HBG28, isolated from the Gauteng Province in South Africa from Illumina sequence data. The introduction provides an overview of the key steps involved with the sequencing and assembly of nematode genomes, as well as addresses challenges such as heterozygosity and contamination. This is followed by a detailed description of the methodology used for establishing an inbred nematode line for genome sequencing, assessing the sequencing data, generating preliminary assemblies with three assembly algorithms, eliminating contaminant sequence reads of a *Xenorhabdus* bacterial symbiont and refining the final assembly with the most consistent assembler evaluated. The resultant genome assembly comprised of 97,040,761 bp in 15,252 scaffolds, with an N50 value of 15,839 bp and a GC-content of 49,49%. The longest scaffold was 647,965 bp in length. Core Eukaryotic Gene Mapping (CEGMA) analysis against a set of 248 highly conserved eukaryotic genes (CEGs) identified 92.34% for complete core genes and 93.95% completeness for partial genes, indicating that the draft genome of *Steinernema* sp. HBG28 was complete.

6.2 Introduction

Nematoda is considered one of the most ubiquitous and biologically-diverse of all animal phyla. The phylum, itself, contains approximately 25,000 recognized nematode species, with as many as 100 million species unaccounted for (Blaxter, 2003; Lambshead, 1993). Aside from their sheer numbers and species richness, nematodes can thrive in nearly every terrestrial and aquatic ecosystem on earth, indicating an underlying genetic plasticity.

By far, one of the best-studied model organisms in biology is the free-living nematode, *Caenorhabditis elegans.* Coincidentally, *C. elegans* was also the first metazoan to be fully sequenced over two decades ago with Sanger sequencing technology (*C. elegans* Sequencing Consortium, 1998). While sequencing the 100 Mb genome proved to be both a costly and labour-intensive endeavour in its entirety, the end-result represented an important milestone in genetic research, and it remains by far the best-assembled and annotated metazoan genome to date. The efforts made on the *C. elegans* genome has greatly influenced subsequent work on several other nematode species.

The study of nematodes has often relied on various approaches for the purpose of identification and taxonomy, such as classical morphology that requires a thorough knowledge of key diagnostic features, or molecular-based detection methods that make use of ribosomal, mitochondrial and satellite DNA (De Ley and Blaxter, 2002; De Ley and Blaxter, 2004; Holterman et al., 2006). While the prospect of using routine whole genome sequencing (WGS) in the field of nematology has always held significant appeal to researchers, it was never exclusively considered, due to Sanger sequencing being a prohibitively expensive and time-consuming technology.

Nevertheless, advancements in DNA sequencing over the last two decades has given rise to high-throughput next-generation sequencing (NGS) technologies, which has revolutionized the field of genomics. Capable of producing large-scale sequence data in under a short timeframe and at a significantly reduced cost, NGS technologies has become widely used in a variety of research projects and is accessible to even small research groups. Unlike in previous years, it is now possible to obtain entire genome sequences for various nematode species of interest, including novel ones; and as sequencing technologies continue to improve, and costs decrease, it is likely that genome sequencing may be performed as a routine part in nematode characterization.

6.2.1 Genome sequencing

With a diversity and abundance of nematode species available, it is practical that nematodes of medical and economic importance be selected for sequencing. Once a species of interest has been identified for sequencing, the basic requirement in any genome project is the generation of a draft genome assembly. The core steps for genomic sequencing includes: i) isolation of genomic DNA, ii) construction of genomic libraries, iii) sequencing and iv) assembly of the sequence reads into contigs and possibly, scaffolds and gene annotation (Dillman et al., 2012).

As one of the earliest sequencing strategies (first generation), Sanger sequencing is still regarded as the gold standard for DNA sequencing. However, the main downsides to this method remain its high cost and low speed in relation to NGS platforms. With the withdrawal of Roche 454 pyrosequencing, Illumina is perhaps the most dominant NGS technology on the market and the preferred platform for most *de novo* sequencing projects worldwide. It offers a cheaper per base cost and has a high

throughput (Utturkar et al., 2015), albeit at the expense of accuracy during base-calling.

Illumina technology, as with several other NGS strategies, relies on a similar approach of whole-genome shotgun sequencing, but without the need for cloning fragmented DNA into plasmid vectors (Anderson, 1981). Briefly, target DNA is isolated and randomly sheared into short fragments, to which specific sequencing adapters are ligated either at one or both ends. The fragments are then hybridized to fixed primers found on the surface of the flow cell, allowing for synthesis of the complement strands through PCR. Any fragments of DNA that are not bound to the flow cell are eventually washed away, leaving behind the attached fragments. Through bridge amplification, small clusters of double-stranded DNA are established from the attached DNA fragments, which are identical in sequence. Each DNA cluster emits a signal when fluorescently-labelled nucleotides are incorporated to the flow cell, also known as the sequencing by synthesis (SBS) step. Specific cameras are then used to detect fluorescence signals and the sequence is automatically determined.

Illumina platforms generate short length reads up to 100 base-pair (bp), with numbers ranging in the millions. Despite the high throughput, the reduced read length of Illumina sequencing can present a challenge when repeat regions are considered, especially if the length of the repeat exceeds that of the read length. Illumina compensates for such a scenario by supporting paired-end sequencing, whereby opposite ends of the same DNA molecule are sequenced to the full read length (Levy and Myers, 2016). Additionally, the most common source of errors with Illumina technology are miscalled bases. Often, the raw sequence data of short-read technologies is prone to an error rate of 1-2%, depending on the choice of NGS platforms used (Kircher et al., 2009; Rothberg et al., 2011). The effects of sequence

errors can be circumvented by sequencing at a higher depth of coverage. Sequencing coverage is a large oversampling of the genome being sequence, taking into account the length of the sequencing reads, the total number of reads and the genome size. With current NGS platforms, a coverage of 60X to 100X coverage is quietly easily attainable, while a 100X coverage is typically required for the generation of a satisfactory genome assembly (Dillman, 2012).

An important consideration during the generations of genomic libraries is whether the purpose is for resequencing (mapping reads to an available reference genome) or *de novo* sequencing (assembling sequence reads to establish an entirely new reference sequence). If it is for the latter, then short-insert, paired-end (PE) libraries can be used, supplemented with long-insert mate-pair (MP) libraries. While PE sequence data is generated from genomic fragments and is mostly without erroneous mergers (chimaeras), the generation of MP libraries includes DNA circularization that can introduce chimeras or false mate pairs (Illumina, 2009). MP data can cater for higher coverage of the reference genome (Illumina, 2012), allowing for the generation of scaffolds from contigs. Although MP libraries take more time to prepare and are costlier to sequence, the information provided is essential for establishing long-range contiguity and resolving spanning repeats within an assembly.

6.2.2 Bioinformatic approaches to a *de novo* genome assembly

The generation of raw sequence data produced by Illumina platforms is merely one aspect of a sequencing project, with the computational (*in silico*) processing and analysis of the sequences representing a significant part. Shifting from millions of short sequence reads to an assembled genome is not an easy task and depends on

nucleotide coverage, read length and read quality of the sequence data. An assembly algorithm must be able to contend with a high volume of raw data using minimal computing resources while handling read errors incurred during sequencing, as well as be able to resolve biological anomalies within a genome such as repeats and heterozygous alleles.

Assembly programs work by stitching together a genome through the alignment of the short-sequence reads to create overlapping regions, which are merged into longer, continuous fragments or contigs, and this process can be performed over several iterations. The most implemented method for assembling short-reads relies on de Brujin graphs (DBG), which rely on the concept of overlapping strings of a defined length *k* (*k*-mers) from initial reads to construct a larger sequence (Pevzner et al., 2001; Zerbino and Birney, 2008). Assembling short sequence reads into a genome can be difficult in instances where repeat sequences are encountered. Incorporating repeat sequences can break the contiguity of an assembly, especially if the length of a repeat exceeds that of the sequence read. Unresolvable repeats can affect the final output, even resulting in a fragmented assembly.

In the absence of a high-quality reference genome, the quality of a *de novo* genome assembly is assessed through several meaningful and useful genomic metrics, which affords a certain degree of confidence and provides a way in which later analyses can be interpreted. Perhaps the most commonly used metric in assessing any genome assembly is the N50 value, which is indicated by the size of the contig that represents half of that genome assembly. The N50 statistic is important as it gives an indication of the potential contiguity of an assembly, however, it does not assert to the underlying accuracy and quality of the genome (Earl et al., 2011). Among the other metrics often reported for a genome assembly include the total genome length, the

number of contigs produced, the GC content. Additionally, the biological accuracy of assembled genomes can also be inferred by determining the coding potential of the genome of interest through the use of the Core Eukaryotic Genes Mapping Approach (CEGMA) tool (Parra et al., 2007), which identifies the most biologically meaningful sequences within the genome based on alignments created by Hidden Markov Models (HMMs). CEGMA utilizes a set of 248 highly conserved eukaryotic genes (CEGs) that are found spanning six evolutionary distinct genomes (*C. elegans*, *Saccharomyces cereviseae*, *Drosophila melanogaster*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana* and *Homo sapiens*). Typically, a high percentage of 'complete' genes corresponds to a better genome assembly. Irrespective of the organism being sequenced or the sequencing technology employed, the main goal of any sequencing project is to produce the most contiguous genome assembly that can then be implemented for downstream genome annotation.

6.2.3 Entomopathogenic nematodes and genomics

The outcome of *de novo* genome assemblies can be essential in the advancement of genomic research, particularly when it involves the use of novel species. High-quality draft genomes can sometimes contain at least 95% of all essential genes, and therefore several questions pertaining to gene evolution, novel gene function and even phylogenetics can be addressed with a genome assembly (Kumar et al., 2012). There is little doubt that the application of genomic sequencing can be extremely beneficial, especially in the field of nematology. Genomics can allow researchers the opportunity of learning more about their nematode of interest, as well as uncovering information on genome structure and nematode evolution.

Since the release of the *C. elegans* genome, the number of draft genome sequences belonging to several free-living and parasitic nematodes has risen significantly in

recent years. To date, there has been a total of 114 nematode genomes reported on the online Wormbase database (https://wormbase.org/; Howe et al., 2016; Lee et al., 2017). A specialized group of nematodes known as entomopathogenic nematodes that show tremendous potential as biological control agents merely account for a small fraction of these sequenced genomes. There have been only six species of entomopathogenic nematodes sequenced thus far (Dillman et al., 2015; Bai et al., 2013).

There still remains a great deal to learn about entomopathogenic nematodes and by adopting a genomic approach, researchers can gain a better understanding of the biology, behaviour, ecology and host specialization of these nematodes. The availability of entomopathogenic nematode genomes can allow for the discovery of useful genes and gene products, particularly of agricultural importance.

6.2.4 Aims and objectives

The main aim of this chapter was to sequence and assemble the genome of a novel entomopathogenic nematode species, *Steinernema* sp. HBG28 indigenous to the grasslands of South Africa using Illumina sequencing technology. To this end, nematodes were inbred prior to sequencing and the resulting sequence data was filtered for possible contaminants to acquire a draft assembly that could later be used for functional genome annotation. Furthermore, a series of assembly algorithms were evaluated and based on the performance of each assembler, the most applicable assembly pipeline was determined for this study. The genome assembly of *Steinernema* sp. HBG28 shall be the focus of this chapter, while the transcriptome assembly and downstream functional annotation will be discussed in Chapter 7.

6.3 Materials and methods

The developed workflow used to create the draft genome of *Steinernema* sp. HBG28 is shown in the Figure 6.1. While the methodology may not have been exact nor resulted in the most optimal results, it did render traceable data for downstream application.

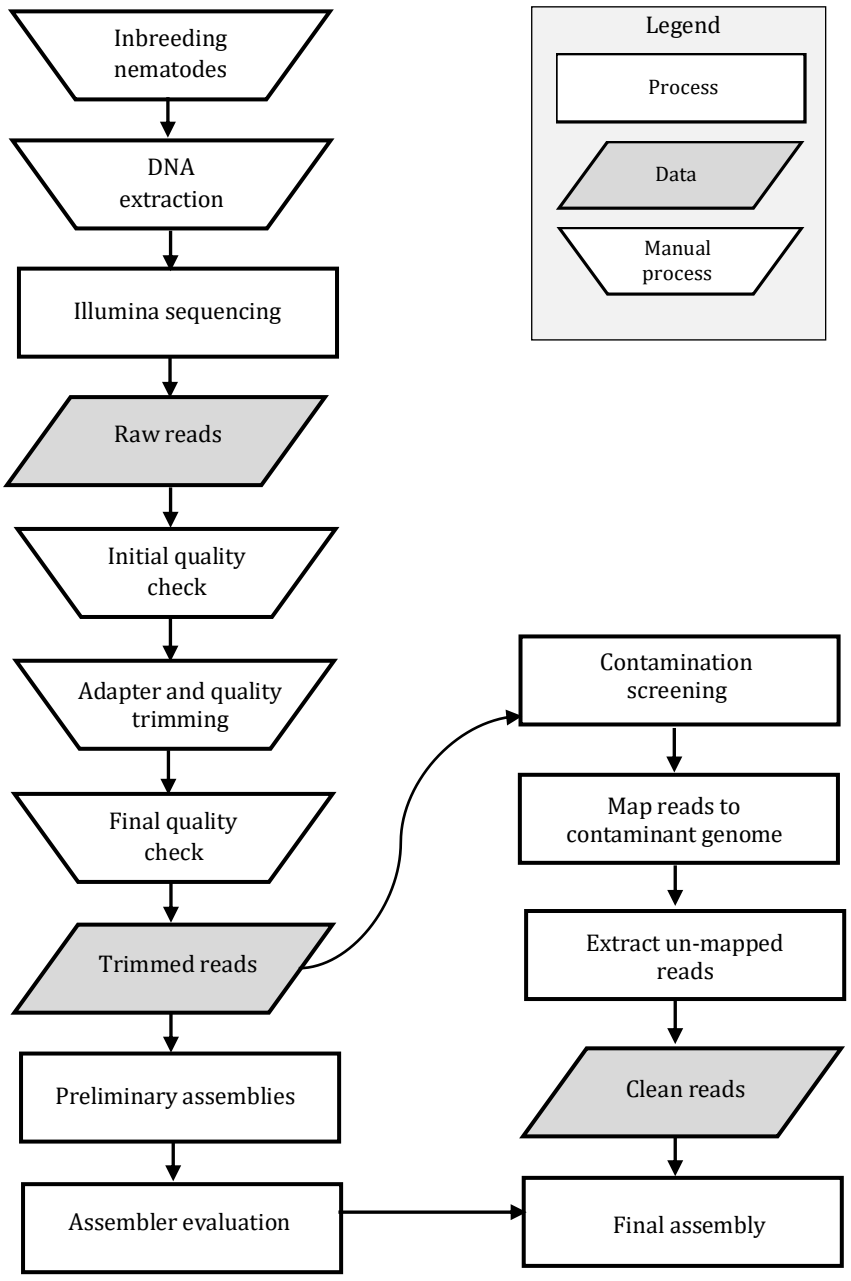


Figure 6.1. Workflow for the genome assembly of *Steinernema* sp. HBG28, adapted from Kumar and Blaxter, 2011.

6.3.1 Genome sequencing of *Steinernema* sp. HBG28

6.3.1.1 Establishing an inbred nematode line

Steinernema sp. HBG28 infective juvenile nematodes were isolated from a locality in Gauteng, South Africa: Suikerbosrand Nature Reserve, (26.4838° S, 28.2293° E). The nematodes were produced *in vivo* in *Galleria mellonella* (L.) larvae on modified White traps as described in detail in Chapter 2 (section 2.3.2.4). Inbreeding was initiated through sibling mating to generate a homozygous population of nematodes for Illumina sequencing. For the starter culture of nematodes, surface-sterilized infective juveniles (500 IJs/ml) were inoculated on a fresh bacterial lawn of *Xenorhabdus khoisanae* and incubated at 25°C for 3-5 days. The starter culture of nematodes, containing a mixed population that was assumed to be in Hardy-Weinberg Equilibrium, was subsampled by the removal of a single pregnant female (i.e. undergoing endotokia matricida) and transferred onto a fresh bacterial lawn that was incubated at 25°C for 3-5 days, representing the first round of inbreeding, as shown in Figure 6.2A. From the nematode subculture of round one, a subsample consisting of a single gravid female was selected and plated onto a fresh bacterial lawn, representing the second round of inbreeding. This process of subsampling an inbred nematode female that was either gravid (i.e. fertilized egg-containing) or pregnant from the previous inbred culture to initiate sibling mating was repeated over seven rounds. This resulted in ~90% homozygosity (Wright, 1921b).

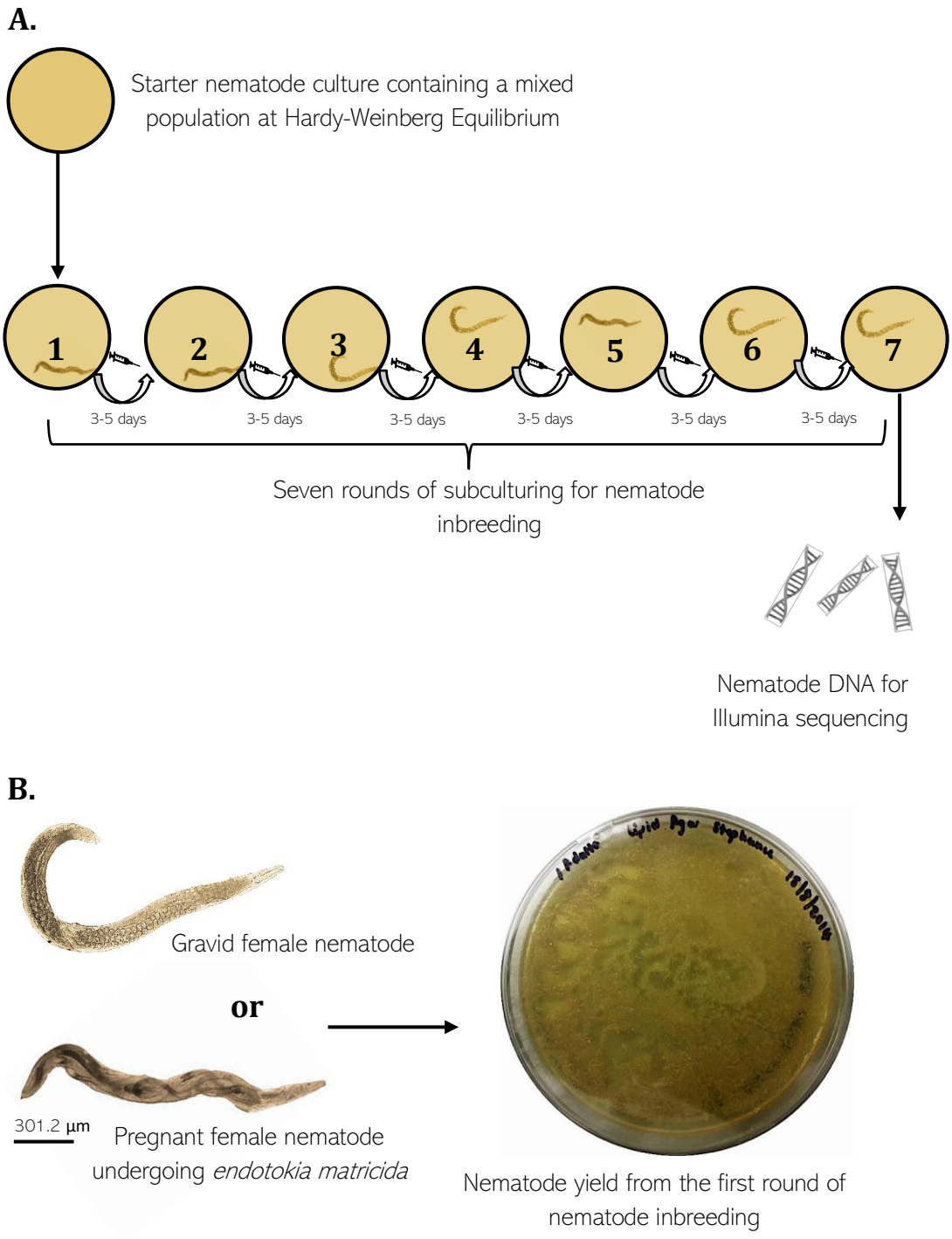


Figure 6.2. Diagram illustrating the culturing of *Steinernema* sp. HBG28. (A): The nematode line was established from a starter culture. For subculturing, a single female entomopathogenic nematode from the previous culture was used. Nematode DNA was isolated for Illumina sequencing after seven rounds. **(B):** either a single gravid (i.e. egg-containing) or pregnant (i.e. undergoing *endotokia matricida*) female nematode was subcultured on lipid agar inoculated with *Xenorhabdus khoisanae.* Initial nematode cultures gave rise to a high yield, as was seen in the first round of nematode inbreeding.

6.3.1.2 Isolation of nematode genomic DNA

The inbred nematodes were pooled and surfaced-sterilized with a 0.1% sodium hypochlorite solution. Total genomic DNA was isolated from all developmental larval stages (egg, L1, L2, L3, L4 and adult) of *Steinernema* sp. HBG28 using the Pure gene® DNA Purification Kit (Gentra Systems 2003) and the Genomic DNA Purification kit (Thermo Scientific), according to the protocols supplied by the manufacturers. All DNA samples were quantified using a Qubit™ 3.0 Fluorometer (Life Technologies) and Nanodrop (ThermoFisher), as a total of 5 µg of DNA required for genome sequencing.

6.3.1.3 Library preparation and genome sequencing

The paired-end genomic library was prepared using the Nextera DNA Sample Prep kit (Illumina) with 2 × 125 bp Illumina SBS version 4 chemistry and a targeted insert size of 180- to 220-bp. Additionally, the mate-pair library was prepared using the Nextera Mate Pair Library Prep kit (Illumina) with 2 × 100 bp chemistry and an insert size of 3,5 kb. All sequencing was performed on the Illumina HiSeq 2500 platform at Agriculture Research Council (ARC) Biotechnology Platform facility in Onderstepoort, Pretoria. Raw sequence data was generated as FASTQ output files.

6.3.1.4 Raw data trimming

Adapter contamination and quality of the paired-end sequence data were assessed using FastQC version 0.11.5 (Andrews, 2010) and using CLC Genomics Workbench version 11 (CLC bio). Nextera transposase adapter sequences were removed from all the reads; and quality trimming was also performed, where bases with a Phred quality score (Q) of less than Q20 were removed from all sequences using Trimmomatic version 0.32 (Bolger et al., 2014), The trimmed reads were re-examined with FastQC to verify successful trimming.

6.3.1.5 Preliminary *de novo* genome assemblies

Preliminary assemblies were generated from the trimmed PE data by evaluating three *de novo* genome assemblers: CLC Genomic Workbench version 11 (CLC bio), Velvet version 1.2.10 (Zerbino and Birney 2008) and Allpaths-LG version 52485 (Gnerre et al. 2011). For CLC Genomic Workbench and Velvet, a variety of assembly parameters such as word size, bubble size and *k*-mer length were tested and those that provided the best assembly statistics for each assembler were selected. For Allpaths-LG, the assembly parameters were set to default, apart from the following options specified: `K=95` and the `HAPLOIDIFY=True`. Contig assemblies were performed using CLC (with a word size of 20 and a bubble size of 50) and Velvet (with a *k*-mer depth of 29) using PE read data only. A scaffold-level assembly was performed using Allpaths-LG with both PE and MP data.

By assessing each assembly, the assembler responsible for producing the most optimal preliminary assembly would be used to assemble the final nematode genome with a clean read set that had been filtered out for contaminant read sequences. The assembly algorithms were assessed according the quality of the assemblies produced, taking into account memory requirements and processing time for each assembly algorithm, and this was achieved in two ways. First, assembly contiguity of the preliminary assemblies was evaluated using various assembly-derived statistics obtained from QUAST version 4.6.1 (Gurevich et al., 2013). Second, the biological accuracy of the preliminary assemblies was determined by identifying which of the competing assemblies captured the most biologically meaningful sequences using CEGMA (Parra et al., 2007).

6.3.1.6 Screening for bacterial contamination

The raw sequence data of the nematode was screened for the presence of bacterial contaminants using a metagenomic program that assigned taxonomic information called Kraken version 1 (Wood and Saizberg, 2014), which identified bacterial sequences. Based on the Kraken output report, the percentage of bacterial contamination present within the sequence reads was determined. A clean dataset of the sequence reads was created for the final genome assembly of *Steinernema* sp. HBG28 by removing contaminated reads belonging to the endosymbiont bacteria, *Xenorhabdus khoisanae* using CLC Genomics Workbench, as it was the least memory-intensive program to use. This was carried out by mapping the PE reads of the nematode to the assembled bacterial genome generated in Chapter 5. Sequence reads that mapped to the bacterial genome were discarded; while those that were not mapped were extracted and used for the final genome assembly of *Steinernema* sp. HBG28.

6.3.1.7 Final assembly

Using Allpaths-LG, the final assembly of the nematode genome was produced from the clean PE dataset that had undergone read-filtering for bacterial sequences. The assembly parameters were set to default, apart from the following options specified: `K=95` and the `HAPLOIDIFY=True`. MP data was used to connect contigs in the scaffolding step. The final genome assembly was also gap-filled. Assembly metrics was generated with QUAST, while the biological accuracy was determined with CEGMA. The average coverage (N) was calculated by the total number of sequenced bases (T) in relation to the genome size of the assembly (G), shown by the formula below:

$$N = \frac{T}{G}$$

where:

N = Average genome coverage

T = total number of bases

G = Estimated genome size

6.4 Results

6.4.1 Nematode inbreeding

The nematodes of *Steinernema* sp. HBG28 were inbred to avoid the recognized issues often associated with heterozygosity for an accurate genome assembly (Blaxter and Koutsovoulos, 2015). The starting culture for inbreeding consisted of a mixed population that was assumed to be in Hardy-Weinberg Equilibrium. With every subsequent subculture of inbreeding, a single inbred female was used from the previous subculture and thereby the degree of homozygosity increased inexorably away from the original Hardy-Weinberg Equilibrium.

On a qualitative level, the visual phenotypic appearance of the inbred progeny changed, and signs of inbreeding depression were visible. Quantitatively, the yield of inbred nematode diminished with each succession, which limited the amount of total genomic DNA extracted for sequencing.

6.4.2 Genome assembly of *Steinernema* sp. HBG28

6.4.2.1 Determination of read quality using FastQC

Whole genome sequencing was performed on an Illumina HiSeq 2500 instrument with a targeted insert size of 180- to -220 bp, generating a total of 51 million read

pairs (~102 million reads in total) in 10.9-gigabases (Gb) of raw sequence data from a single DNA sample of *Steinernema* sp. HBG28. As there had not been any *Steinernema* species sequences made available at the time of sequencing, the nematode, *C. elegans* was used as a benchmark for the expected genome size. With a raw sequence yield of 10,9 Gb produced during sequencing, this corresponded to an average genome coverage of about 109X when assuming a relative genome size of 100 Mb for *Steinernema* sp. HBG28.

The PE sequences were assessed prior to the *de novo* assembly of the nematode genome to identify the potential presence of adapter sequences and to determine the overall quality of the sequence reads using FastQC. While the sequence quality was satisfactory, the reads still contained Nextera Transposase sequences, as indicated in Figure 6.3. Thus, adaptor sequences and low-quality bases were removed using Trimmomatic, resulting in a total yield of 47,884,316 PE reads and 10.3 Gb of raw sequence data. The trimmed reads were re-assessed with FastQC to ensure successful adapter trimming, as shown in Figure 6.4.

