Genome analysis of an entomopathogenic nematode belonging to the genus *Oscheius* and its insect pathogenic bacterial endosymbiont.

PhD

Tiisetso Elizabeth Lephoto

A thesis submitted to the Faculty of Science under the school of Molecular and Cell Biology in fulfilment for requirements for Doctor of Philosophy Degree.

February 2016

Microbiology and biotechnology department, school of molecular and cell biology, University of the Witwatersrand, Braamfontein 2050, Johannesburg, South Africa.

DECLARATION

I Tiisetso Elizabeth Lephoto, (student number: 0713627e) am a student registered for the degree of Doctor of Philosophy (PhD) in the academic year 2015
I hereby declare the following:

- I am aware that plagiarism is wrong and declare that this thesis is written in my own words and have referenced and cited all sources used.
- This research and all laboratory experiments were done by me
- A complete thesis was written by me.
- 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 submitted this thesis for Doctor of Philosophy Degree at the University of the Witwatersrand, Johannesburg
- This thesis has not been submitted before for any degree in any other University 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.

Name: Tiisetso Elizabeth

Surname: Lephoto

Date: 01 February 2016

Signature: Lephoto

Supervisor: Prof Vincent Miles Gray

DEDICATION

I dedicate this work to my loving parents Mr David Mabuti Lephoto and Mrs Monica Mamosesedi Lephoto who always put my education first and always support me in everything I do. Day, you have been a father that I believe God beautifully made for us, you are the best and I love you. Mama, I love you, thank you for your unconditional love, your warm heart, your spirit and thank you for standing by me always. I'm where I am today because of your team work as my parents and May God bless you and keep you strong. To my partner, my pillar of strength, my love Mr Lebang Nong; thank you for your outmost support and motivation which kept me going through this challenging and yet fruitful journey, I love you so much. To my brothers, Mr Kutlo Lephoto, Mr Justice Teboho Lephoto and Mr Emmanuel Baloyi, thank you for all you have done for me as your only sister whom you protect so much, I love you all.

Psalm 7: 17

I will give thanks to the LORD because of his righteousness and will sing praise to the name of the LORD Most High.

Psalms 107:1

O give thanks to the LORD, for he is good: for his mercy endures for ever.

Philippians 4:13

I can do all things through Christ which strengthens me.

Thank you God

You are a great God.

ABSTRACT

The use of synthetic chemical pesticides has several negative implications for the Agricultural industry, which include the development of resistance to the insecticides, crop contamination and the killing of non-target insects. This has brought many scientists in the field of nematology and entomology to investigate biological control agents which can help solve identified challenges and these biocontrol agents have also included entomopathogenic nematodes. The majority of entomopathogenic nematodes species that have been isolated belong to *Heterorhabditids* and *Steinernematids* which act as vectors for insect pathogenic bacteria species belonging to the genera, *Photorhabdus* and *Xenorhabdus*, respectively. However, other species of nematodes, one of which includes a strain of *Caenorhabditis briggsae*, have also been shown to act as a vector for an insect pathogenic strain of *Serratia marcescens*. *Oscheius* sp. TEL-2014 EPNs have been observed to act as vectors for insect pathogenic bacteria belonging to the genus *Serratia*. In this study a novel insect pathogenic *Serratia* sp. strain *TEL* was isolated from the gut of infective juveniles belonging to a species of *Oscheius* sp. TEL-2014. Next generation sequencing of the bacteria was conducted by generating genomic DNA paired-end libraries with the Nextera DNA sample preparation kit (Illumina) and indexed using the Nextera index kit (Illumina). Paired-end (2 × 300 bp) sequencing was performed on a MiSeq Illumina using the MiSeq reagent kit v3 at the Agricultural Research Council Biotechnology Platform. Quality control and adapter trimming was performed and the genome was assembled using SPADES. 19 contigs were generated with an average length of 301767 bp and N_{50} of 200,110 bp. The genome of the *Serratia sp. TEL* was found to be 5,000,541 bp in size, with a G+C content of 59.1%, which was similar to that of other *Serratia* species previously identified. Furthermore, the contigs were annotated using NCBI Prokaryotic Genome Automatic Annotation Pipeline. Features of the annotated genome included protein encoding sequence or genes, rRNA encoding genes, tRNA encoding genes, ncRNA sequences and repeat regions. 4,647 genes were found and 4,495 were protein-coding sequences (CDS). The genome contains 36 pseudo genes, 2 CRISPR arrays, 13 rRNA genes with five operons (5S, 16S, 23S), 88 tRNAs genes, 15 ncRNA sequences and 9 frameshifted genes. Several genes involved in virulence, disease, defense, stress response, cell division, motility and chemotaxis were identified. This genome sequence will allow for the investigation of identified genes and that will be critical in furthering the understanding of the insect pathogenicity of *Serratia sp*. strain *TEL*. Furthermore, it will provide additional genomic insights about the insect-nematode interactions and thus help us improve their ability to be used as biological control agents in agricultural industries. *Oscheius* sp. TEL-2014 was tested for its entomopathogenicity and it was found that this species was able to infect and kill two model insects *Galleria mellonella* and *Tenebrio molitor*. This new nematode

species brought 100% mortality within 72 h post-exposure in *G. mellonella* and whereas, within 96 hours in *T. molitor*. Following morphometrics analysis of *Oscheius* sp. TEL-2014 it was concluded that this nematode is described as a novel entomopathogenic nematode species based on its morphometrics and 18S rRNA gene sequence originality. Whole genome sequencing of *Oscheius* sp. TEL-2014 inbred lines (7 and 13) was performed using Illumina Hiseq sequencing system and paired ends library preparation protocol. Sequencing reads assembled on Velvet resulted in generation of 75965 contigs (line 7) and 53190 contigs (line 13). Gene prediction tools showed that proteins involved in gene expression and DNA replication are present in *Oscheius* sp. TEL-2014. The draft genome of *Oscheius* nematodes will support the improvement and initiation of further studies intended to help us understand the molecular and metabolic processes in this genus.

RESEARCH OUTPUTS

ORIGINAL PUBLICATIONS

Publications forming part of PhD thesis

Lephoto TE, Featherston J, Gray VM. 2015. Draft whole-genome sequence of *Serratia* sp. strain TEL, associated with *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae) isolated from a grassland in South Africa. Genome Announc 3(4):e00747-15. doi:10.1128/genomeA.00747-15.

Lephoto TE and Gray VM. 2015. Genome sequencing and annotation of *Serratia* sp. strain TEL. Genomics data 6:54-56.

MANUSCRIPT ACCEPTED FOR PUBLICATION

Tiisetso E. Lephoto, Phelelani Mpangase, Shaun Aron, Vincent M. Gray, Whole genome sequence of *Oscheius* sp. TEL-2014 entomopathogenic nematodes isolated from South Africa, *Genomics Data* (2016), doi:10.1016/j.gdata.2016.01.017

MANUSCRIPT SUBMITTED FOR PUBLICATION

A novel species of entomopathogenic nematode *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae) isolated from South Africa

MANUSCRIPT IN PREPARATION FOR SUBMISSION TO JOURNAL

Investigating the pathogenicity of surface sterilised entomopathogenic nematode *Oscheius* sp. TEL-2014 and isolation of *Serratia* sp. Strain *TEL-2014* found in South Africa

Whole genome sequencing project of *Oscheius sp. TEL-2014*

CONFERENCE PRESENTATIONS AND AWARDS

Selected delegate for University of the Witwatersrand 7th Cross-faculty graduate symposium 2016

Poster presentation

Title: Genomics of a novel entomopathogenic nemaode

Authors: Tiisetso E. Lephoto and Vincent M. Gray

Australia Agricultural bioscience international conference 2015

Poster supported by oral presentation

Title: Genome sequencing and annotation of *Serratia* sp. strain TEL

Authors: Tiisetso E. Lephoto and Vincent M. Gray

8th Annual agriculture research symposium 2015

Oral presentation

Title: The molecular biology of entomopathogenic nematodes behaviour

Authors: Tiisetso E. Lephoto and Vincent M. Gray

Award: Best crop production oral presentation award

Selected for Molecular Bioscience Research Thrust research day December 2015

Poster presentation

Title: Whole genome sequencing and annotation of *Serratia* sp TEL and molecular biology of entomopathogenic nematodes

Authors: Tiisetso E. Lephoto and Vincent M. Gray

Award Nominations Gauteng Biotech Fundi Awards November 2015

Young Reseacher category

Best postgraduate category

Capacity builder category

University of the Witwatersrand 6th Cross-faculty graduate symposium 2014

Poster presentation

Title: The molecular biology of entomopathogenic nematodes behaviour

Authors: Tiisetso E. Lephoto and Vincent M. Gray

7th Annual agriculture research symposium 2014

Poster presentation

Title: The molecular biology of entomopathogenic nematodes behaviour

Authors: Tiisetso E. Lephoto and Vincent M. Gray

Award: Best New and upcoming Researcher award

Gauteng Biotech Fundi Awards 2014

Award: Best young reacher in biotechnology

Falling Wall Lab Johannesburg conference 2013

Oral presentation

Title: Isolation of entomopathogenic nematodes with the potential to be used as biological control agents in Agricultural industries

Authors: Tiisetso E. Lephoto and Vincent M. Gray

Award: 2nd place price winner for Best science and innovative research project in South Africa/Africa

University of the Witwatersrand 5th Cross-Faculty Annual Symposium 2013

Poster presentation

Title: The molecular biology of entomopathogenic nematodes behaviour

Authors: Tiisetso E. Lephoto and Vincent M. Gray

Molecular Bioscience Research Thrust research day 2013

Poster presentation

Title: Isolation of entomopathogenic nematodes with the potential to be used as biological control agents in Agricultural industries

Authors: Tiisetso E. Lephoto and Vincent M. Gray

5th Annual agriculture research symposium 2012

Poster presentation

Title: Isolation, identification and characterization of entomopathogenic nematodes and desiccation tolerance of cultured EPNs with a potential to be used as biological control agents in agricultural industries

Authors: Tiisetso E. Lephoto and Vincent M. Gray

University of the Witwatersrand 4th Cross-faculty Annual Symposium 2012

Poster presentation

Title: Isolation, identification and characterization of entomopathogenic nematodes

Authors: Tiisetso E. Lephoto and Vincent M. Gray

Molecular Bioscience Research Thrust research day 2012

Poster presentation

Title: Isolation, identification and characterization of entomopathogenic nematodes

Authors: Tiisetso E. Lephoto and Vincent M. Gray

OTHER ACHIEVEMENTS AND AWARDS WHILE REGISTERED FOR PHD

- Mail and Guardian Greening the future special mention award for *Yes we are moving* (Community outreach project-founded by Tiisetso Elizabeth Lephoto)
- Top 20 Inspirational Youth in South Africa 2015 http://www.youthvillage.co.za/2015/06/top-20-inspirational-youth-in-south-africa-2015-part2/
- Postgraduate Merit Award by Wits University 2013 to 2015
- National Research Foundation Scare Skills and innovation PhD scholarship 2013 to 2015
- Golden Key International Honour Society member 2014
- One of the 100 Brightest Young Minds in South Africa delegate
- One Young World summit Ambassador 2013.
- Delegate at the 11th Annual Nelson Mandela Lecture 2013

ACKNOWLEDGEMENTS

My deepest gratitude goes to my supervisor Prof Vincent Miles Gray for giving me a life changing opportunity to work in his research group, thank you Prof for the best supervision, support and knowledge.

Thanks to Gauteng Department of Agriculture and Rural Development (GDARD) (GDARO 12 TS) for funding the research project.

Thanks also for the Innovation Doctoral Scholarship from the NRF National Research Foundation (NRF) with the grant number [SFH1208147793].

Wits Postgraduate Merit Award (PMA) from the University Witwatersrand.

Thanks to the Agricultural Research Council (ARC) Biotechnology platform for Illumina technology sequencing services.
A special thank you goes to Mr Jonathan Featherston for his assistance with bioinformatics

Thanks to the Microscopy and Microanalysis Unit at University Witwatersrand for assistance with microscopy

Thanks to the Sydney Brenner Institute for Molecular Bioscience for assistance with bioinformatics. A special thank you goes to Mr Phelelani Mpangase and Mr Shaun Aron for their great help with bioinformatics. Thank you so much Prof Scott Hazelhurst for allowing us to use the computing resources.

~Thank you family, friends and collegues~

~Most of all, thank you God~

Contents

CHAPTER 1

Literature Review

1. Introduction

1.1 Entomopathogenic nematodes and taxonomy

Nematodes are microscopic worms commonly known as eelworms or roundworms (Flint *et al*, 1998). Entomopathogenic nematodes belong to the Animal Kingdom, Phylum Nematoda, Class Chromadorea, Order Rhabditida, Family Rhabditinae which is a species rich (Blaxter *et al.,* 1998). Together with the Phyla Mollusca and Arthropoda, nematodes are one of the most abundant groups of animals. As invertebrates they are multicellular, triploblastic, pseudocoelomic, non-segmented, bilaterally symmetrical, colourless, slender worms which exist almost everywhere in nature. Anatomically nematodes have a tubular shaped body structure which is covered with a tough, elastic, flexible and complex outer cuticular integument. The cuticle consists of many layers of proteinaceous fibres (Flint *et al*, 1998). Beneath the cuticle is a hypodermis and a layer of longitudinal muscles (they do not have circular muscular system). The longitudinal muscles surround the pseudocoelomic body cavity which functions as a hydrostatic skeleton. The action of the longitudinal muscle against the hydrostatics skeleton results in the generation of the typical whip-lash motion of nematodes. The nervous system consists of a circumpharyngeal nerve ring with ventral and dorsal nerve cords running the length of the body. Nerves extending from the circumpharyngeal ganglia also enervate the mouth, lips and anterior sense organs.

Nematodes are very diverse, with entomopathogenic nematodes (EPNs) receiving a lot of interest in nematological and entomological studies because of their ability to infect and kill insects (Hatting *et al*., 2009). EPNs reside naturally in the soil and are obligate parasites of a wide variety of different insect species. They have evolved a symbiotic relationship with insect pathogenic bacteria (Malan *et al*., 2006). There are three genera of EPNs which have been identified and reported so far; *Heterorhabditis*, *Steinernema*, and the newly discovered *Oscheius* genus. Their ability to kill insects is due to their symbiotic association with pathogenic bacteria belonging to the genera *Photorhabdus*, *Xenorhabdus*, and *Serratia*, respectively (Abebe *et al*, 2010; Adams *et al*., 2006; Torres-Barragan *et al*, Pervez *et al*, 2013).

EPNs have a broad host range and express several differences with respect to the types of host they infect, symptoms they cause, reproduction and infectivity (Alia *et al*., 2014). Other variations are seen in their ability to tolerate desiccation, osmotic stress, extreme temperature tolerance, ultra violet radiation tolerance and response to lack of oxygen (Grewal *et al*, 2006).

1.2 EPNs as biological control agents

The application of synthetic chemical pesticides to control insect pests in agricultural industries is linked with various disadvantages which include crop contamination, non-specific killing of beneficial insects, chemical resistance development in insects and human health related issues identified since the 1960's and 1970s (Smart, 1995). In order to minimise chemical contamination of crops and nonspecific insect death, biological control agents of insects were identified and studies to isolate and characterise more species are presently being conducted. EPNs were found to be the most suitable natural enemies of problematic insects because they impose no risk to humans and other related vertebrates (Athanassiou *et al*., 2010; Campos-Herrera *et al*., 2012). EPNs may be used for controlling insect pests in small-scale and large-scale farming, forestry and diverse agricultural industries and even in home gardens. The Agricultural Research Council (ARC) in South Africa has identified several problematic insect pests which destroy leaves and stems of plants. For example, the maize aphid feeds on leaves and thus causing damage of the plant. This is mainly observed in summer and winter rainfall seasons throughout the country. Another example identified by the ARC is the aphid which parasitizes wheat harvested under irrigation in the entire country and in dry lands of Western Cape during the rainfall season. These bugs are still controlled by chemical insecticides but alternatively, EPNs were tested for their potential to kill them. *Heterorhabditis bacteriophora* was identified as one of the biological control agents which may be to control *Ceratitis capitata* however, the EPNs potential were tested only under laboratory conditions and found to be also effective in causing mortality of these bugs (Malan *et al*, 2009). *Heterorhabditis zealandica* was tested for its ability to control *Planococcus citri,* the *citrus mealybug* (Sonnica *et al*, 2012).

EPNs are selected as biological agents based on their ability to locate, distinguish and invade insect hosts in the soil using foraging strategies such as ambushing (on soil surface) and cruising (into the soil) (Koppenhofer *et al*, 2006). Information on the foraging behaviour of nematodes is thus also useful in developing and improving their application in environments and industries where they are

needed as biocontrol agents, also bearing in mind appropriate application technology (Dillion *et al*, 2006; Shapiro-Ilan *et al*, 2006).