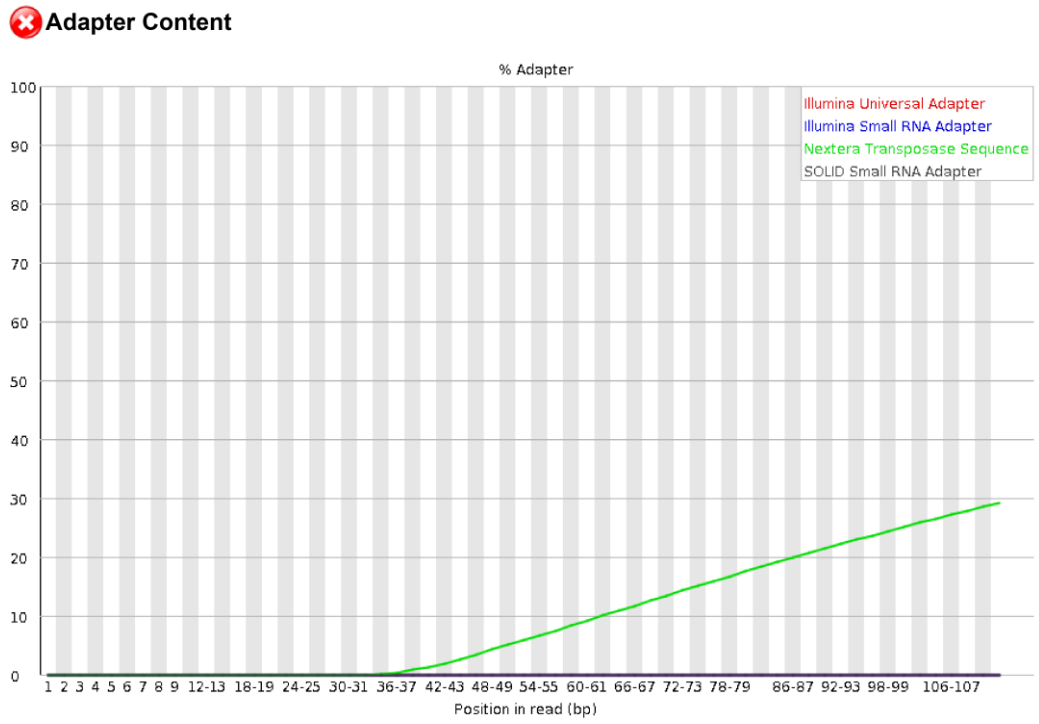


Figure 6.3: FastQC graphical representation of adapter content before trimming with Trimmomatic, flagged as "fail".

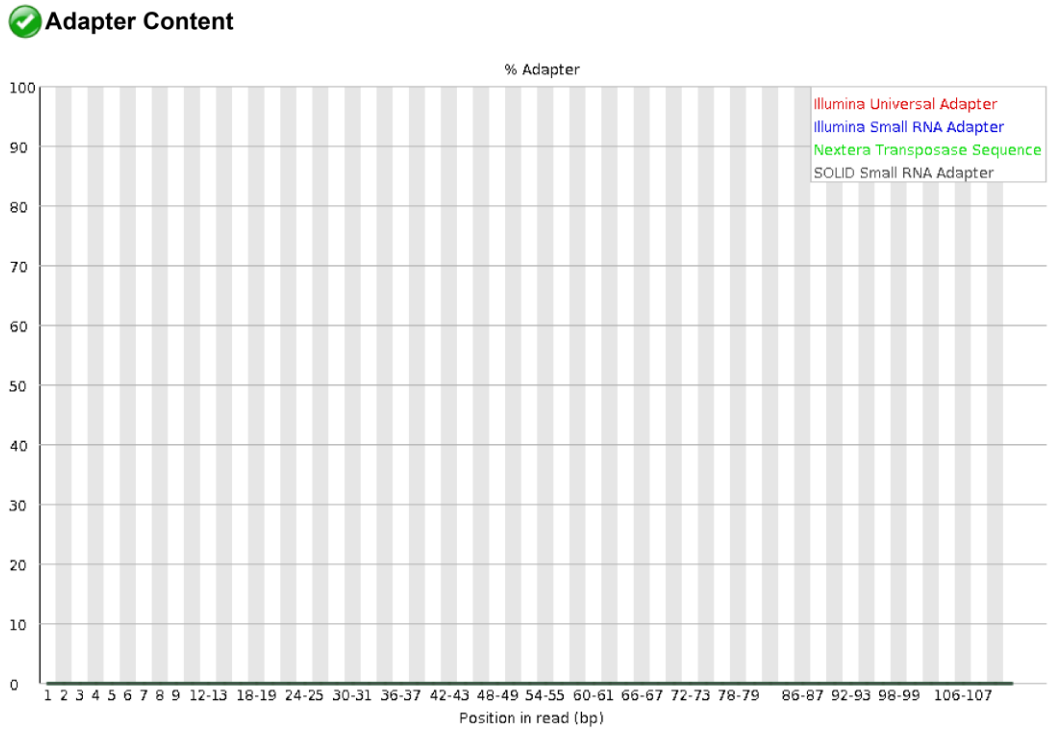


Figure 6.4. FastQC graphical representation of adapter content after trimming with Trimmomatic, flagged as "passed".

6.4.2.2 Analysis of trimmed data using CLC Genomics Workbench

A pre-sequence analysis was performed on the trimmed sequence data using CLC Genomics Workbench, which included length distribution, GC-content, ambiguous base-content and quality distribution (see the Appendix section for full report). The quality statistics were also determined for each sequence read with the inclusion of Phred quality scores, which was indicative of read quality. Phred scores can range from 1 to 40 and logarithmically relates to the probability of an incorrect base being called (Ewing and Green, 1998). Approximately 98.79% of the PE sequences had a Phred score of above Q20 and in the range of Q20 to Q37 (Figure 6.5), while a small percentage of reads fell below a Phred score of Q20, indicating poor quality. Low quality bases were later removed from the PE reads. Additionally, the content of ambiguous bases was determined for the trimmed sequence reads. According to the results, a low percentage (0.56%) of ambiguous bases was observed (Figure 6.6).

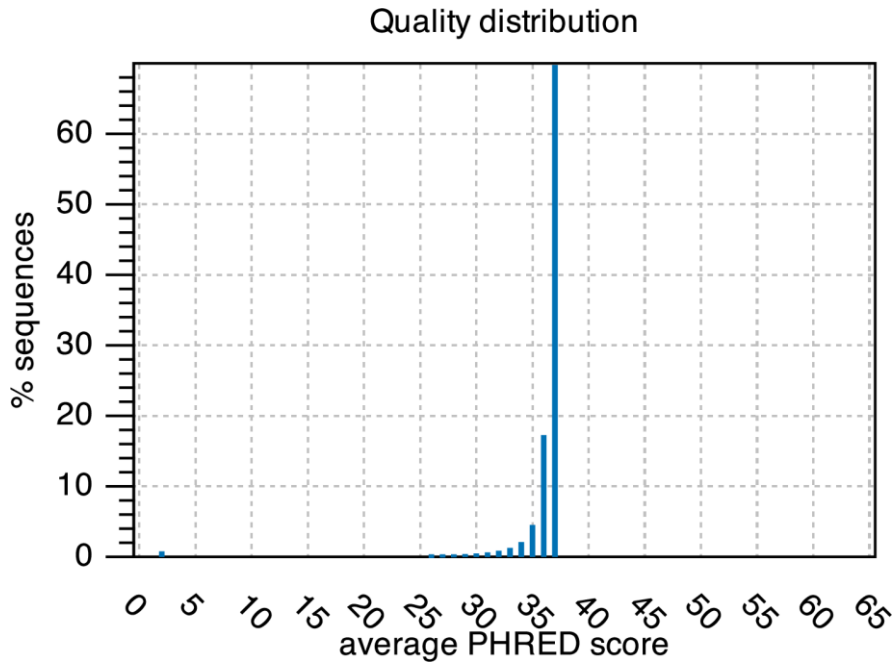


Figure 6.5. Phred quality score (Q) distribution, generated with CLC Genomic Workbench. The distribution of average Phred quality score (x-axis) is plotted against the percentage of sequences (y-axis) for Illumina paired-end reads.

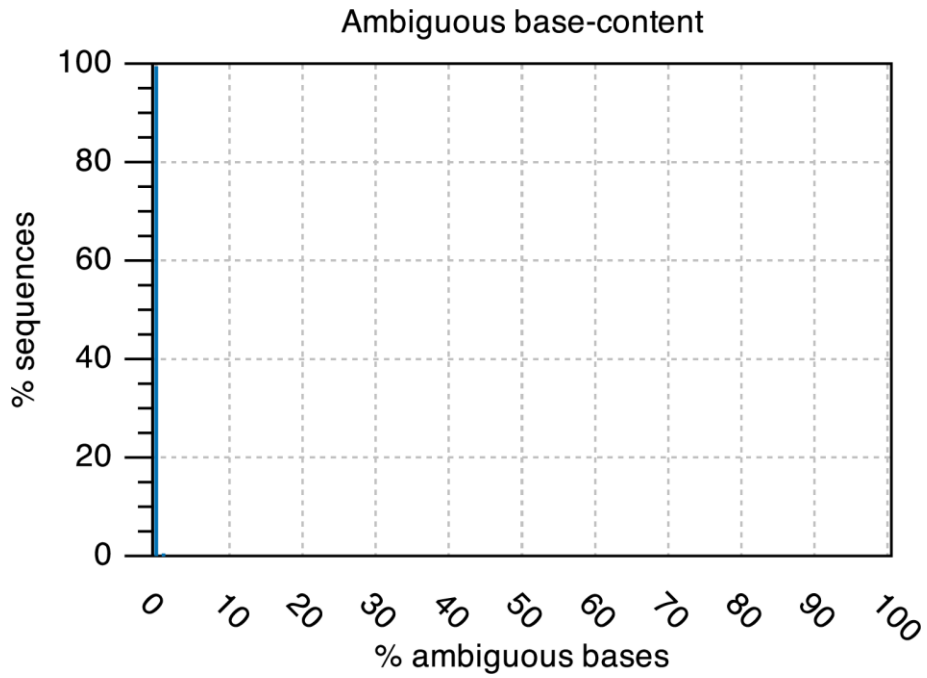


Figure 6.6. Ambiguous base-content generated with CLC Genomic Workbench. The percentage of ambiguous bases (x-axis) is plotted against the percentage of sequences (y-axis) for Illumina paired-end reads.

6.4.2.3 Evaluation of preliminary assemblies

For *Steinernema* sp. HBG28, the genome metrics of each preliminary assembly were determined and used in selecting the most appropriate assembler. The data in Table 6.1 highlights key metrics obtained for each assembly and describes the total number of contigs generated by each assembler, the size of the assembled genomes (in Mb), the N50 length, the GC content and the CEGMA completeness percentage. The analysis with CEGMA revealed that the preliminary assemblies were largely complete: 94.76% for the CLC assembly, 90.73% for the Velvet assembly and 92.34% for the Allpaths-LG assembly.

Table 6.1 Comparison of preliminary assemblies for *Steinernema* sp. HBG28

	CLC	VelvetK29	Allpaths-LG
Total number of contigs	74,257	31,936	**18,219**
Largest contig (bp)	**1,133,054**	199,049	437,220
N50 length (bp)	8,728	7,387	**14,715**
GC content (%)	**49.9%**	50.63%	51.40%
Total length (bp)	148,500,299	112,005,499	105,225,773
CEGMA completeness	**94.76%**	90.73%	92.34%

Note: Values in bold font represent the best metric in that row.

The N50 measurements were relatively low for the preliminary assemblies of CLC and Velvet, at 8,728 and 7,387 bp, respectively. The Allpaths-LG assembly produced the largest N50 size of 14,715 bp, which was at least two-fold larger than the N50 values of the CLC and Velvet assemblies. In addition, the Allpaths-LG preliminary assembly also had the least number of contigs and a total assembly length that was relatively consistent with the expected size for the nematode genome (i.e. ~100 Mb of C*. elegans*).

In favour over the Velvet and Allpaths-LG assemblies, the CLC assembly had the higher CEGMA completeness percentage of 94.76%, despite having the highest number of assembled contigs. Together with the low N50 measurement observed, this indicated that the CLC assembly had the poorest contiguity overall. The CLC preliminary assembly also spanned 148 Mb, greatly exceeding the estimated genome size of *C. elegans* (~100 Mb). The genome length can usually provide some indication of redundancy within an assembly, especially when it is not within the expected size range, as was observed with the CLC assembly.

Lastly, the Velvet algorithm produced the least optimal assembly and was found to have the lowest N50 value and the lowest CEGMA completeness relative to the other assemblies. While underperforming on most genome metrics, Velvet did have a

total length that was relatively better than that of CLC. Based on these analyses, it was determined that the Allpaths-LG assembler performed the best overall, producing the most contiguous and biologically-accurate preliminary assembly.

6.4.2.4 Removal of symbiont contamination

As *Steinernema* sp. HBG28 nematodes were grown on bacterial lawns of their associated bacterial symbiont, there was the potential that contigs generated from sequencing the nematode genome would contain contaminating sequence reads from *X. khoisanae*. To avoid the possibility of bacterial reads being incorporated into the nematode genome assembly, which, in turn, could lead to inaccurate downstream analyses, a clean dataset consisting of solely of nematode sequence reads was required in order to produce a genome assembly for *Steinernema* sp. HBG28. The raw sequence PE data of the nematode was screened using Kraken. Kraken searches revealed that 8.33% of all reads contained evidence of bacterial contamination. The next step was to separate the bacterial sequence reads from those belonging to the nematode and, in the process, obtain a clean dataset from which a new genome assembly could be generated. Since the genome of *X. khoisanae* had already been sequenced and assembled during the course of this research project (See Chapter 5), the sequence data for the nematode was aligned to the bacterial genome, using a feature in CLC Genomics Workbench that allowed for reads to be mapped against a reference genome. In doing so, the reads that mapped to the genome were filtered to yield 6,925,433 reads. The remaining reads were collected, resulting in 88,843,199 reads, which were used in the reassembly of the nematode genome.

6.4.2.5 Final draft assembly of *Steinernema* sp. HBG28

Unlike the preliminary assembly step where the purpose was to evaluate different genome assemblers and determine the best possible assembly, the aim of the re-assembly was to produce the final assembly of the nematode genome at scaffold-level, without the presence of contaminant reads belonging to the symbiont, *X. khoisanae*.

After producing the preliminary assemblies, Allpaths-LG was selected as the most consistent assembler and, as such, was used to re-assemble the final genome assembly of *Steinernema* sp. HBG28 (Table. 6.2). The final draft genome assembly of *Steinernema* sp. HBG28 spanned a total of 97,040,761 bp in 15,252 scaffolds at 112X sequence coverage, with a N50 scaffold size of 15,839 bp and a GC-content of 49.49%. Of the 248 highly-conserved CEGs, 233 (93.95%) was detected as complete copies and 241 (97.18%) was found in complete or partial forms in the assembly of *Steinernema* sp. HBG28. The removal of contaminant bacterial reads from the raw genome sequence data resulted in a marginally improved genome assembly, as indicated by the genome and biological accuracy metrics. Additionally, the genome assembly metrics for six previously sequenced entomopathogenic nematode species (i.e. *H. bacteriophora*, S. *carpocapsae*, *S. scapterisci*, *S. monticolum*, *S. feltiae* and *S. glaseri*) (Bai et al., 2013; Rougon-Cardoso et al., 2016; Dillman et al., 2015) were compared against the current assembly, as reported in Table 6.2.

The cumulative scaffold length of the final assembly with Allpaths-LG was generated with QUAST. The curve of the Allpaths-LG assembly is gentle, implying shorter sequences overall, as indicated in Figure 6.7. The end point of the curve reflects the total assembly span and the number of scaffolds in the final assembly.

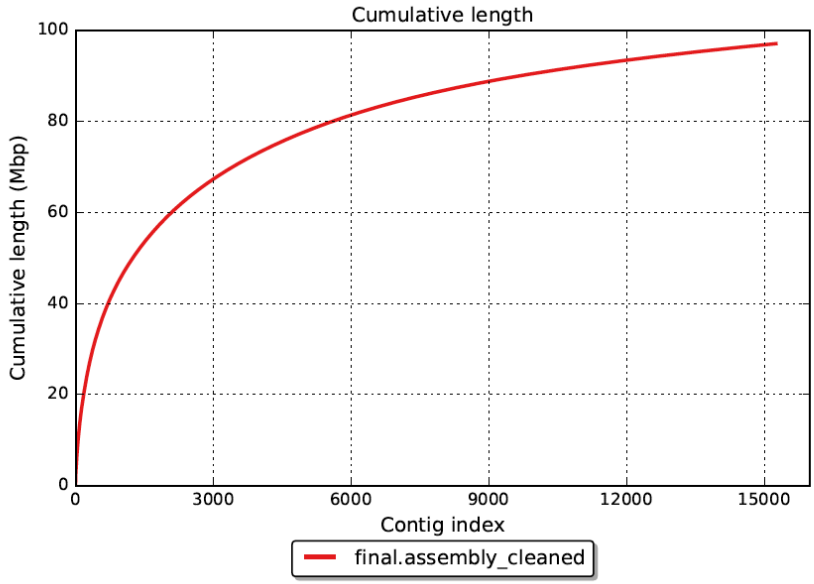


Figure 6.7. Cumulative scaffold length curve for *Steinernema* sp. HBG28 re-assembly as generated with QUAST. For the final assembly with clean nematode reads, the number of scaffolds or contigs ordered in length from largest to smallest (x-axis) is plotted against the cumulative assembly size (y-axis).

Table 6.2 Comparison of the final assembly metrics for *Steinernema* sp. HBG28

	***Steinernema* sp. HBG28**	***H. bacteriophora*[1]**	***S. carpocapsae* Breton[2]**	***S. carpocapsae* All[3]**	***S. scapterisci*[3]**	***S. monticolum*[3]**	***S. glaseri*[3]**	***S. feltiae*[3]**
Number of scaffolds	15,252	1,263	347	1,578	2,877	14,331	7,515	5,839
Largest scaffold (bp)	647,965	2,228,510	8,793,593	1,722,607	1,149,164	110,081	339,094	1,470,990
N50 length (bp)	15,839	312,328	1,245,171	299,566	90,783	11,556	37,444	47,472
GC content (%)	49.49	32.2	45.67	45.53	47.98	42.01	47.63	46.99
Total length (Mbp)	97.04	77.0	84.6	85.66	79.4	89.3	92.9	82.4
CEGMA completeness	93.95	-*	99.2%	98.67	97.13	96.68	97.13	97.57

[1] *Heterorhabditis bacteriophora* sequenced by Bai et al. (2013).

[2] *Steinernema carpocapsae* sequenced by Rougon-Cardoso et al. (2016).

[3] *Steinernema* spp. sequenced by Dillman et al. (2015).

6.5 Discussion

A specialized group of nematodes known as entomopathogenic nematodes have garnered considerable interest over the years due to their capability in biological control (Burnell, 2002; Glazer, 2014; Glazer, 2015). Surprisingly, despite their agricultural and ecological importance, genome information on these nematodes is still greatly lacking. With over 70 recognized species of *Steinernema* and 25 species of *Heterorhabditis* having been described, there are only six sequenced entomopathogenic nematode genomes to date, which includes five *Steinernema* species (*S. carpocapsae*, *S. glaseri*, *S. feltiae*, *S. scapterisci*, *S. monticolum*) and one *Heterorhabditis* species (*H. bacteriophora*) (Dillman et al., 2015; Rougon-Cardoso et al., 2016; Bai et al., 2013). Among entomopathogenic nematodes, the *Steinernema* genus is unique because nematodes belonging to this genus can undergo both amphimictic reproduction (Wang and Bedding, 1996) and *endotokia matricida*. Also, as steinernematids serve a dual-role both as parasites to insects and hosts to *Xenorhabdus* bacteria, *Steinernema* nematodes make promising models to study mutualism and parasitism. Sequencing the genomes of entomopathogenic nematodes can provide important knowledge about the mechanistic underpinnings of their biology, as well as discover the genes involved in parasitism – the implications of which can be significant where their field efficiency as biological control agents is concerned.

This chapter introduces a first genome assembly of a novel *Steinernema* species indigenous in the grasslands of South Africa and is, in fact, among the first *Steinernema* species in South Africa to have its genome sequenced. The aim of this chapter was to sequence the genome of *Steinernema* sp. HBG28 and, in doing so, produce a draft assembly of the nematode using Illumina short-read sequencing.

6.5.1 The genome project of *Steinernema* sp. HBG28

The genome project of *Steinernema* sp. HBG28 was initiated in 2014 with the goal to sequence and assemble local *Steinernema* species that had not been previously sequenced. A research project focused solely on whole genome sequencing of native entomopathogenic nematodes had never previously been conducted in South Africa. Obtaining the genome of local entomopathogenic nematode species could offer important insight into the genetic mechanisms underlying the nematodes' responses to local environments. Coincidentally, the *Steinernema* species isolated during this research project was, in fact, a novel species at the time, belonging to a new phylogenetic group called the 'Cameroonian' clade (Nthenga et al., 2014) and, therefore, served as an excellent candidate for genomic sequencing.

The nematode genome assembly was generated entirely from Illumina short-read sequence data, and while this may appear challenging for a complex eukaryotic genome, the strategy of using mate-pair sequence data, along with paired-end reads did improve genome completeness and sequence representation. The following sections describe some of the challenges faced during sequencing of *Steinernema* sp. HBG28 and the results observed after the implementation of an assembly pipeline tailored towards this study.

6.5.1.1 Considerations prior to genome sequencing

The quality of sequence data is generally determined by the integrity of the starting material. Quality control is pivotal for the preparation of a robust genomic library and subsequent sequencing. As such, several practical considerations were taken into account before commencing with genome sequencing. In particular, the extraction of

high-quality DNA in a sufficient amount was of the utmost importance. Poor yields of DNA can often interfere with the library preparation, especially at various stages where quality control steps are necessary.

Furthermore, the implications of zygosity within the nematode genomic sample was also considered prior to sequencing. As nematodes often occur in large population sizes when propagated under laboratory conditions, their genomes are subject to high levels of heterozygosity. This poses a challenge as nematode individuals are often pooled in order to obtain high quantities of genomic DNA for sequencing. As a consequence of pooling nematodes, the genetic variability of the DNA sample increases, which has implications in the downstream processing of sequence data. In particular, sequencing highly heterozygous genomes can lead to assembly errors as some assembly programs fail to make the distinction between homozygous and heterozygous regions. Multiple assembly paths can then arise, which are not always easy to resolve (Pryszcz and Gabaldón, 2016), leading to highly fragmented assemblies with significantly larger genome sizes than expected or contigs with questionable homology (Del Angel et al., 2018). For the purpose of this study, nematodes of *Steinernema* sp. HBG28 were inbred to reduce the overall heterozygosity within the nematode population and, in turn, mitigate possible downstream errors during the assembly process (Dillman et al., 2012). The inbreeding procedure used in this study entailed a series of related-sibling matings between brother-sister progeny from a single inbred parent for each subculture, resulting in ~90% homozygosity (Wright, 1921b). The final nematode population of *Steinernema* sp. HBG28 from which DNA was isolated for genome sequencing was homozygous with respect to all allelic genetic loci on the homologous chromosomes (See Figure 1; Wright, 1921a). In conformity to the inbreeding model formulated by

Wright (1921a), the homozygosity of alleles would increase inexorably with each generation of mating between sibling progeny derived from a single inbred female. The gradual decline in nematode fecundity and the eventual extinction of the inbred line of *Steinernema* sp. HBG28 were consistent with the effects of inbreeding depression, resulting from either reduced fitness or an accumulation of deleterious mutations (Deborah and Willis et al., 2007).

6.5.2 *De novo* assembly of a *Steinernema* genome

6.5.2.1 Genome sequencing with Illumina technology

The genomic assembly pipeline implemented in this study included stringent quality control steps that involves the removal of low-quality bases and adapter sequences, multiple assemblies with various programs and parameters, contamination screening, error correction and a final re-assembly of a genome sequence.

The main issue encountered when attempting to sequence the genome of *Steinernema* sp. HBG28 was the quality of the extracted DNA, which hindered the preparation of the genomic libraries, and, in turn, affected the sequencing. As a consequence, several DNA extract samples were prepared. Inbred nematodes also produced a fewer offspring over time, which affected the amount of DNA obtained.

6.5.2.2 Adapter and quality trimming

The next approach, once Illumina sequencing had been carried out, was to perform an initial quality check on the raw genome sequence data derived from the inbred DNA samples to determine if any issues had arisen during the sequencing run. The PE sequences were examined for overall quality and the presence of adapter sequences.

Low-quality bases were removed and resulted in the majority of the reads (i.e. 98.79%) having a Phred score of above Q20. This was regarded as satisfactory in terms of quality, as it corresponded to a per base error rate of 0.01. The presence of Nextera Transposase adapter sequences was also detected, which prompted their removal from the reads as to avoid the occurrence of possible chimeric contigs during the assembly process.

6.5.2.3 Genome quality of preliminary assemblies

There is no standardized method for assessing the quality of a genome assembly. Typically, the scaffold or contig N50 metric can be a useful, though flawed, indicator of sequence contiguity. A larger N50 measurement is often thought to be better; however, assemblies can be manipulated by filtering out shorter sequences from the set of assembled sequences to produce higher N50 values, or even by inaccurately merging contigs and scaffolds. For this reason, it is practical to consider other metrics when comparing various assembly outputs, such as the size of the assembled genomes and the total number of contigs. Ideally, the most optimal assembly would have a cumulative sequence length that is in line with expectations. An assembly should also contain as few and as long contigs as possible in order to obtain the most complete and uninterrupted sequences for downstream gene prediction (Dillman, 2012). The coding potential of competing assemblies can also be used in the assessment process using CEGMA to identify which of the 248 core eukaryotic genes (CEGs) are found within a particular assembly.

The preliminary assemblies generated by CLC Genomics Workbench, Velvet and Allpaths-LG resulted in genomes between 105 and 148 Mb in size with various contig sizes, GC content percentages, N50 measurements and CEGMA completeness

percentage values (Table 6.1). Allpaths-LG performed the best on the following genome metrics: number of contigs, N50 length and total assembly length. The Allpaths-LG assembly consisted of the least number of contigs (i.e. 18,219), the highest contig N50 length (i.e. ~14 kb), and the most representative genome length of 105 Mb than those of the CLC and Velvet assemblies.

One irregularity observed amongst the preliminary assemblies was that of the CLC assembly. While this assembly had the largest CEGMA completeness percentage (i.e. 94.76%), the total span was 148 Mb, which seemed very unlikely given that *C. elegans* is about 100 Mb. This could possibly have indicated redundancy within the assembly generated by CLC, possibly due to reads erroneously joined together into contigs. The presence of contaminants may have also been a contributing factor for the poor assembly, resulting in possible chimerism of reads.

6.5.2.4 Screening of bacterial contamination

There is always a risk of contamination when working with genomic material and this can pose a serious complication in any genome project. Contamination can arise from either being introduced pre-DNA extraction or from existing organisms present within the sequencing sample (Del Angel et al., 2018). In instances involving nematode-bacterium symbioses, it can often be difficult to separate endosymbionts (and their DNA) from their respective nematode host, especially in host-directed genome projects (Sumar and Blaxter, 2011). As a result, the recovery of genomic sequence data from nematode samples can often be accompanied by sequences of bacterial origin. The presence of symbiont sequences can ultimately contaminate the final genome if incorporated into the assembly. Understandably, the use of axenic nematode is one approach in dealing with symbiont contamination, however, axenic

cultures can be difficult to establish and can result in poor nematode yields and, in turn, limited the amount of DNA for sequencing.

For this reason, the detection and removal of contaminants was a necessary step in the assembly process. As *Steinernema* sp. HBG28 nematodes were grown on a lawn of their own associated symbiont, *Xenorhabdus khoisanae* during the inbreeding procedure, it was anticipated that raw genome sequence data would include sequences derived from the bacterial symbiont. Thereby, with the use of CLC Genomics Workbench, the sequence reads of *Steinernema* sp. HBG28 were aligned to various bacterial genomes. The Kraken output, based solely on the presence of short *k*-mers, indicated that only a small percent of the sampled reads (i.e. 8.33%) were bacterial in nature. Through the direct mapping of nematode sequence data to the reference genome of the bacterial symbiont, *X. khoisanae*, all unmapped reads were extracted to use in the re-assembly of the nematode genome.

6.5.2.5 Final draft assembly of *Steinernema* sp. HBG28

A final assembly with the filtered read set was performed using the Allpaths-LG algorithm. The final assembly also included scaffolding with MP data. It was anticipated that with the elimination of bacterial contamination and the use of MP data, an improved genome assembly would be achieved for *Steinernema* sp. HBG28. However, with the exception of the N50 measurement, the genome metrics of the final assembly had only marginally improved, when compared to the initial Allpaths-LG assembly.

Among the *Steinernema* genomes, *Steinernema* sp. HBG28 was the largest in size. Earlier work on the *Steinernema* genomes by Dillman (2012) indicated much larger

genome sizes than those later published (Dillman et al., 2015). It is not uncommon for refinements to be made in the assembly process, leading to the generation of several assemblies internally; however, only one assembly sequence is reported for publications and sequencing projects. Based on previously published work by Dillman et al. (2015) and Rougon-Cardoso et al. (2016), the genomes of *Steinernema* nematodes are found to be in the range of 79 and 93 Mbp long. In the present study, the larger genome size of *Steinernema* sp. HBG28 may have been an artefact of the genome assembly process itself. Another consideration is that despite the precautionary steps taken to remove bacterial contaminating sequences, it is possible that the *Steinernema* sp. HBG28 assembly still contained contamination to some degree, leading to a genome size that was larger than expected.

The N50 statistic is often used as a determinant of assembly quality, primarily because a high value signifies better representation of the genome by virtue of having longer and possibly intact sequences within the assembly. When comparing the N50 statistic of the *Steinernema* sp. HBG28 assembly to those of previously reported for *Steinernema*, it was found to be significantly lower (Table 6.3), especially when compared to the N50 value of 1,245,171 bp observed in the *S. carpocapsae* strain Breton assembly (Rougon-Cardoso et al., 2016). While an N50 value of 15,839 bp for the *Steinernema* sp. HBG28 genome assembly may not be ideal, it is, in fact, larger than the average gene size reported for the genome of *S. carpocapsae* strain Breton (i.e. a gene size of 2,681 bp), indicating that full length genes were likely to have been sequenced and were present within the *Steinernema* sp. HBG28 assembly. The CEGMA analysis also indicated that 93.95% (complete) and 97.18% (partial/complete) of the 248 highly-conserved core eukaryotic genes were detected in the genome sequence of *Steinernema* sp. HBG28, indicating that the final genome assembly was complete.

6.6 Conclusion

The genome assembly produced in this chapter was complete; however, the assembly-derived features of *Steinernema* sp. HBG28 was not comparable to other recently assembled *Steinernema* genomes. Moreover, while the N50 value observed in this study was relatively small, this does not negate the fact that there is biologically-relevant information contained within the genome assembly of *Steinernema* sp. HBG28; and questions pertaining to nematode parasitism and host-seeking behaviour can still be addressed. For future work, the genome assembly could be improved upon by refining the assembly process, evaluating other assemblers and including more stringent steps for the screening and removal of bacterial DNA contamination.

6.7 References

1. Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Research*, **9:**3015–3027.
2. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
3. Bai, X. D., Adams, B. J., Ciche, T. A., Clifton, S., et al. (2013). A lover and a fighter: The genome sequence of an entomopathogenic nematode *Heterorhabditis bacteriophora*. *Plos One*, **8:**e69618.
4. Blaxter, M. and Koutsovoulos, G. (2015). The evolution of parasitism in Nematoda. *Parasitology*, **142:**S26–S39.
5. Blaxter, M. L. (2003). Molecular systematics: counting angels with DNA. *Nature*, **421:**122–124.
6. Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30:**2114–2120.
7. Burnell, A. (2002). Genetics and Genetic Improvement. In: Gaugler, R., editor. Entomopathogenic Nematology. New York: CAB International; 2002. pp. 241–264.
8. Charleworth, D. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, **10:**783–796.
9. De Ley, P. and Blaxter, M. L. (2004). A new system for Nematoda: combining morphological characters with molecular trees and translating clades into ranks and taxa. *Nematology Monographs and Perspectives*, **2:**633–653.
10. De Ley, P. and Blaxter, M. L. (2002). Systematic position and phylogeny. In: The Biology of Nematodes, D. L. Lee, ed., London: Taylor and Francis, pp. 1–30.

11. Del Angel, V. D., Hjerde, E., Sterck, L., Capella-Gutierrez, S., et al. (2018). Ten steps to get started in Genome Assembly and Annotation. *F1000Research*, **7:**148.
12. Dillman, A. R. (2012). Host Seeking and the Genomic Architecture of Parasitism Among Entomopathogenic Nematodes, Doctoral dissertation. California Institute of Technology, Pasadena, California.
13. Dillman, A. R., Mortazavi, A. and Sternbeg, P. W. (2012). Incorporating Genomics into the Toolkit of Nematology. *Journal of Nematology*, **44:**191-205.
14. Dillman, A. R., Chaston, J. M., Adams, B. J., Ciche, T. A., et al. (2012). An entomopathogenic nematode by any other name. *PLoS Pathogens*, **8:**e1002527.
15. Dillman, A. R., Macchietto, M., Porter, C. F., Roger, A., et al. (2015). Comparative genomics of *Steinernema* reveals deeply conserved gene regulatory networks. *Genome Biology*. **16:**200.
16. Earl, D., Bradnam, K., St. John, J., Darling, A., et al. (2011). Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Research,* **21:**2224–2241.
17. Ewing, B. and Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, **8:**186–194.
18. Glazer, I. (2014). Genetic Improvement and Breeding of EPN: The Race for the "Super Nematode". *Journal of Nematology*, **46:**168–168.
19. Glazer, I. (2015). Improvement of Entomopathogenic Nematodes: A Genetic Approach. In: Campos-Herrera R, editor. Nematode Pathogenesis of Insects and Other Pests. Springer; 2015. pp. 29–55.
20. Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F. J., et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **1084:**1513-1518.

21. Gurevich, A., Saveliev, V. Vyahhi, N. and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29:**1072–1075.
22. Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., et al. (2006). Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Molecular Biology and Evolution*, **23:**1792–1800.
23. Howe, K. L., Bolt, B. J., Cain, S., Chan, J., et al. (2016). WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Research*, **44:**D774-D780.
24. Illumina. (2009). Mate Pair Library v2 Sample Preparation Guide For 2–5 kb Libraries.
25. Illumina. (2012). Data Processing of Nextera® Mate Pair Reads on Illumina Sequencing Platforms.
26. Kircher, M., Stenzel, U. and Kelso, J. (2009). Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biology*, **10:**R83.
27. Kumar, S. and Blaxter, M. L. (2011). Simultaneous genome sequencing of symbionts and their hosts. *Symbiosis*, **55:**119–126.
28. Kumar, S., Koutsovoulos, G., Kaur, G. and Blaxter, M. (2012). Towards 959 nematode genomes. *Worm*, **1:**42-50.
29. Lambshead, J. (1993). Recent developments in marine benthic biodiversity research. *Oceanis*, **19:**5–24.
30. Lee, R. Y. N., Howe, K. L., Harris, T. W., Arnaboldi, V. et al. (2017). WormBase 2017: molting into a new stage. *Nucleic Acids Research*, **46:**D869–D874.
31. Levy. S. E. and Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, **17:**95–115.
32. Nthenga, I., Knoetze, R., Berry, S., Tiedt, L. R., et al. (2014). *Steinernema sacchari* n. sp. (Rhabditida: Steinernematidae), a new entomopathogenic nematode from South Africa. *Nematology*, **16:**475–494.

33. Parra, G., Bradnam, K. and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23:**1061–1067.

34. Pevzner, P. A., Tang, H. and Waterman, M. S. (2001). An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences*, **98:**9748–9753.

35. Pryszcz, L. P. and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, **44:**e113.

36. Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz J., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, **475:**348–352.

37. Rougon-Cardoso, A., Flores-Ponce, M., Ramos-Aboites, H. E., Martínez-Guerrero, C. E., et al. (2016). The genome, transcriptome, and proteome of the nematode *Steinernema carpocapsae*: evolutionary signatures of a pathogenic lifestyle. *Scientific Reports*, **6:**37536.

38. The C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, **282:**2011–2046.

39. Utturkar, S. M., Klingeman, D. M., Bruno-Barcena, J. M., Chinn, M. S., et al. (2015). Sequence data for *Clostridium autoethanogenum* using three generations of sequencing technologies. *Scientific Data*, **2:**150014.

40. Wang, J. and Bedding, R. A. (1996). Population development of *Heterorhabditis bacteriophora* and *Steinernema carpocapsae* in larvae of *Galleria mellonella*. *Fundamental and Applied Nematology*,**19:**363–367.

41. Wood, D. E. and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, **15:**R46.

42. Wright, S. (1921a). Systems of Mating. II. Systems of Mating. II. the Effects of Inbreeding on the Genetic Composition of a Population. *Genetics*, **6:**124–143.
43. Wright, S. (1921b). Systems of mating. V. General considerations. *Genetics*, 6: 167–178.
44. Zerbino, D. R. and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using *de Bruijn* graphs. *Genome Research*, **18:**821–829.

Chapter 7

Genome characterization of a novel *Steinernema* entomopathogenic nematode species

7.1 Abstract

This chapter describes the genome annotation and subsequent analyses of a draft genome of the entomopathogenic nematode, *Steinernema* sp. HBG28, isolated from the Gauteng Province of South Africa. The present study is unique as it represents the first genomic characterization of a novel nematode species using Illumina sequencing, in addition to molecular and morphological analyses. Structural genome annotation predicted 35,869 gene models in the 97.04 Mb genome of *Steinernema* sp. HBG28. On average, genes were 1,637 base pairs (bp) in length with the shortest and longest gene lengths being 201 and 32,894 bp, respectively. Genes comprised a total of 123,244 exons and 87,375 introns. Repeat detection identified 5.07% of the *Steinernema* sp. HBG28 draft genome as repetitive, with a total of 40,684 repeats. A total of 1,281 putative tRNA genes were predicted, of which 75 were regarded as pseudogenes and 1194 genes were determined to decode for the standard 20 amino acids; and 43 rRNA genes were identified in the genome. Approximately 64% of predicted-coding sequences were functionally annotated using BLASTP, 69% were annotated with InterProScan and 46.5% annotated with Gene Ontology (GO) terms. Additionally, 48.1% and 15% were functionally classified using KEGG and KOG classifications, respectively. A total of 3,307 single-copy orthologues were common to *Steinernema* sp. HBG28, *S. carpocapsae*, S. *feltiae*, *S. glaseri*, *S.*

***monticolum* and *S. scapterisci.* Phylogenomic analysis indicated that *Steinernema* sp. HBG28. was closely related to *S. monticolum* and *S. feltiae*. Protein domain analyses of *Steinernema* sp. HBG28 confirmed a role in the virulence of insect hosts beyond serving as merely a vector and offering a protective environment for its associated bacterial symbiont, *X. khoisanae*. The identification of a variety of proteases, protease inhibitors, venom-like proteins and GPCRs, as well as key genes involved in important pathways implicated in stress resistance and longevity in this study, could serve as candidate targets for genetic manipulation.**

7.2 Introduction

Reductions in crop yields at a global-scale can often be attributed to insect herbivory, resulting in significant economic and agricultural losses. The consumption of major grain crops by insect pests, for example, accounts for 5 to 20% of yield losses (Deutsch et al., 2018). In efforts to control various agricultural pests, chemical pesticides have been used; however, their application can pose a serious risk to human health and surrounding wildlife, in addition to affecting the environment by contaminating groundwater, adversely affecting soil fertility and reducing beneficial soil flora (Aktar et al., 2009).

Entomopathogenic nematodes (EPNs) from the family of Steinernematidae and Heterorhabditidae have garnered considerable attention because they feature the ability to induce rapid insect death to a wide range of agriculturally-important insect pests. Due to this ability, these specialized nematodes can be exploited as biological control agents. They also serve as attractive models to gain understanding into the mechanisms behind their pathogenicity and their symbiotic association to insect-

pathogenic bacterial members of the genera *Xenorhabdus* and *Photorhabdus*. *Steinernema* nematodes are found to exclusively associate with *Xenorhabdus* spp. and it has been found that insect death is, for the most part, orchestrated by the associated bacterial symbiont. As a consequence, the main focus of most research has often been on the virulence effects of the bacteria, instead of the nematode itself (Thaler et al., 1998; Caldas et al., 2002; Herbert and Goodrich-Blair, 2007). Nonetheless, there is overwhelming evidence to suggest that entomopathogenic nematode plays a more active role in the pathogenicity of insects, aside from being merely vectors for their pathogenic bacteria (Binda-Rossetti et al., 2016; Toubarro et al., 2009).

In spite of the tremendous potential held by entomopathogenic nematodes, their inconsistency in the field has prevented these nematodes from being fully exploited as biological control agents. For decades, researchers have attempted to enhance the efficacy of entomopathogenic nematodes in controlling various insect pests during field application, through genetic approaches such as artificial selection and breeding of these nematodes (Lu et al., 2016). Artificial selection has often relied on the collection of novel species of entomopathogenic nematodes that are acclimated to local environmental conditions and insect pests. The selection and breeding of entomopathogenic nematodes to improve upon their inherent insect control ability has mainly been reliant on the identification and manipulation of key traits (Burnell, 2002; Glazer, 2014; Glazer, 2015). Some of these traits encompass enhanced tolerance to fluctuating temperatures and desiccation conditions, in addition to improving in host-seeking behaviour, nematode pathogenicity and withstanding nematicide exposure (Lu et al., 2016). Also, traits improved in this manner are subject to instability or the eventual loss of the selected trait(s) (Glazer, 2014), as a result of

a reduction in the overall fitness and inbreeding depression of the entomopathogenic nematodes being mass-produced (Bilgrami, 2006; Chaston, 2011).

Genomic characterization through genome sequencing, assembly and annotation of entomopathogenic nematodes can provide a framework for understanding nematode biology and nematode-bacterium symbiosis; in addition to identifying important genes linked to insect pathogenicity, nematode parasitism and signal transduction (Bai et al., 2013; Dillman et al., 2015). The genetic information gathered from the genomes of entomopathogenic nematodes can facilitate new research, leading to the genetic enhancements of these nematodes and, thereby improve their field efficacy overall (Lu et al., 2016).

7.2.1 Aims and objectives

The main aim of this chapter, after producing a draft genome assembly for a *Steinernema* species indigenous to the grasslands of South Africa, was to perform genome annotation using the MAKER2 genome annotation pipeline. To this end, the genome was then functionally annotated and comparative genomic analyses was performed. The predicted protein sequences of *Steinernema* sp. HBG28 were also analysed in this study.

7.3 Materials and methods

7.3.1 Transcriptome sequencing

7.3.1.1 Isolation of nematode RNA

Total RNA was isolated from a pool of individual infective juvenile nematodes (*Steinernema* sp. HBG28) using the ZR Tissue and Insect RNA MicroPrep Kit (Zymo

Research), according to the protocol supplied by the manufacturer. Isolated RNA samples were quantified using a Qubit™ 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA).

7.3.1.2 Library preparation and sequencing

The RNA-seq library was prepared with the TruSeq stranded-mRNA Library Prep kit (cat # RS-122-2101). Transcriptome sequencing was performed on an Illumina HiSeq 2500 instrument (2 × 125 bp) at ARC Biotechnology Platform facility in Onderstepoort, Pretoria.

7.3.1.3 Adapter and quality trimming of reads

Quality assessment of reads prior to trimming was performed using FastQC version 0.11.5 (Andrews, 2010). Adapter-content and quality trimming were performed using Trimmomatic version 0.32 (Bolger et. al. 2014). Quality trimming was performed so that the Phred quality score (Q) of each base was at least Q20 and above. The trimmed reads were assessed for a final time using FastQC to verify successful trimming.

7.3.1.4 Transcriptome assembly

The assembly of a *de novo* transcriptome was generated with Bridger version r2014-12-01 (Chang et al., 2015) using ~116 million reads. For the transcriptome assemblies, two *k*-mer lengths (i.e. *k*=25 and *k*=31) were tested. The quality metrics of the *de novo* transcriptomes was determined using TransRate version 1.0.3 (Smith-Unna et al., 2016), whereby paired-end reads were aligned to the *de novo* transcriptome assemblies produced for both *k*-mer lengths (i.e. 25 and 31).

7.3.2 Genome annotation

7.3.2.1 Gene prediction and annotation using the MAKER2 pipeline

Automated genome annotation was performed with the MAKER2 version 2.31.3 pipeline (Holt and Yandell 2011), using a combination of transcriptome-based prediction generated by the Bridger assembler, curated protein sequences from the SWISS-PROT database (Boutet et al. 2016) and homology-based data of *S. carpocapsae*, *S. scapterisci*, S. *monticolum*, *S. glaseri* and *S. feltiae* (Dillman et al., 2015) and *C. elegans* from Wormbase (https://wormbase.org/; Howe et al., 2016; Lee et al., 2017) as evidence. Protein-coding genes were predicted on the repeat-masked assembly with two different *ab initio* gene finders, Augustus version 3.02 (Stanke et al. 2004) and SNAP version 2013-11-29 (Korf 2004) using a two-pass (iterative) workflow for MAKER2. Automated annotation with MAKER2 was carried out for three rounds. The Annotation Edit Distance (AED) metric was used in the assessment of the annotation quality. A cumulative histogram plot of AED scores was produced with MAKER2. Gene models were produced in GFF3 format by MAKER2 and were used to calculate gene prediction metrics such as the lengths of exons, introns, genes, transcripts and protein-coding genes (CDS).

7.3.1.2 Repeat-masking

Identification of repeat families was performed with RepeatModeler (open 1.0.11) (Smit and Hubley, 2008-2015) on the unannotated draft assembly by building consensus models of putative repeats. RepeatMasker (open 4.0.7) (Smit et al., 2013; 2015) was used, which searched for all known repeat sequences against the consensus repeat library that was built. The repeats were included in the GFF3 output file generated with MAKER2.

7.3.1.3 RNA annotation

As MAKER2 did not include the functionality to annotate tRNA isotypes, tRNAscan-SE version 1.3.1 (Lowe and Eddy, 1997) was used to identify putative tRNAs within the draft genome assembly of *Steinernema* sp. HBG28. Covariance models from the Rfam database version 14.0 (Nawrocki et al. 2015) was used to examine the un-masked genome assembly to detect rRNA genes using the Infernal software version 1.1.2 (Nawrocki and Eddy, 2013). In addition, RNAmmer version 1.2-core (Lagesen et al., 2007) was also used in rRNA prediction.

7.3.1.4 Functional annotation of protein-coding sequences

The predicted protein-coding sequences were assigned gene function based on the best hits retrieved during a sequence similarity search against the GenBank non-reductant (NR) protein database using BLASTP (Altschul et al. 1997) with an E-value cut-off $1x10^{-5}$ in Blast2GO suite version 5.2.4. InterProScan was also used to identify matches against the GENE3D, SMART, Pfam, SUPERFAMILY, CDD, TIGRFAMs and PANTHER databases. Gene ontology (GO) annotation was assigned to the predicted sequences. The KASS (KEGG Automatic Annotation Server) was used to assign Kyoto encyclopedia of genes and genomes (KEGG) pathways to the predicted proteins of *Steinernema* sp. HBG28 through BLASTX with an E-value cut-off of $1x10^{-5}$, via the online GhostKOALA version 2.0 server (Kanehisa et al., 2016; https://www.kegg.jp/ghostkoala/). Eukaryotic orthologous groups (KOG) classifications were assigned to the predicted protein sequences of *Steinernema* sp. HBG28 using the KOG database (Tatusov et al., 2003) to deduce putative gene function of *Steinernema* sp. HBG28 through the WebMGA portal (Wu et al., 2011), with an E-value cut-off of $1x10^{-5}$.

7.3.1.5 Protein domain analyses

To evaluate the prevalence of protein domains in the proteomes of *Steinernema* sp. HBG28, *S. carpocapsae* strain All and *S. glaseri* (Dillman et al., 2015), *Heterorhabditis bacteriophora* (Bai et al., 2013; McLean et al., 2018), *Necator americanus* (Tang et al., 2014) and *Dictyocaulus viviparus* (McNulty et al., 2016), protein domain analyses were performed using HMMER3 version 3.1 b (Zhang and Wood, 2003) against the Pfam domain database with an E-value cut-off of $1x10^{-5}$, according to the approach by Dillman et al. (2015).

7.3.1.6 Detection of proteases and protease inhibitors

To detect putative peptidases (also termed proteases or proteinases) and peptidase inhibitors, protein-coding sequences were subjected a MEROPS batch BLASTP sequence similarity search, with an E-value cut-off of 1e-5 (Rawlings et al., 2012).

7.3.1.7 Gene orthology analyses

The OrthoMCL pipeline version 2.09 (Li et al. 2003) was used to cluster proteins into families of orthologous genes among *S. carpocapsae, S. scapterisci, S. monticolum, S. feltiae and S. glaseri* (Dillman et al., 2015), with a Markov Clustering (MCL) inflation value of 1.5 and an E-value of $1x10^{-5}$. The output results were subsequently visualised using OrthoVENN (Wang et al., 2015).

7.3.1.8 Phylogenetic analysis

Putative single-copy orthologs were identified for six *Steinernema* spp. using from OrthoMCL and were further explored by constructing a phylogenetic tree. Each single-copy orthologue consisted of six proteins with each amino acid sequence derived from

a different *Steinernema* species (also known as singletons). These protein sequences were aligned separately with MUSCLE (Edgar 2004) and subsequently trimmed with trimAl version 1.3 (Capella-Gutierrez et al. 2009). The trimmed alignments were concatenated using a Perl-derived script, FASTconCAT (Kück and Meusemann 2010). The most appropriate phylogenetic substitution model was determined using IQ-Tree version 1.6.8 (Nguyen et al., 2015), where a general matrix (JTT) (Jones et al., 1992), empirical AA frequencies from the data (+F), allowing for a proportion of invariable sites (+I) and a discrete Gamma model with default 4 rate (+G4) (Yang, 1994) were selected as the best-fit model. The maximum likelihood tree was constructed with IQ-TREE with bootstrap support of 100 using the model parameters, JTT, +F, +I and +G4.

7.4 Results

7.4.1 Transcriptome assembly

Using a *k*-mer length of 25 and 31, two *de novo* transcriptome assemblies of *Steinernema* sp. HBG28 were generated with the Bridger assembler to provide organism-specific transcript/gene evidence for genome annotation using the automatic MAKER2 pipeline. The TransRate tool was utilized in assessing the transcriptome assemblies by aligning the mRNA paired-end reads to the generated *de novo* transcriptomes. Of the two Bridger transcriptome assemblies, the 25-mer assembly (*k*=25) was selected for annotation purposes as it rendered slightly better assembly metrics when compared to the 31-mer assembly (*k*=31). This was indicated by several assembly metrics, shown in Table 7.1, which include the total number of contigs produced, the number of contigs over 1000 bases in length, the mean contig length, as well as the N90, N50, N30, N10 cumulative metrics. It was also found that both transcriptome assemblies produced the same good mappings' percentage of

76%, which reflected the accuracy of how well the read pairs mapped to the transcripts produced in the assembly.

Additionally, assembly scores, calculated by TransRate, were used in assessing the transcriptome assemblies. The assembly score, based on the geometric mean of all contig scores and the input read data, provides insight into the quality and completeness of any transcriptome assembly. An assembly score of 1, the maximum possible score, indicates perfect read mapping to the assembled contigs, whilst a score of 0 is the minimum score achievable, indicative of poor read mapping. As such, the assembly scores of the 25- and 31-mer transcriptome assemblies produced in this study resulted in 0.2408 and 0.2536, respectively. Although, the 31-mer assembly resulted in a better assembly score, the 25-mer assembly was selected based on the other optimal metrics previously mentioned.

Table 7.1 Comparison of transcriptome assembly metrics for *Steinernema* sp. HBG28.

Features	BridgerK25	BridgerK31
No. of contigs (transcripts)	63,335	61,582
Smallest contig length (bp)	201	201
Largest contig length (bp)	29,067	21,458
No. of bases in assembly	71,171,292	66,659,717
Mean contig length (bp)	**1,123.73**	1082.45
No. over 1kb	**23,425**	21,924
Mean ORF %	64.69	**66.31**
N90 (bp)	**442**	415
N50 (bp)	**1,954**	1,911
N30 (bp)	**3,022**	2,947
N10 (bp)	**5,525**	5,262
GC %	49	49
Good mappings %	**76**	**76**
TRANSRATE ASSEMBLY SCORE	0.2408	**0.2536**
good contigs	**58,803**	57,675
good contigs %	93	**94**

Note: Values in bold font represent the best metric in that row.

7.4.2 Genomic structural annotation

7.4.2.1 Validation of the MAKER2 annotation

Annotation quality can be assessed through the Annotation Edit Distance (AED), which determines the goodness of fit of any given annotation in relation to the aligned evidence. The AED can range between 0 and 1, where an AED of 0 denotes perfect support with the available evidence, and an AED of 1 signifying a lack of support for the predicted gene models (Eilbeck et al., 2009).

The genome annotation of *Steinernema* sp. HBG28 had satisfactory levels of evidence supporting the gene models generated with MAKER2. The cumulative

histogram plot of AED scores (Figure 7.1) indicated that of the predicted *Steinernema* sp. HBG28 gene products, approximately 75% of proteins were supported by aligned evidence, with an AED score less than 0.5. Only about 25% of proteins were without evidence with an AED score less than 1, but these proteins did possess Pfam-A domains.