1.3 EPNs life cycle

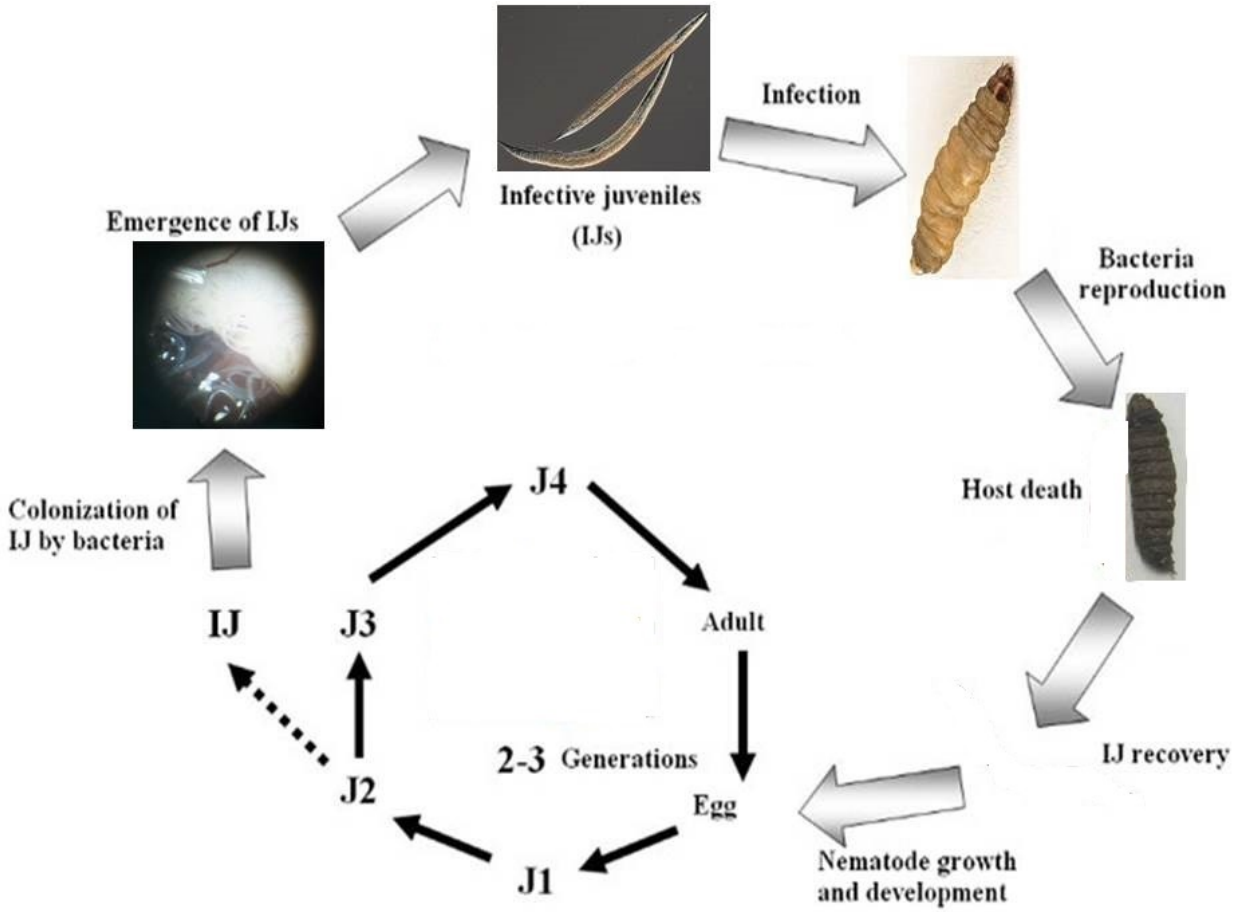


Fig. 1 Characteristic EPNs life cycle

The infective juveniles (IJs) of EPNs are free-living, soil dwelling and capable of infecting and killing susceptible insects. IJs enter the insect host through the mouth, anus or any other natural opening and then release symbiotic bacteria into the hemocoel of the insect (Adams *et al*., 2006). During the infection process it has been observed that *Steinernema* IJs mature into either feeding males or amphimictic females, while *Heterorhabditis* and *Oscheius* IJs develop into hermaphrodites later producing males, hermaphroditic females, and amphimictic females in the second generation (Nguyen *et al*, 1996; Pervez *et al*, 2013; Zhang *et al*., 2008). These pathogenic bacteria belonging to the *Enterobacteriaceae* secrete toxins and targeted immune depressors that suppress the insect's immune system resulting in death with 24-48 hours (Adams *et al*., 2006). Due to favourable

nutrition conditions, nematodes continue to proliferate inside the insect cadaver feeding on bacteria and later emerge following nutrient depletion into the soil to search for new hosts to infect (Sandhu *et al*, 2006; Shapiro-Ilan *et al*., 2012).

1.4 *Oscheius* genus

In 1976 the *Oscheius* genus was established as a subgenus of *Rhabditis* (Andrassy, 1976) and *Oscheius insectivorus* was identified as a type of species falling under the *Oscheius* genus. The *O. insectivorus* species was however not reported to have a symbiotic relation with pathogenic bacteria. The *Oscheius* genus was also put into a compartment with the family *Rhabditidae* (Sudhaus, 1976). As the taxonomic investigations continued, the *Oscheius* genus was acknowledged as an independent genus (Tabassum *et al*, 2002) and later (Ali *et al*, 2011) supported the recognition of *Oscheius* as an autonomous genus; however the evidence to these findings were not clearly stated. Examples of described *Oscheius* species include *O. maqbooli* (Tabassum *et al*, 2002), *O. shamimi* (Tahseen *et al*, 2006), *O. carlianonsis* (Weinin *et al*, 2012) and *O. amsactae* (Ali *et al*, 2011). Not all nematodes belonging to the *Oscheius* genus are described as entomopathogenic (Pervez *et al*, 2012).

Non-EPNs *Oscheius* species have been used in several laboratories for vulval developmetal studies (Sommer, 2000). *Oscheius* species were found to be different from *Caenorhabditis* species which are closely related *C. briggsae*. One of the reasons highlighted for the dissimilarities include observations made on variants in vulva patterning and development due to variations that occurred at genetic level (Delattre *et al*,2001) as seen in *Oscheius* sp. CEW1 (Sommer, 2000).

The *Oscheius* sp. CEW1 was identified as the third *Oscheius* species used for developmental biology and genetics studies alongside *C. elegans* and *Pristionchus pacificus* (Sommer, 2000). This information also supports some differences in morphological features and phylogenetic history as shown in chapter 2 of the research. Even though the *Oscheius* genus shares some common morphological and genetic features with *C. elegans*, this study further aims to introduce a novel *Oscheius* species, characterise it and take first steps at identifying genetic features by utilizing next generation sequencing and genome annotation. It is important to also bear in mind that *C. elegans* are free-living soil nematodes but not entomopathogenic like nematodes belonging to *Heterorhabditis* and *Steinernema* which have a free-living infective juvenile stage (Shapiro-Ilan *et al*, 2014). It has been stated that some nematodes within the *Oscheius* genus also have this free-

living infective juvenile stage which carries pathogenic bacteria belonging to the *Serratia* genus and therefore qualifying these worms as entomopathogenic. However, note that not all entomopathogenic *Oscheius* nematodes have symbiotic relations with *Serratia* bacteria only. For example *Oscheius Carolinensis* was found to be associated with four bacteria, *Serratia marcescens* found on its cuticle and caused the nematodes to be entomopathogenic. The *Enterococcus mundti* was found to also cause pathogenicity however, *Achromobacter xylosoxidans* was unable to cause death in the model insects tested. Lastly *Providencia rettgeri* was found mainly in mechanically homogenised surface sterilized nematodes and also caused some mortality of model insects (Torres-Barragan *et al*, 2011). In quest to study properties of the novel nematode species isolated in this study the entomopathogenic potential of *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae), was tested on two model insects larvae *Galleria mellonella* and *Tenebrio Molitor* also referred to as the greater wax moth larvae and yellow wheat larvae, respectively. Outcomes of the investigation are discussed in chapter 2 and the EPNs association with bacteria is described in chapter 3.

Recent discoveries and reports show that a small number of *Oscheius* species have been described as EPNs and examples include *O. carlianonsis* (Weinin *et al*. 2010), *O. siddiqii, O. niazii* (Tabassum *et al*, 2010) and *O. amsactae* (Ali *et al*. 2011) as listed in (Pervez *et al*, 2012). In our study, we also report a new *Oscheius* species and found to have association with insect pathogenic *Serratia*.

The genus *Serratia* comprises of several species which have varying lifestyles and thus exist as insect pathogens, human pathogens, some are free-living in the soil, and some are associated with plants and fungi. *Serratia* species may further be classified as opportunistic pathogens and obligate intracellular endosymbionts (Burke *et al*, 2011). The majority of species belonging to *Serratia* were found to produce powerful extracellular enzymes such as lipases, proteases, nucleases and hemolysin (Braun *et al*, 1980). Similar active extracellular enzymes were identified in insect pathogenic bacteria belonging to the genus *Xenorhabdus* and *Photorhabdus* associated with entomopathogenic nematodes *Steinernema* and *Heterorhabditis*, respectively (Burnell *et al*, 2000;Bennett *et al*, 2005).

Serratia have made successful relationships with some nematodes belonging to *Caenorhabditis* and *Oscheius* species (Abebe *et al*, 2010; Abebe *et al*, 2011). For example, *Serratia* species SCBI associated with *Caenorhabditis briggasae* isolated from South Africa in the KwaZulu Natal province was found to have entomopathogenic potential as it caused mortality of the insect larvae

Galleria mellonella. Another example includes *Serratia marcescens* DB11 which was identified and established as pathogenic to invertebrates (Abebe-Akele *et al*, 2015).

Our study took a genomics approach to explore whole genomic DNA of *Serratia* isolated from *Oscheius* species collected from Suikerbosrand Nature Reserve near Johannesburg in South Africa. Next generation sequencing was used to sequence the genome of the bacteria and nematodes and several bioinformatics tools were employed for genome quality control, assembly and annotation.

1.5 The EPNs genome

While nematodes have a fairly basic and simple anatomical structure they have a similar number of protein coding genes (approximately 20 000) when compared with the more anatomically complex vertebrates. They have been reported to contain many novel genes which are peculiar to their specific genera and this diversity at molecular level has become apparent across all the sequenced genomes of the different nematode genera including *Heterorhabdistis*, *Steinernema* and *Oscheius*. *Caenorhabditis elegans* has 6 chromosomes and its full or complete genome sequence size is 100.2 Mb (Mitreva *et al*, 2005). Previously, a cDNA-sequencing project of the EPNs *Heterorhabditis bacteriophora* was reported in 2006 in which 1246 expressed sequence tags were found and annotated resulting in 1072 useful ESTs and further categorised by function using *Caenorhabditis elegans* as a reference genome (Hao *et al*,2010; Sandhu *et al*, 2006).

1.6 Why next generation sequencing?

Sequencing is simply defined as finding the sequence of DNA or RNA molecules. In 1958, Fred Sanger sequenced proteins and later in 1980 he successfully sequenced RNA. This is how the Sanger sequencing method got enormous attention in the scientific market and elicited interest to sequence genomic DNA of several organisms such as tomatoes, mice, fruit flies and the human whole genome. It took about 11 years to complete sequencing the human genome at the high cost of 2.7 billion dollars using Sanger sequencing. The diagram below shows a typical Sanger sequencing workflow which was widely used to sequence short genomic regions and whole genomes. The method is still used today and it is effective for sequencing of short fragments of DNA.

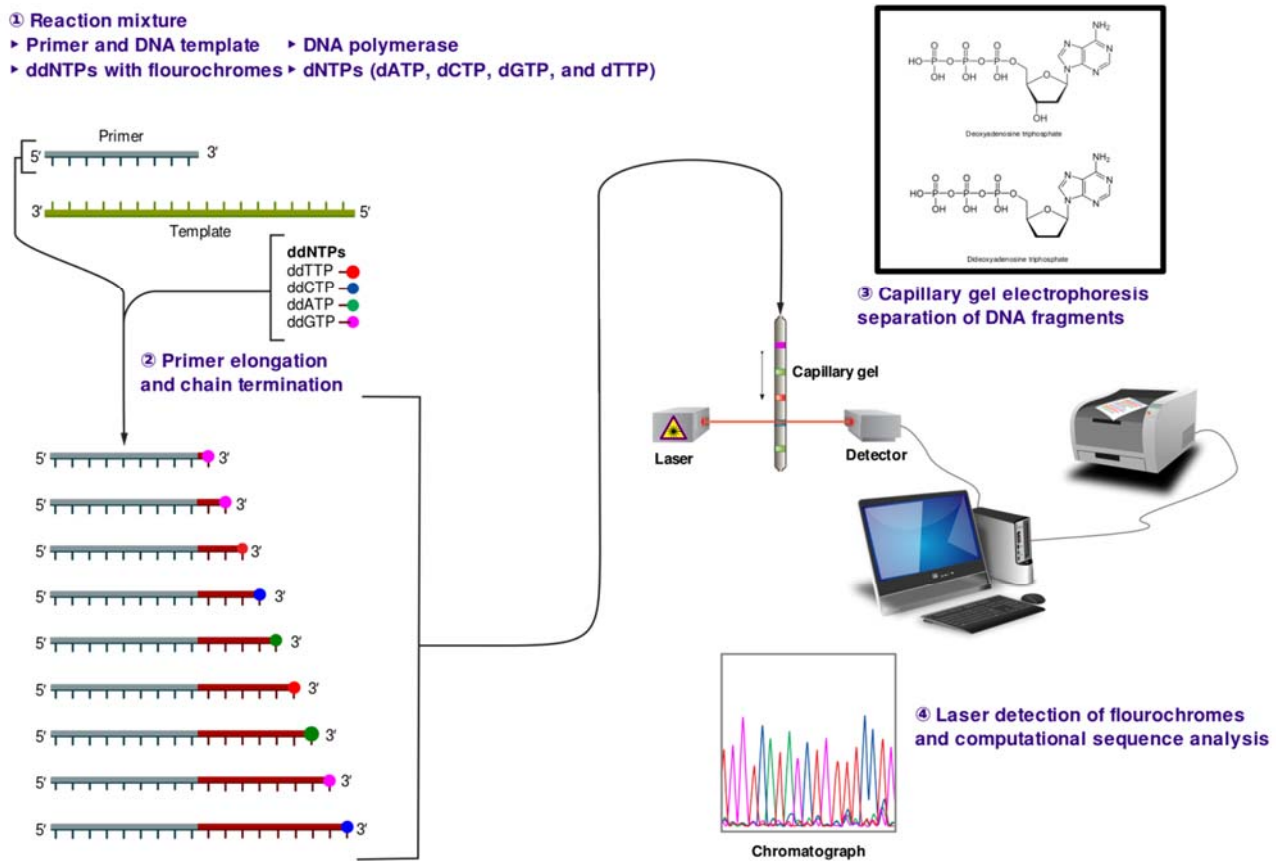


Fig. 2 Overview of Sanger sequencing workflow (image courtesy of en.wikipedia.org)

Today next generation sequencing (NGS) tools have made it possible to sequence human genomes in about 24 hours. Next generation sequencing has the capacity to process millions of reads in parallel rather than the Sanger methods which processes 96 sequence reads at a time. This has great cost implications because NGS can be performed at low costs and quicker. For example, 454 by the company Roche can sequence approximately 1 million reads per run with a read length of up to 1000bp and Hiseq by Illumina can sequence nearly 0.5 billion reads per run each with a read length of 100-150bp in just a few hours or weeks depending on the size of the genome.

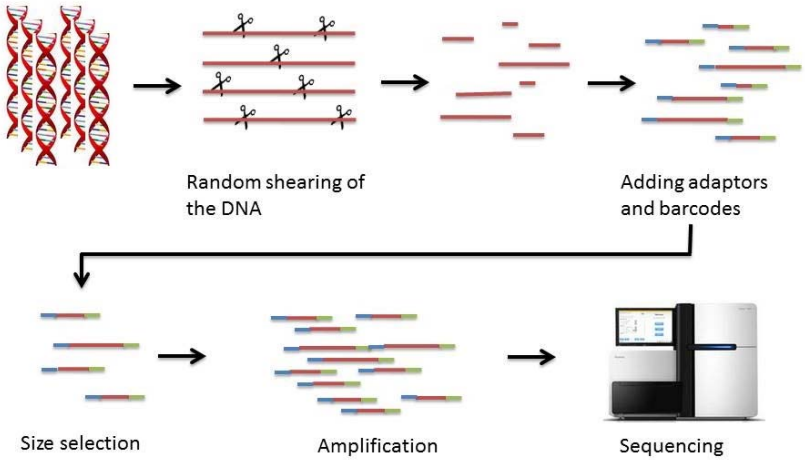


Fig. 3 Next generation sequencing general workflow.

Each system works differently, but they have some similar steps which include shearing of target DNA into small pieces, binding individual fragments to a solid surface, amplifying each molecule into a cluster, copying one base at a time and detecting various signals for A, C, T, and G bases as seen in fig. 3.

The power to sequence large genomes at affordable costs and less time motivated our study to take a genomics approach and sequence whole genomes of entomopathogenic nematodes and their associated symbiotic bacteria. In our research, we therefore generated the following aims and objectives:

1) To isolate, identify and characterise a novel nematode species
2) To isolate and identify bacteria associated with the nematode
3) Morphological analysis of nematodes using light microscopy and scanning electron microscopy
4) To perform phylogenetic analysis of the nematodes' 18S rDNA region
5) To perform phylogenetic analysis of the bacteria 16S rDNA region
6) To test the nematodes pathogenicity
7) To investigate the location of the pathogenic bacteria, whether they dwell in the nematode or on the surface of the nematode and test efficacy of surface sterilised nematodes
8) To perform inbreeding of nematodes
9) To sequence the whole genomic DNA of an insect pathogenic strain of bacteria associated with the identified nematodes species
10) To sequence whole genomic DNA of novel nematodes species
11) To assemble the bacteria genome

12) To assemble the nematodes genome
13) To mask repeats on sequenced and assembled nematodes genome
14) To annotate the genome of bacteria
15) To annotate the genome of nematodes using bioinformatics tools for gene prediction

The above aims and objectives were designed bearing in mind that bioinformatics will be extremely essential in whole genome data analysis and determining the meaning of the results obtained. Sequencing may be seen as a read-out for diverse types of experiments which are conducted in the laboratory. The challenge for some biologists is learning how to use informatics tools for NGS and also choosing suitable tools for analysis. Bioinformatics tools change promptly whereby new file formats, new data types and new methods are developed to cater for people in the science community who demand high-quality computational tools for sequencing, genome assembly and annotation. Bioinformatics tools are needed for aligning sequenced reads to available reference genome(s), de novo assembly of the sequence, quality control and genome annotation.

In this study, it was anticipated that de novo assembly methods will be used because one of the objectives included sequencing of a novel nematode species and its bacterial symbioint. De novo sequencing encompasses assembling overlapping reads to form contiguous sequence of DNA and it is usually performed if there is no available reference genome or genomic information available. Examples of assembly tools include SPADES for prokaryotic genomes and VELVET, SOAP DE Novo and ALLPATHS usually used for eukaryotic genomes.

Prior genome assembly, a critical step is taken to check the quality of the sequenced reads mainly because it is expected that sequencing machines may make mistakes especially on the ends of the sequence. Quality control tools were developed to ensure that sequenced data is of good quality to ensure meaningful downstream analysis. An additional step called trimming may be used were necessary to clip off poor quality reads. For example, FASTQC tool is generally used for quality control and it may be used to rescue bad quality data and thus avoiding having to sequence the genome again. Once the sequences are of good quality, they may be annotated to identify features on the genome using specific annotation tools and biological knowledge already existing in explicit genome databases. Sequence features which may be revealed by annotations tools include transcribed regions (mRNA, tRNA and rRNA), structural regions (promoter, introns and exons), conserved regions and repeats.

Gene finding or gene prediction is also important in this study to help us predict exon structures for the primary transcript of a gene. In less complicated gene prediction process, genes are simply

identified and may be linked to their specific function if the information is readily available in annotation databases. Examples tools include Gene finder, Augustus and Glimmer and may be used define coding segments of the sequenced genome.

This research embarked on a journey to isolate and identify entomopathogenic nematodes and characterise them using molecular and morphological methods, identify bacteria associated with the nematodes and test them for their insect pathogenic capabilities. Multiple sequence alignment and phylogenetic analysis of the nematode and bacteria was conducted using 18S rDNA and 16S rDNA sequences respectively to study evolutionary relations of the isolates. Powerful and new sequencing technology and methods were used to sequence whole genomes of the identified nematodes and symbiotic bacteria, followed by bioinformatics applications to analyse the sequenced genome. Nematodes were also surface sterilised to confirm the location of pathogenic. Outcomes of this study may have implications on the possibility of using this nematode species as an effective biological control agent; however more studies will need to be conducted to come to such conclusions.