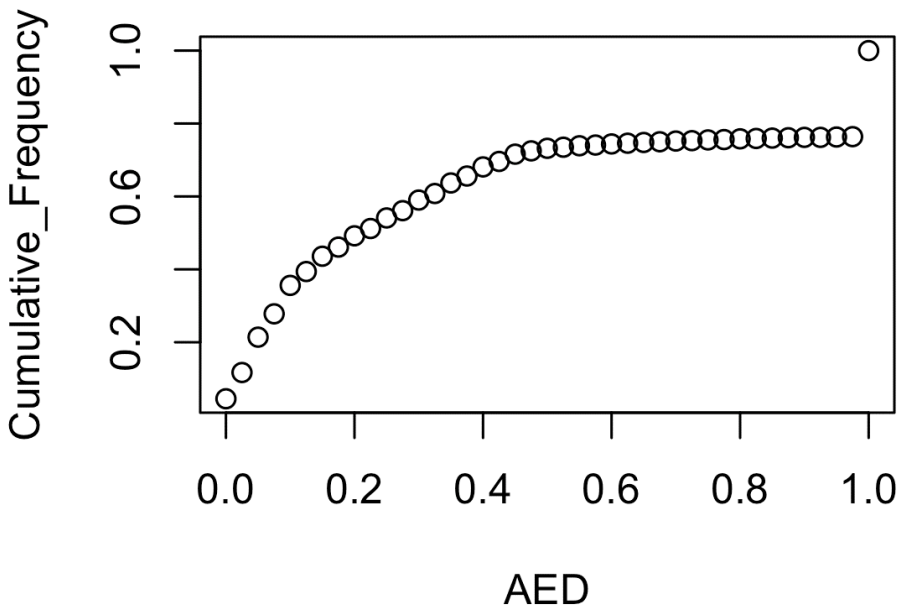


Figure 7.1. Cumulative distribution frequency curve of annotation edit distance (AED) scores shown for the MAKER2 gene models.

7.4.2.2 Gene model prediction metrics

The MAKER2 annotation pipeline predicted 35,869 gene models and transcripts in the annotated *Steinernema* sp. HBG28 genome. Other genomic features that were annotated in the *Steinernema* sp. HBG28 draft genome include the lengths of exons, introns, genes and transcripts (Figure 7.2). On average, genes and transcripts were 1,637 bp in length with the shortest and longest gene lengths being 201 and 32,894 bp, respectively. Genes comprised a total of 123,244 exons and 87,375 introns (Table 7.2). The total exon length was 46 Mbp (373 bp on average) and the total intron length

was 12 Mbp (147 bp on average). Similarly, the average CDS length was 1,175 bp, with the shortest and longest gene lengths spanning up to 201 and 27,141 bp, respectively.

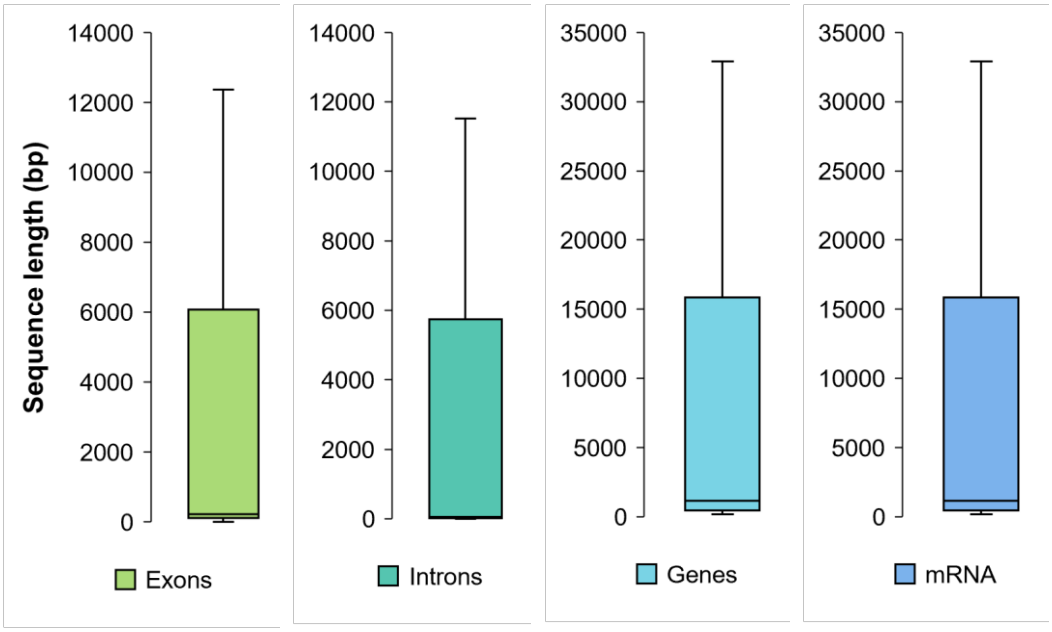


Figure 7.2. Box and whisker plots indicating the length distribution of genomic features. Genomic features were determined through the automated annotation pipeline of MAKER2 and comprise the lengths of exons, introns, genes, mRNA and CDS in bp.

7.4.2.3 Repeat detection

The MAKER2 pipeline included repeat detection using the algorithm, RepeatMasker, which masked repeats within the genome assembly. In summary, repeat detection identified 5.07% of the *Steinernema* sp. HBG28 draft genome as repetitive, with a total of 40,684 repeats. Unclassified repeats constituted the bulk of the repeat content with 26,900 repeats (65%), followed by 11,475 simple sequence repeats (SSRs) with (28%), 1,186 low complexity region repeats (5%), 282 DNA transposons (1%) and 104 long terminal repeats (LTRs) (0.3%), as shown in Figure 7.3. Long and short

interspersed nuclear elements (LINEs and SINEs) were rare, making up only 0.1 and 0.2% of the repeat content, respectively.

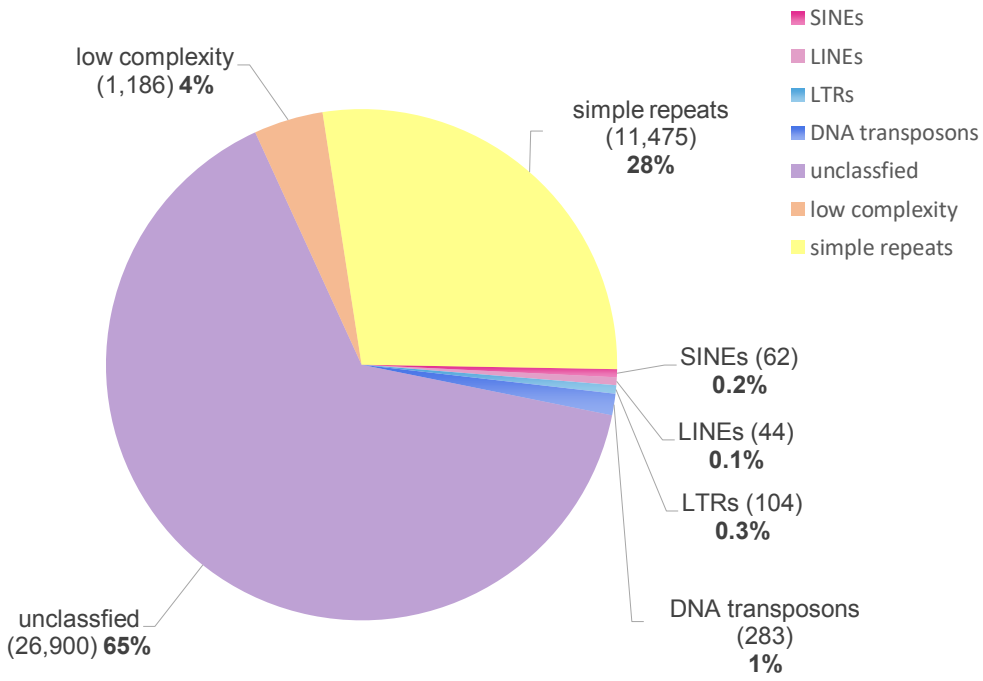


Figure 7.3. Repeat content of the *Steinernema* sp. HBG28 genome. Relative proportions are shown, with long and short interspersed (LINEs and SINEs), long terminal repeats (LTRs), DNA transposon elements, low complexity, simple and unclassified repeats identified using RepeatMasker. Unclassified repeats were the most abundant with 66%, followed by simple repeats with 28%, low complexity regions with 5% and DNA transposons with 1%. LINEs and SINEs were the least abundant repeats.

Table 7.2 Comparison of gene model statistics of the draft genome of *Steinernema* sp. HBG28. and five other *Steinernema* spp.[1]

Features	*Steinernema* sp. HBG28	*S. carpocapsae* strain Breton[1]	*S. carpocapsae* strain All[2]	*S. scapterisci*[2]	*S. monticolum*[2]	*S. glaseri*[2]	*S. feltiae*[2]
Estimated assembly size (in Mbp)	97.04	84.6	85.6	79.4	89.3	92.9	82.4
Number of predicted genes	35,869	16,333	28,313	31,378	36,007	34,143	33,459
Number of predicted transcripts	35,869	-*	31,944	33,149	38,381	37,120	36,434
Number of exons	123,244	81,665	141,565	125,512	144,028	136,572	133,836
Number of introns	87,375	-*	113,252	94,134	108,021	102,429	100,377
Mean gene length (bp)	1,637	2,681	2,030	1,842	1,604	1,855	1,730
Mean exon length (bp)	373	222.37	212	224	217	216	220
Mean intron length (bp)	147	145.44	194	153	161	218	154
Mean number of exons per gene	3.4	6	5	4	4	4	4
Mean number of introns per gene	2.4	-*	4	3	3	3	3
Repeat content (%)	5.07	7.46	7.46	2.75	10.49	5.34	6.70
tRNAs	1,281	1097	-	-	-	-	-
rRNAs	43	40	-	-	-	-	-

[1] *Steinernema carpocapsae* strain Breton sequenced by Rougon-Cardoso et al. (2016).

[2] *Steinernema* spp. sequenced by Dillman et al. (2015).

*Not specified.

7.4.2.4 RNA annotation

Transfer RNA (tRNA) genes were detected within the genome assembly. A total of 1,281 putative tRNA genes were predicted, of which 75 were regarded as pseudogenes and 1194 genes were determined to decode for the standard 20 amino acids, as shown in Figure 7.4. The most abundant tRNA isoform was glycine (131), followed by arginine (99) and leucine (98). In addition, 43 rRNA genes were identified in the genome of *Steinernema* sp. HBG28. The genome was also enriched in mir-1, mir-10, mir-124, mir-2, mir-252, mir-46, mir-71, mir-8, mir-81, mir-86 and mir-9 (microRNAs), comprising 14 micro-RNA hairpins.

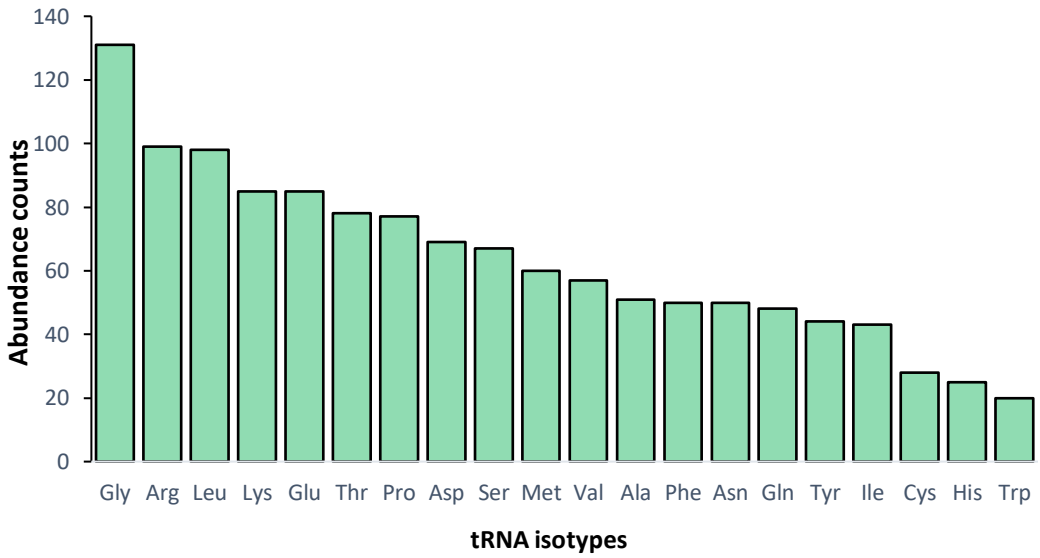


Figure 7.4. Abundance of putative tRNA isotypes within the *Steinernema* sp. HBG28 genome assembly, as predicted by tRNAscan-SE (Lowe and Eddy, 1997).

7.4.3 Genomic functional annotation

7.4.3.1 Gene ontology (GO) based distribution of protein

Of the 35,869 predicted-coding sequences, 23,085 (64%) had significant BLASTP matches to proteins in GenBank non-redundant protein database and 24,910 (69%)

were annotated with InterProScan. Using the Blast2GO program, GO annotation were retrieved for 16,683 (46.5%) sequences, resulting in 39,945 GO terms assignments. The GO terms associated with biological processes, molecular functions and cellular components were assigned generic terms at level 2, 3 and 4.

This resulted in 28,195 CDS involved in biological processes with 18,867 and 18,611 sequences implicated in molecular functions and cellular components, respectively. At level 2, the most prevalent GO terms associated with biological processes were metabolic process genes (8,231), followed by cellular process genes (7,270) and biological regulation genes (3,206) (Figure 7.5). Among the genes annotated to various molecular functions, the largest categories were catalytic activity (8,011), followed by binding activity (7,011) and transporter activity (1,576). The distribution according to cellular components largely comprised genes involved in the cell (3,831), membrane (3,736) and protein-containing complex components (2,902).

At level 3, the most frequent GO terms categories associated with biological processes were organic substance metabolic process genes (8,915), followed by primary metabolic process genes (8,285) and nitrogen compound metabolic process genes (8,127) (Figure 7.6). For molecular functions, the largest categories were organic cyclic compound binding (6,608), followed by heterocyclic compound binding (6,608) and ion binding (4,821) (Figure 7.7). The distribution based on cellular components largely included genes involved in intracellular function (4,006), intracellular part function (3,926) and intrinsic components of the membrane (2,867) (Figure 7.8).

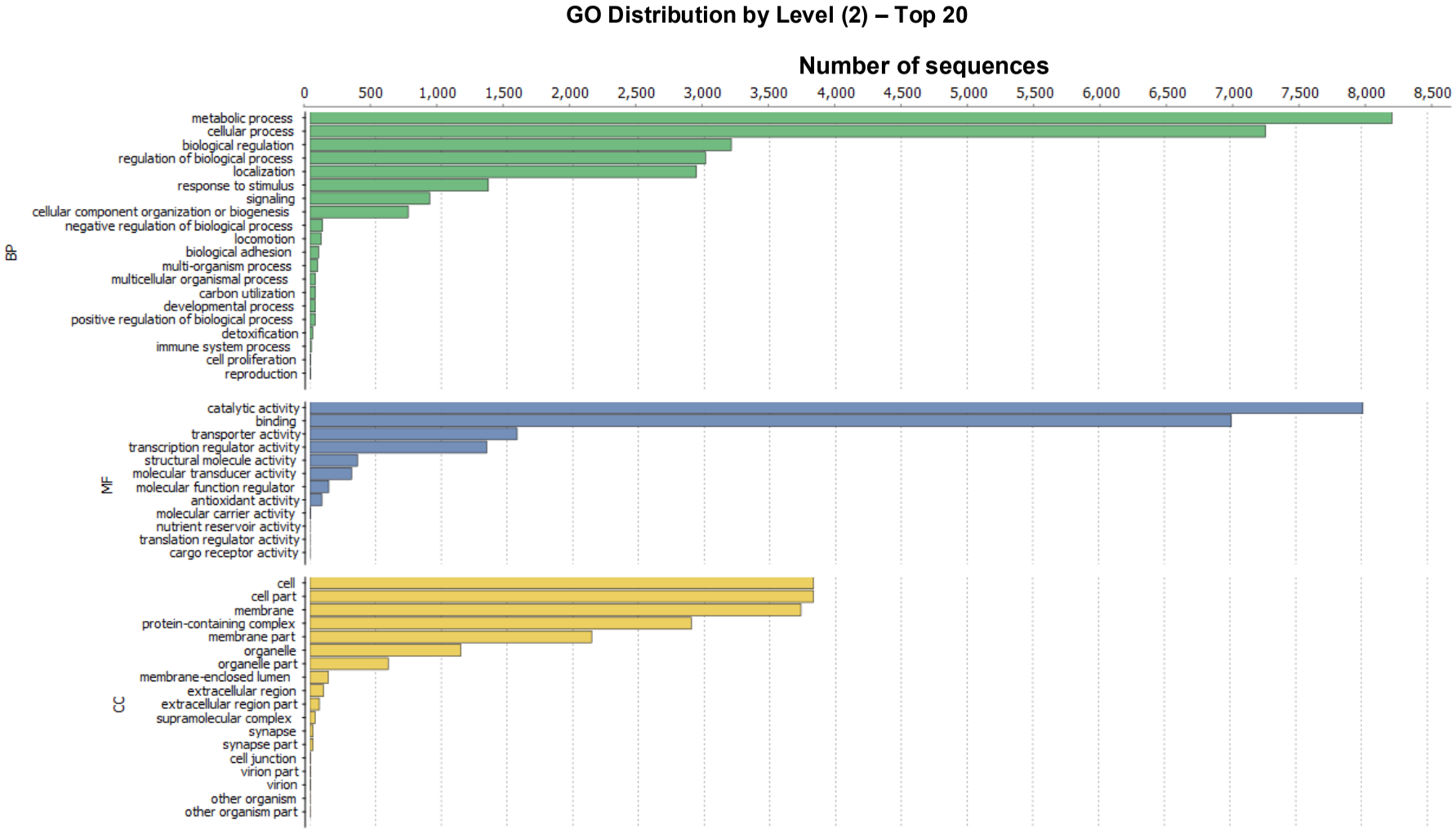


Figure 7.5. Gene Ontology (GO) annotation of *Steinernema* sp. HBG28 protein-coding sequences in the genome assembly. Results are summarized in three main categories: cellular components, molecular functions and biological processes. A total of 16,683 protein-coding genes have been assigned GO terms at level 3.

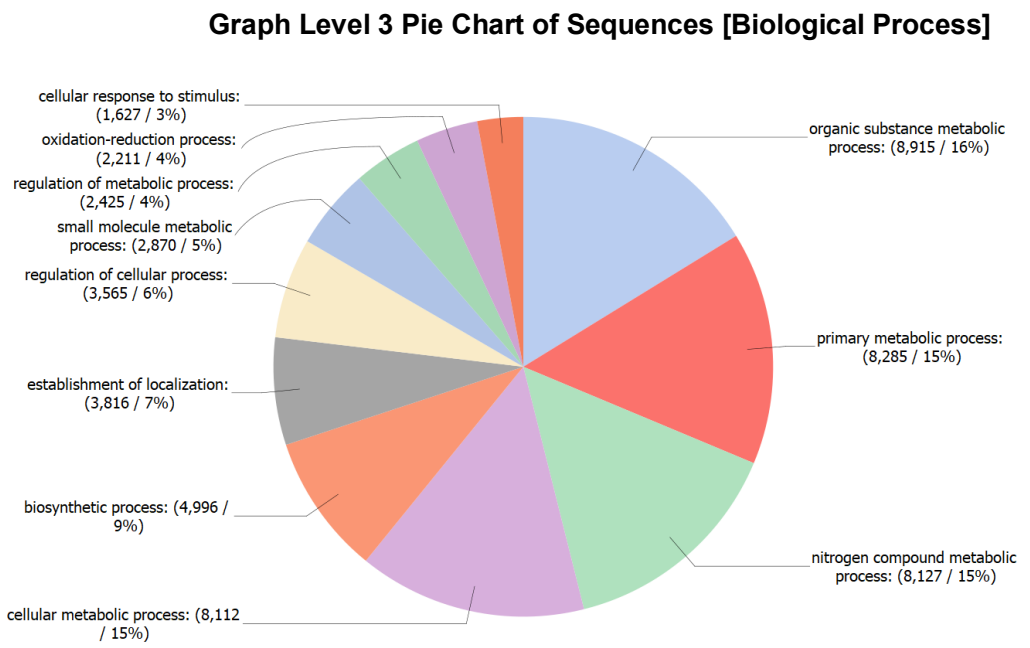


Figure 7.6. Biological Process GO term distribution of *Steinernema* sp. HBG28. The pie chart represents various biological processes several protein-coding genes involved in individual biological process with their distribution ratio also indicated.

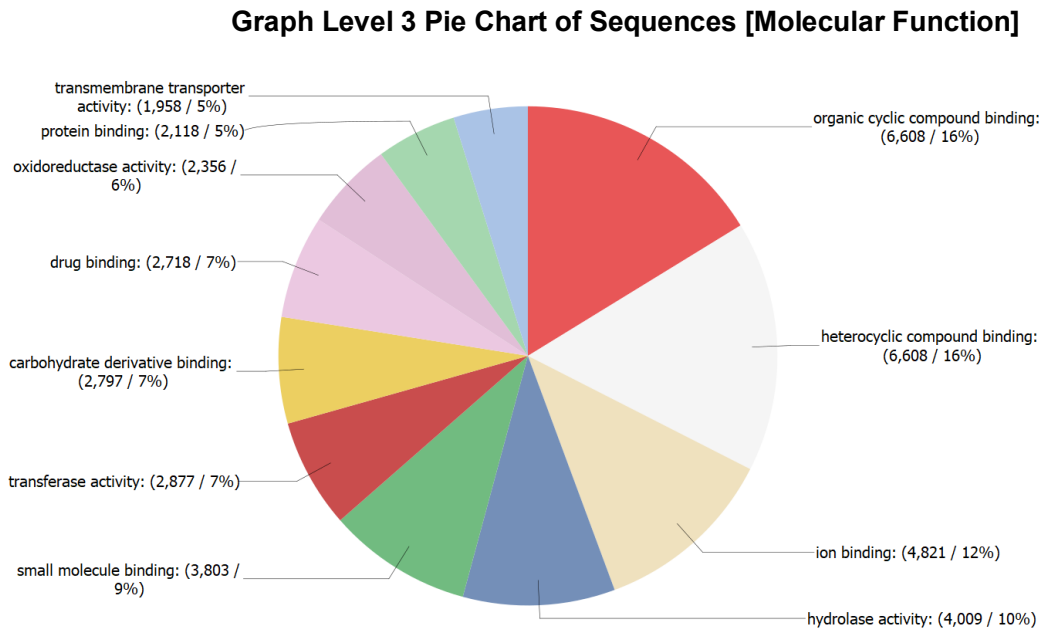


Figure 7.7. Molecular Function GO term distribution of *Steinernema* sp. HBG28. The pie chart represents the various molecular functions in which protein-coding genes are involved.

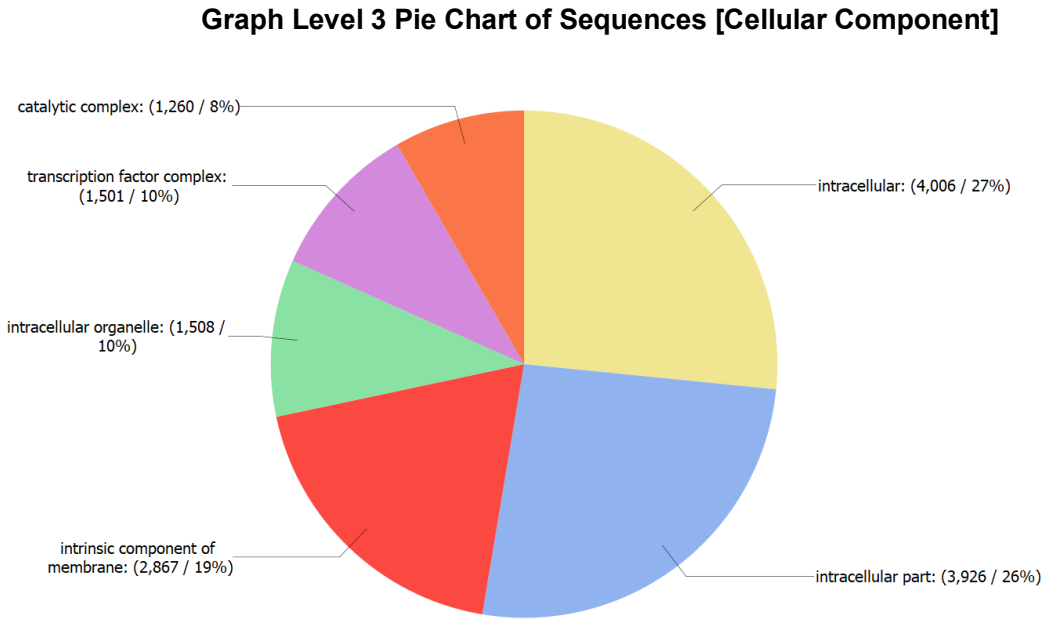


Figure 7.8. Cellular Component GO term distribution of *Steinernema* sp. HBG28. The pie chart represents various cellular component processes several protein-coding genes involved in individual cellular component processes with their distribution ratio also indicated.