2 References

Abebe-Akele F, Tisa L.S, Cooper V.S, Hatcher P.j, Abebe E and Thomas W.K, 2015. Genome sequence and comparative analysis of a putative entomopathogenic *Serratia* isolated from *Caenorhabditis briggsae*. BMC Genomics, 16:531.

Abebe E, Akele F.A, Morrison J, Cooper V, Thomas W.K, 2011. An insect pathogenic symbiosis between a *Caenorhabditis* and *Serratia*. Virulence, 2(2):158–61.

Abebe E, Jumba M, Bonner K, Gray V, Morris K, Thomas W.K, 2010. An entomopathogenic *Caenorhabditis briggsae*. J Exp Biol, 213(18):3223–9.

Adams, B.J, Fodor A, Koppenhöfer H.S, Stackebrandt E, Stock S.P, and Klein M.G, 2006. Biodiversity and systematics of nematode – bacterium entomopathogens. Biol. Control, 37:32–49.

Ali S.S, Pervez R, Andrabi R, Sharma R, Verma V, 2011. *Oscheius amsactae* n. sp. (Nematoda:Rhabditida), a necromenic associate of red hairy caterpillar, *Amsacta moori* (*Lepidoptera:Arctiidae*) from Kanpur, India. Arch Phytopathol Plant Prot, 44(9):871–881.

Alia V.I.K, Harma G.A.S, Lan D.A.I.S.H, and Anguly S.U.G, 2014. Biocontrol Potential of Steinernema thermophilum and its Symbiont Xenorhabdus indica Against Lepidopteran Pests : Virulence to Egg and Larval Stages. Journal Nematology, 46:18–26.

Andrassy I, 1976. Evoluation as a basis for the systematization of nematodes. London: Pitman.

Athanassiou C.G, Kavallieratos N.G, Menti H, and Karanastasi, 2010. Mortality of Four Stored Product Pests in Stored Wheat When Exposed to Doses of Three Entomopathogenic Nematodes. J. Econ Entomol, 103(3):977–984.

Bennett H.P.J, Clarke D.J, 2005. The pbgPE operon in *Photorhabdus luminescens* is required for pathogenicity and symbiosis. J Bacteriol, 187(1):77–84.

Braun V, Schmitz G, 1980. Excretion of a protease by *Serratia marcescens*. Arch Microbiol, 124(1):55–61.

Burke G.R, Moran N.A, 2011. Massive Genomic Decay in *Serratia* symbiotica, a Recently Evolved Symbiont of Aphids. Genome Biol Evol, 3:195–208.

Burnell A.M, Stock S.P, 2000. *Heterorhabditis, Steinernema* and their bacterial symbionts - lethal pathogens of insects. Nematology, 2:31–42.

Campos-Herrera R, Barbercheck M., Hoy C.W, and Stock S.P, 2012. Entomopathogenic nematodes as a model system for advancing the frontiers of ecology. J. Nematol, 44:162–176.

Delattre M and Fe´ lix M, 2001. Polymorphism and evolution of vulval precursor cell lineages within two nematode genera, *Caenorhabditis* and *Oscheius.* Current Biology, 11:631–643.

Dillon A.B, Ward D, Downes M.J, Griyn C, 2006. Suppression of the large pine weevil *Hylobius abietis* (L.) (*Coleoptera:Curculionidae*) in pine stumps by entomopathogenic nematodes with different foraging strategies. Biological Control, 38: 217–226.

Flint M.L and Dreistadt S.H, 1998. Natural enemies' handbook, the illustrated guide to biological pest control 2-35.

Grewal P.S, Bornstein-Forst S, Burnell A.M, Glazer I, Jagdale G.B, 2006. Physiological, genetic, and molecular mechanisms of chemoreception, thermobiosis, and anhydrobiosis in entomopathogenic nematodes. Biological Control, 38: 54–65.

Hatting J, Stock P.S, and Hazir S, 2009. Diversity and distribution of entomopathogenic nematodes (*Steinernematidae, Heterorhabditidae*) in South Africa. J. Invertebr. Pathol, 102:120–128.

Hao Y, Montiel R, Abubucker S, Mitreva M, and Simões N, 2010. Transcripts analysis of the entomopathogenic nematode *Steinernema carpocapsae* induced in vitro with insect haemolymph. Molecular & Biochemical Parasitology, 169:79–86.

Koppenhofer A.M and Fuzy E.M, 2006. Effects of soil type on infectivity and persistence of the entomopathogenic nematodes *Steinernema scarabaei*, *Steinernema glaseri*, *Heterorhabditis zealandica*, and *Heterorhabditis bacteriophora*. Journal of Invertebrate Pathology, 92:11–22.

Malan A.P, Nguyen K.B, and Addison M.F, 2006. Entomopathogenic nematodes (*Steinernematidae* and *Heterorhabditidae*) from the southwestern parts of South Africa. African Entomol, 12:65–69.

Malan A.P and Manrakhan A, 2009. Susceptibility of the Mediterranean fruit fly (*Ceratitis capitata*) and the Natal fruitfly (*Ceratitis rosa*) to entomopathogenic nematodes. Journal of Invertebrate Pathology, 100:47–49.

Mitreva, M., Blaxter, M.L, Bird, D.M, and McCarter, J.P, 2005. Comparative genomics of nematodes. Trends Genet, 21:573–581.

Niekerk S and Malan A.P, 2012. Potential of South African entomopathogenic nematodes (*Heterorhabditidae* and *Steinernematidae*) for control of the citrus mealybug, *Planococcus citri* (*Pseudococcidae*). Journal of Invertebrate Pathology, 111:166–174.

Nguyen K.B and Smart G.C, 1996. Identification of entomopathogenic nematodes in the *Steinernematidae* and *Heterorhabditidae*. Nematology, 28: 286-300.

Pervez R, Eapen S.J, Devasahayam S and Jacob T.K, 2012. A new species of entomopathogenic nematode *Oscheius gingeri* sp. n. (Nematoda: *Rhabditidae*) from ginger rhizosphere, Archives of Phytopathology And Plant Protection, 46:526-535.

Sandhu S.K, Jagdale G.B, Hogenhout S.A, and Grewal P.S, 2006. Comparative analysis of the expressed genome of the infective juvenile entomopathogenic nematode, *Heterorhabditis bacteriophora*. Molecular & Biochemical Parasitology, 145:239–244.

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

Shapiro-Ilan, D.I., Han, R., and Dolinksi, C. (2012). Entomopathogenic nematode production and application technology. J. Nematol. *44*, 206–217.

Shapiro-Ilan D.I., Han R and Qiu X, 2014. Chapter 10 – Production of Entomopathogenic Nematodes. Invertebrates and Entomopathogens,321–355.

Smart, G.C. (1995). Entomopathogenic nematodes for the biological control of insects. Journal of Nematology. 27: 529-534.

Sommer R.J, 2000. Comparative genetics: A third model nematode species. Current Biology 2000, 10:R879–R881

Sudhaus W, 1976. Verg leichende untersuchngen assignment untersuchngen to phylogeny, systematic , ecology , biology and ethology of Rhabditidae (Nematoda). Zoologica, 43:1–229.

Tabassum K.A, Shahina F, 2002. *Oscheius maqbooli* n. sp. and observations on three known rhabditid species (Nemata: *Rhabditida*) from sugarcane fields of Balochistan, Pakistan. Pakistan J Nematol, 20:1–21.

Tabassum K.A, Shahina F, 2010. *Oscheius siddiqii* and *O. niazii*, two entomopathogenic nematode species from Pakistan with observations on *O. shamimi*. Int J Nematol, 20(1):75–84.

Tahseen Q, Nisa S.U, 2006. Embryology and gonad development in *O. shamimi* sp. n. (Nematoda:*Rhabditida*). Nematol, 8:211–221.

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.

Weimin Y.E, Barragan A.T, Cardoza Y.Z, 2010. *Oscheius carolinensis* n. sp. (Nematoda: *Rhabditidae*), a potential entomopathogenic nematode from vermicompost. Nematol. 12 (1):121–135.

CHAPTER 2

Oscheius sp. TEL-2014 (Nematoda: Rhabditidae), a potential entomopathogenic nematode isolated from a grassland in South Africa

Abstract

Oscheius sp. TEL-2014 was recoverd from a soil sample collected from grassland at the Suikerbosrand Nature Reserve near Johannesburg in South Africa using the *Galleria* bait method. Morphological studies with light and scanning electron microscopy, also including molecular analyses of the 18S rDNA gene was performed and revealed that this a novel species, described herein as *Oscheius* sp. TEL-2014. Amplification and Sanger sequencing of the 18S ribosomal DNA placed the nematode in genus *Osheius*. The 18S rDNA was deposited on NCBI Genbank and assigned the accession number, KM492926. The new species is characterised by various traits including its original 18S rDNA sequence, six lips and didelphic reproduction. The adult male was diagnosed by the presence of six lips, labial papillae on the head and amphidial openings. Males were also characterised by the presence of spicules towards their posterior end and short blunt tails. Females were distinguished by a didelphic reproductive system, a vulva opening, a body with irregular ridges and long pointed tails. The males closely resemble *O. andrassyi* but were shorter in length (vs. L= 1601µm), the pharynx length was longer (vs.L=191 µm), lines (vs. = 4) and male body posture was slightly curvilinear in some adults. The females resembled *O. carolinensis* but had shorter body length (vs.L=1728 µm) and a longer pharynx (vs.L=247 µm). The infective juvenile body width was thin and had a cuticle sheath. Entomopathogenicity of this nematode was confirmed against model insect hosts *Galleria mellonela* and *Tenebrio molitor* larvae. The infective juveniles of this new species are able to invade both model insect larvae and cause mortality within 24-72 hours after inoculation under laboratory conditions. These infective juveniles act as vectors for an insect pathogenic bacteria causing death of model insects used.

Keywords: Entomopathogenic nematodes; *Oscheius* species, description, molecular, 18S rDNA sequencing, morphology

2. Introduction

The majority of entomopathogenic nematodes (EPNs) species that act as vectors for insect pathogenic bacteria belong to the families of *Steinernematidae* and *Heterorhabtidae*. These EPNs have evolved special symbiotic relationships with enterobacteria belonging to families of *Xenorhabdus* and *Photorhabdus*, respectively. The recently identified nematodes belonging to the genus *Oscheius* such as *Oscheius carolinensis* have been reported to act as vectors for insect pathogenic bacteria belonging to the genus *Serratia* (Torres-Barragan *et al*, 2011). In 2013, four species belonging to the genus *Oscheius* such as *O. carlianonsis*, *O. siddiqii*, *O. niazii* and *O. amsactae* were reported as EPNs (Pervez *et al*, 2013). Further studies on *O. carolinensis* showed that this nematode is associated with insect pathogenic strain of *Serratia marcescens*, which was found to be localized on the cuticle of the nematodes and subsequently identified as the main bacteria causing entomopathogenecity. *Providencia rettgeri*, mainly found in the ground-up homogenate of surface sterilized nematodes also contribute to insect pathogenicity of *O. carolinensis* (Torres-Barragan *et al*, 2011).

Like steinernematids and heterorhabditids the above species from the genus *Oscheius* also behave like EPNs in that they are capable of foraging for and locating their insect hosts in soil. Their symbiotic relationships with insect pathogenic bacteria are similar to that of steinernematids and heterorhabditids. The insect pathogenicity of certain strains of *Serratia marcescens* have been well established, and they are able to cause mortality within 24-48 hours of invading the hosts (Ye *et al*., 2010). These nematodes appear to have all the attributes of conventional EPNs and could also be used as biological control agents of insect pests; because they proliferate at a high rate, are cost effective in terms of production and they are not harmful to humans and plants (Torres-Barragan *et al*, 2011).

Other features that entomopathogenic *Oscheius* nematodes share with steinernematid and heterorhabditids include similar life cycle stages such as the non-feeding infective juveniles that is able to function as a vector for insect pathogenic bacteria belonging to the genus *Serratia*. These dauer larvae are able to withstand unfavourable conditions and are also referred to as third stage infective juveniles (IJs) because they exhibit a second stage cuticle which enables them to persist in the soil through anhydrobiosis until they infect new host and continue their life cycle (Ye *et al*., 2010). Unlike the other nematode developmental

stages that take place within the infected insect host, the IJs have the capacity to resist and tolerate harsh environmental conditions (Ma *et al*, 2013).

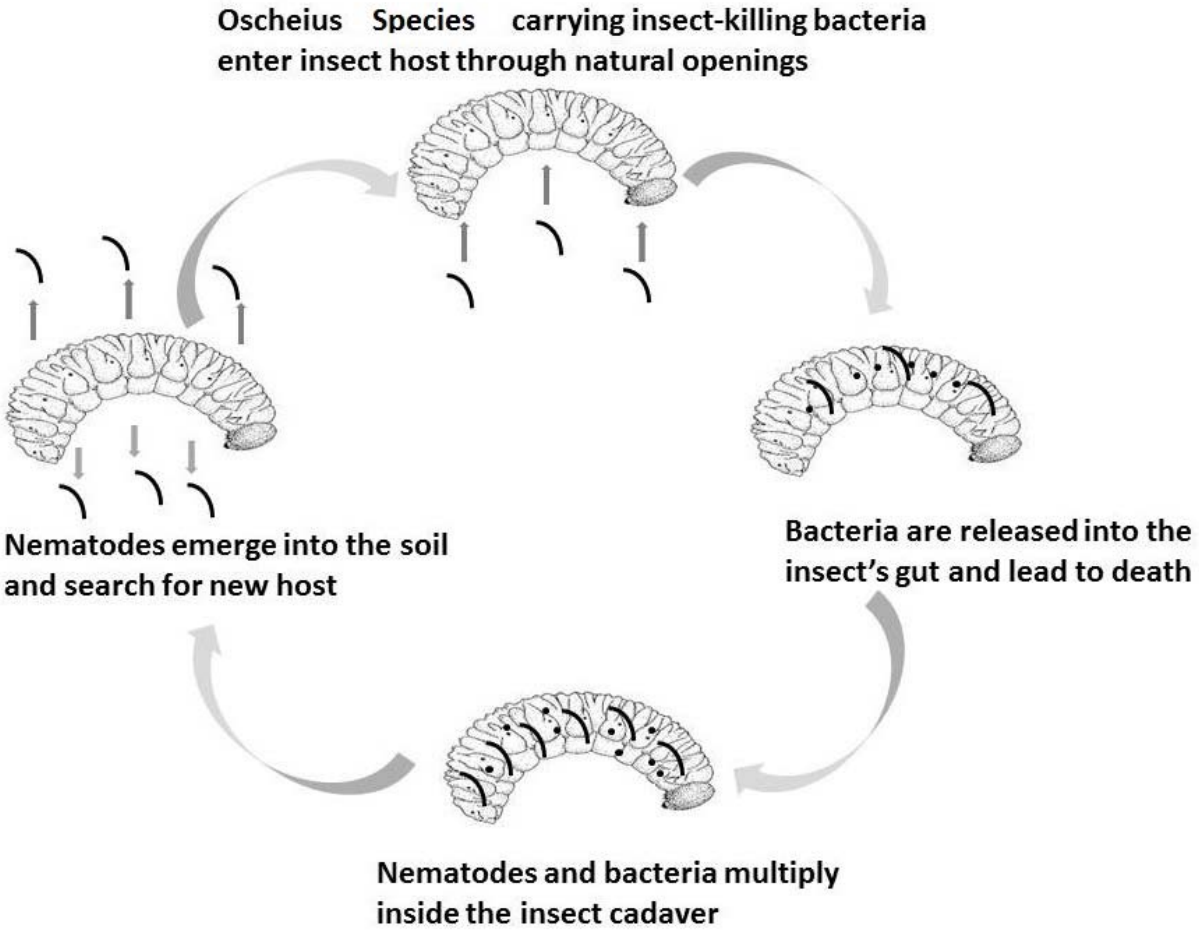


Fig. 1 A typical life cycle of entomopathogenic nematodes.

Here we report a novel entomopathogenic nematodes species *Oscheius* sp. TEL-2014 recently isolated from Suikerbosrand Nature Reserve near Johannesburg in South Africa. Light and electron scanning microscopy was employed for morphological characterisation of the adult nematodes and infective juveniles. Its 18S rDNA sequence was deposited into NCBI GenBank and assigned this accession number KM492926.

2. Materials and methods

2.1 Isolation of the nematodes

Soil samples were collected from Suikerbosrand Nature Reserve near Johannesburg in South Africa. They were collected to a depth of 15 cm and were placed into 1 L rectangular plastic tubs. The EPNs were isolated by baiting the soil collected with ten *Galleria mellonella*

larvae per tub (Bedding and Akhurst 1975). Infective juveniles that emerged from dead insect larvae were collected by using a modified White trap method (Kaya and Stock 1997).

2.2 Morphological analysis by Olympus light microscope

The Olympus light microscope was used to photograph the internal structures of adult males, adult females and IJs. Nematodes taken from White traps and lipid agar plates were killed in 60 °C hot water. Nematodes were placed in a Petri dish in 1ml distilled water and 3-4ml of triethnolamine formaldehyde (TAF) was added and left at room temperature for 24 hours, (recipes for the solutions used are in Appendix 1). TAF was replaced with TAF with double-strength TAF and store at 4°C to relax nematodes for one hour. TAF was added again to allow fixative to infiltrate for at least 24 hours and then most of the fixative was removed. This was then followed by processing nematodes with pure glycerine where fixed nematodes were transferred to a clean Petri dish containing 0.5ml of solution I which was allowed to slowly evaporate and stored at room temperature for 12 hours in an uncovered Petri dish. Solution II was added to the Petri dish with nematodes and left partially opened to allow for slow ethanol evaporation and thereafter it was placed in an oven preheated to 40°C for 3 hours making sure that nematodes do not dry out. Nematodes were then mounted on glass slides carefully to avoid crushing them before analysis under the microscope.

2.3 Morphological analysis by scanning electron microscopy

Lipid agar plates

Lipid agar plates contain tributyrin which is a substrate. If bacteria cultured on these plates are able produce the exoenzyme lipase it will have the potential to digest or hydrolyze the triglycerides/lipids in the tributyrin. This process results in development of clear zones around the growing colonies on the media. Lipid agar plates where used because they were ideal for growing lawns of symbiotic bacteria which were used to support growth of the nematodes. This method was also used to confirm that the isolated bacteria are a symbiont of the *Oscheius* nematodes by stimulating nematodes growth and reproduction. The objective of using this medium was also to allow IJs to grow into adults nematodes under suitable conditions. Lipids and glucose are some of the most imperative nutrients necessary for nematode growth, reproduction and bacterial proliferation (Gil *et al.,* 2002). This is why the lipid agar media used was composed of honey (glucose), yeast extract, nutrient agar, cod liver

oil (lipids), and $MgCl_2.6H_2O$ as seen Appendix 1. Some of these ingredients were also used in lipid media in studies conducted by (Shapiro-Ilan *et al*, 2002).