7.4.3.2 KEGG classification

For KEGG classification of *Steinernema* sp. HBG28, KEGG orthology (KO) identifiers were assigned to the 17,400 annotated protein-encoding genes (48.1%) using the KEGG database. Of which, 10,334 KO assignments of gene products in biochemical pathways are displayed in Figure 7.9. It was found that KEGG pathways associated with metabolism had the highest representation, with a larger number of *Steinernema* sp. HBG28 sequences associated with 'carbohydrate metabolism' (603) and 'amino acid metabolism' (505) pathways. In the environmental information processing category, 'signal transduction' was, by far, the highest represented with 1,230 assigned genes alone. The transforming growth factor beta (TGF-β) signalling pathway in *Steinernema* sp. HBG28 included 27 gene products (i.e. 27 KO gene identifiers), which were represented by 20 protein-coding gene sequences (Figure 7.10). Additionally, there were 36 gene products (i.e. 36 KO gene identifiers) implicated in the longevity regulating pathway, also under organismal systems, which were represented by 32 protein-coding gene sequences (Figure 7.11). Both these signalling pathways have been founded to play crucial roles in in dauer development, stress resistance, lifespan and innate immunity in *Caenorhabditis elegans.*

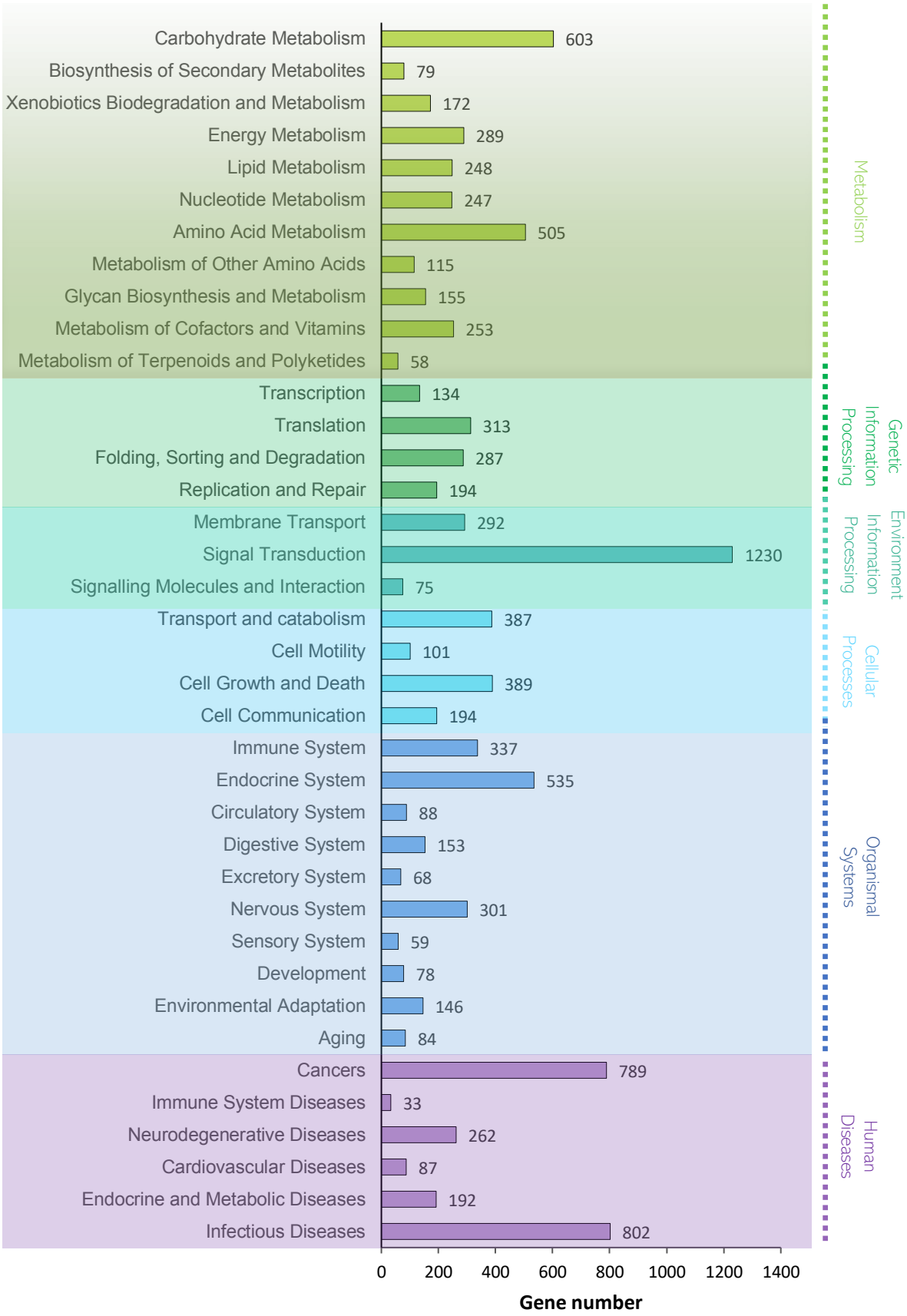


Figure 7.9. KEGG functional classification (A-Z) of *Steinernema* sp. HBG28 proteins.

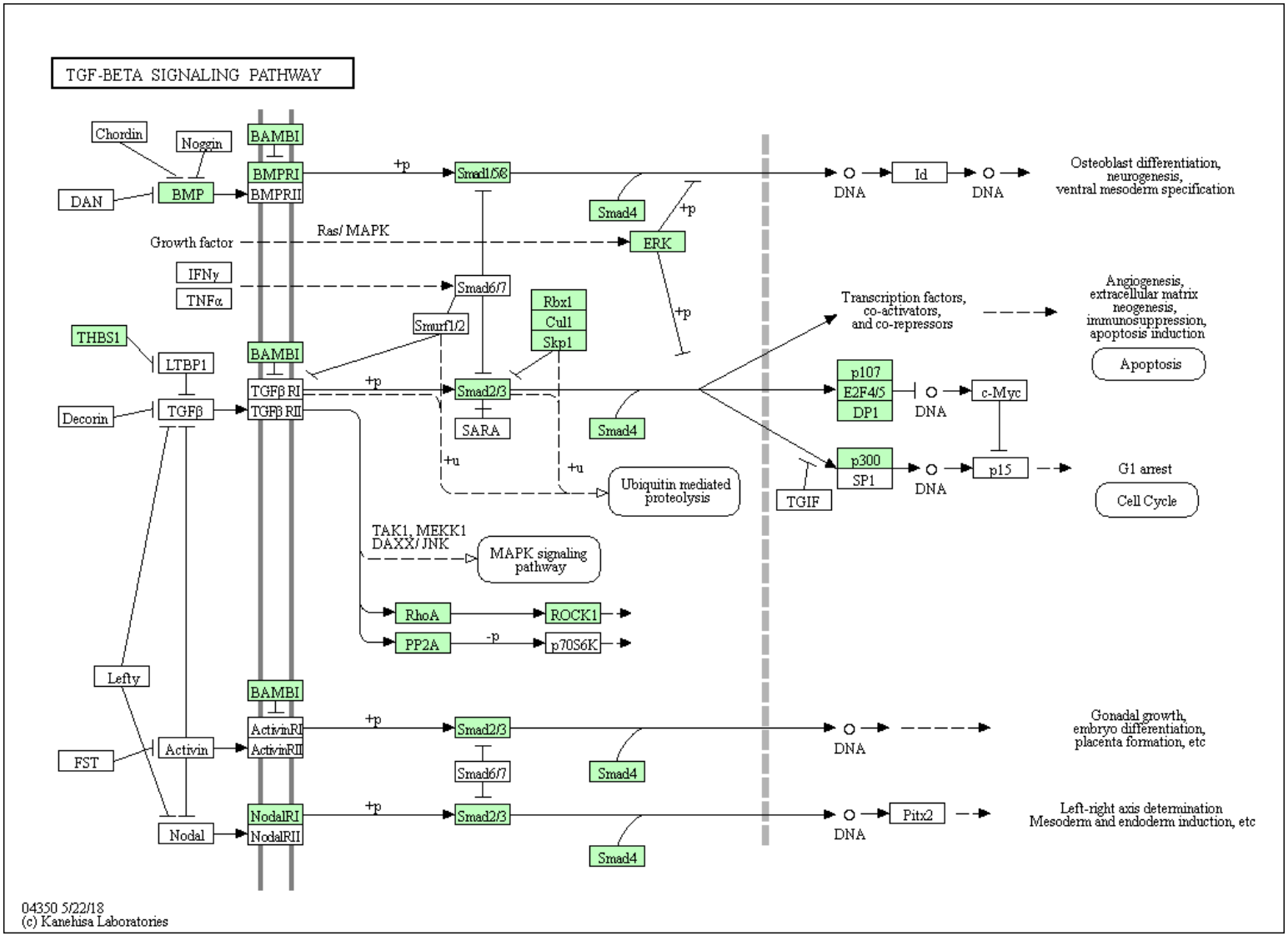


Figure 7.10. TGF-β pathway with gene products found in the annotation of *Steinernema* sp. HBG28, shown in green.

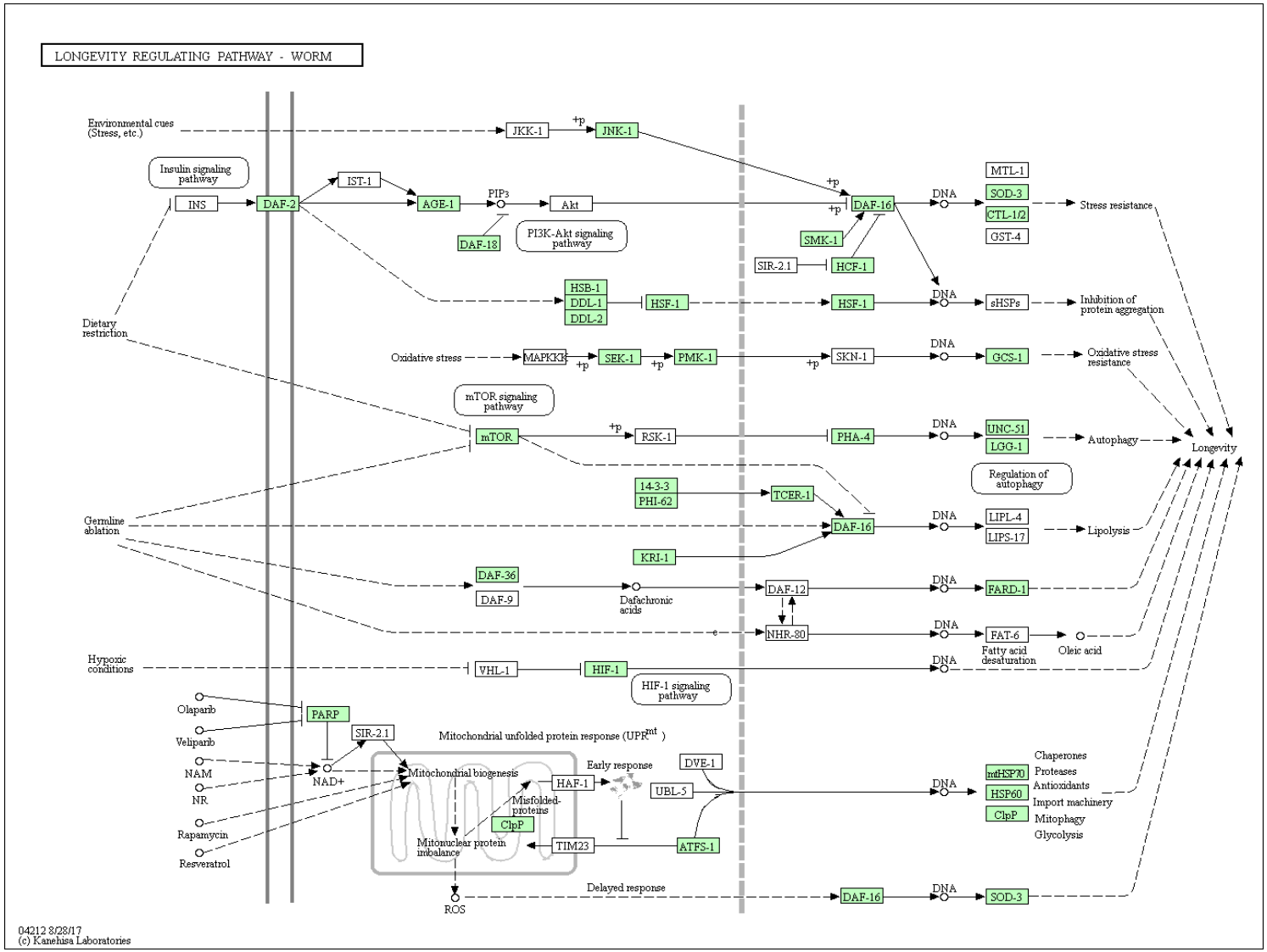


Figure 7.11. Longevity regulating pathway with gene products found in the annotation of *Steinernema* sp. HBG28, shown in green.

7.4.3.3 KOG functional classification

The functional classification of *Steinernema* sp. HBG28 proteins was predicted by performing KOG analysis. Of the 35,869 predicted protein-encoding genes annotated, 5,384 (15%) were assigned to twenty-five KOG categories (Figure 7.12). Among the assigned KOGs, general function prediction was the most frequent category with 687 genes, followed by the categories, unknown function and signal transduction mechanisms with 651 and 539 genes, respectively. Cell motility and nuclear structures were the least frequent categories, consisting of 10 and 33 genes, respectively.

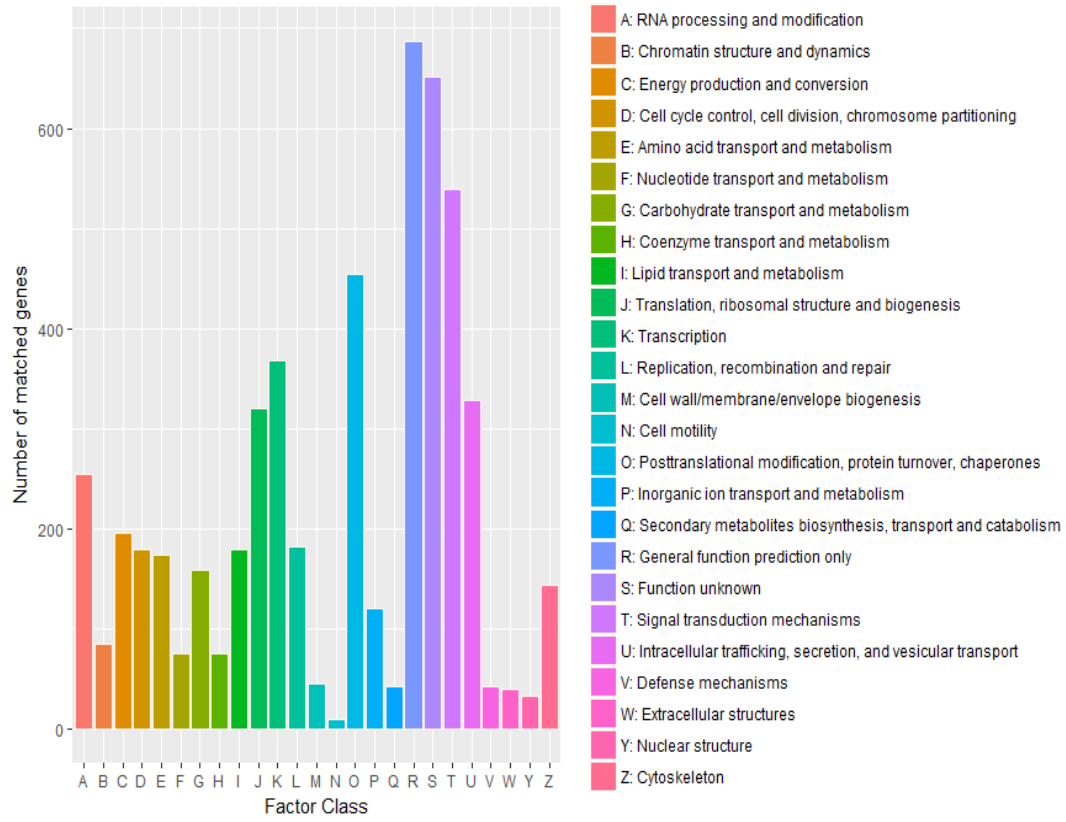


Figure 7.12. KOG functional classification (A-Z) of Steinernema sp. HBG28 proteins. KOG functional categories descriptions are: (A) RNA processing and modification; (B) Chromatin structure and dynamics; (C) Energy production and conversion; (D) Cell cycle control, cell division and chromosome partitioning; (E) Amino acid metabolism and transport; (F) Nucleotide transport and metabolism; (G) Carbohydrate transport and metabolism; (H) Coenzyme transport and metabolism; (I) Lipid transport and metabolism; (J) Translation, ribosomal structure and biogenesis; (K) Transcription; (L) Replication, recombination and repair; (M) Cell wall/membrane/envelope biogenesis; (N) Cell motility; (O) Posttranslational modification, protein turnover, chaperones; (P) Inorganic ion transport and metabolism; (Q) Secondary metabolites biosynthesis, transport and catabolism; (R) General function prediction only; (S) Function unknown; (T) Signal transduction mechanisms; (U) Intracellular trafficking, secretion and vesicular transport; (V) Defence mechanisms; (W) Extracellular structures; (Y) Nuclear structure; (Z) Cytoskeleton.

7.4.3.4 Protein domain abundance

Protein domains were identified through HMMER scanning against the Pfam database. An analysis of protein domain abundance of various protein families showed similarities and prominent differences between the entomopathogenic nematodes *Steinernema* sp. HBG28 and *Heterorhabditis bacteriophora*, as indicated in Figure 7.13. The *Steinernema* sp. HBG28 genome revealed a great abundance of domains associated with trypsin, trypsin inhibitor, nematode fatty acid retinol binding protein, Shk-domain like proteins, among others when compared to the genome of *H. bacteriophora.* By contrast, the *H. bacteriophora* genome was abundant in reverse transcriptase, the Sri family G-protein-coupled receptors (GPCRs), winged helix-turn helix, putative peptidase (DUF1758), Pao retrotransposon peptidase and putative peptidase (DUF1758) domains, among others (Figure 7.13).

Protein domains of particular interest are shown in Table 7.3. A striking difference between the genome of *Steinernema* sp. HBG28 relative to that of *H. bacteriophora* was the high abundance of trypsin inhibitor and trypsin domains (Figure 7.13 and Table 7.3). This observation in *Steinernema* sp. HBG28 was hardly surprising, given the utility of proteases and protease inhibitors by *Steinernema* nematodes in insect immune suppression and host tissue degradation (Allen and MacDonald, 1998), however it was interesting that the *H. bacteriophora* genome had few trypsin inhibitor and trypsin domains.

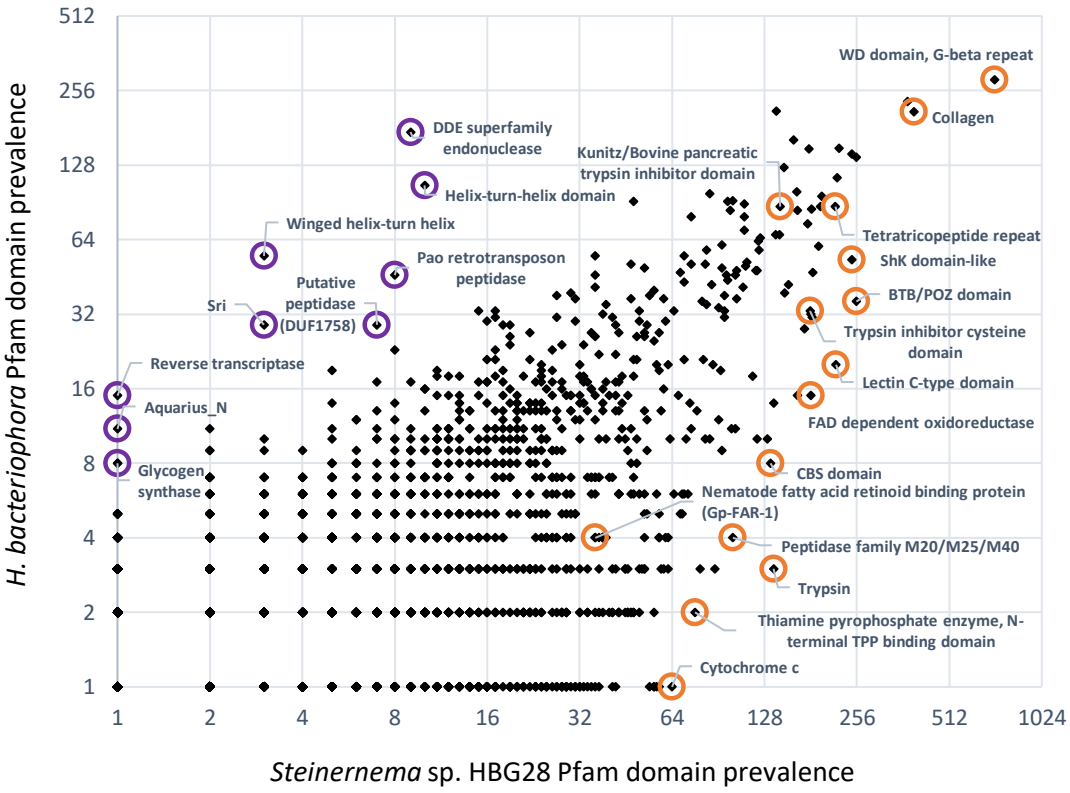


Figure 7.13. A scatterplot showing the comparison of Pfam protein domain prevalence between *H. bacteriophora* and *Steinernema* sp. HBG28. The occurrence of protein domains in equal proportions in both species appear on the diagonal axis, while those protein domains that are more abundant in *H. bacteriophora* and *Steinernema* sp. HBG28 are displayed in the upper left and lower right quadrants, respectively. The most divergent and abundant protein domains relative to *Steinernema* sp. HBG28 have been highlighted in purple, while those that are most abundant in *Steinernema* sp. HBG28 have been highlighted in orange.

Additionally, *Steinernema* sp. HBG28 also had comparable levels of the chymotrypsin family peptidase-S1 domain to that *S. carpocapsae* but was entirely absent in *H. bacteriophora.* The kunitz/bovine pancreatic trypsin inhibitor and EB module domains were also found to be more abundant in the *Steinernema* sp. HBG28 genome than in the *H. bacteriophora* genome (Table 7.3). It has been proposed that the kunitz/bovine pancreatic trypsin inhibitor and EB domain play a critical role in cuticle moulting in *C. elegans* (Page et al., 2006) and cuticle formation (Stepek et al.,

2010) in several nematode species. Additionally, the nematode fatty acid retinol binding protein domain was found to be more enriched in the *Steinernema* genomes compared to the *Heterorhabditis* genome (Table 7.3). Similarly, the Shk-domain-like proteins were also found in high abundance in *Steinernema* sp. HBG28 and was at least four-fold more abundant than in the *H. bacteriophora* genome. These protein domains, having been implicated in nematode pathogenicity, appear to be conserved among various parasitic nematodes (Lu et al., 2017).

Table 7.3. Protein domains of interest in *Steinernema* sp. HBG28

Protein domain families of interest	*Steinernema* sp. HBG28	*S. carpocapsae* strain All	*H. bacteriophora*
Trypsin	137	135	3
Trypsin Inhibitor like cysteine domain	180	266	33
Chymotrypsin family peptidase-S1	3	1	0
Kunitz/bovine pancreatic trypsin inhibitor domain	144	209	87
EB module	163	261	100
Nematode fatty acid retinoid binding protein (Gp-FAR-1)	36	62	4
Shk domain-like	247	311	53

7.4.3.5 G-protein-coupled receptors

Several olfactory receptors are known to be GPCRs and facilitate the initial step in the transduction of chemical stimuli (Skoufos et al., 2000). The GPCR protein domains were of particular interest in this study due to their obvious involvement in host-seeking behaviour in entomopathogenic nematodes. The prevalence of GPCR families in the genome of *Steinernema* sp. HBG28 were identified through the Pfam database and compared to the entomopathogenic nematodes, *S. carpocapsae*, *H. bacteriophora* and *S. glaseri* that adopt various foraging strategies. It should be noted that over the

course of this study, *Steinernema* sp. HBG28 was observed to exhibit intermediate-type foraging behaviour upon isolation and culturing, utilizing both ambusher and cruiser foraging approaches, however this behaviour was not been fully investigated. By contrast, *S. carpocapsae* has been shown to employ a foraging strategy distinctive to an ambusher, while *H. bacteriophora* and *S. glaseri* both adopt approaches that mimic cruiser-type foraging (Bal and Grewal, 2015; Shapiro-Ilan et al., 2012).

It was found that the serpentine GPCRs in the genomes of *Steinernema* sp. HBG28 (intermediate-type forager) and *S. carpocapsae* strain All (ambusher-type forager) (Dillman et al., 2015) were at least three-fold more abundant than that of *H. bacteriophora* (Bai et al., 2013) (cruiser-type forager). Remarkably, the genome of *S. glaseri* had the highest abundance of serpentine GPCR domains (1,300), which was at least two-times more abundant than *Steinernema* sp. HBG28 and *S. carpocapsae* strain All. This was not unexpected, given that *S. glaseri* is a recognized cruiser-type forager, while of *Steinernema* sp. HBG28 and *S. carpocapsae* utilize intermediate-type and ambusher-type foraging strategies, respectively. It was, however, interesting that *H. bacteriophora*, had only 278 serpentine domains, which was at least six-times less abundant than that of *S. glaseri*, given that they both utilize cruiser-type foraging. Additionally, the genomes of *S. carpocapsae* strain Breton (Rougon-Cardoso et al., 2016) and *H. bacteriophora* (McLean et al., 2018) were screened and found to only possess 13 and 63 serpentine family domains, respectively. This was interesting, given that the annotations of these genomes were recently improved upon.

When comparing the prevalence of GPCR serpentine protein domains in *Steinernema* sp. HBG28 to those of *S. carpocapsae* strain All , *H. bacteriophora* (Bai et al., 2013) and *S. glaseri*, the Srw family of serpentine GPCRs was found in comparable

levels in all four genomes, being most abundant in *Steinernema* sp. HBG28. Furthermore, the Srb family of serpentine GPCRs was only detected in the *Steinernema* sp. HBG28 genome and was absent in all entomopathogenic nematode genomes compared within this analysis.

When comparing the prevalence of specific families of serpentine GPCRs within the *Steinernema* genomes used in this analysis, based on foraging strategy alone, it would seem that *S. glaseri* requires a high abundance of serpentine domains (Srh, Srd, Str, Srx, Srsx, class ab, Srv, Srt) for cruiser-foraging, while *S. carpocapsae* requires the Srbc and Sre families of GPCRs as an ambush-type forager. Similarly, when comparing all four entomopathogenic nematode genomes, the Sra family of GPCRs occurred only in the genomes of *H. bacteriophora* and *S. glaseri* and appears to be cruiser-specific. Additionally, the genome of *H. bacteriophora* contained the highest abundance of Sri serpentine domains, followed by *S. glaseri* and *Steinernema* sp. HBG28; and the Srh serpentine domains was found to be more abundant in the intermediate and cruiser foragers when compared to the ambush forager, *S. carpocapsae*. This would suggest that the Sra serpentine domains are essential for host-seeking in *Heterorhabditis* and *Steinernema* nematodes where only cruiser-type foraging is exploited, while the Sri and Srh serpentine GPCRs are required for both cruiser- and intermediate-type foraging strategies.

Also included in this analysis were two vertebrate parasites, *N. americanus* and *D. viviparous* that belong to the order Strongylida. The Srw, Srsx, ab class, Srt families of serpentine GPCRs were found in comparable levels in the genomes of *H. bacteriophora, N. americanus* and *D. viviparus*, suggesting that these specific GPCRs may be highly conserved amongst these nematodes within the order Strongylida.

Table 7.4. Serpentine GPCR protein domains identified by Pfam. Percentages are included (in brackets).

All GPCR families identified by Pfam	*S.*HBG28	*S. carpo* All[1]	*H. bacterio*[2]	*H. bacterio* BRAKER1[3]	*S. carpo* Breton[4]	*S. glaseri*[1]	*N. america*[5]	*D. vivi*[6]
Serpentine type 7TM GPCR chemoreceptor Srh	74 (11)	59 (7)	79 (28)	29 (46)	0 (0)	248 (21)	4 (3)	0 (0)
Serpentine type 7TM GPCR chemoreceptor Srd	12 (2)	21 (3)	10 (4)	1 (2)	0 (0)	40 (3)	13 (9)	3 (2)
Serpentine type 7TM GPCR chemoreceptor Str	12 (2)	20 (3)	10 (4)	0 (0)	0 (0)	48 (4)	12 (8)	1 (1)
Serpentine type 7TM GPCR chemoreceptor Sri	3 (0)	0 (0)	29 (10)	6 (10)	0 (0)	7 (1)	1 (1)	0 (0)
Serpentine type 7TM GPCR chemoreceptor Srx	183 (28)	167 (21)	31 (11)	7 (11)	5 (38)	235 (20)	23 (15)	33 (22)
Serpentine type 7TM GPCR chemoreceptor Srw	26 (4)	22 (3)	23 (8)	13 (21)	1 (8)	25 (2)	18 (12)	17 (11)
Serpentine type 7TM GPCR chemoreceptor Srsx	81 (12)	135 (17)	21 (8)	2 (3)	1 (8)	172 (14)	22 (14)	22 (15)
Srg family chemoreceptor	17 (3)	12 (2)	7 (3)	0 (0)	0 (0)	25 (2)	2 (1)	0 (0)
Serpentine type 7TM GPCR chemoreceptor Srbc	15 (2)	22 (3)	1 (0)	0 (0)	0 (0)	16 (1)	4 (3)	13 (9)
Serpentine type 7TM GPCR chemoreceptor ab (Srab)	58 (9)	75 (10)	19 (7)	1 (2)	4 (31)	100 (8)	15 (10)	27 (18)
Sre G protein-coupled chemoreceptor	43 (6)	47 (6)	13 (5)	0 (0)	1 (8)	39 (3)	8 (5)	2 (1)

Serpentine type 7TM GPCR chemoreceptor Srv	20 (3)	15 (2)	14 (5)	1 (2)	1 (8)	25 (2)	9 (6)	11 (7)
Serpentine type 7TM GPCR chemoreceptor Srt	117 (18)	193 (24)	18 (6)	3 (5)	0 (0)	216 (18)	15 (10)	14 (9)
Serpentine type 7TM GPCR chemoreceptor Sra	0 (0)	0 (0)	3 (1)	0 (0)	0 (0)	1 (0)	6 (4)	4 (2)
Serpentine type 7TM GPCR chemoreceptor Srb	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1)
Serpentine receptor-like protein, class xa	1 (0)	1 (0)	0 (0)	0 (0)	0 (0)	3 (0)	0 (0)	0 (0)
Total Serpentine GPCR Domains	**663**	**789**	**278**	**63**	**13**	**1,200**	**152**	**148**

[1] *Steinernema carpocapsae* strain All (*S. carpo*) and *S. glaseri* sequenced by Dillman et al. (2015).

[2] *Heterorhabditis bacteriophora* (*H. bacterio*) sequenced by Bai et al. (2013).

[3] *Heterorhabditis bacteriophora* BRAKER1 (*H. bacterio*) sequenced byMcLean, et al. (2018).

[4] *Steinernema carpocapsae* strain Breton (SC) sequenced by Rougon-Cardoso et al. (2016).

[5] *Necator americanus* (*N. america*) sequenced by Tang et al. (2014).

[6] *Dictyocaulus viviparus* (*D. vivi*) sequenced by McNulty et al. (2016).

7.4.3.6 Proteases

When a batch BLASTP search was performed against the MEROPS protease database, a total of 864 proteases were identified in *Steinernema* sp. HBG28. The proteases were assigned into five key categories according to their catalytic type: aspartic (6.13%), cysteine (19.1%), metallo- (32.4%), serine (34.02%), and threonine (6.9%), as shown in Table 7.5. The total number of proteases in *Steinernema* sp. HBG28 (864) resembled the numbers previously reported for *S. monticolum* and *S. feltiae* at 881 and 831, respectively and these findings were consistent, given the close phylogenetic relationship shared among *Steinernema* sp. HBG28, *S. monticolum* and *S. feltiae.* Notably, the serine proteases and metalloproteases constituted the largest protease categories in *Steinernema* sp. HBG28, with 280 and 294 genes spread over 53 and 29 families, respectively. Among these, prolyl aminopeptidase (S33), chymotrypsin-like (S01A), prolyl oligopeptidase-like (S9), as well as isoaspartyl dipeptidase-like (M38) and astacin-like (M12A) represented the most abundant subdivisions/family types of serine proteases and metalloproteases. Based on the comparisons of proteases present in *Steinernema* spp. in Table 7.5, the expansion of metalloproteases and serine proteases in the *Steinernema* genus is quite evident, and confirms the findings made by Dillman et al., (2015).

Table 7.5 Protease and protease inhibitor abundance summary. Percentages are included (in brackets).

Protease type	*Steinernema* sp. HBG28	*S. carpocapsae*[1]	*S. scapterisci*[1]	*S. monticolum*[1]	*S. glaseri*[1]	*S. feltiae*[1]
Aspartic	53 (6)	41 (6)	66 (8)	67 (8)	39 (5)	40 (5)
Cysteine	165 (19)	124 (19)	147 (17)	122 (14)	116 (15)	147 (18)
Metallo	280 (32)	214 (33)	268 (31)	307 (35)	314 (42)	306 (37)
Serine	294 (34)	246 (38)	347 (40)	349 (40)	253 (33)	301 (36)
Threonine	60 (7)	27 (4)	36 (4)	32 (4)	29 (4)	34 (4)
Unassigned	12 (1)	7 (1)	7 (0)	4 (0)	5 (1)	4 (0)
Total	**864**	**654**	**871**	**881**	**756**	**831**
Inhibitors	168	146	175	153	133	214

[1] Summary of protease and protease inhibitors in *Steinernema* spp. by Dillman et al. (2015).

7.4.3.7 Gene orthology

The predicted protein sequences of *Steinernema* sp. HBG28 was used in the identification of single-copy orthologues by comparing the predicted proteomes of four other entomopathogenic nematode species of the same genus: *S. carpocapsae*, *S. feltiae S. glaseri* and *S. monticolum* (Dillman et al., 2015) using OrthoMCL (Li et al. 2003). Comparative genome analysis of orthologous genes was performed with OrthoVenn (Wang et al., 2015), where a Venn diagram was plotted using 127,496 sequences from all four nematode species, which were grouped to display overlapping orthologous protein clusters (Figure 7.14).

A total of 28,775 proteins in the *Steinernema* sp. HBG28 proteome were assigned into 15,919 orthologous gene clusters. Among the orthologous clusters, 2,648

contained 9,570 genes that were specific to *Steinernema* sp. HBG28 and 8,382 clusters were common to all five species, indicative of their conservation within these steinernematids. There were 18,041 orthologous clusters that contained at least two species with 3,633 single-copy gene clusters containing one gene for each of the five *Steinernema* species. Additionally, 109 clusters were confined to the entomopathogenic nematodes, *Steinernema* sp. HBG28, *S. carpocapsae* and *S. feltiae*, 273 gene clusters were confined to *Steinernema* sp. HBG28, *S. monticolum* and *S. feltiae*; and 80 gene clusters were confined to *Steinernema* sp. HBG28, *S. carpocapsae* and *S. glaseri*.

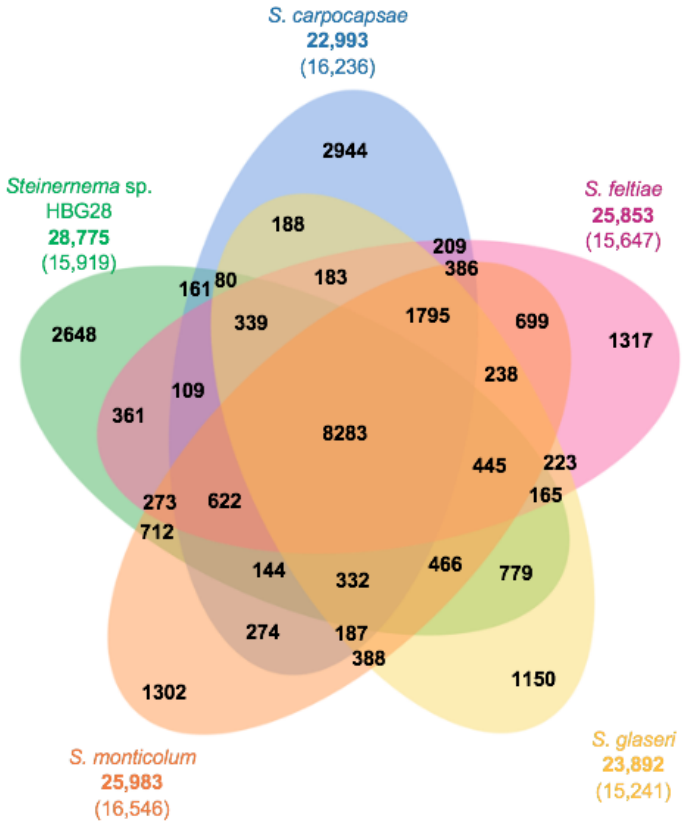


Figure 7.14. Venn diagram plotted by OrthoVenn indicating the distribution of protein families (with orthologous protein clusters indicated in parenthesis) among the predicted proteomes of five *Steinernema* species, *Steinernema* sp. HBG28, *S. carpocapsae*, *S. monticolum*, *S. feltiae* and *S. glaseri* (Dillman et al., 2015).

7.4.3.8 Phylogenomic relationships of *Steinernema* spp.

The relationships among *Steinernema* sp. HBG28 and five other *Steinernema* spp. (*S. carpocapsae*, *S. feltiae S. glaseri*, *S. monticolum* and *S. scapterisci*) were evaluated using a supermatrix of single-copy orthologues (n=3,308) as identified by OrthoMCL. A maximum likelihood phylogeny tree was constructed with a bootstrap support of 100 (Figure 7.15). The relationships recovered in the phylogenomic analysis strongly supported the phylogenetic relationships observed among the selected *Steinernema* spp. based on ITS regions of partial 18S rDNA sequences.

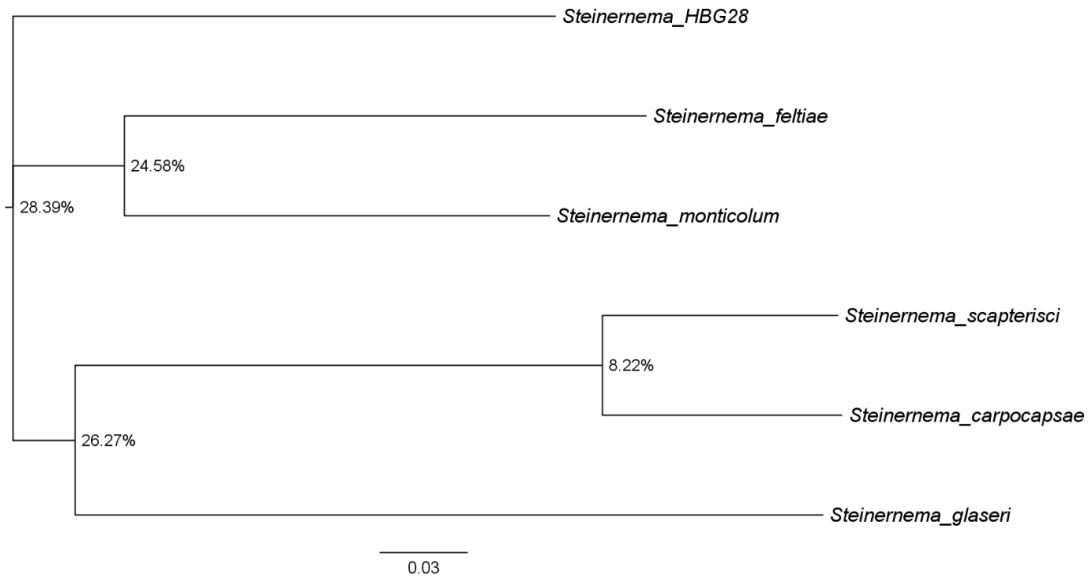


Figure 7.15. Maximum likelihood tree built using IQ-Tree from the alignment of 3,308 strict one-to-one protein orthologues from *Steinernema* genomes under the JTT + F + I + G4 amino acid substitution model. The analysis was performed with bootstrap replicates of 100 and included 1,091,649 aligned, trimmed and concatenated amino acid residues. An outgroup taxon was not included in this analysis.

7.5 Discussion

This chapter introduces the genome annotation of *Steinernema* sp. HBG28, a species indigenous in the grasslands of South Africa, which is among the first *Steinernema* species to be sequenced and annotated in South Africa.

The aim of this chapter was to annotate the genome of *Steinernema* sp. HBG28 and, in doing so, explore the genomic architecture of this entomopathogenic nematode and making comparisons, where necessary, to previously published entomopathogenic nematode genomes. The genome sequence of *Steinernema* sp. HBG28 revealed several gene products putatively implicated in nematode pathogenicity and survival that could be exploited for possible field application in controlling insect pests indigenous to the grassland biome of South Africa.

7.5.1 Genomic structural annotation

7.5.1.1 Gene model prediction metrics

In this study, automated genome annotation was carried out with the MAKER2 pipeline using evidence in the form of the transcriptome assembly of *Steinernema* sp. HBG28, the Swiss-Prot database and protein homology data of previously sequenced *Steinernema* genomes by Dillman et al. (2015) and *C. elegans* from Wormbase (https://wormbase.org/.

When comparing the gene model statistics of *Steinernema* sp. HBG28 to those of previously annotated *Steinernema* genomes by Dillman et al. (2015), features such as the predicted gene number, exon number, intron number and mean gene length were mostly comparable (Table 7.2). This, however, was not the case when similar

comparisons were made with *S. carpocapsae* strain Breton by Rougon-Cardoso et al. (2016). In fact, the most notable difference among all the *Steinernema* genomes compared was the total number of predicted genes. While the number of genes predicted in *Steinernema* sp. HBG28 was comparable to those of previously annotated genomes of the same genus by Dillman et al. (2015), it was significantly lower than the 16,333 predicted genes reported for *S. carpocapsae* strain Breton by Rougon-Cardoso et al. (2016), as shown in Table 7.2.

When only considering the genome of *S. carpocapsae* strain All by Dillman et al. (2015) and that of *S. carpocapsae* strain Breton by Rougon-Cardoso et al. (2016), there is a striking difference in that the number of predicted genes reported was considerably lower for *S. carpocapsae* strain Breton. As indicated in the study by Rougon-Cardoso et al. (2016), steps which could have been taken to reduce the heterozygosity in *S. carpocapsae* strain All genome were not reported on by Dillman et al. (2015); and if such steps were not taken, this may have affected the genome assembly. Furthermore, only one *ab initio* gene finder was utilized for gene prediction of the *S. carpocapsae* strain All genome with *C. elegans* selected as the model species (Dillman et al., 2015). Given that *Caenorhabditis* diverged later than *Steinernema* (Douzery et al., 2004), the use of *C. elegans* gene models may have resulted in an over-estimation rather than an under-estimation of predicted genes within *S. carpocapsae*. By comparison, the study performed by Rougon-Cardoso et al. (2016) made use of a near isogenic line for *S. carpocapsae* strain Breton, various sequencing platforms and an annotation approach that included HMM training on *S. carpocapsae* gene structures, HMM homology-based predictions using *C. elegans* and *Brugia malayi* gene models and *ab initio* gene predictions.

In the present study, the high number of predicted genes in *Steinernema* sp. HBG28 may have been a consequence of using proteomes derived from the *Steinernema* spp. sequenced by Dillman et al. (2015) as protein evidence in the MAKER2 pipeline. As this pipeline is evidence-driven, the aforementioned protein sequences have influenced results of the genome annotation for *Steinernema* sp. HBG28. For future work, it would be interesting to compare the genome annotation of *Steinernema* sp. HBG28 in this study with that using the *S. carpocapsae* Breton proteome as homology-based protein evidence in the MAKER2 pipeline.

7.5.1.2 Repeat detection

Repeat detection identified 5.07% of the *Steinernema* sp. HBG28 draft genome as repetitive and was comparable to the repeat content reported for other *Steinernema* nematode genomes (2.75% in *S. scapterisci*, 5.34% in *S. glaseri*, 6.70% in *S. feltiae*, 7.46% and 6.99% in *S. carpocapsae* and 10.49% in *S. monticolum*) (Dillman et al., 2015; Rougon-Cardoso et al., 2016). Additionally, *Steinernema* sp. HBG28 consisted of 11,475 simple sequence repeats (SSRs), which was significantly higher than the 3,794 SSRs found in *H. bacteriophora* (Bai et al., 2013). It has been found that these repeat sequences are prone to a high mutation rate and as a result can serve as genetic markers in resolving relationships among closely-related species (Bowcock et al., 1994).

7.5.1.3 RNA annotation

A total of 1,194 predicted tRNAs, 75 tRNA pseudogenes and 43 rRNAs were identified in the genome of *Steinernema* sp. HBG28. Rougon-Cardoso et al. (2016) reported similar estimates of 1,097 tRNAs and 40 rRNAs in the *S. carpocapsae* genome. By comparison, the genome of *H. bacteriophora* consisted of a total of 254 tRNA genes

and 1 tRNA pseudogene (Bai et al., 2013), which is dramatically lower than the tRNA genes and tRNA pseudogenes predicted in *Steinernema* sp. HBG28 and *S. carpocapsae*. In addition, a total of 14 micro-RNA hairpins were detected in *Steinernema* sp. HBG28, which was less than that reported by Rougon-Cardoso et al. (2016).

7.5.2 Genomic functional annotation

7.5.2.1 KEGG pathway and KOG classification

The KEGG pathways associated with environmental information processing had the highest representation for signal transduction in *Steinernema* sp. HBG28, with 1,230 assigned genes alone. Analysis of the functional categories and the distribution of protein-coding genes with assigned KOG functions indicated that while the most frequent KOG categories in the *Steinernema* sp. HBG28 genome were genes in general prediction or unknown functions, a significant number of genes were revealed to have direct impact on signal transduction.

Steinernema sp. HBG28 was found to possess 32 genes associated with the longevity regulating pathway – which can be linked to stress resistance against various environmental cues and increased longevity of nematodes in a soil environment. The gene, *INS* within the insulin signalling pathway has been implicated in regulating chemotaxis and thermotaxis learning in *C. elegans* (Tomioka et al., 2006; Kodama et al., 2006), whilst the genes, *daf-2*, *age-1*, *daf-18* and *daf-16* found in the longevity regulating pathway (Figure 7.11) have been shown to also contribute to thermotaxis, chemotaxis, odorant association and sensory integration learning (Murakami et al., 2005; Lin et al., 2010; Kauffman et al., 2010; Byrne et al., 2014). As the dauer-stage of *C. elegans* nematodes is comparable to steinernematid and

heterorhabditid infective juveniles, it is believed that the genes and gene products of these pathways perform a similar function in entomopathogenic nematodes.

The mechanisms involved in signalling and communication govern the way in which nematodes perceive their external environment, triggering adaptive changes that result in alterations in gene expression. It is hardly surprisingly that *Steinernema* sp. HBG28 has a large repertoire of genes that catered towards signal transduction, given that infective juveniles are inherently found in complex soil environments, which they must effectively navigate and interpret in order to seek out insect hosts.

Interestingly, dauer development is another example of how signal transduction can influence nematode parasitism, more so as it is a key feature for infective larval development prior to and during host infection (Crook, 2014, Hotez et al., 1993). While some parasitic nematode species, such as *Parastrongyloides trichosuri* (Grant et al., 2006) and *Strongyloides* spp. (Yamada et al., 1991; Viney, 1996) can circumvent the arrested dauer stage entirely and develop directly into adults, infective larvae of parasitic and entomopathogenic nematodes must be able to precisely control their activation upon entry and exit of a host (Crook, 2014). The morphological and behavioural similarities that exist between third-stage dauer larvae (L3s) of free-living nematodes and arrested third-stage infective larvae (iL3) of parasitic nematodes can be considered as a fundamental pre-adaptation to parasitism (Rogers and Sommerville, 1963; Hawdon and Schad, 1991; Hotez et al., 1993). The transition between dauer and adult development in *C. elegans* nematodes, for example, is regulated by the GPCR, insulin/IGF- and TGFβ-like signalling pathways (Butcher et al., 2007; Kimura et al., 1997; Ogg et al., 1997). Similarly, the functions of the GPCR and insulin/IGF- signalling pathways in parasitic nematodes, which are broadly

conserved, serve to also regulate larval development (Hunt et al., 2016; Stoltzfus et al., 2012a; Stoltzfus et al., 2012b; Tissenbaum et al., 2000). The TGFβ-like signalling pathway, on the other hand, appears to be fundamentally different in parasitic nematodes, having been modified for functions other than dauer regulation (Crook, 2014). It has been suggested that one such function may include the modulation of the host immune response (Gomez-Escobar, 2000, McSorley et al., 2010).

Evidently, signal transduction plays an important role in the parasitic lifestyle that *Steinernema* sp. HBG28 has adopted, particularly in the location of potential hosts, host infection and dauer formation – all of which are mediated by the recognition of various host-emitted sensory cues. The gene products found in these pathways could serve as possible candidates for genetic enhancement for increased field efficacy through the improvement of stress resistance and longevity of infective juvenile nematodes for pest control (Lu et al., 2016).

7.5.2.2 Protein domain abundance

In a genomic analysis, whereby the Pfam database was used in assigning protein domain families to sequences of the *Steinernema* sp. HBG28 proteome, proteins involved in parasitism were considered. The genome of *Steinernema* sp. HBG28 was found to be enriched in a variety of proteases and protease inhibitors domains, which was not surprising given the reliance of a variety of proteases and protease inhibitors by entomopathogenic nematode in host invasion and immune suppression and/or evasion. In particular, trypsin and trypsin inhibitor domains were found in great abundance in the genome of *Steinernema* sp. HBG28. Proteases (peptidase) and protease (peptidase) inhibitors are involved in various biological processes, ranging from development, immune suppression and tissue digestion by entomopathogenic

nematodes (Cooper and Eleftherianos, 2016; Kanost and Clarke, 2005). These molecules also facilitate nematode invasion of the insect host and the subsequent breakdown of host tissues (Brivio et al., 2005). Highly virulent strains of entomopathogenic nematodes have been shown to produce excreted or secreted products (ESP) containing a greater number of proteases when compared to other strains that are less virulent (Simões et al., 2000).

The detection of the chymotrypsin family peptidase-S1 (Sc-CHYM) domain in the *Steinernema* sp. HBG28 genome was of particular relevance in the present study. This trypsin-like serine protease serves as a virulence factor and has been shown to interfere with the activation of the prophenoloxidase pathway in insect hosts by targeting immune recognition proteins (Balasubramanian et al., 2009). As a result, the melanization response is hampered and the invading nematode entirely avoids encapsulation by haemocytes of the insect. In a previous study by Balasubramanian et al. (2009), infective juveniles of *S. carpocapsae* were shown to have increased expression levels of the chymotrypsin-like protease upon infection of *Galleria mellonella* larvae, thereby resulting in the suppression of prophenoloxidase and essentially affecting haemocyte recognition and aggregation. In the same study, infective juveniles of *H. bacteriophora* were promptly recognized by haemocytes of *G. mellonella*, suggesting a lack of the chymotrypsin-like protease. Notably, the observations by Balasubramanian et al. (2009) corroborate the findings made in the present study whereby the chymotrypsin family peptidase-S1 protein domain was absent in the *H. bacteriophora* genome while being detected in the *Steinernema* sp. HBG28 genome. This could also account for the lack of black melanized spots presented on the cuticle of insect hosts during *Heterorhabditis* nematode infections (ffrench-Constant et al., 2003).

Another protein domain identified as the kunitz/bovine pancreatic trypsin inhibitor (Figure 7.13 and Table 7.3) was also detected in *Steinernema* sp. HBG28. *S. carpocapsae* and *Steinernema* sp. HBG28 were found to contain 209 and 144 protein domains of this particular inhibitor, respectively compared to 87 in *H. bacteriophora*. The kunitz/bovine pancreatic trypsin inhibitor domain, which typically associates with a nematode-specific EB domain, has been linked to maintaining cuticle structure in a wide range of nematodes (Stepek et al., 2010). This trypsin inhibitor has also been implicated in suppressing haemocyte aggregation and even preventing encapsulation reactions of foreign bodies by targeting recognition proteins of the insect immune system, however it fails to prevent the initial activation of prophenoloxidase in insect hosts (Cooper and Eleftherianos, 2016).

The nematode fatty acid- and retinol-binding (FAR) protein domain was also detected in *Steinernema* sp. HBG28 and was at least two-fold less than that found in *S. carpocapsae* and nine-fold more abundant when compared to *H. bacteriophora*. It would seem that FAR proteins are expanded in the *Steinernema* genomes, as Dillman et al. (2015) has elucidated and it is proposed that these proteins are paramount in a wide variety of parasitic lifestyles exhibited by nematodes, such as those involving vertebrates, insects and plants (Kennedy et al., 2013; Hao et al., 2010; Iberkleid et al., 2013). It is thought that FAR proteins work by sequestering host-derived retinoids and influence immune evasion or suppression (Garofalo et al., 2002; Kennedy et al., 2013), albeit the exact mechanistic function is not fully understood. Dillman et al. (2015) has also suggested that FAR proteins in entomopathogenic nematodes, particularly in the *Steinernema* genus may interact with eicosanoids – fatty acids, thereby interfering with immune signalling in insects (Campos et al., 2014; Lawrence

et al., 2002; Stanley and Miller, 2006). Interference with eicosanoid production, which has been considered to be the main insect defence against parasitic nematodes has been linked to a reduction in the encapsulation response of the insect host (Castillo et al., 2011; Carton et al., 2002).

Among the abundance of proteases and protease inhibitors, a toxin-related protein domain (i.e. the Shk-domain-like proteins) was especially abundant in *Steinernema* sp. HBG28; and was also found to be significantly reduced in the genome of *H. bacteriophora.* In a study by Lu et al. (2017), it was shown that axenic *S. carpocapsae* infective juveniles were lethal, inducing insect death through the release of various venom proteins, including Shk domain-containing proteins. It was further suggested that this protein is involved in immune suppression and appears to be conserved among *Steinernema* entomopathogenic nematodes, as well as other vertebrate-parasitic nematodes (Lu et al., 2017).

The general trend observed in the protein domain analysis was the sheer abundance of proteases, protease inhibitors and venom-like proteins in *Steinernema* sp. HBG28 relative to the genome of *H. bacteriophora.* This study served to highlight the significant difference between *Steinernema* and *Heterorhabditis* nematodes in terms of their contributions towards nematode virulence and the extent to which these nematodes rely on their symbiotic bacteria in inducing insect death. This genomic analysis may also reflect that *H. bacteriophora* and perhaps other *Heterorhabditis* species are largely dependent on their *Photorhabdus* bacteria for the inhibition of insect immunity, host tissue metabolism and host-killing. In contrast, *Steinernema* nematodes appear to have a more dynamic role in immune suppression, evasion and pathogenicity beyond merely serving as vectors for their *Xenorhabdus*

counterparts. Steinernematids can be lethal to the insects they infect, even without the support of their associated bacteria (Han and Ehlers, 2000; Lu et al., 2017).

7.5.2.3 Prevalence of G-protein-coupled receptors

Olfactory receptors in nematodes play an important role in chemoreception, many of which are also known to be GPCRs (Skoufos et al., 2000). The domain prevalence of GPCRs families was of significant relevance due to their likely contribution in foraging and host-seeking behaviour by entomopathogenic nematodes (Thomas and Robertson, 2008). As such, the genome of *Steinernema* sp. HBG28 was screened for GPCRs and was found to be enriched in the serpentine GPCR domains. Of the entomopathogenic nematodes that were analysed, it was found that *S. glaseri*, a cruiser-type forager, possessed a higher abundance of serpentine GPCR domains than *Steinernema* sp. HB28 and *S. carpocapsae*, which detect insect hosts by utilizing intermediate-type and ambusher-type foraging strategies, respectively.

In a study performed by Srinivasan et al. (2013) where the presence of serpentine protein domains was compared among various nematodes, it was found that the free-living nematodes, *Panagrellus redivivus* and *C. elegans* possessed an greater abundance of serpentine protein domains than the parasitic nematodes, *Ascaris suum*, *Brugia malayi*, and *Trichinella spiralis*. It was suggested that free-living nematodes require a larger repertoire of GPCRs to perceive the complex environments in which they occupy in order to elicit appropriate responses to various stimuli, safeguard against predation and locate food (Robertson and Thomas, 2006), whereas parasitic nematodes may require far less GPCR domains as they inhabit more specialized environments within their hosts (Srinivasan et al., 2013).