EPNs were suspended with sterile distilled water and 1ml of EPNs was collected directly from lipid agar plates and transferred into 1.5ml Eppendorf tubes and heat-killed at 80°C for 5 minutes. EPNs were rinsed with Ringer solution three times with 5 minutes between each rinse. EPNs were then fixed in 8% glutaraldehyde overnight (glutaraldehyde 25% EM grade, diluted in Ringers solution). EPNs were further rinsed with distilled water three times and dehydrated with 30, 50, 70, 90, 95, and 100% ethanol at 10 minutes interval sequentially. Each sample was allowed to air dry overnight and mounted on SEM stubs, coated with carbon and palladium, then scanned using FEI QUANTA 200 scanning electron microscope fitted with a digital camera.

2.4 Measurements

Fresh IJs and adult nematodes prepared using the method in 2.3 were placed on clean slides then covered gently with cover slips. The IJs were then viewed under an Olympus light microscope connected to a digital camera and images were captured. 10 males, 10 Females and 10 IJs were measured. Morphometric table was generated based on the mean and standard deviation statistical calculations.

The following abbreviations were used:

BL Body length
GD Greatest diameter
AEP Distance from anterior end to excretory pore
PL Pharynx length
CL Corpus length
IL Isthmus length
PVD Pharyngeal-intestinal valve diameter
TBD Terminal bulb diameter
AGL Length of anterior gonad
PGL Length of posterior gonad
RAL Rectal length
ABD Anal body width
TL Tail length

SL Spicule Length

CT Cuticle thickness

2.5 Molecular identification of EPNs by 18S rDNA Sanger sequencing

Genomic DNA was extracted from nematodes using the protocol from the Puregene® DNA Purification Kit, Gentra systems 2003. PCR amplification of 18S rDNA ITS region was performed. The following universal primers were used: forward primer; TW81 (5'-GCGGATCCGTTTCCGTAGGTGAACCTGC -3', Tm: 71.94 ºC), and reverse primer; AB28 (5'- GCGGATCCATATGCTTAAGTTCAGCGGGT-3', Tm: 68.87 ºC). Initial denaturation before cycling: 94°C for 5 minutes followed 25 cycle amplification series: denaturation at 95°C for 60 seconds, annealing at 64°C for 60 seconds, extension at 72°C for 120 seconds and final extension after cycling: 72°C for 10 minutes. PCR products were sequenced using Sanger sequencing method at Inqaba Biotechnical Industries (Pty) Ltd; South Africa using the above PCR primers. NCBI nucleotide database BLASTn tool was used identify the unknown sequences of EPNs. The 18S rDNA sequence for *Oscheius* sp. TEL-2014 (KM492926) was edited to derive the consensus using Bioedit sequence alignment editor. Phylogenetic analyses was done using the following: *Oscheius carolinensis* (FJ547241), *Oscheius* sp. MCB (KF684370), *Heterorhabditidoides chongmingensis* (KF500235), *Oscheius myriophilus* strain JU1386 (KP792651), *Oscheius* sp. KAT-2015 (KR119081), *Rhabditis* sp. Tumian-2007 (EU273598), *Oscheius* sp. BW282 (AF082994), *Oscheius insectivore* (AF083019), *Oscheius* sp. Pak.S.N.10 (KT878513), *Oscheius* tipulae CEW1 (KP792649), *Oscheius chongmingensis* Tumian (EU273598), *Oscheius dolichura* LDY30 (M355811), *Oscheius guentheri* SB133 (EU195996), *Oscheius dolichuroides* DF5018 (AF082998), *Oscheius* sp.DF5000 (AF082995) were obtained from NCBI GenBank database and *Caenorhabditis elegans* (Z92784) sequence was used to root the tree. Sequences were aligned first using MUSCLE on MEGA 6 software. The evolutionary history of the aligned sequences was centred on the analysis of 18S rDNA ITS region inferred by using the Maximum Likelihood method based on the Tamura-Nei model in MEGA6. The bootstrap consensus tree inferred from 1000 replications and the tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches).

2.6 Entomopathogenecity of *Oscheius sp. TEL-2014*

Nematode entomopathogenicity was evaluated using sixth instar stage of *G. mellonella* and *T. molitor,* respectively. For evaluated nematode infectivity *G. mellonella* and *T. molitor* larvae were each placed in 90 mm plastic Petri dishes containing sterile 40g of sandy loam soil, replicated four times for each larval species with each plate containing five larvae. The soil had an initial moisture content of 8% (w/w) and each plate was inoculated with 100 IJs suspended in 1 ml of sterile distilled water. The time interval for the onset of larval mortality after inoculation with IJs was monitored and recorded. Insect cadavers were placed on White-traps to monitor and recover the emerging IJs.

3. Results and discussion

Oscheius sp. TEL-2014

(Fig. 2-4)

Measurements

(See Table 1)

Description

Infective juveniles and adults of *Oscheius* sp. TEL-2014 have a finely annulated cuticle covering the body as shown in Fig. 2 and Fig 3. Scanning electron microscopy images in Fig. 5 show 6 unconnected lips with labial papillae and amphidial opening. The lateral field has irregular ridges which are visible on the anterior and posterior regions of the male and female adult nematodes. The corpus is tubular and occupies 30-40% of the pharynx length. The terminal bulb is present in males and females including IJs and has well-developed pharyngeal-intestinal valve. The isthmus forms 10–15% of pharynx length.

Male

The body is straight and some curved upon fixation. The testis is present on the posterior arm gonad. The body length is shorter than the female and it is covered with a thick cuticle layer.

Genital papillae were not observed. Spicules are thin and the tails are short and have blunt ends.

Female

The body of the adult female straightened after fixation. The reproductive system is didelphic. The uterus is well-developed. Endotokia matricida or intra-uterine birth causing maternal death was observed in most of the mature adult females.

Juvenile

Body is thin and covered by a thin cuticle layer. The pharynx was clearly visible comprising corpus, isthmus, terminal bulb and underdeveloped pharyngeal-intestinal valve. Reproductive structures are not observed.

Comparison and remarks

The female body length is longer than the male. The pharynx is longer in females than in males. In both sexes the pharynx is comprised of a corpus and an isthmus. The males closely resemble *O. andrassyi* but differs in smaller length (vs. L= 1601μm), the pharynx length was longer (vs.L=191 μm lines (vs. = 4) and male body posture is slightly curvilinear in some adults (Pervez *et al*, 2012). The females resemble *O. carolinensis* but differ in smaller body length (vs.L=1728 μm and a longer pharynx (vs.L=247 μm (Pervez *et al*, 2012).

3.1 Morphological analysis by compound light microscopy and Olympus light microscopy

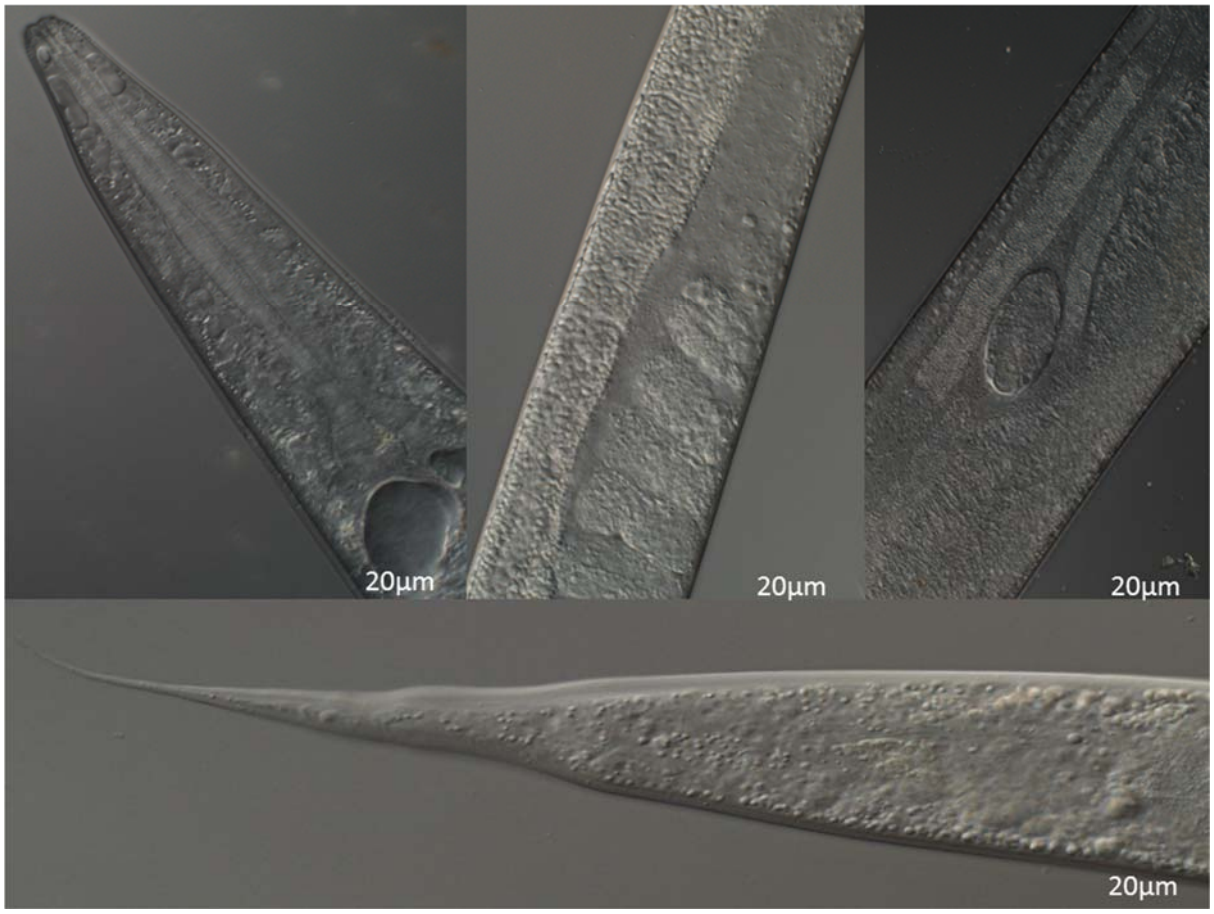


Fig. 2 Light photograph of the *Oscheius* sp. TEL-2014: Left: anterior region showing the pharynx of a female nematode, centre: mid-body showing IJs in the 3 fold stage of development and right: Mid-body showing the egg and some infective juveniles and bottom: posterior region showing the tail. A BX 63 Olympus light microscope was used under DIC settings.

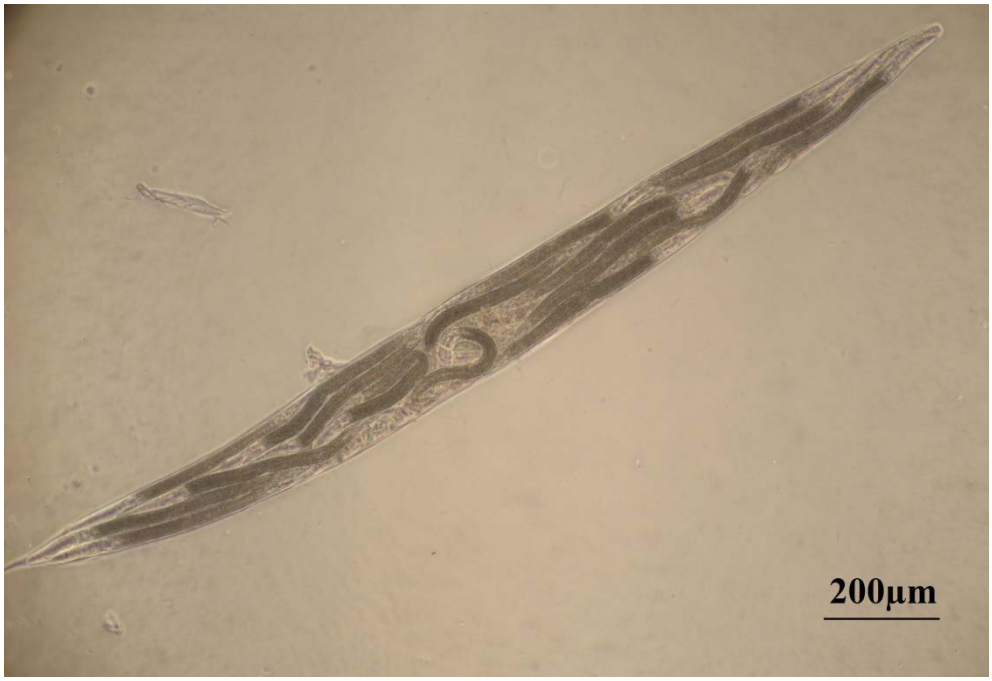


Fig. 3 Photographs of an *Oscheius* sp. TEL-2014 adult female showing A- Endotokia matricida which is an intra-uterine birth instigating maternal death (Stefan-Andreas and Ralf-Udo, 1999)

3.2 Morphological analysis by scanning electron microscopy

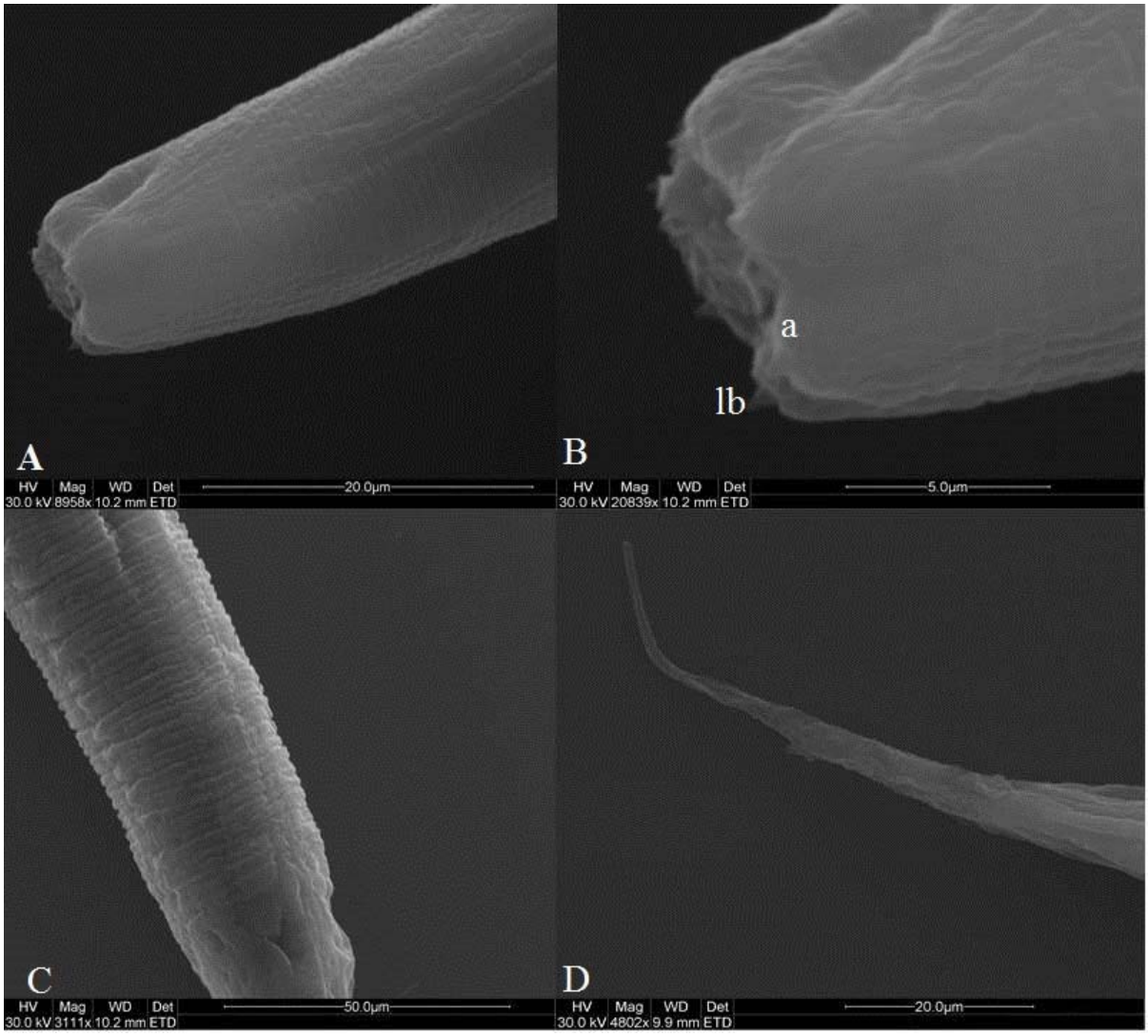


Fig. 4 Morphology of male *Oscheius* sp. TEL-2014 nematode. A: Anterior showing lateral ridges, B: Head showing six lips, labial papillae (lp) and amphidial opening (a), C: posterior of male showing the lateral ridges and D: male tail.

3.3 Measurements of *Oscheius* sp. TEL-2014

Measurements of various characters of the males, females and IJs of *Oscheius* sp. TEL-2014 are given in Table 1.

Table 1

Morphometrics of infective-stage juveniles of *Oscheius* sp. TEL-2014 and all measurements are done in µm presented as mean +/- SD (range).

Character	**Holotype male**	**Male**	**Female**	**Infective juvenile**
n	1	10	10	10
Body length	1874.45	1844.185+/-70.619(1689.25-1954.14)	3807.671+/-5709.163(1996.58-20056.22)	510.71+/19.179 (488.02-535.36)
Greatest diameter	108.63	108.364+/0.582(107.12-109.32)	158.768+/-2.098(154.25-161.24)	44.185+/2.279(40.3-46.96)
AEP	40.92	40.198+/0.531(39.24-40.99)	48.791+/-0.689(47.69-49.89)	10.547+/0.288(9.99-10.9)
Pharynx length	156.44	155.447+/1.432(152.87-157.9)	169.457+/-0.798(168.36-170.97)	164.917+/2.747(160.3-168.45)
Corpus length	79.88	79.321+/1.323(76.33-81.65)	86.438+/-0.744 (85.56-87.96)	66.426+/-4.844(60.1-71.36)
Isthmas length	76.56	75.955+/-1.071(74.11-78.3)	82.64+/-1.00(82.64-81.64)	42.347+/-2.142(40-45.67)
PVD	52.69	52.103+/-1.091(50.44-54.3)	60.332+/-1.038(58.9-62.54)	14.708+/0.717(13-15.39)
TBD	58.62	58.268+/0.997(56.12-60.1)	66.411+/-0.986(64.59-68.1)	21.478+/-1.46(19.3-23.48)
AGL	356.55	356.719+/-0.96(355.98-358.69)	412.335+/-1.521(409.2-414.89)	154.61+/-11.141(140.2-165.95)
PGL	352.46	351.914+/-1.379(350.14-355.3)	409.88+/-5.08(409.88-405.08)	164.915+/3.582(160.2-169.25)
RAL	76.96	76.242+/1.345(74.21-79.36)	88.468+/0.968(87.25-90.57)	31.399+/-1.321(29.9-33.9)
Distal gonad length	79.68	79.129+/-1.194(76.22-80.55)	90.637+/-1.464(89-93.89)	41.342+/- 1.291(39.2-42.9)
ABD	46.25	46.566+/-0.895(45.12-48.3)	54.846+/-1.404 (53.98-58.31)	34.135+/1.323(32.5-36.45)
Tail length	60.5	60.556+/-2.041(55.9-63.7)	87.673+/-1.222(86.89-90.58)	21.96+/-1.262(19.99-23.51)
Spicule	49.62	48.555+/-0.774(47.02-49.62)		
CT	28.56	8.324+/-0.432(7.6-8.9)	7.551+/-0.682(6.84-8.97)	3.56+/-1.75(3.25-10.00)

Table 2

Comparative morphometrics of O. sp TEL-2014 with other known species of Oscheius (all measurements in μm).

	***O.* sp TEL-2014**	***O. shamimi***	***O. columbiana***	***O. necromena***	***O. amsactae* n. sp.**
Female					
L	1996.58-20056.22	760–1524	923–1805	830–1500	658.1–786.1
b		4.2–6.3	5.2–8.0	4.2–6.3	4.1–4.9
c		6.8–13.8	8.3–10.0	9.7–13.9	8.9–12.1
Stoma length	39.24-40.99	19–23	21–28	14–18	15.8–18.1
Stoma width		4.6–5.7	4.5	4.5	3.5–3.9
abd	53.98-58.31	21–32	22–38	45	15.8–17.4
Male					
L	1689.25-1954.14	938–1118	665–1163	671–950	594–804
b		5.1–5.5	3.9–5.4	4.3–5.0	4.0–5.1
c		25.1–31.1	13–16	11.5–17.6	10.7–17.8
Spicule length	47.02-49.62	53–67	42–68	34–44	30.8–35.5
abd	45.12-48.3	22–28	16–24	12–23	13.4–16.5

Table 3

General morphology of entomopathogenic nematodes *Heterorhabditis* adopted from (Nguyen and Smart, 1996) compared with *Oscheius* sp. TEL-2014.

Heterorhabditis		*Oscheius* (In this paper)	
Adult Female and hermaphrodites	Adult Males	Adult Females	Adult Males
Anterior region (Head)	Posterior region	Anterior region	
oesophagus without metacorpus, and enlarged basal bulb	Single testis; Bursa peloderan usually has 9 pairs of genital papillae	Six lips, 6 labial papilla, and amphidial opening present. Pharynx with corpus, isthmas,terminal bulb and pharyngeal-intestinal valve	
Posterior region (Tail)	Posterior region	Posterior region	Posterior region
Pointed tail, females have ovaries, hermaphrodites have ovotestis and amphidelphic have a vulva median	Spicules present, sometimes curved, paired or separated. Blunt tail in males	Pointed tail, vulva present, ovaries present and vulva flap	Spicules present, sometimes curved, blunt tail and short body.
Additional characteristics		Additional characteristics	
Gubernaculum present			
Fat bodies		Fat bodies	
Infective juvenile (3rd stage)		Infective juvenile (3rd stage)	
Pointed tail, 2 cuticular sheaths covering the IJ, mouth and anus closed, immature oesophogus. Symbiotic bacteria in intestine present		Head has a mouth and a pharynx with underdeveloped pharyngeal-intestinal valve Symbiotic bacteria in intestine present	

3.4 Molecular identification of EPNs by 18S rDNA Sanger sequencing

The 18S rDNA sequence was assigned the accession KM492926 by Genbank, National Centre of Bioinformatics Information (NCBI).

Direct link to deposited data:

http://www.ncbi.nlm.nih.gov/nuccore/KM492926

3.5 Phylogenetics

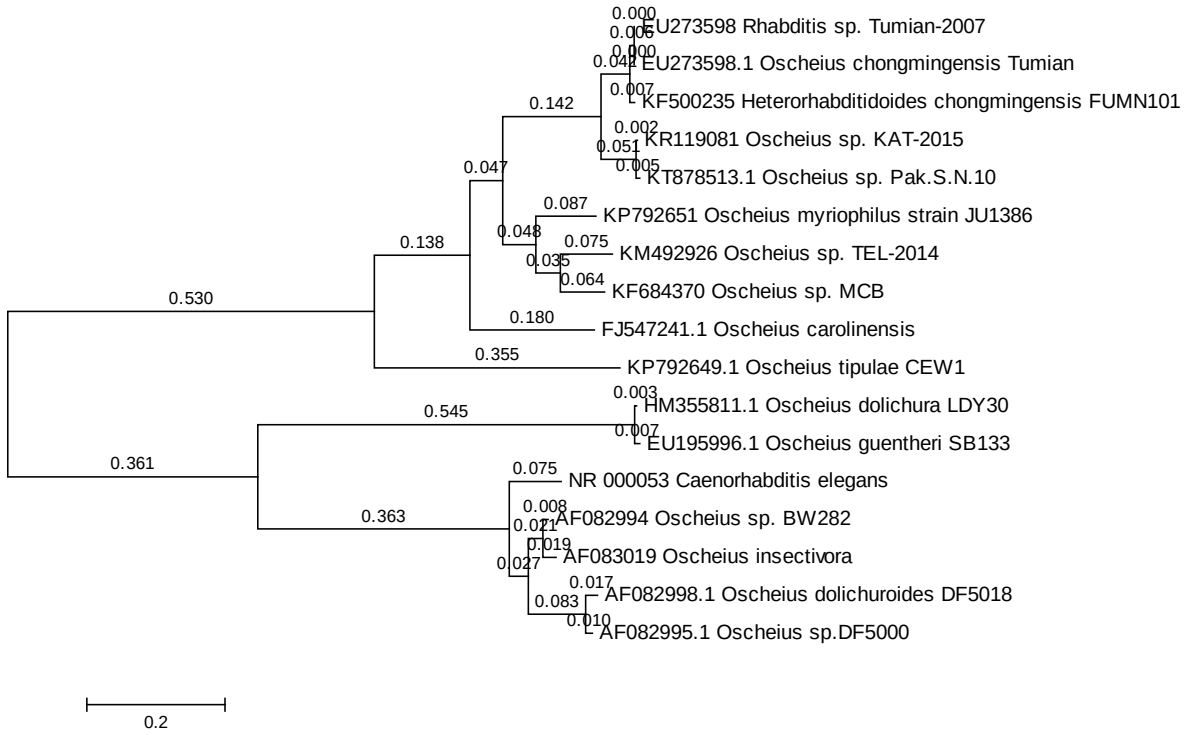


Fig. 5 The evolutionary history of several species of *Oscheius* was centred on the analysis of 18S rDNA ITS region inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The bootstrap consensus tree inferred from 1000 replications and tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches). Evolutionary analyses were conducted in MEGA6.

Table 4

Time to host death, time to nematode emergence after host death, and numbers of adults and juveniles emerged from host after inoculation with *Oscheius* sp. TEL-2014 nematodes (100 per host larva). Each value represents the mean +/- standard error

Host name	Time to host death	Time to nematode emergence after host death	IJs emerged
G. mellonella	62.4+/-9.6	26+/-2	842+/-252
T. molitor	51.8+/-6.0	46+/-4	2450+/-56

Mortality rate was greater in *G. mellonella* than in *T. molitor*. However, in a study conducted by (Torres-Barragan *et al*, 2011) they found that *O. carolinensis* took longer to lead to mortality of *Galleria* however; juveniles yield from this species was higher

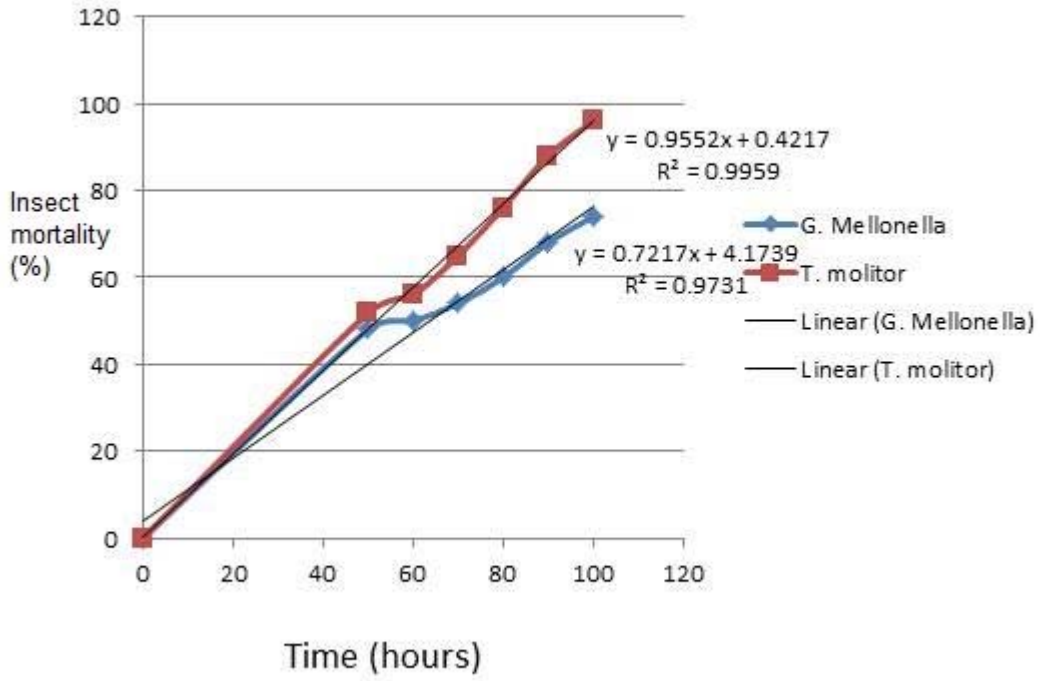


Fig. 6 Percent mortality caused by *Oscheius* sp. TEL-2014 on two different species of insects, *G. mellonella* and *T. molitor*.

In addition, this new species is capable of killing *G. mellonella* and *T. molitor* within 48–96 hours under laboratory conditions. Results indicated that, *Oscheius* sp.TEL-2014 is pathogenic to both model insects. The nematodes started killing insect larvae within 48 hours causing about (47-50%) mortality. Furthermore, they brought 100% mortality within 72 h post-exposure in *G. mellonella* and whereas, within 96 hours in *T. molitor.*

Oscheius sp. TEL-2014 is was described as a novel entomopathogenic nematode species based on its morphometrics and 18S rRNA gene sequence originality. Further studies need to be conducted to investigate entomopathogenecity mechanisms and efficacy in greater detail so that the nematode may be assigned or recommended as biopesticide for controlling specific problematic insects especially in Agricultural industries.

4. References

Bedding R.A, Akhurst R.J, 1975. A simple technique for the determination of insect parasitic rhabditid nematodes in soil. Nematol, 21:109–110.

Gil G.H, Choo H.Y and Gaugler R, 2002. Enhancement of entomopathogenic nematode production in *in vitro* liquid culture of *Heterorhabditis bacteriophora* by batch fed culture with glucose supplementation. Appl Microbiol Biotechnol, 58:751-755.

Kaya H.K, Stock S.P, 1997. Techniques in insect Nematology. In: Lacey LA, editor. Techniques in Insect Pathology. London: Academic Press 822–831.

Ma J, Chen S, Moens M, De Clercq P, Li X and Han R, 2013. Characterization in biological traits of entomopathogenic nematodes isolated from North China. Journal of Invertebrate Pathology, 114:268–276.

Nguyen K.B and Smart G.C, 1996. Identification of entomopathogenic nematodes in the *Steinernematidae* and *Heterorhabditidae*. Nematology, 28: 286-300.

Pervez R, Eapen S.J, Devasahayam S and Jacob T.K, 2012. A new species of entomopathogenic nematode *Oscheius gingeri* sp. n. (Nematoda: *Rhabditidae*) from ginger rhizosphere, Archives of Phytopathology And Plant Protection, 46:526-535.

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.

Stefan-Andreas J and Ralf-Udo E, 1999. *Endotokia matricida* in hermaphrodites of *Heterorhabditis* spp. and the effect of the food supply. Nematology, Vol. 1(7—8):717-726.

Sturm T, College C, Aptos C.A, 2014. Lipid agar. American society for microbiology Microbe Library. Online web: http://www.microbelibrary.org/library/laboratory-test/3968-lipid-agar

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.

Ye W, Torres-Barragan A and Cardoza Y.J, 2010. *Oscheius carolinensis* n. sp. (Nematoda: Rhabditidae), a potential entomopathogenic nematode from vermicompost. Nematology, 12:121–135.

CHAPTER 3

Comparison of *Oscheius sp. TEL-2014* 18S rDNA sequence to other entomopathogenic nematodes famalies using sequence alignment tool

Abstract

The 18S ribosomal DNA internal transcribed spacer region of entomopathogenic nematodes *Oscheius* sp. TEL-2014 isolated from South Africa was compared to other nematodes species belonging to the following five genera, *Oscheius, Steinernema, Heterorhabditi, Rhabtiditis* and *Caenorhabditis.* Comparative analyses and alignment of the 18S rDNA ITS region shows that *Oscheius* sp. TEL-2014 does not share 100% sequence similarity with other *Oscheius* species and other nematodes in related genera. In support of this observation, the percentage identity matrix shows that the sequences are not 100% similar to each other and thus suggesting variation amongst the *Oscheius, Steinernema* and *Heterorhabditis* species. Molecular Phylogenetic analysis by Maximum Likelihood methods between these nematodes amplified by TW81 forward primer and AB28 primer shows that *Oscheius* sp. TEL-2014 is closely related to *Oscheius myriophilus* and branches from different arms compared to the rest of the isolates. These observations support the uniqueness of the *Oscheius* sp. TEL-2014 18S rDNA sequence and thus provide evidence that this isolate is indeed a novel species.

Keywords: Entomopathogenic nematode, *Oscheius* sp. TEL-2014, sequence alignment, 18S rDNA, phylogeny and nucleotide composition

1. Introduction

Entomopathogenic nematodes (EPNs) are soil dwelling and found in diverse habitants (Eustachio *et al*, 2008). They have a broad host range and thus have great potential as biological control agents for several economically important insect crop pests (Grenier *et al*, 1995). Within the *Steinernema*, *Heterorhabditis* and *Oscheius* genera, there are numerous species which may share similarities and some differences in terms of infectivity, host range, infection symptoms or symptomology, structure, reproduction, environmental conditions preference, defence and infection mechanisms (Bastidas *et al*, 2014; Campos-Herrera *et al*, 2012). Slight sequence variations in the 18S rDNA may also be correlated to the other

genetic differences arising from speciation amongst these nematodes which are linked to unique species specific adaptive and morphological characteristics.

EPNs are found in diverse environments throughout the world (Grenier *et al*, 1995) and their classification has been simplified by employing molecular identification techniques such as the PCR amplification and sequencing of the internal transcribed spacers between 18S and 28S ribosomal DNA. The coding regions of 18S and 28S rDNA are highly conserved among EPNs species however the ITS regions contain genetic polymorphism arising from mutations such as deletions, insertions and translocations (Hillis *et al*, 1991). The nucleotide sequences of the ITS regions can be used as reliable genetic finger prints to identify nematodes species. Multiple alignments of the 18S rDNA ITS regions provide quantitative genetic or evolutionary distance data for establishing the phylogenetic affinities for EPNs species belonging to the same genus. In this study the alignment of 18S rDNA sequence data for EPNs selected species belonging *Oscheius*, *Steinernema*, *Heterorhabditis, Rhabtiditis* and *Caenorhabditis* genera was undertake establish the phylogenetic relationship of the newly isolated *Oscheius sp. TEL-2014*. From the phylogenetic analysis the inferred evolutionary relationship between the new *Oscheius* species with the other EPNs species was demonstrated.

2. Materials and methods

2.1 18S rDNA Sequence alignment

The NCBI accession numbers and species names for the sequences used were: *Oscheius* sp. TEL-2014 (KM492926), *Oscheius carolinensis* (FJ547241.1), *Oscheius myriophilus* JU1386 (KP792651), *Oscheius* sp. MCB (KF684370), *Oscheius* sp. KAT-2015 (KR119081), *Rhabditis* sp. Tumian-2007 (EU273598), *Heterorhabditidoides chongmingensis* FUMN101 (KF500235), *Oscheius* sp. BW282 (AF082994), *Oscheius insectivore* (AF083019), *Caenorhabditis elegans* (NR_000053), *Heterorhabditis bacteriophora* (KJ938576) and *Steinernema khoisanae* (KM275351).

The multiple sequence alignment was done using MEGA 6. Parameters used include DNA weight matrix set to IUB, with a gap open of 10, gap extension of 0.20, gap distance of 5, no gap end was included and the clustering used was neighbour joining.

ClustalW2 multiple sequence alignment tool was used to align *Oscheius* sp. TEL-2014 and other nematodes. Clustal W2 alignment tool is suitable for medium alignments. Slow Pairwise alignment was used with the following parameters: DNA weight matrix set to IUB, gap open of 10 amd gap extension of 0.1. The output options were set to Clustal w/ numbers or FASTA format and the order of the alignment was set to align. MEGA6 was employed to determine the base frequencies for each sequence as well as an overall average, to show the nucleotide composition.

2.2 Phylogenetic analysis

Sequences were aligned by using MUSCLE on Mega 6 before generating phylogenetic trees. Maximum Likelihood method based on the Tamura-Nei model was used to construct the trees.

2.3 Percentage identity matrix

Percentage identity matrix was created using Clustal2.1 web based and accessible on http://www.ebi.ac.uk/Tools/msa/

3. Results and Discussion

3.1 Nucleotide composition

Table 3

Nucleotide composition of the 18S rDNA of *Oscheius* sequences

Accession number, organism	T(U)	C	A	G	Total
KM492926, *Oscheius sp. TEL-2014*	30.0	24.1	19.1	26.8	800.0
FJ547241, *Oscheius carolinensis*	28.0	24.6	19.6	27.9	1176.0
KP792651, *Oscheius myriophilus* JU1386	28.6	24.5	19.7	27.2	1018.0
KF684370, *Oscheius* sp. MCB	29.4	24.4	18.8	27.5	816.0
KR119081, *Oscheius* sp. KAT-2015	30.4	22.6	21.1	25.9	858.0
AF082994, *Oscheius* sp. BW282	26.2	21.1	26.4	26.3	1709.0
AF083019, *Oscheius insectivore*	26.6	20.8	26.4	26.2	1715.0
Avg.	28.0	22.8	22.5	26.7	1156.0

The nucleotide composition shows that the 18S rDNA of *Oscheius* sequences are of varying length and contain unequal numbers of the adenines, thymines, cytocines and guanines.