When screening the genomes of *H. bacteriophora* (Bai et al., 2013 and McLean et al., 2018), it was determined that while *H. bacteriophora* possessed a variety of serpentine family domains, it was far less abundant when compared to its *Steinernema* counterparts (Table 7.4). This was interesting, given that *Heterorhabditis* infective juveniles like those of *Steinernema*, can occupy complex soil environments for long periods of time before encountering a suitable host, and as such, would likely require a larger complement of GPCRs to enable them to navigate and respond appropriately to signals emanating from their soil environment. Furthermore, it was expected that as *H. bacteriophora* exclusively adopts a cruiser-type foraging strategy to seek out its hosts in the soil, its genome was likely to contain a higher abundance of serpentine GPCRs, especially in relation to an ambush-type forager (*S. carpocapsae*) and intermediate-type forager (*Steinernema* sp. HBG28). For additional comparison, the genomes of *N. americanus* and *D. viviparous* were screened for serpentine GPCR domains as these obligate vertebrate parasites are representative of Strongyloidea, to which *Heterorhabditis* nematodes share most recent common ancestry (Bai et al., 2013). Notably, while the genomes of *N. americanus* and *D. viviparous* contained fewer serpentine GPCR domains than those observed in the *H. bacteriophora* genome, it was interesting that all three genomes had relatively comparable levels of GPCR domains, particularly with the Srw, Srsx, ab class and Srt families of serpentine GPCRs. This may be taken to indicate some degree of functional conservation with respect to these specific serpentine domains among nematodes in the order Strongylida.

This study does raise an important question about the prevalence of serpentine GPCR domains within other *Heterorhabditis* genomes, particularly in relation to foraging and host seeking behaviour. At present, the knowledge surrounding this topic is very sparse, and until more genomes within the genus are sequenced, our

understanding remains limited for the time being. Gaining further insight into the utilization of GPCRs in host-seeking strategies by entomopathogenic nematodes could improve their effectiveness as biological control agents. GPCRs could also serve as gene targets for genetic manipulation by enhancing the host-seeking capabilities of entomopathogenic nematodes under field conditions.

7.5.2.4 Proteases in *Steinernema* sp. HBG28

The secreted proteases of entomopathogenic nematodes are crucial for host invasion (Abuhatab et al., 1995), the breakdown of host tissues (McKerrow et al., 2006) and suppression of insect immunity (Balasubramanian et al., 2009). As *Steinernema* nematodes can be fatal to insect host targets, even in the absence of their associated bacteria (Lu et al., 2017), they are prone to high proteolytic activity. In the present study, a total of 864 proteases were detected in *Steinernema* sp. HBG with serine proteases and metalloproteases constituting a large part of these proteases with implications of their roles in nematode virulence. Dillman et al. (2015) have demonstrated that steinernematids have a greater abundance of proteases than any other nematode genomes reported. The number of proteases identified in *Steinernema* sp. HBG28 were relatively consistent with these findings. It is thought that the high volume of proteases observed in the *Steinernema* genus could account for the wide host range observed among these nematodes.

7.5.2.5 Phylogenomic analysis of *Steinernema* sp. HBG28

The relationships among the six *Steinernema* spp. were evaluated whereby the predicted proteomes of these taxa were included in a phylogenomic analysis. *Steinernema* sp. HBG28 was shown to have orthologous genes in common with other

Steinernema nematodes that have been sequenced. Using a supermatrix of 3,308 strictly orthologous genes, the phylogenomic analysis presented in this study was consistent with the findings of the taxonomic placement of *Steinernema* sp. HBG28 in relation to the other *Steinernema* species sample derived from the phylogenetic analysis involving the ITS regions of the conserved 18S rRNA gene (see Chapter 2). The relationships recovered confirmed the relatedness of *S. monticolum* and *S. feltiae* to *Steinernema* sp. HBG28.

7.6 Conclusion

The genome annotation and subsequent analyses of *Steinernema* sp. HBG28, a *Steinernema* species indigenous to the grasslands of South Africa, confirmed a role in the virulence of insect hosts beyond serving as merely a vector and offering a protective environment for its associated bacterial symbiont, *X. khoisanae.* The identification of a variety of proteases, protease inhibitors, venom-like proteins and GPCRs, as well as key genes involved in important pathways implicated in stress resistance and longevity in this study can serve as candidate targets for genetic manipulation. For future work, the genome annotation of *Steinernema* sp. HBG28 can certainly be improved upon, particularly to also include the protein evidence of *S. carpocapsae* strain Breton in the MAKER2 pipeline. A comparison between the genome annotation for *Steinernema* sp. HBG28 results based solely on previous data by Dillman et al. (2015) and Rougon-Cardoso et al. (2016), respectively, would be interesting.

7.7 References

1. Abuhatab, M., Selvan, S. and Gaugler, R. (1995). Role of Proteases in Penetration of Insect Gut by the Entomopathogenic Nematode *Steinernema Glaseri* (Nematoda, Steinernematidae). *Journal of Invertebrate Pathology*, **66:**125–130.
2. Aktar, W., Dwaipaya, S. and Chowdhury, A. (2009). Impact of pesticides use in agriculture: their benefits and hazards. Interdisciplinary *Toxicology*, **22:**1-12.
3. Allen, J. E., and MacDonald, A. S. (1998). Profound suppression of cellular proliferation mediated by the secretions of nematodes. *Parasite Immunology*, **20:**241–247.
4. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25:**3389–3402.
5. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at:
6. Bai, X. D., Adams, B. J., Ciche, T. A., Clifton, S., et al. (2013). A lover and a fighter: The genome sequence of an entomopathogenic nematode *Heterorhabditis bacteriophora*. *Plos One*, **8:**e69618.
7. Bal, H. K. and Grewal, P. S. (2015). Lateral Dispersal and Foraging Behaviour of Entomopathogenic Nematodes in the Absence and Presence of Mobile and Non-Mobile Hosts. *PLoS One*, **10:**e0129887.
8. Balasubramanian, N., Hao, Y.J., Toubarro, D., Nascimento, G., et al. (2009). Purification, biochemical and molecular analysis of a chymotrypsin protease with prophenoloxidase suppression activity from the entomopathogenic nematode *Steinernema carpocapsae*. *International Journal for Parasitology*, **39:**975–984.
9. Bilgrami, A. L., Gaugler, R., Shapiro-Ilan, D. I. and Adams, B. J. (2006). Source of trait deterioration in entomopathogenic nematodes *Heterorhabditis*

bacteriophora and *Steinernema carpocapsae* during *in vivo* culture. *Nematology*, **8:**397–409.

10. Binda-Rossetti, S., Mastore, M., Protasoni, M. and Brivio, M. F. (2016). Effects of an entomopathogen nematode on the immune response of the insect pest red palm weevil: Focus on the host antimicrobial response. *Journal of Invertebrate Pathology*, **133:**110–119.
11. Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30:**2114–2120.
12. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., et al. (2016). UniProtKB/Swiss-Prot, the manually annotated section of the uniprot knowledgebase: how to use the entry view. *Methods in Molecular Biology*. **1374:**23–54.
13. Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., et al. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368:**455–457.
14. Brivio, M.F., Mastore, M. and Pagani, M. (2005). Parasite-host relationship: A lesson from a professional killer. *Invertebrate Survival Journal*, **2:**41–53.
15. Burnell, A. (2002). Genetics and Genetic Improvement. In: Gaugler, R., editor. Entomopathogenic Nematology. New York: CAB International; 2002. pp. 241–264.
16. Butcher, R. A., Fujita, M., Schroeder, F. C. and Clardy, J. (2007). Small-molecule pheromones that control dauer development in *Caenorhabditis elegans*. *Nature Chemical Biology*, **3:**420–422.
17. Butcher, R. A., Fujita, M., Schroeder, F. C. and Clardy, J. (2007). Small-molecule pheromones that control dauer development in *Caenorhabditis elegans*. *Nature Chemical Biology*, **3:**420–422.

18. Byrne, A. B., Walradt, T., Gardner, K. E., Hubbert, A., et al. (2014). Insulin/IGF1 signaling inhibits age-dependent axon regeneration. *Neuron*, **81:**561–573.
19. Caldas, C., Cherqui, A., Pereira, A. and Simões, N. (2002). Purification and characterization of an extracellular protease from *Xenorhabdus nematophila* involved in insect immunosuppression. *Applied and Environmental Microbiology*, **68:**1297–1304.
20. Campos, M. L., Kang, J. H. and Howe, G. A. (2014) Jasmonate-triggered plant immunity. *Journal of Chemical Ecology*, **40:**657–675.
21. Capella-Gutiérrez, S., Silla-Martínez, J. M. and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25:**1972–1973.
22. Carton, Y., Frey, F., Stanley, D. W., Vass, E., et al. (2002). Dexamethasone inhibition of the cellular immune response of *Drosophila melanogaster* against a parasitoid. *The Journal of Parasitology*, **88:**405–407.
23. Castillo, J. C., Reynolds, S. E. and Eleftherianos, I. (2011). Insect immune response to nematode parasites. *Trends in Parasitology*, **27:**537–547.
24. Chang, Z., Li, G., Liu, J., Zhang, Y., et al. (2015). Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biology*, **161:**30.
25. Chaston, J. M., Dillman, A. R., Shapiro-Ilan, D. I., Bilgrami, A. L., et. al. (2011). Outcrossing and crossbreeding recovers deteriorated traits in laboratory cultured *Steinernema carpocapsae* nematodes. *International Journal for Parasitology*, **41:**801–809.
26. Cooper, D. and Eleftherianos, I. (2016). Parasitic Nematode Immunomodulatory Strategies: Recent Advances and Perspectives. *Pathogens*, **5:**58.
27. Crook, M., (2014). The dauer hypothesis and the evolution of parasitism: 20 years on and still going strong. *International Journal for Parasitology*, **44:**1–8.

28. Deutsch, C. A., Tewksbury, J. J., Tigchelaar, M., Battisti, D. S. et al., (2018). Increase in crop losses to insect pests in a warming climate. *Science*, **361:**916–919.
29. Dillman, A. R. (2012). Host Seeking and the Genomic Architecture of Parasitism Among Entomopathogenic Nematodes, Doctoral dissertation. California Institute of Technology, Pasadena, California.
30. Dillman, A. R., Macchietto, M., Porter, C. F., Roger, A., Williams, B., et al. (2015). Comparative genomics of *Steinernema* reveals deeply conserved gene regulatory networks. *Genome Biology*, **16:**200.
31. Douzery, E. J., Snell, E. A., Bapteste, E., Delsuc, F. et al. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proceedings of the National Academy of Sciences of the United States of America*, **101:**15386–15391.
32. Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5:**113.
33. Eilbeck, K., Moore, B., Holt, C. and Yandell, M. (2009). Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics*, **10:**67.
34. ffrench-Constant, R., Waterfield, N., Daborn, P., Joyce, S., et al., (2003). *Photorhabdus*: towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiology Reviews*, **26:**433-456.
35. Garofalo, A., Klager, S. L., Rowlinson, M. C., Nirmalan, N, et al. (2002). The FAR proteins of filarial nematodes: secretion, glycosylation and lipid binding characteristics. *Molecular and Biochemical Parasitology*, **122:**161–170.
36. Glazer, I. (2014). Genetic Improvement and Breeding of EPN: The Race for the "Super Nematode". *Journal of Nematology*, **46:**168–168.

37. Glazer, I. (2015). Improvement of Entomopathogenic Nematodes: A Genetic Approach. In: Campos-Herrera R, editor. Nematode Pathogenesis of Insects and Other Pests. Springer; 2015. pp. 29–55.
38. Golden, J. W. and Riddle, D. L. (1982). A pheromone influences larval development in the nematode *Caenorhabditis elegans*. *Science*, **218:**578–580.
39. Gomez-Escobar, N., Gregory, W. F. and Maizels, R. M. (2000) Identification of *tgh-2*, a filarial nematode homolog of *Caenorhabditis elegans daf-7* and human transforming growth factor beta, expressed in microfilarial and adult stages of *Brugia malayi*. *Infection and Immunity*, **68:**6402–6410.
40. Grant, W. N., Stasiuk, S., Newton-Howes, J., Ralston, M., et al. (2006). *Parastrongyloides trichosuri*, a nematode parasite of mammals that is uniquely suited to genetic analysis. *International Journal of Parasitology*, **36:**453–466.
41. Han, R. and Ehlers, R. U. (2000). Pathogenicity, development, and reproduction of *Heterorhabditis bacteriophora* and *Steinernema carpocapsae* under axenic *in vivo* conditions. *Journal of Invertebrate Pathology*, **75:**55–58.
42. Hao, Y. J., Montiel, R., Abubucker, S., Mitreva, M., et al. (2010). Transcripts analysis of the entomopathogenic nematode *Steinernema carpocapsae* induced *in vitro* with insect haemolymph. *Molecular Biochemical Parasitology*, **169:**79–86.
43. Hawdon, J. M. and Schad, G. A. (1991). Albumin and a dialyzable serum factor stimulate feeding in vitro by third-stage larvae of the canine hookworm *Ancylostoma caninum. Journal of Parasitology*, **77:**587–591.
44. Herbert, E. E. and Goodrich-Blair, H. (2007). Friend and foe: the two faces of *Xenorhabdus nematophila*. *Nature Reviews Microbiology*, **5:**634–646.
45. Holt, C. and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, **12:**491.

46. Hotez, P., Hawdon, J. and Schad, G. A. (1993). Hookworm larval infectivity, arrest and amphiparatenesis: the *Caenorhabditis elegans Daf-c* paradigm. *Parasitology Today*, **9:**23–26.

47. Howe, K. L., Bolt, B. J., Cain, S., Chan, J., et al. (2016). WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Research*, **44:**D774-D780. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. http://www.repeatmasker.org (2008).

48. Hunt, V. L., Tsai, I. J., Coghlan, A., Reid, A. J., et al. (2016). The genomic basis of parasitism in the *Strongyloides* clade of nematodes. *Nature Genetics*, **48:**299–307.

49. Iberkleid, I., Vieira, P., Engler, J. D., Firester, K, et al. (2013). Fatty acid-and retinol-binding protein, Mj-FAR-1 induces tomato host susceptibility to root-knot nematodes. *Plos One*, **8:**e64586.

50. Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, **8:**275–282.

51. Kanehisa, M., Furmichi, M., Tanabe, M., Sato, Y., et al. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, **45:**D353–D361.

52. Kanost, M. R. and Clarke, T. Proteases. In: Gilbert LI, Iatrou K, Gill S, editors. Comprehensive Molecular Insect Science, vol. 4. Oxford: Elsevier; 2005. p. 247–66.

53. Kauffman, A. L., Ashraf, J. M., Corces-Zimmerman, M. R., Landis J. N., et al. (2010). Insulin signaling and dietary restriction differentially influence the decline of learning and memory with age. *PLoS Biology*, **8:**e1000372.

54. Kennedy, M. W., Corsico, B., Cooper, A. and Smith, B. O. The unusual lipid-binding proteins of nematodes: NPAs, nemFABPs and FARs. In: Kennedy, M, W, Harnett,

W., editors. Parasitic nematodes: molecular biology, biochemistry, and immunology. Wallingford: CABI; 2013. p. 397–412.

55. Kennedy, M. W., Corsico, B., Cooper, A. and Smith, B. O. The unusual lipid-binding proteins of nematodes: NPAs, nemFABPs and FARs. In: Kennedy MW, Harnett W, editors. Parasitic nematodes: molecular biology, biochemistry, and immunology. Wallingford: CABI; 2013. p. 397–412.
56. Kimura, K. D., Tissenbaum, H. A., Liu, Y. and Ruvkun, G. (1997). daf-2, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science*, **277:**942–946.
57. Kimura, K. D., Tissenbaum, H. A., Liu, Y. and Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science*, **277:**942–946.
58. Kodama, E., Kuhara, A., Mohri-Shiomi, A., Kimura K. D., et al. (2006). Insulin-like signalling and the neural circuit for integrative behaviour in *C. elegans*. *Genes & Development,* **20:**2955–2960.
59. Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, **5:**59.
60. Kück, P. and Meusemann K. (2010). FASconCAT: Convenient handling of data matrices. *Molecular Phylogenetics and Evolution*, **56:**1115–1118.
61. Lagesen, K. et al. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, **35:**3100–3108.
62. Lawrence, T., Willoughby, D. A. and Gilroy, D. W. (2002). Anti-inflammatory lipid mediators and insights into the resolution of inflammation. *Nature Reviews. Immunology*, **2:**787–795.
63. Lee, R. Y. N., Howe, K. L., Harris, T. W., Arnaboldi, V., et al. (2017). WormBase 2017: molting into a new stage. *Nucleic Acids Research*, **46:**D869–D874.

64. Li, L., Stoeckert, C. J. and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13:**2178–2189.

65. Lin, C. H., Tomioka, M., Pereira, S., Sellings, L., Iino, Y., et al. (2010). Insulin signaling plays a dual role in *Caenorhabditis elegans* memory acquisition and memory retrieval *The Journal of Neuroscience: the official journal of the Society for Neuroscience*, **30:**8001–8011.

66. Lowe, T. M. and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acid Research*, **25:**955-964.

67. Lu, D., Baiocchi, T. and Dillman, A. R. (2016). Genomics of Entomopathogenic Nematodes and Implications for Pest Control. *Trends in Parasitology*, **32:**588-598.

68. Lu, D., Macchietto, M., Chang, D., Barros, M. M., et al. (2017). Activated entomopathogenic nematode infective juveniles release lethal venom protein. *PLOS Pathogens*, **13:**e1006302.

69. McKerrow, J. H., Caffrey, C., Kelly, B., Loke, P., et al. (2006). Proteases in parasitic diseases. *Annual Review of Pathology*, **1:**497–536.

70. McLean, F., Berger, D., Laetsch, D. R., Schwartz, H. T., et. al. (2018). Improving the annotation of the *Heterorhabditis bacteriophora* genome. *GigaScience*, **7:**giy034.

71. McNulty, S. N., Strübe, C., Rosa, B. A., Martin, J. C., et al. (2016). *Dictyocaulus viviparus* genome, variome and transcriptome elucidate lungworm biology and support future intervention. *Scientific Reports*, **6:**20316.

72. McSorley, H. J., Grainger, J. R., Harcus, Y., Murray, J., et al. (2010). *daf-7*-related TGF-beta homologues from Trichostrongyloid nematodes show contrasting life-cycle expression patterns. *Parasitology*, **137:**159–171.

73. Murakami, H., Bessinger, K., Hellmann J. and Murakami S. (2005). Aging-dependent and -independent modulation of associative learning behaviour by

insulin/insulin-like growth factor-1 signal in *Caenorhabditis elegans*. *The Journal of Neuroscience: the official journal of the Society for Neuroscience*, **25:**10894–10904.

74. Nawrocki, E. P. and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29:**2933–2935.
75. Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., et al., (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*, **43:**11.
76. Ogg, S., Paradis, S., Gottlieb, S., Patterson, G. I., et al. (1997). The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature*, **389:**994–999.
77. Page, A. P., McCormack, G. and Birnie, A. J. (2006). Biosynthesis and enzymology of the *Caenorhabditis elegans* cuticle: identification and characterisation of a novel serine protease inhibitor. *International Journal for Parasitology*, **36:**681–689.
78. Rawlings, N. D., Barrett, A. J. and Bateman, A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, **40:**D343–D350.
79. Robertson, H. M. and Thomas, J. H. (2006). The putative chemoreceptor families of *C. elegans*, *WormBook*, ed. The *C. elegans* Research Community, WormBook, http://www.wormbook.org
80. Rogers, W. P. and Sommerville, R. I. (1963). The infective stage of nematode parasites and its significance in parasitism. *Advances in Parasitology*, **1:**109–177.
81. Rougon-Cardoso, A., Flores-Ponce,M., Ramos-Aboites, H. E., Martínez-Guerrero, C. E., et al. (2016). The genome, transcriptome, and proteome of the nematode *Steinernema carpocapsae*: evolutionary signatures of a pathogenic lifestyle. *Scientific Reports*, **6:**37536.

82. Simões, N., Caldas, C., Rosa, J. S., Bonifassi, E., et al. (2000). Pathogenicity caused by high virulent and low virulent strains of *Steinernema carpocapsae* to *Galleria mellonella*. *Journal of Invertebrate Pathology*, **75:**47–54.

83. Skoufos, E., Marenco, L. Nadkarni, P. M, Miller, P. L., et al. (2000). Olfactory Receptor Database: a sensory chemoreceptor resource. *Nucleic Acids Research*, **28:**341–343.

84. Smit, A. F. A. and Hubley, R. RepeatModeler Open-1.0.

85. Smit, A. F. A., Hubley, R. and Green, P. RepeatMasker Open-4.0 http://www.repeatmasker.org (2013).

86. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., et al. (2016). TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Research*, **268:**1134–1144.

87. Srinivasan, J., Dillman, A. R., Macchietto, M. G., Heikkinen, L., et al. (2013). The Draft Genome and Transcriptome of *Panagrellus redivivus* Are Shaped by the Harsh Demands of a Free-Living Lifestyle. *Genetics*, **193:**1279–1295.

88. Stanke, M., Steinkamp, R., Waack, S. and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research*, **32:**W309–W312.

89. Stanley, D. and Miller, J. (2006). Eicosanoids in invertebrate immunity: an *in vitro* approach. *In Vitro Cellular & Developmental Biology*, **42:**5a–a.

90. Stepek, G., McCormack, G. and Page, A. P. (2010). The kunitz domain protein BLI-5 plays a functionally conserved role in cuticle formation in a diverse range of nematodes. *Molecular and Biochemical Parasitology*, **169:**1–11.

91. Stoltzfus, J. D., Massey, H. C. Jr., Nolan, T. J., Griffith, S. D. et al. (2012a). *Strongyloides stercoralis age-1*: a potential regulator of infective larval development in a parasitic nematode. *PLoS ONE*, **7:**e38587.

92. Stoltzfus, J. D., Minot, S., Berriman, M., Nolan, T. J., et al. (2012b). RNAseq analysis of the parasitic nematode *Strongyloides stercoralis* reveals divergent regulation of canonical dauer pathways. *PLoS neglected tropical diseases*, **6:**e1854.

93. Tang, Y. T., Gao, X., Rosa, B. A., Abubucker, S., et. al. (2014). Genome of the human hookworm *Necator americanus*. *Nature Genetics*, **46:**261-269.

94. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **11:**4-41.

95. Thaler, J. O., Duvic, B., Givaudan, A. and Boemare, N. (1998). Isolation and entomotoxic properties of the *Xenorhabdus nematophilus* F1 lecithinase. *Applied and Environmental Microbiology*, **64:**2367–2373.

96. Thomas, J. H., and Robertson, H. M. (2008). The *Caenorhabditis* chemoreceptor gene families. *BMC Biology*, **6:**42.

97. Tissenbaum, H. A., Hawdon, J., Perregaux, M., Hotez, P., et al. (2000). A common muscarinic pathway for diapause recovery in the distantly related nematode species *Caenorhabditis elegans* and *Ancylostoma caninum*. *Proceedings of the National Academy of Sciences of the United States of America*, **97:**460–465.

98. Tomioka, M., Adachi, T., Suzuki, H., Kunitomo, H., et al. (2006). The insulin/PI 3-kinase pathway regulates salt chemotaxis learning in *Caenorhabditis elegans*. *Neuron*, **51:**613–625.

99. Toubarro, D., Lucena-Robles, M., Nascimento, G., Costa, G., et al. (2009). An apoptosis-inducing serine protease secreted by the entomopathogenic nematode *Steinernema carpocapsae*. *International Journal for Parasitology*, **39:**1319–1330.

100. Viney, M. E. (1996). Developmental switching in the parasitic nematode *Strongyloides ratti*. *Proceedings of the Royal Society of London Series B*, **263:**201–208.

101. Wang, Y., Coleman-Derr, D., Chen, G. and Gu, Y. Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, **43:** W78–W84.

102. Wu, S., Zhu, Z., Fu, L., Niu, B., et al. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, **12:**444.

103. Yamada, M., Matsuda, S., Nakazawa, M. and Arizono, N. Species-specific differences in heterogonic development of serially transferred free-living generations of *Strongyloides planiceps* and *Strongyloides stercoralis*. *Journal of Parasitology*, **77:**592–594.

104. Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution*, **39:**306–314.

105. Zhang, Z. and Wood, W. I. (2003). A profile hidden Markov model for signal peptides generated by HMMER. *Bioinformatics*, **19:**307-308.

Chapter 8

Final thoughts and conclusions

The study of entomopathogenic nematodes in South Africa has often relied on conducting extensive surveys in various provinces and, thereby, making use of classical morphology and molecular-based detection methods for the purpose of nematode identification. Genomic sequencing is the next and most palpable step in the advancement of entomopathogenic nematode research. Genomics can allow researchers the opportunity of learning more about their nematode of interest, as well as uncovering information on genome structure and nematode evolution. While processing sequence data can pose both a computational and bioinformatics challenge, the genomic information held within an assembled genome can far exceed research expectations, leading to novel findings. To date, there are only six *Steinernema* nematode genomes (Dillman et al., 2015; Rougon-Cardoso et al., 2016) and twenty-four *Xenorhabdus* bacterial genomes that have been sequenced (https://www.ncbi.nlm.nih.gov/taxonomy/626). In the present study, a novel *Steinernema* nematode and its associated *Xenorhabdus* symbiont were isolated and characterized using a genomic approach.

The first part of this study was aimed at recovering an indigenous entomopathogenic nematode species, which could later serve as a potential candidate for genomic sequencing. In June of 2014, a survey was carried out in the grasslands of the Suikerbosrand Nature Reserve in Gauteng, South Africa, which, at the time, led to the recovery of a novel *Steinernema* species. The genome project of *Steinernema* sp.

HBG28 was initiated in the same year with the goal to sequence and assemble its genome. A research project focused solely on genome sequencing of native entomopathogenic nematodes within South Africa had never previously been undertaken.

The next part of this study was to isolate and identify the bacterial symbiont associated with *Steinernema* sp. HBG28, which was later recognized as a strain of *Xenorhabdus khoisanae* (Ferreira et al., 2013). It was observed that *X. khoisanae* strain MCB exhibited a specialized form of bacterial translocation on solid agar, referred to as swarming. While swarming behaviour is not uncommon among several bacterial species, including those of *Xenorhabdus*, it has never been previously observed among phase II variants of *Xenorhabdus* bacteria and was thought to occur only among phase I variants of *Xenorhabdus* (Giuvaden et al., 1995). The findings within this study have shown that this is not the case, as both phenotypic variants of *X. khoisanae* strain MCB demonstrated swarming behaviour on solid agar. This is the first report of bacterial swarming in *Xenorhabdus* bacteria undergoing phenotypic variation.

To gain insight into the biology of *X. khoisanae* strain MCB, genomic sequencing was performed. Additionally, as *X. khoisanae* has been recovered from several local *Steinernema* nematode hosts in South Africa, its highly prevalent nature made it an ideal model for studying nematode-mutualist interactions. The draft genome assembly of *X. khoisanae* strain MCB revealed several findings. Perhaps, the most notable of which was the large number of genomic islands present within its genome. These genomic islands were enriched in genes related to novel functions, phage remnants and virulence; and are likely to have been acquired through horizontal gene transfer (HGT).