Table 4

Nucleotide composition of the 18S rDNA of *Oscheius* sequences and other related species belonging to unique genera

Accession number: Organism	T(U)	C	A	G	Total
NR000053, *Caenorhabditis elegans*	27.4	20.4	25.7	26.5	1754.0
KR119081, *Oscheius* sp. KAT-2015	30.4	22.6	21.1	25.9	858.0
KP792651, *Oscheius myriophilus* JU1386	28.6	24.5	19.7	27.2	1018.0
KM492926, *Oscheius sp. TEL-2014*	30.0	24.1	19.1	26.8	800.0
KM275351, *Steinernema khoisanae*	35.7	17.4	22.3	24.6	826.0
KJ938576, *Heterorhabditis bacteriophora*	29.1	19.9	26.0	25.0	849.0
KF684370, *Oscheius* sp. MCB	29.4	24.4	18.8	27.5	816.0
KF500235, *Heterorhabditidoides chongmingensis* FUMN101	32.3	22.3	18.7	26.7	815.0
FJ547241, *Oscheius carolinensis*	28.0	24.6	19.6	27.9	1176.0
EU273598, *Rhabditis* sp. Tumian-2007	30.6	22.3	20.2	26.9	1017.0
AF083019, *Oscheius insectivore*	26.6	20.8	26.4	26.2	1715.0
AF082994, *Oscheius* sp. BW282	26.2	21.1	26.4	26.3	1709.0
Avg.	28.9	21.9	22.7	26.5	1112.8

There are differences noted especially on the *Caenorhabditis elegans* and *Oscheius* sp. TEL-2014 in terms of sequence length and nucleotide arrangement. This may support previous investigations on comparative genetics, evolution and phylogenetics of the nematodes (Sommer, 2000). Reasons for the differences observed may be further investigated in future studies which will help us to understand why the variations are present.

3.2 18S rDNA sequence alignment

Fig. 1 A portion of the alignment of 18S rDNA region between *Oscheius* species and other nematodes using Muscle alignment tool on Mega 6. Green: Adenine, Red: Thymine, Blue: Cytosine and Purple: Guanine. The sequences are in the order specified in the methods section 2.1.

Fig. 2 A portion of the alignment of 18S rDNA ITS region between *Oscheius* species and other nematodes using Muscle alignment tool on Mega 6. Green:Adenine, Red: Thymine, Blue: Cytosine and Purple: Guanine. The sequences are in the order specified in the methods section 2.1.

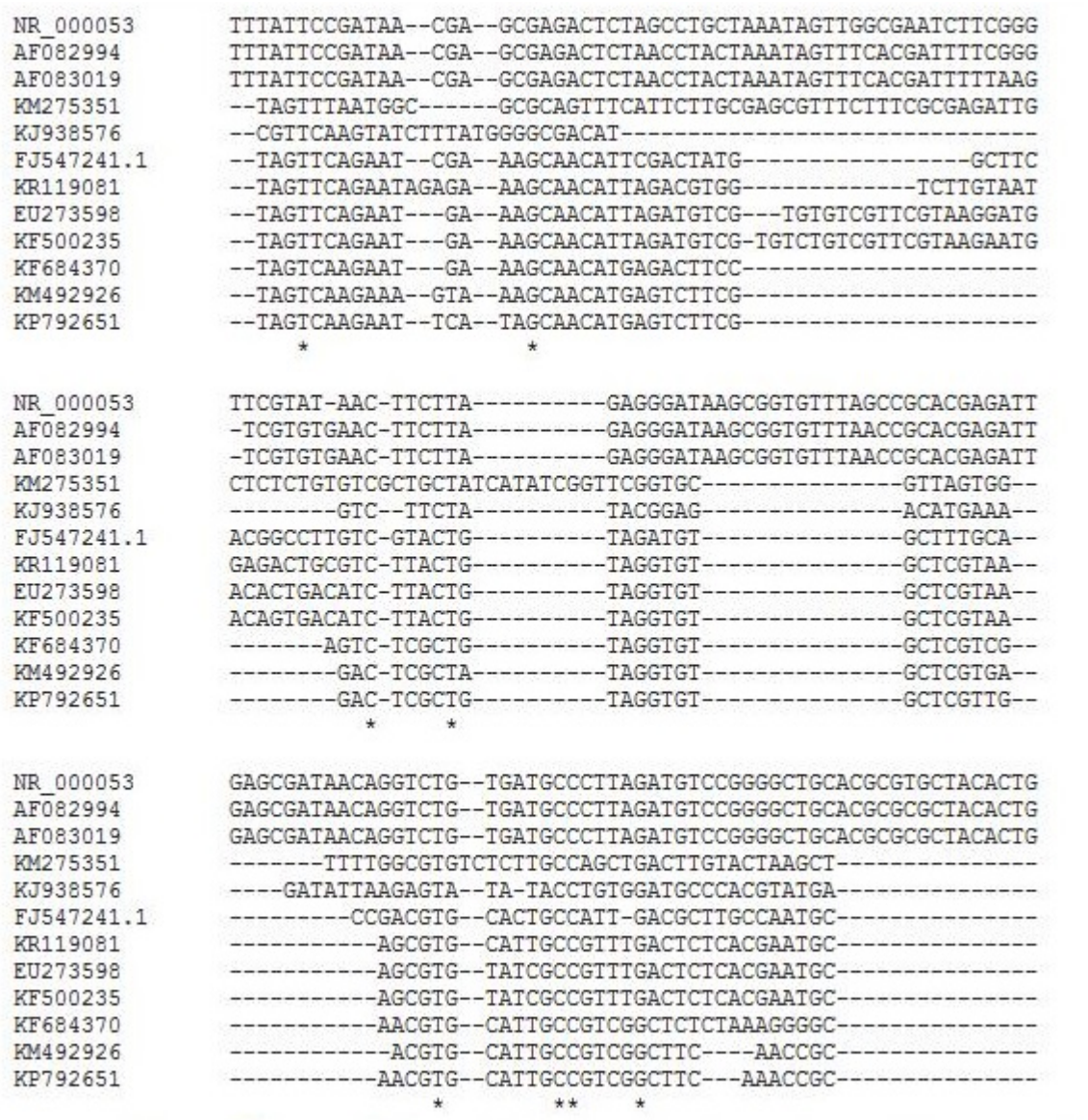


Fig. 3 Alignment of the 18S rDNA sequences for species *Oscheius* sp. TEL-2014 (KM492926), *Oscheius carolinensis* (FJ547241.1), *Oscheius myriophilus* JU1386 (KP792651), *Oscheius* sp. MCB (KF684370), *Oscheius* sp. KAT-2015 (KR119081), *Rhabditis* sp. Tumian-2007 (EU273598), *Heterorhabditidoides chongmingensis* FUMN101 (KF500235), *Oscheius* sp. BW282 (AF082994), *Oscheius insectivore* (AF083019), *Caenorhabditis elegans* (NR_000053), *Heterorhabditis bacteriophora* (KJ938576) and *Steinernema khoisanae* (KM275351). Nucleotide sequences used for comparison were taken from have NCBI GenBank database. CLUSTAL multiple sequence alignment by MUSCLE (3.8). * (asterisk) indicates positions which have a single, fully conserved residue, - (hyphen) indicates gaps or no match between the sequences.

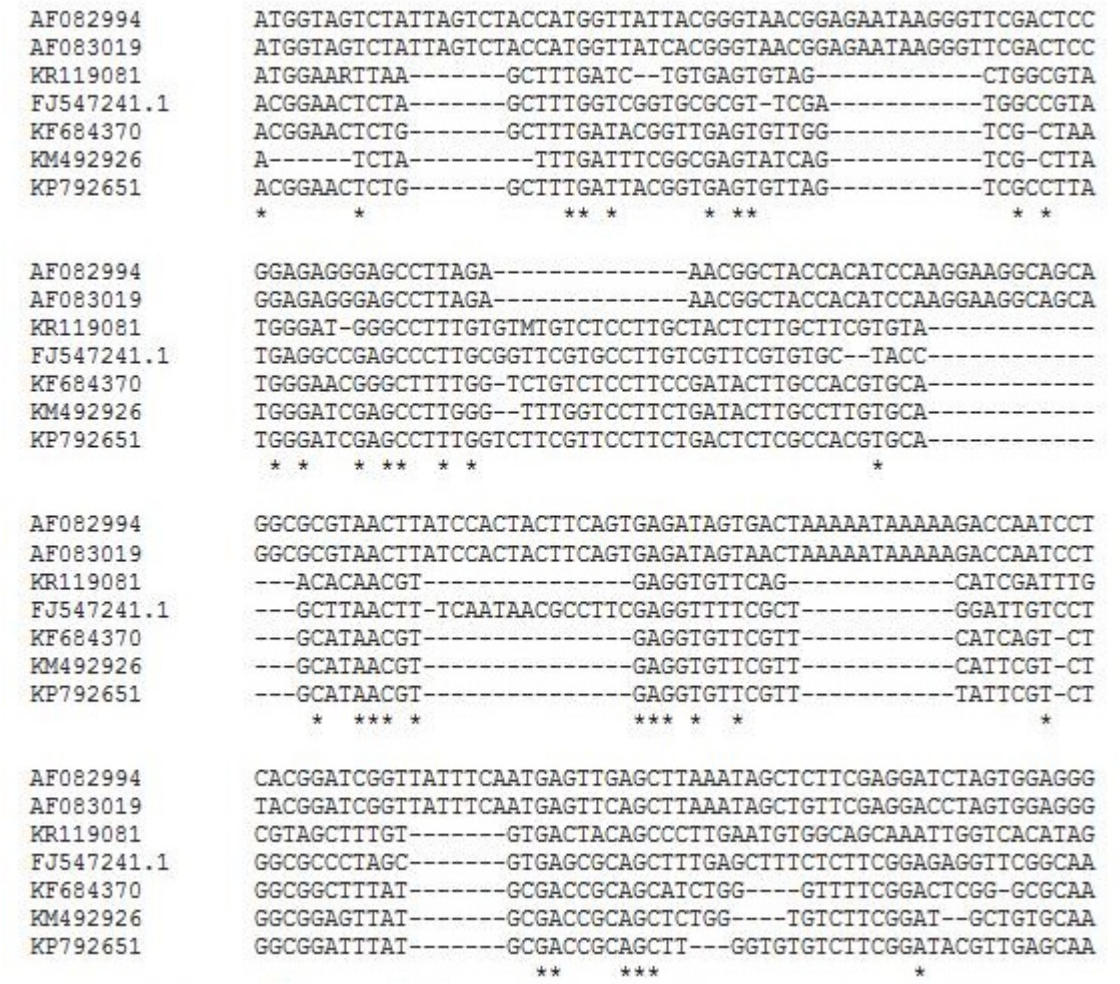


Fig. 4 Alignment of the 18 S rDNA sequences for 7 species *Oscheius* nematodes. These nucleotide sequences were taken from have NCBI GenBank database: *Oscheius carolinensis* (FJ547241.1), *Oscheius myriophilus* JU1386 (KP792651), *Oscheius* sp. MCB (KF684370), *Oscheius* sp. KAT-2015 (KR119081), *Oscheius* sp. BW282 (AF082994), *Oscheius insectivore* (AF083019) to be compared to *Oscheius* sp. TEL-2014 (KM492926). CLUSTAL multiple sequence alignment by MUSCLE (3.8). Diverse matrices are chosen by Clustal W online software by EMBL-EBI as the alignment proceeds, depending on the different divergences at each stage (Larkin *et al*, 2007).

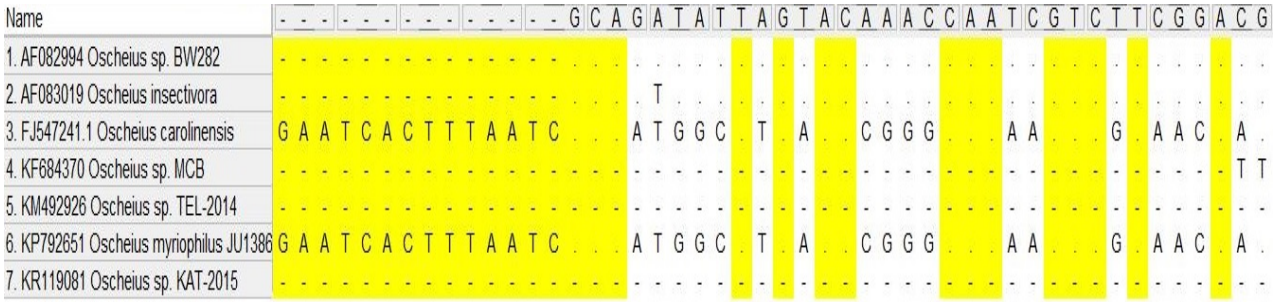


Fig. 5 Representative portion of conserved regions within the species of *Oscheius, Heterorhabditis, Steinernema, Rhabditis and Caenorhabditis*

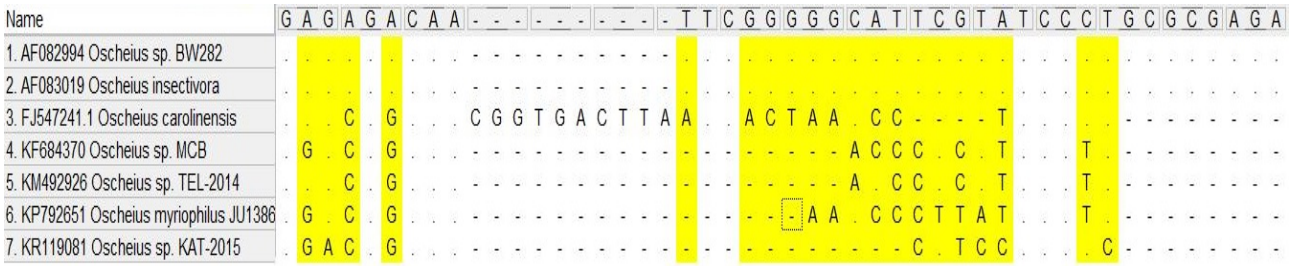


Fig. 6 Representative portion of variable regions within the species of *Oscheius, Heterorhabditis, Steinernema, Rhabditis and Caenorhabditis*

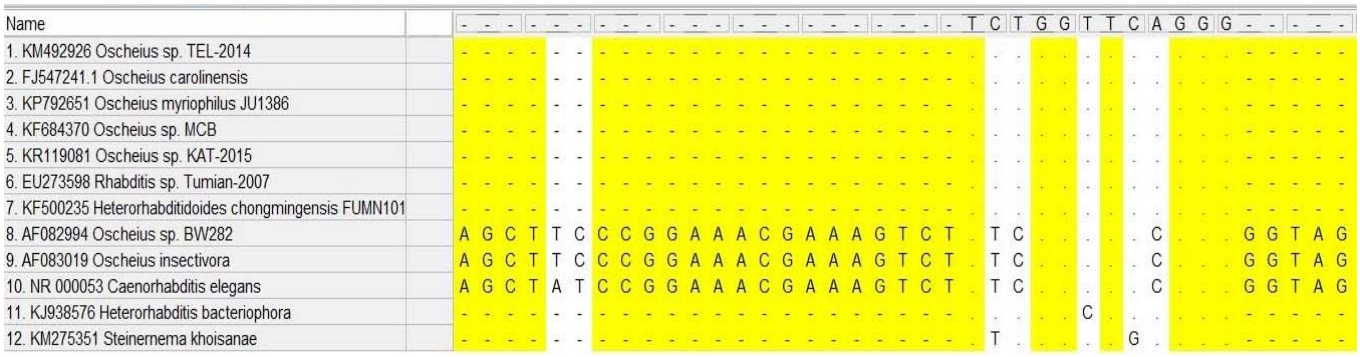


Fig. 7 Representative portion of conserved regions within the species of *Oscheius, Heterorhabditis, Steinernema, Rhabditis and Caenorhabditis*

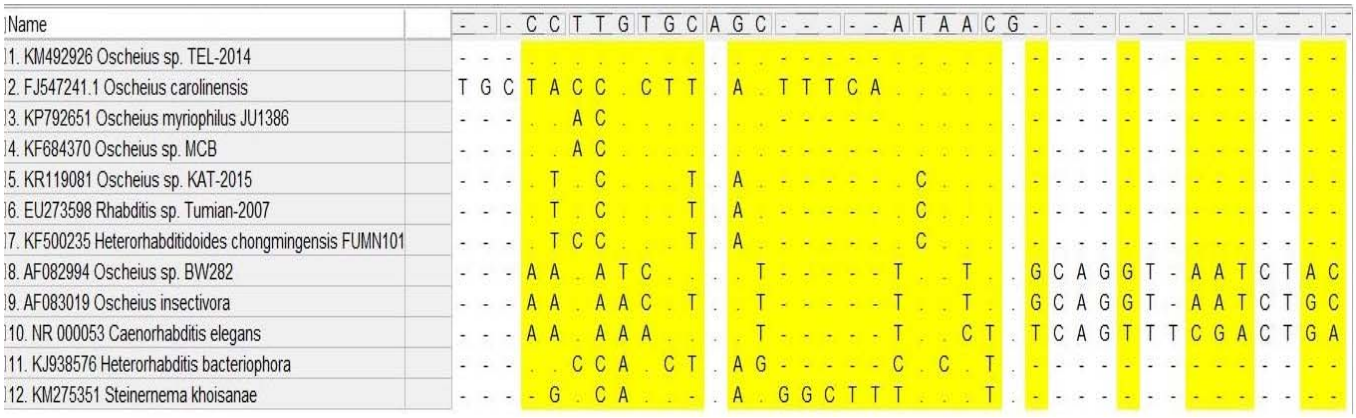


Fig. 8 Representative portion of variable regions within the species of *Oscheius, Heterorhabditis, Steinernema, Rhabditis and Caenorhabditis*

3.3 Phylogenetic analysis

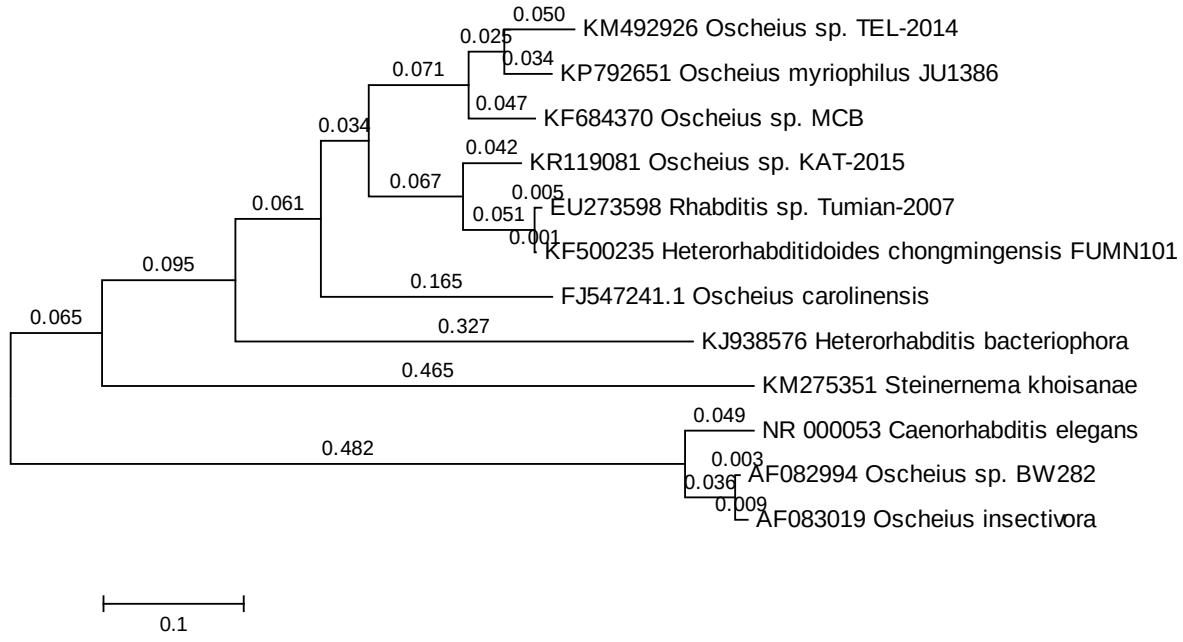


Fig. 9 the evolutionary history of species of *Oscheius, Heterorhabditis, Steinernema, Rhabditis and Caenorhabditis* was centred on the analysis of 18S rDNA ITS region was inferred by using the Maximum Likelihood method based on the Tamura-Nei model

3.4 Percentage identity

1: NR_000053	100.00	85.43	85.18	43.13	44.31	45.03	44.89	45.08	45.81	47.89	51.14	48.49
2: AF082994	85.43	100.00	97.48	42.22	44.08	46.81	46.21	47.03	46.69	49.88	51.39	51.63
3: AF083019	85.18	97.48	100.00	41.77	44.20	46.71	46.21	46.94	46.94	49.88	51.01	50.95
4: KM275351	43.13	42.22	41.77	100.00	49.53	50.48	51.67	53.02	52.91	55.87	53.94	55.03
5: KJ938576	44.31	44.08	44.20	49.53	100.00	60.10	62.85	64.53	61.46	62.39	62.54	63.70
6: FJ547241.1	45.03	46.81	46.71	50.48	60.10	100.00	67.95	72.23	68.90	69.91	71.45	74.35
7: KR119081	44.89	46.21	46.21	51.67	62.85	67.95	100.00	85.42	85.55	76.09	74.87	75.06
8: EU273598	45.08	47.03	46.94	53.02	64.53	72.23	85.42	100.00	97.89	77.15	74.50	79.89
9: KF500235	45.81	46.69	46.94	52.91	61.46	68.90	85.55	97.89	100.00	76.81	74.22	76.45
10: KF684370	47.89	49.88	49.88	55.87	62.39	69.91	76.09	77.15	76.81	100.00	83.06	85.03
11: KM492926	51.14	51.39	51.01	53.94	62.54	71.45	74.87	74.50	74.22	83.06	100.00	87.88
12: KP792651	48.49	51.63	50.95	55.03	63.70	74.35	75.06	79.89	76.45	85.03	87.88	100.00

Fig 11 Percent Identity Matrix - created by Clustal2.1

1: AF082994	100.00	97.48	50.37	48.40	52.36	54.05	51.69
2: AF083019	97.48	100.00	50.85	48.79	52.74	54.57	51.91
3: KR119081	50.37	50.85	100.00	68.26	76.52	73.94	74.78
4: FJ547241.1	48.40	48.79	68.26	100.00	73.70	75.91	77.80
5: KF684370	52.36	52.74	76.52	73.70	100.00	83.91	85.91
6: KM492926	54.05	54.57	73.94	75.91	83.91	100.00	88.24
7: KP792651	51.69	51.91	74.78	77.80	85.91	88.24	100.00

Fig 12 Percent Identity Matrix - created by Clustal2.1

Nematodes genome regions just like most eukaryotic genomes may have skewed nucleotide composition due to mutations or presence of repeats.

4. Conclusion

The sequence 18S rDNA alignments support that *Oscheius* sp. TEL-2014 is a novel species based on the originality of its 18S rDNA sequence which does not have 100% sequence identity with other nematodes belonging to *Oscheius, Heterorhabditis, Steinernema, Rhabditis and Caenorhabditis* used for sequence comparison in the study.

The investigation demonstrated that even if nematodes species belong to the same or related genus, the 18S rDNA sequence may contain variations. Factors such as deletions, translocations and insertions may have contributed to the genetic variations as seen in fig. 6 and 8 showing variable regions however fig. 5 and 7 showed conserved regions within the species of *Oscheius, Heterorhabditis, Steinernema, Rhabditis and Caenorhabditis* suggesting a degree of sequence identity. Other factors that may influence these differences include climate type and soil type the nematodes reside in (Griffiths *et al*, 2006).

The 18S rDNA sequence may be useful for phylogenetic studies (Zhu *et al*, 1998). *Oscheius* sp. TEL-2014 was found to be closely related to *Oscheius myriophilus* and not closely related to *Oscheius* sp. BW282, *Oscheius insectivore*, *Caenorhabditis elegans* based on phylogenetic analyses and observations made on sequence alignments.

5. References

Bastidas B, Portillo E and San-Blas E, 2014. Size does matter: The life cycle of *Steinernema* spp. in micro-insect hosts. Journal of Invertebrate Pathology, 121:46–55.

Campos-Herrera R, Barbercheck M, Hoy C.W, and Stock S.P, 2012. Entomopathogenic nematodes as a model system for advancing the frontiers of ecology. J. Nematol, 44:162–176.

Tarasco E, Mracek Z, Nguyen K.B and Triggiani O, 2008. Steinernema ichnusae sp. n. (Nematoda: Steinernematidae) a new entomopathogenic nematode from Sardinia Island (Italy). Journal of Invertebrate Pathology, 99:173–185.

Grenier E, Laumond C and Abad P, 1995. Characterization of a species-specific satellite DNA from the entomopathogenic nematode *Steinernema carpocapsae*. Molecular and Biochemical Parasitology, 69:93-100.

Griffiths B.S, Donn S, Neilson R and Daniell T.J, 2006. Molecular sequencing and morphological analysis of a nematode community. Applied Soil Ecology, 32:325–337.

Hillis D.M, Dixon M.T, 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. The Quarterly review of biology, 66 (4): 411-453.

Larkin M.A, Blackshields G, Brown N.P, Chenna R, McGettigan P.A, McWilliam H, Valentin F, Wallace I.M, Wilm A, Lopez R, Thompson J.D, Gibson T.J and Higgins, 2007. ClustalW and ClustalX version 2. DGBioinformatics, 21: 2947-2948.

Malan A.P, Nguyen K.B, and Addison M.F, 2006. Entomopathogenic nematodes (Steinernematidae and Heterorhabditidae) from the southwestern parts of South Africa. African Entomol, 12:65–69.

Mracek Z, Qi-zhi and Nguyen K.B, 2009. *Steinernema xueshanense* n. sp. (Rhabditida, Steinernematidae), a new species of entomopathogenic nematode from the province of Yunnan, southeast Tibetan Mts., China. Journal of Invertebrate Pathology, 102:69–78.

Nguyen K.B, Puza V and Mracek Z, 2008. *Steinernema cholashanense* n. sp. (Rhabditida, Steinernematidae) a new species of entomopathogenic nematode from the province of Sichuan, Chola Shan Mountains, China. Journal of Invertebrate Pathology, 97:251–264.

Powers T.O, Todd T.C, Burnell A.M, Murray P.C, Fleming C.C, Szalanski A.L, Adams B.A, and Harris T.S, 1997. The rDNA Internal Transcribed Spacer Region as a Taxonomic Marker for Nematodes. J. Nematol, 29,(4):441–450.

Sommer R.J, 2000. Comparative genetics: A third model nematode species. Current Biology 2000, 10:R879–R881.

Zhu X, Gasser R.B and Chilton N.B, 1998. Differences in the 5.8s rDNA sequences among ascarid nematodes. International Journal for Parasitology, 28:617-622.

CHAPTER 4

Investigating the pathogenicity of surface sterilised entomopathogenic nematode *Oscheius* sp. TEL-2014

Abstract

An entomopathogenic nematode *Oscheius* sp. TEL-2014 was recently isolated from a grassland in Suikerbosrand Nature Reserve near Johannesburg in South Africa. The nematode has a symbiotic relationship with an insect pathogenic bacteria identified as *Serratia sp.* strain TEL which is a gram negative, rod shaped Enterobacteria. The pathogenicity of *Oscheius* sp. TEL-2014 following surface sterilization at varying time periods, 0 min (sterile water treatment), 30 min, 60 min, 90 min, 120 min, 150 min and 180 min in 0.1% sodium hypochlorite solution was investigated. *Serratia* sp. TEL were isolated from surface sterilised nematode's haemocoel and tested without the nematode for its entomopathogenecity. Whole genome sequencing, assembly and annotation of *Serratia* sp. TEL was performed and 19 contigs were generated with an average length of 301 767 bp and N_{50} of 200 110 bp. The genome of the *Serratia* sp TEL was found to be 5 000 541 bp in size and contained a total of 4 618 coding regions and 103 for RNAs. This nematode was able to invade *Galleria mellonella* larvae, releasing bacteria into the larvae's gut, causing 20% mortality within 24 hours after being surface- sterilised for 30-60 minutes. Within72 hours 100% larval mortality was observed in all NaClO treated nematodes. Furthermore; bacterial growth rate was measured for 24 hours and optical density (OD) readings were taken every hour using a spectrophotometer at 600nm wavelength. After 7 hours of culturing the bacteria in Luria Broth, the media changed from yellow to red due to phase I bacteria releasing pigments. Additional studies are needed to investigate compounds secreted by *Serratia* sp. TEL which contributes to the entomopathogenecity of the nematode. This information will increase our understanding of the potential of *Oscheius* sp. TEL-2014 as an effective biological control agent.

Keywords: Entomopathogenic nematodes; *Serratia, Oscheius*, insect pest, pathogenicity, biological control

1. Introduction

Entomopathogenic nematodes (EPNs) are soil dwelling insect-killing microscopic worms which have become popular in pest management projects across the world (Torres-Barragan *et al*, 2011). In their guts, they carry insect pathogenic symbiotic enterobacteria that are able to produce lethal toxins that can kill an insect within 48-72hours after infection, (Malan *et al*, 2006). EPNs belonging to *Heterorhabditis*, *Steinernema*, and the *Oscheius* genus have a symbiotic association with *Photorhabdus*, *Xenorhabdus*, and *Serratia* bacteria respectively (Stuart *et al*, 2006; Torres-Barragan *et al*, 2011; Zhang *et al*, 2008). Not all species of bacteria belonging to the genus *Serratia* have been shown to have fatal effects on insects (Kwak *et al*, 2015). *S. marcescens* is an example of an insect pathogenic nematode belonging to the *Serratia* genus which is very effective and able to rapidly kill insects (Abede *et al* 2010; Ishii *et al*, 2014).

The free-living non-feeding stage of EPNs refered to as injective juveniles (IJs) or dauer stage are able to invade the body of susceptible insects by entering through the mouth, anus or by any of the other natural openings. The pathogenic bacteria may be classified according to external larval symptoms they cause during infection after being released by the nemotode into the hemocoel of the infected insects (Park *et al*, 2004). Once they have been released into the hemocoel they excrete toxins and immune depressors, the latter are involved in blocking the activity of phospholipase A2 of the insect (Ji *et al*, 2004). IJs then proliferate because of the nutritional benefits within the cadaver and later emerge into the soil in search of new hosts to infect, (Serwe-Rodriguez *et al*, 2004).

A study on *Oscheius* nematodes revealed that *Serratia marcescens* bacteria were found on the cuticle of *Oscheius Carolinensis* and mainly contributed to their entomopathogenicity by killing the fourth-instar *Helicoverpa zea* larvae. The second bacterial species identified in their investigation was *Enterococcus mundti* also residing on the cuticle of the nematodes and was found to also cause pathogenicity however *Achromobacter xylosoxidans* was unable to cause death in the model insects they used. *Providencia rettgeri* was found primarily in crushed surface sterilized nematodes and also caused some mortality of model insects (Torres-Barragan *et al*, 2011).

In our study we investigate the pathogenicity of surface sterilised entomopathogenic nematode *Oscheius* sp. TEL-2014 and thus identify bacterial species which may be involved in causing pathogenicity. Based on previous reports, *Serratia* bacteria is found on the surface of *Oscheius* nematodes and thus if the nematodes are surface sterilised, their ability to cause insect larvae death will decline or discontinue. However, it must be taken into account that other bacterial species which may belong to different family or genus may also be involved in the entomopathogenicity as seen in aforementioned reports which showed that the combination of *P. rettgeri* with *S. marcescens* instigated death of insect larvae.

We therefore hypothesize that *Serratia* bacteria are found on the surface of infective juveniles based on observations made in prior studies which tested the hypothesis of *Serratia*'s location 'in' or 'on' the IJs. In our study we similarly test the hypothesis on the newly isolated nematode species.

2. Materials and methods

2.1 Isolation of bacteria from sterile and homogenised nematodes (Kaya *et al*, 1997)

Fresh infective juveniles were collected from White-traps and transferred into clean Falcon tubes and allowed to settle. The excess water was removed and the settled EPNs were re-suspended in 10ml of 0.1% sodium hypochlorite (NaClO) and left for 1 hour. The IJs were transferred into 10ml fresh 0.1% NaClO solution and left for a further 3 hours. IJs were then rinsed twice with Ringer's solution under a laminar flow hood. Sterile IJs were carefully suspended in 1ml of sterile nutrient broth in a micro tube and crushed with a small sterile plastic pestle. The homogenate was aseptically transferred into a sterile McCartney bottles with about 10ml of Nutrient Broth. Bacteria were grown in the dark on a shaker for 24 hours at 25°C. In order to allow colony growth the bacteria from the Nutrient Broth was streaked on McConkey and NBTA streak plates and incubated at 25°C for 24hours in the dark. Pink or red colonies were isolated and plated on fresh MacConkey and NBTA streak plates and the pure cultures were used for further analysis.

2.2 Pathogenecity of surface sterilised nematodes and isolation bacteria using selective media

Fresh infective juveniles (IJs) were collected from White traps and sterilised with 5ml of 0.1% sodium hypochlorite solution for 0min (control, treated with sterile distilled water), 30min, 60min, 90min, 120min, 150min and 180 min in separate 10ml Falcon tubes. IJs in each Falcon tube were then rinsed twice with sterile distilled water. Sterile sandy loam soil was distributed into 28 x 90mm Petri dishes, 4 for each treatment, with 10% soil moisture and inoculated with 500 IJs immersed in 1ml of sterile distilled water. Ten *Galleria Mellonella* larvae were added in each petri dish and incubated at 25°C. Mortality percentage was recorded daily and experiments were repeated 4 times. Dead larvae were dry in the air in empty Petri dishes for 24 hours at room temperature and before been placed on White traps to test if any IJs could be recovered from larvae originally infected with sterilized nematodes, i.e sterilized with different concentrations of sodium hypochlorite. Furthermore, surface-sterilised nematodes for each sterilization time period were each transferred into sterile 1.5 ml micro-centrifuge tube and homogenised under a laminar flow using a sterile plastic pestle. One ml of fresh nutrient broth in a sterile 1.5 ml micro-centrifuge tube was inoculated with 100µl of the homogenate. The homogenate was then transferred 10 ml sterile Nutrient Broth medium in a sterile McCartney bottles. The bacteria were allowed to multiply in the dark on a shaker at 130rpm for 24 hours at 25°C. After 24 hours, the bacterial were streaked onto NBTA and MacConkey agar plates and incubated in the dark for 24 hours at 25°C.

2.3 Pathogenicity of Infective juveniles which are not surface sterilised and identification of associated bacterial isolates using Sanger sequencing and Illumina 16S rDNA-based metagenome sequencing

Colonies were obtained from nematodes which newly emerged from a cadaver which was sterilised with 100% ethanol and flame. The water in the White-traps was sterilised and IJs were collected aseptically under a laminar fume hood. 16S rDNA-based metagenome analysis of IJs which were not surface sterilised was performed to detect any bacterial species on the surface of the nematode. Bacteria identified were separately cultured and used to inoculate healthy insect larvae. To confirm the identity of each isolate, colony PCR was employed to amplify the 16S rDNA, followed by Sanger sequencing was used to sequences the conserved region. Sequencing results were subjected to NCBI BLAST to find the identity of the unknown isolated colonies. Once the colonies or bacterial isolates were identified, they were tested for their pathogenicity.

BD BBL Prompt inoculation system was used. Prompt inoculation tubes and inoculation wands were aseptically removed from the kit box. The tip of the wand was held perpendicular to the agar surface and 3 bacterial isolates colonies which were greater than 1mm in diameter were touched. As reference, the tip of the wand is 2mm in diameter. For colonies less than 1mm, ten colonies were touched with the tip, avoiding dragging the tip and penetrating the agar. The colonies for each isolate were transferred to a sterile tube by sealing the wand into the tubes provided which contained saline. The tubes with colonies were agitated with on a vortex for 10 seconds to release the bacterial cells from the wand tip. When the colonies were not properly released from the wand, the tip was left in the solution for 5 minutes and then agitated again using a vortex. Each tube contained approximately $1.5x10^8$ CFU/mL (colony forming units per mL). Insect mortality was tested using: bacterial isolate no. 1, isolate no. 2 and isolate no. 3. *G. mellonella* 10 larvae were inoculated with 100µl of $1.5x10^8$ CFU/mL using a sterile syringe. The larvae were put into clean Petri dishes and stored at room temperature, approximately 25°C and monitored for mortality within 24 hours.

2.4 Selection of bacterial from MacConkey Agar and NBTA plates and confirmation of endosymbionts

Phase I bacteria were obtained by selecting green or blue-green colonies from NBTA agar plates and red colonies from MacConkey agar plates (selective media recipe in appendix 2). The selected colonies were incubated in 1.0 ml sterile Nutrient Broth in 1.0 ml sterile Eppendorf tubes for 4 hours at 25°C. Lipid agar plates were inoculated with the 1.0 ml Nutrient Broth culture and incubated at 25°C for 24hours until a bacterial lawn was established. After the serial dilution of sterilized IJs their numbers were counted in 100 µl droplets under a light microscope by a dilution series. One ml aliquots of sterile IJs was added to Lipid Agar bacterial lawns at concentration of about 50-100 IJs/ml. IJs which grew and multiplied on the bacterial lawns in the lipid agar plates confirmed that the unknown bacterial isolated from the sterilized IJs could support nematode growth.

2.5 Genomic extraction from bacteria

The bacterial genomic DNA was extracted using ZR Bacterial DNA Kit, # D6005. Bacteria were grown on NBTA and Mc Conkey plates. An isolated bacterial colony was picked and suspended in a ZR BashingBeadTM Lysis tube and agitated at maximum speed for 5minutes. This aided in gently breaking down the bacterial wall in order to release nucleic material.