The detection of two chemotaxis genes, *cheA* and *cheY* linked specifically to flagellar motility in *X. khoisanae* strain MCB has implications that these genes may serve regulatory roles in bacterial swarming, especially during phase variation and may even attribute to the high prevalence of *X. khoisanae* observed among several *Steinernema* spp. in South Africa (Dreyer et al., 2017). It can further be proposed that this unique swarming ability in *X. khoisanae* bacteria may allow for mismatched associations with non-cognate, local *Steinernema* hosts. For instance, when *X. khoisanae* is taken up by a non-cognate *Steinernema* host, this may lead to conditions that induce phenotypic variation in the bacterial symbiont. Consequently, swarming may potentially facilitate intestinal colonization of a non-cognate *Steinernema* nematode by phenotypic variants of *X. khoisanae.* As such, the swarming behaviour in *X. khoisanae* strain MCB, as well as the genes implicated in its regulation would require further study.

The latter part of the study was aimed at exploring the genomic architecture of the isolated entomopathogenic nematode, *Steinernema* sp. HBG28. First, this was done by establishing an inbred nematode line prior to DNA extraction and genomic sequencing. Once a draft genome of *Steinernema* sp. HBG28 had been produced, genome annotation and subsequent comparative analyses were performed. In summary, the genomic characterization of *Steinernema* sp. HBG28 confirmed that *Steinernema* nematodes have a mutual role in insect pathogenicity along with their bacterial symbiont through the identification of a variety of proteases, protease inhibitors, venom-like proteins.

The detection of a significant number of genes directly involved in signal transduction was proposed to play an important role in the parasitic lifestyle that

Steinernema sp. HBG28 has adopted with respect to the location of potential hosts, host infection, dauer formation, stress resistance and longevity. Furthermore, this genome represents a valuable resource in facilitating ongoing efforts in entomopathogenic nematology research, as well as providing a framework for understanding nematode biology and nematode-bacterium symbiosis.

For future work, adjustments can be made in the genome assembly process of *Steinernema* sp. HBG28 to account for possible bacterial contamination in the form of sequence reads and a larger genome size. This may improve the overall quality of the genome assembly. A new genome annotation can be performed with the MAKER2 pipeline to include new homology-based protein evidence. A comparison between the genome annotation results based on previous data by Dillman et al. (2015) and Rougon-Cardoso et al. (2016) for *Steinernema* sp. HBG28 would be interesting.

8.1 References

1. Dillman, A. R., Macchietto, M., Porter, C. F., Roger, A., et al. (2015). Comparative genomics of *Steinernema* reveals deeply conserved gene regulatory networks. *Genome Biology*. **16:**200.
2. Dreyer, J., Malan, A. and Dicks, L. M. T., (2017). Three Novel *Xenorhabdus-Steinernema* Associations and Evidence of Strains of Switching Between Different Clades. *Current Microbiology*, **74:**938–942.
3. Ferreira, T., van Reenen, C. A., Endo, A., Spröer, C., et al. (2013). Description of *Xenorhabdus khoisanae* sp. nov., the symbiont of the entomopathogenic nematode *Steinernema khoisanae*. *International Journal of Systematic and Evolutionary Microbiology*, **63:**3220–3224.
4. Givaudan, A., Baghdiguian, S., Lanois, A. and Boemare, N. (1995). Swarming and swimming changes concomitant with phase variation in *Xenorhabdus nematophilus*, *Applied and Environmental Microbiology*, **61:**1408-1413.
5. Rougon-Cardoso, A., Flores-Ponce,M., Ramos-Aboites, H. E., Martínez-Guerrero, C. E., et al. (2016). The genome, transcriptome, and proteome of the nematode *Steinernema carpocapsae*: evolutionary signatures of a pathogenic lifestyle. *Scientific Reports*, **6:**37536.

Appendices

Supplementary material A

CLC assembly report for initial bacterial genome assembly.

Supplementary material B

FASTQC assembly report for nematode genome assembly.

Table of contents

1. nb28_S4_L001_R1_001 (paired) trimmed (paired) merged assembly summary report

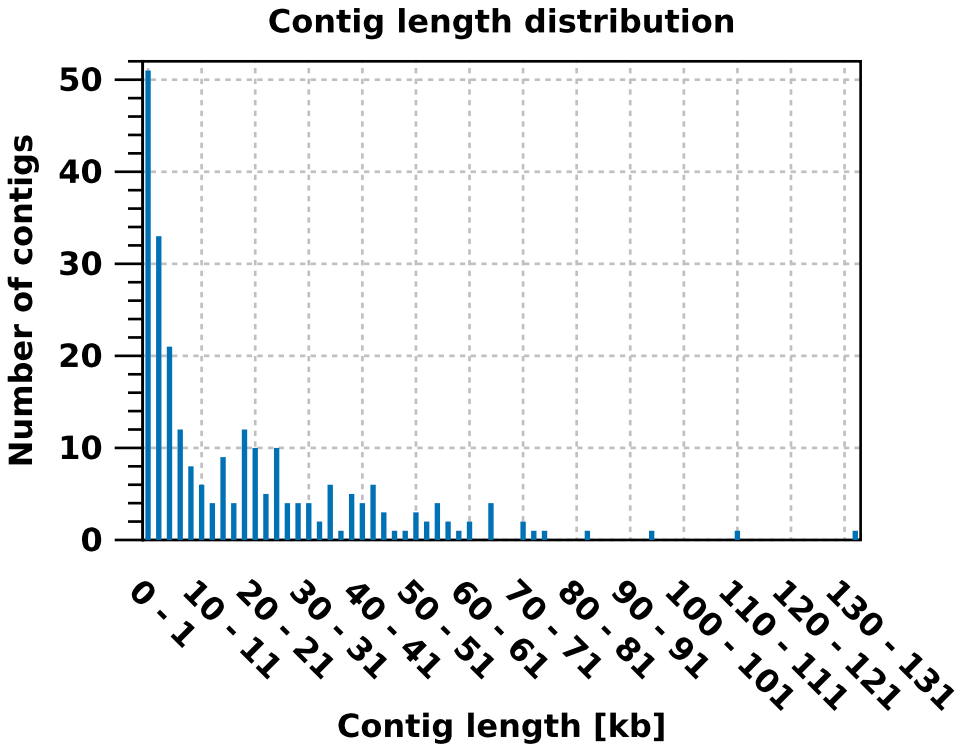
1.1 Nucleotide distribution

Nucleotide	Count	Frequency
Adenine (A)	1,311,579	28.0%
Cytosine (C)	1,015,049	21.7%
Guanine (G)	1,023,096	21.9%
Thymine (T)	1,328,225	28.4%
Any nucleotide (N)	3,630	0.1%

1.2 Contig measurements (including scaffolded regions)

N75	22,557
N50	39,550
N25	56,171
Minimum	443
Maximum	133,423
Average	18,578
Count	252

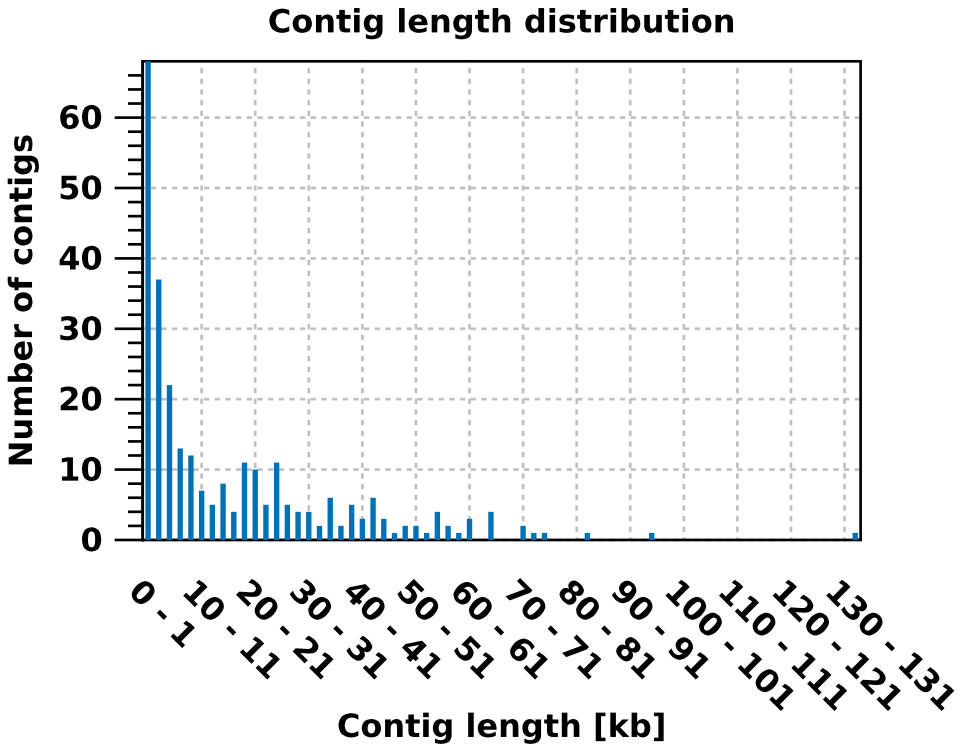
Total	4,681,579



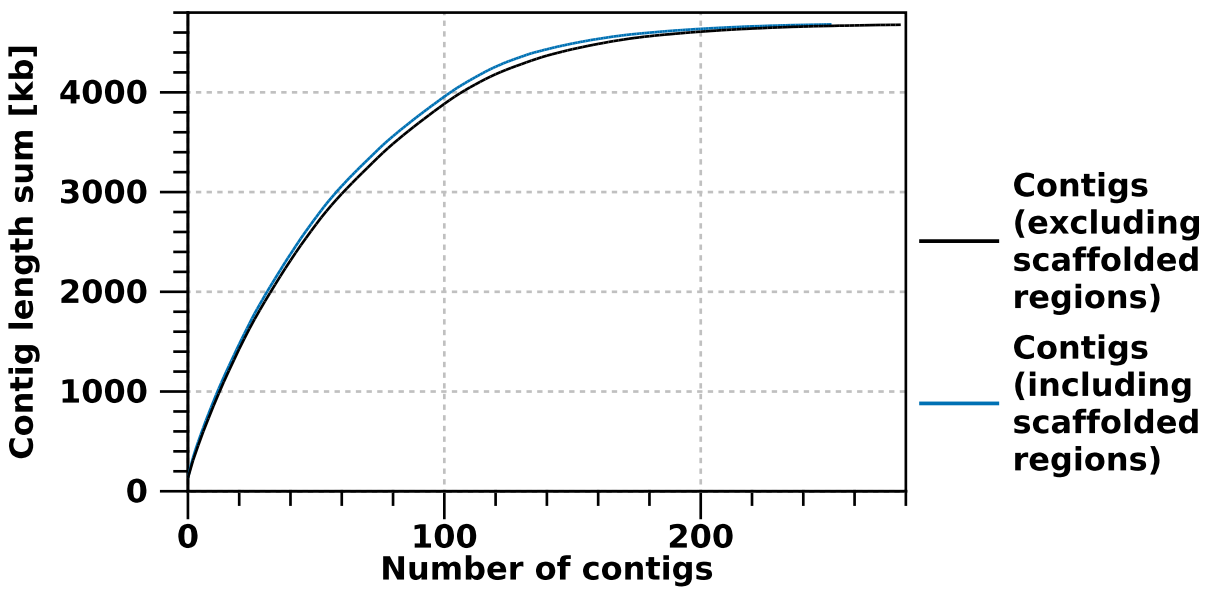
1.3 Contig measurements (excluding scaffolded regions)

N75	21,660
N50	38,897
N25	55,563
Minimum	126
Maximum	133,423
Average	16,707
Count	280

Total	4,677,935



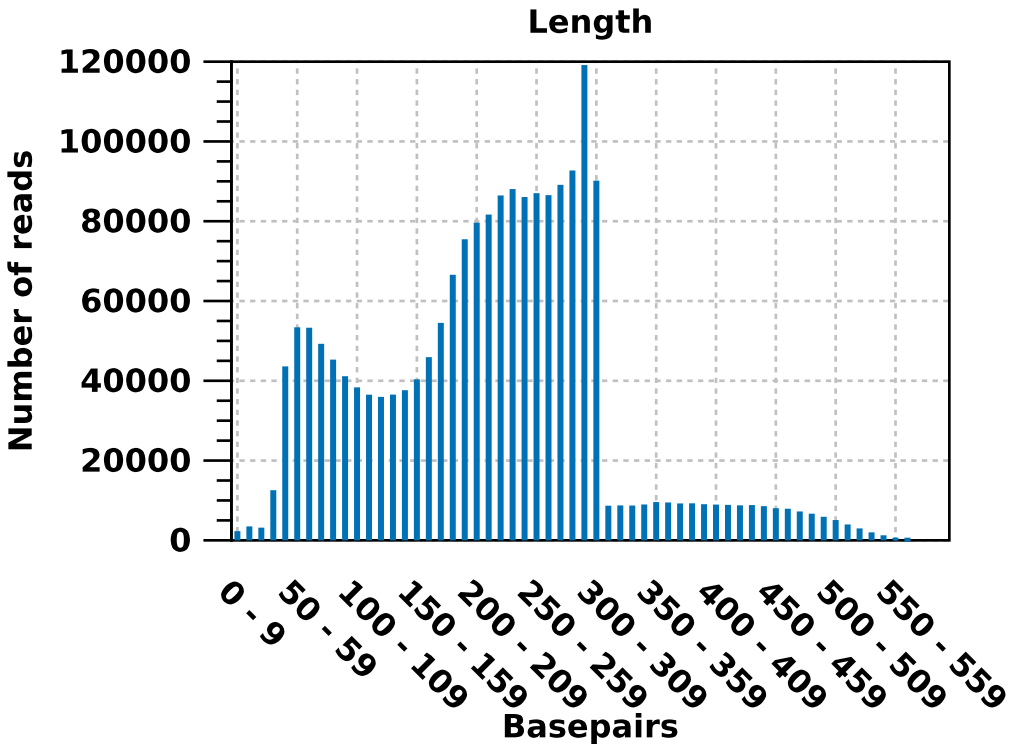
1.4 Accumulated contig lengths



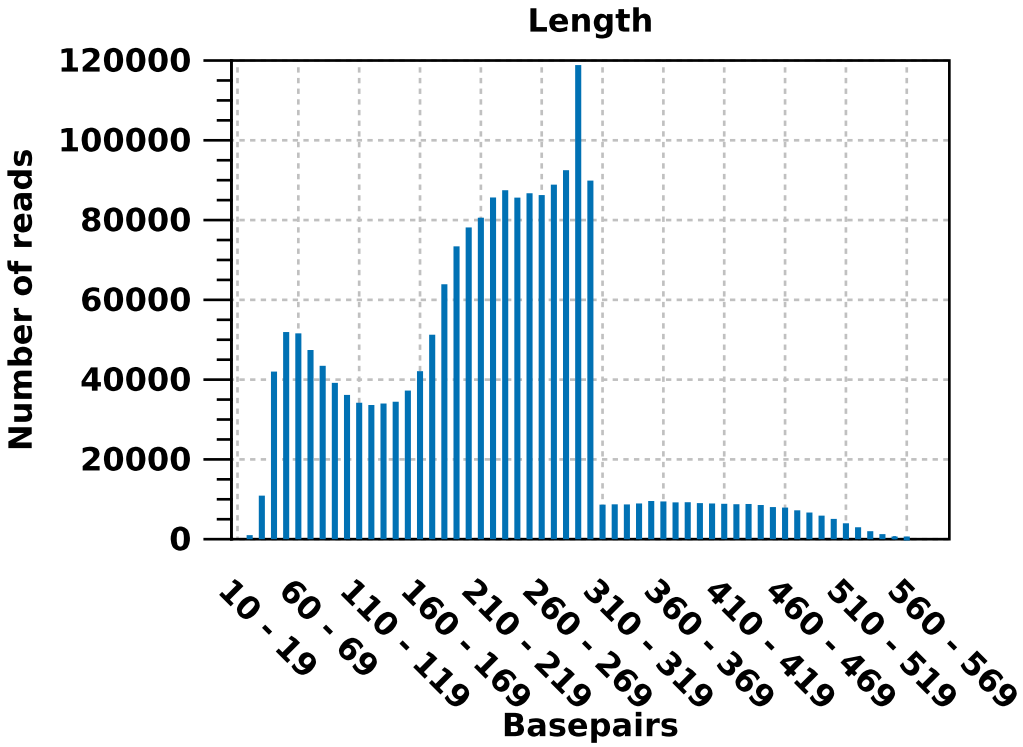
1.5 Summary statistics

	Count	Average length	Total bases
Reads	1,940,231	216.45	419,965,962
Matched	1,886,496	219.13	413,382,308
Not matched	53,735	122.52	6,583,654
Contigs	252	18,577	4,681,579
Reads in pairs	762,528	651.86	
Broken paired reads	432,569	214.11	

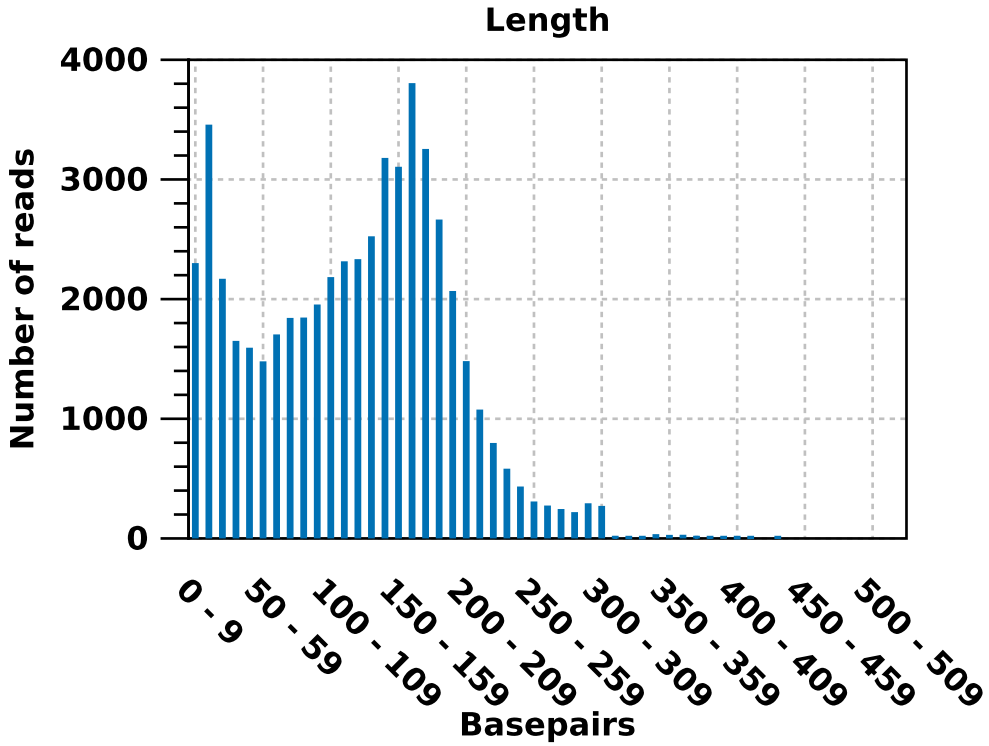
1.6 Distribution of read length



1.7 Distribution of matched read length



1.8 Distribution of non-matched read length



1.9 Paired reads distance distribution

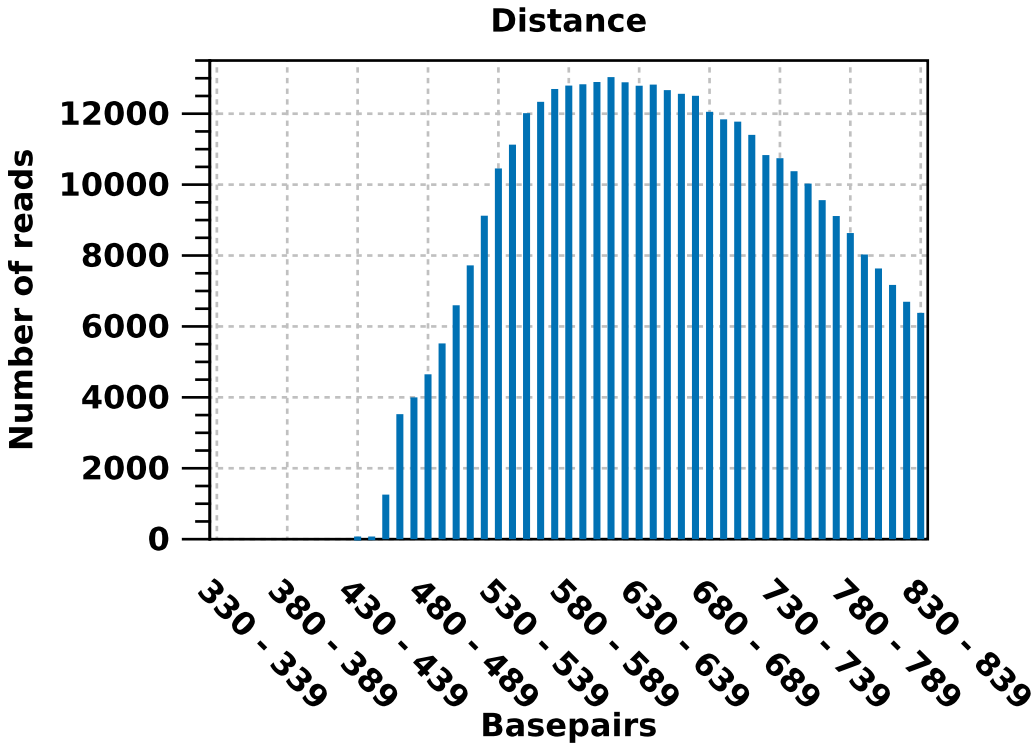


Table of contents

1. nematodeadapters_qualitytrimmed (paired) assembly summary report

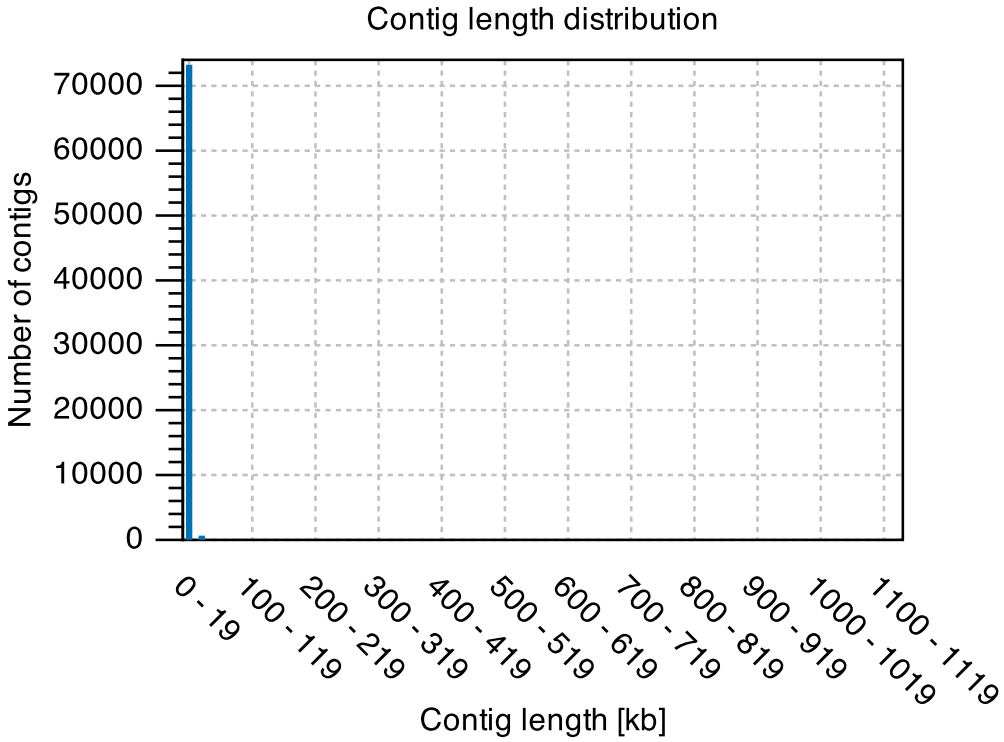
1.1 Nucleotide distribution

Nucleotide	Count	Frequency
Adenine (A)	37,137,012	25.0%
Cytosine (C)	36,984,105	24.9%
Guanine (G)	37,097,080	25.0%
Thymine (T)	37,191,137	25.0%
Any nucleotide (N)	90,965	0.1%

1.2 Contig measurements (including scaffolded regions)

N75	2,017
N50	8,728
N25	37,008
Minimum	200
Maximum	1,133,054
Average	2,000
Count	74,257

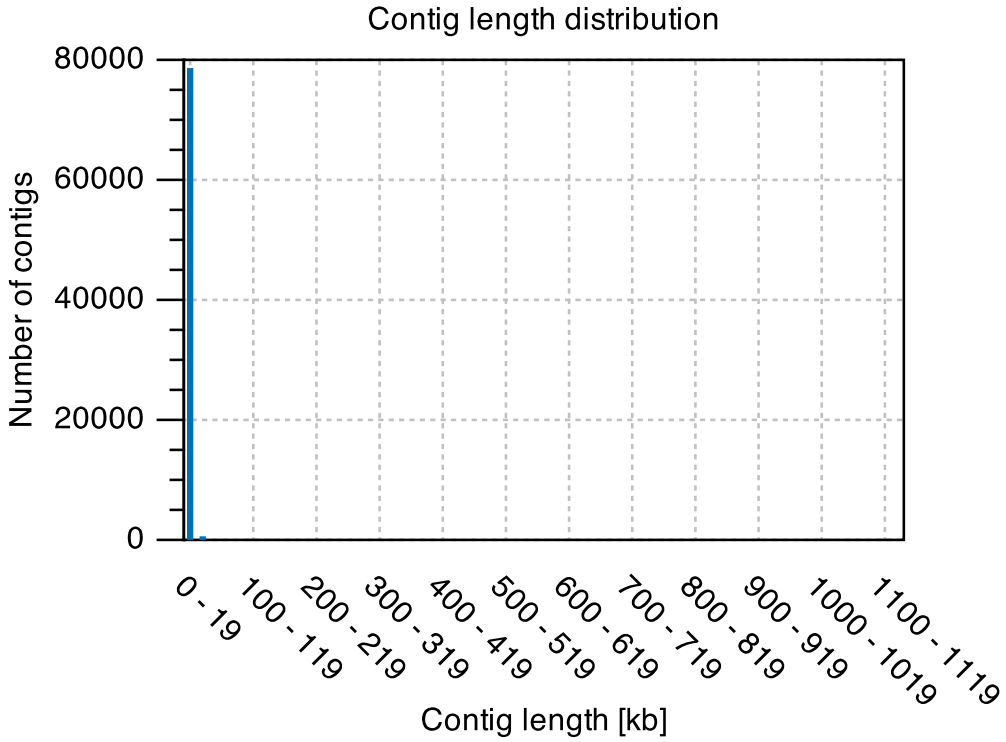
Total	148,500,299



1.3 Contig measurements (excluding scaffolded regions)

N75	1,841
N50	7,572
N25	32,255
Minimum	120
Maximum	1,133,054
Average	1,865
Count	79,572

Total	148,409,334



1.4 Accumulated contig lengths