The ZR BashingBeadTM Lysis tube was centrifuged at 10 000 rpm for 1 minute. Up to 400µl of the supernatant was transferred into a Zymo-Spin TM IV Spin Filter in a collection tube and centrifuged at 7000 rpm for 1 minute. 1200µl of Fungal/ Bacterial DNA binding buffer was added to the filtrate in the collection tube and transferred 800µl of the mixture to a Zymo-SpinTM II column in a collection tube and centrifuged at 10000rpm for 1 minute. The flow was discarded from the collection tube and 200µl DNA Pre-Wash Buffer was added to the Zymo-SpinTM II column in a new collection tube and centrifuged at 10000rpm for 1 minute. 500µl of Bacterial DNA Wash Buffer was added to the Zymo-SpinTM II column and centrifuged at 10000rpm for 1 minute. The Zymo-SpinTM II Column was transferred to a clean 1.5 ml microcentrifuge tube and added 100µl DNA Elution Buffer directly to the column matrix. The tube was centrifuged at 100000rpm for 30 seconds to release the DNA from the matrix. The DNA was stored at 4°C to be used for further analysis.

2.6 PCR amplification of the 16S rDNA region and sequencing

The 16S rDNA region was amplified using 25 µl of PCR master mix, 22 µl Nuclease free water, 1 µl Forward primer EUB968, 1 µl Reverse primer UNIV1382 and 1 µl DNA concluding a total volume of 50µl in a PRC tube. EUB968 forward primer5'-ACGGGCGGTGTGTC-3' Tm (°C) =62 and UNIV1382 reverse primer 5'-AACGCGAAGAACCTTAC-3' Tm (°C) =66 were used. Samples were mixed gently, making sure that no air bubbles were formed as they can inactivate the enzymes (DNA polymerase) in the master mix. The following conditions were used to amplify the 16S rDNA region of the bacterial isolate, initial denaturation before cycling 95 ºC for 3 minutes, 35 cycle amplification series: denaturation at 94ºC for 30 seconds, annealing at 60ºC for 45 seconds, extension at 72ºC for 90 seconds and final extension after cycling: 72ºC for 7 minutes. Agarose gel electrophoresis was used to confirm the quality of PCR. Amplified DNA fragments were sent to Inqaba Biotechnical Industries (Pty) Ltd; South Africa for 16S rDNA sequencing using the same primers employed in the PRC. For identification purposes, a basic local alignment tool by the national centre for biotechnology information (NCBI-BLAST) was used to search for sequences which have a high affinity and maximum identity percentage to the unknown bacterial species.

2.7 Whole genome sequencing of the identified bacterial symbiont

Genomic DNA paired-end libraries were generated with the Nextera DNA sample preparation kit (Illumina) and indexed using the Nextera index kit (Illumina). Paired-end (2 × 300 bp) sequencing was performed on a MiSeq Illumina using the MiSeq reagent kit v3 at the Agricultural Research Council Biotechnology Platform. Quality and adapter trimming was performed with Fastq-mcf toolkit. The genome was assembled using SPADES and the contigs were annotated using NCBI Prokaryotic Genome Automatic Annotation Pipeline.

2.8 Sequence alignment and Phylogenetic analysis

Serratia sp. Strain *TEL-2014* was edited to remove unknown bases using Bioedit sequence alignment editor. 10 more sequences were obtained from NCBI GenBank database where *Photorhabus* and *Xenorhabdus* sequences were used to root the tree. Sequences were aligned first using MUSCLE on MEGA 6 software. The gap penalties were set to the following: gap open= -400 and gap extend = 0. The evolutionary history of the aligned sequences was centred on the analysis of 18S rDNA ITS region inferred by using the Maximum Likelihood method based on the Tamura-Nei model conducted in MEGA6. The bootstrap consensus tree inferred from 1000 replications and tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches).

2.9 *In vitro* bacterial culture for determining the growth curve

A 500ml flask containing Luria Broth was inoculated with a single bacterial colony grown on NBTA agar and incubated at 25°C on an orbital shaker for 130rpm for 24 hours. The absorbency was measured at 600nm hourly for 24 hours.

2.10 Gram staining

Heat-fix smears on microscope slides were prepared using colonies grown on NBTA agar. The slide was flooded with crystal violet for 60 seconds. Excess dye was poured off the slide and washed gently under running tap water and drained the slide against a paper towel.

The smears were exposed to Gram's iodine for 60 seconds and washed with tap water and drained carefully without blotting. The slide was then washed with 95% alcohol for 30 seconds and rinsed with tap water at the end of the 30 seconds to stop the decolourization. The slides were counterstained with 0.25% safranin for 30 seconds, washed

and dried. The slides were then viewed under a light microscope at 100X oil immersion objective.

2.11 Gram reaction of *Serratia* sp. TEL

A potassium hydroxide (KOH) String Test was done on *Serratia sp.* strain *TEL.* A drop of KOH was placed on a slide and mixed with colonies from on NBTA and McConkey agar plates which were incubated for 24 hour. The reaction was evaluated after 1 minutes by using a sterile toothpick to check if the colonies create a stringy reaction or not. A gram-negative reaction is indicated by stringy mixture while gram-positive cells will not produce a viscous mixture.

2.12 Spore stain of *Serratia* sp. TEL

Spore stains are usually done for crystal producing, Gram-positive bacteria. Smears of the bacteria were prepared on the slide, allowed to air dry and heat fixed. The smears were flooded with malachite green and placed over a beaker containing boiling water for 2-3 minutes as demonstrated on figure 1. The stain was not allowed to boil or evaporate. The slides were removed, allowed to cool and rinsed under tap running water. The smears were stained with safranin for 30 seconds and washed with tap water. The slides were blot dry and examined under oil immersion.

Fig. 1 Spore staining procedure

2.13 Statistical analysis

Descriptive statistics were used to calculate the mean and standard deviation of the IJs emerged from one dead larvae and time for IJs emergence values

3. Results and Discussion

3.1 Isolation of bacteria from sterile and homogenised nematodes (Kaya *et al*, 1997)

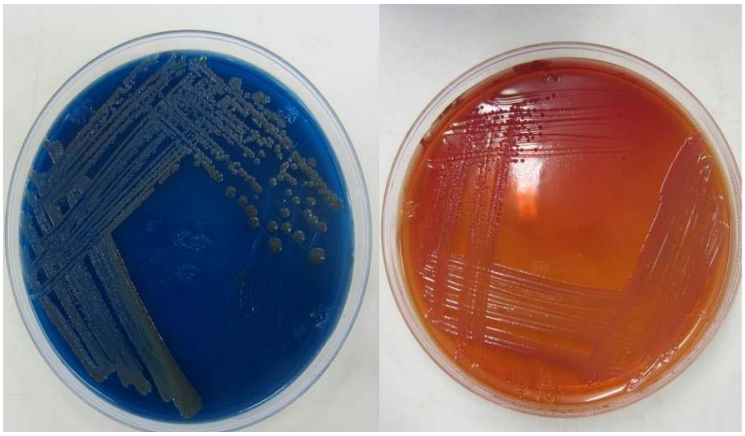


Fig. 2 Phase I colonies obtained from sterile and homogenised *Oscheius sp. TEL-2014*nematodes. Left: green circular colonies on NBTA selective media. Right: red and pink circular colonies when grown on McConkey media and inculated at 25°C for 24-96hours

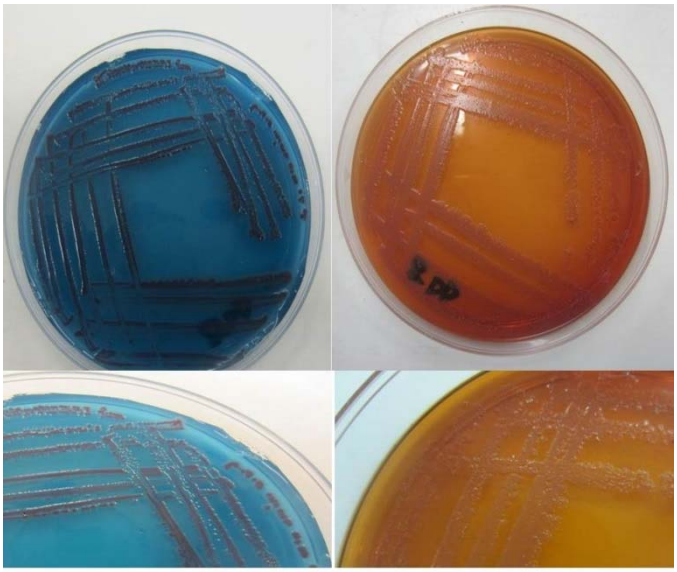


Fig. 3 Bacteria isolated from the hemocoel of sterile nematodes showing phase II variants on NBTA (left) and McConkey agar plate (right).

MacConkey agar was used to isolate gram negative bacteria with a potential to ferment lactose, thus dropping the pH of the media resulting in the growth of red/pink colonies which may be suspected to be associated with the EPNs (Kaya *et al*, 1997). Bromothymol Blue agar was used to also isolate gram negative bacteria and which have the potential to absorb bromothymol blue thus also resulting in blue- green colonies with a red centre. Bacteria can occur in two phases, phase I and phase II which may enable one to identify the desired bacterial species and characterise them on by simple streaking on to MacConkey and NBTA agar plates which are greatly used in selecting bacteria according to lactose fermentation and ability to absorb bromothymol blue and reduction of triphenyltetrazolium chloride (TTC) which results in a red colour of colony surrounded with clear zones seen mostly in *Xenorhabdus* species (Kaya *et al*, 1997).

Table 1

Summary of the isolated bacteria from EPNs characterised according to phase variants of described by (Kaya *et al*, 1997)

	Photorhubdus	*Xenorhabdus*	*Serratia* sp. TEL (In this chapter)
Phase I: McConkey colony phase, colour and shape	pink and red, circular colonies	pink and irregular and some circular colonies	Red and pink, circular colonies
Phase I NBTA colony phase, colour and shape	green/red colonies surrounded by a clear zone	green/red colonies surrounded by a clear zone	green colonies surrounded by a clear zone
Phase II: McConkey colony phase, colour and shape	Yellow/brown flat and irregular colonies	Yellow/brown flat and irregular colonies	Yellow/brown irregular colonies
Phase II: NBTA colony phase, colour and shape	Reddish-brown colonies	Reddish-brown circular and some irregular colonies	Red irregular colonies and some circular

In this study, it has been observed that *Serratia* sp. TEL forms green circular colonies when grown on NBTA selective media and, red and pink circular when grown on McConkey media and inculated at 25°C for 24-96hours. After 96 hours, bacterial colonies started to enter phase II/secondary phase whereby red irregular colonies grew on NBTA and then only light pink, almost cream white irregular colonies grew on McConkey. These observations were interesting and so far there has not been any other report for *Serratia* associated with EPNs going into a secondary phase as with *Xernorhabdus* and *Photorhabdus*. Characteristics of *Xernorhubdus* and *Photorhubdus* which are bacterial symbionts of the entomopathogenic nematodes *Heterorhabditis* and *Steinernema* are summarised in Table 1.

3.2 Molecular identification

The 16S rDNA had high affinity to a novel *Serratia* species which was registered on GenBank NCBI as *Serratia* sp. TEL allocated the accession number KP711410. The nematode *Oscheius* sp. TEL-2014 from which the bacteria was isolated from has the accession KM492926 on Genbank.

Direct links to deposited data:

http://www.ncbi.nlm.nih.gov/nuccore/KP711410

http://www.ncbi.nlm.nih.gov/nuccore/KM492926

3.3 Whole genome sequencing, assembly and annotation

19 contigs were generated with an average length of 301767bp and N_{50} of 200,110 bp. The genome of the *Serratia sp. TEL* was found to be 5,000,541-bp in size, with a G+C content of 59.1%. Prokaryotic Genome Automatic Annotation Pipeline was used to annotate the genome as described in (Lephoto *et al*, 2015). Moreover the functional annotation was carried out by RAST (Rapid Annotation using Subsystem Technology) hosted by Fellowship for Interpretation and several subsystems were identified (Lephoto and Gray 2015). The virulence, disease and defence subsystem was also found and contains 119 genes which may contribute to pathogenicity of the bacteria carried by the entomopathogenic nematode *Oscheius sp. TEL-2014*. Further investigation of this subsystem may provide some insights on the mechanism of infection highlighting proteins involved in pathogenicity.

3.4 Pathogenicity of surface-sterilised nematodes

Oscheius sp. TEL-2014 that had been surface sterilized for at varying time periods, 0min (control, treated with sterile distilled water), 30min, 60min, 90min, 120min, 150min and 180 min in 0.1% sodium hypochlorite solution resulted in 100% mortality of *G. mellonella* larvae. 100% mortality of the larvae was also achieved with the control treatment where the nematodes were not surface sterilised. However, as the length of surface sterilisation increased, the percentage of larvae in which IJs were recovered gradually dropped (Table 2). The remaining cadaver where dissected all were found to have IJs. Even if surface sterilised IJs had the ability to penetrate into the *G. mellonella*, nematodes performance declined as the exposure time in 0.1% NaClO increased.

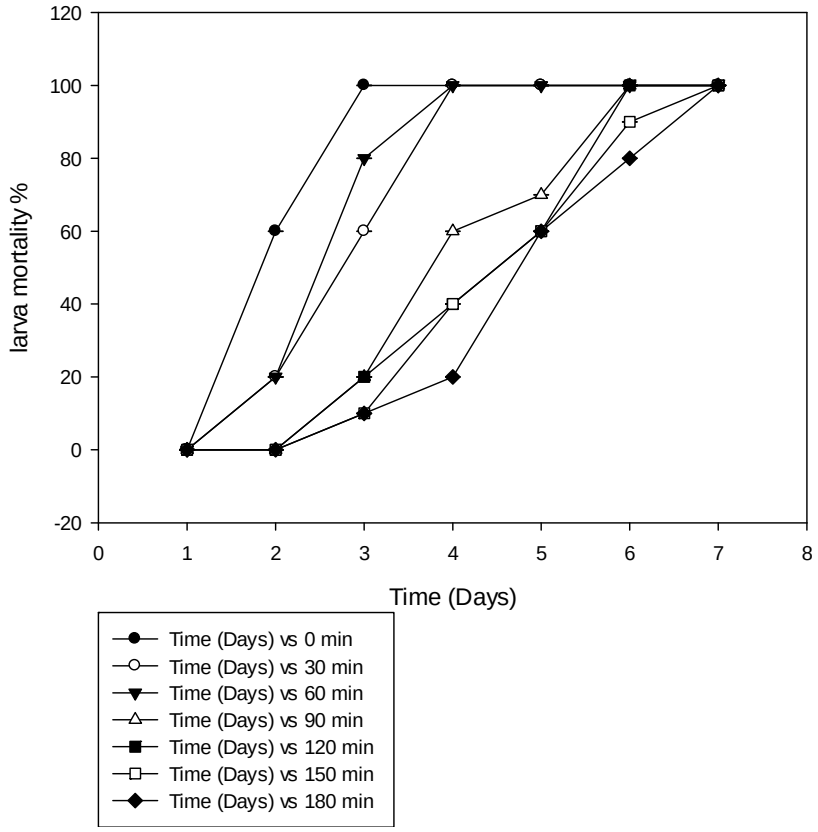


Fig. 4 *Galleria mellonella* percentage mortality caused by *Oscheius* sp. TEL-2014 treated with 0.1% sodium hypochlorite solution for different periods of time.

This *Oscheius* sp. TEL-2014 was able to invade *Galleria mellonella* larvae, release the bacteria into the larvae's gut and cause 20% mortality within 24 hours when surface-sterilised for 30-60 minutes and 100% larvae mortality is observed within 72 hours. However, as the length of surface-sterilisation time increases (90-180) minutes, the 10-20% mortality is observed in 48 hours. 100% larvae mortality is observed by day 6 when surface-

sterilised for 90-120 minutes and by day 7 for 150-180 minutes. These results show that the virulence of the bacteria decreases with an incline in the amount of time the nematodes are exposed to 0.1% sodium hypochlorite solution. In a study conducted in 2011, *Oscheius carolinensis* IJs taken from cadavers surface-sterilised for 1 minute using 1% NaClO and used to infect *Galleria mellonella* larvae and in their results, 100% insect mortality was reported (Torres-Barragan *et al*, 2011). These results prove bacteria involved in the mortality of the insects are found inside the nematodes not on the surface of the nematodes. This implies that the hypothesis stated in the investigation is rejected based on the results obtained.

Table 2

Time to host death, time to nematode emergence after host death, and numbers of adults and juveniles emerged from host after inoculation with *Oscheius* sp. TEL-2014 nematodes (100 per host larva). The IJs emerged from one dead larvae and time for IJs emergence values represent the mean +/- standard error.

Treatment (length of nematode surface sterilisation in minutes)	IJs Emerged from dead larva	Percentage (%) of larvae in which IJs were recovered	Time for IJ emergence (in days)
0	850+/-76.7	100	2+/-0.05
30	720 +/-54.1	100	2+/-0.049
60	680 +/-28.3	100	4+/-0.197
90	590 +/-24.4	70	4+/-0.2
120	225+/-15.5	40	5+/-0.32
150	65+/-10.2	30	6+/-0.34
180	50+/-8.12	10	7+/-0.35

Our findings are not supported by some of the previous studies done on *Oscheius Carolinensis* which had relations with four bacteria.

In our study, surface sterilised nematodes performed well in causing mortality of *G. mellonella* insect larvae while in previous studies surface sterilised *O. carolinensis* nematodes did not successfully cause mortality of target insect hosts.

The combination of *P. rettgeri* with *S. marcescens* caused mortality of insect larvae however it was reported that further studies need to be conducted in order to confirm and describe the relationship and role of between *O. carolinensis* and *S. marcescens* (Torres-Barragan *et al*, 2011)

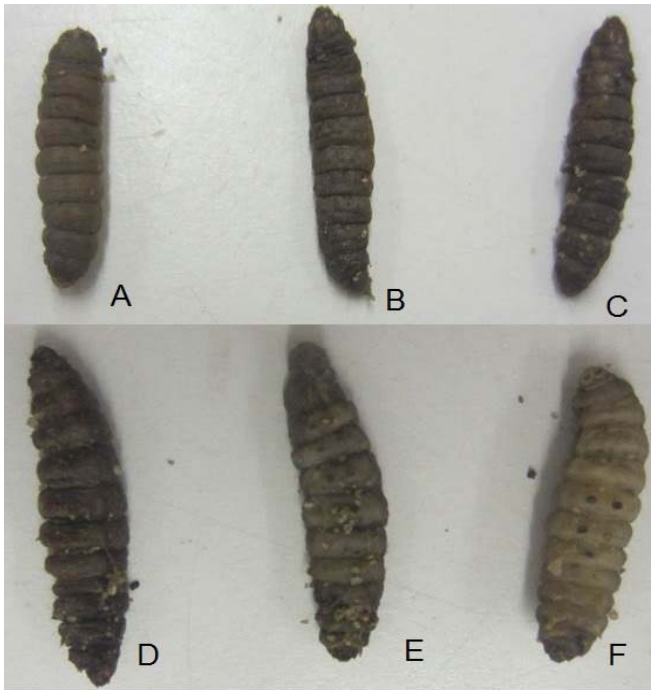


Fig. 5 Infection symptoms recorded for each treatment after 24 hours of inoculating *G. mellonella* larvae surface-sterilised *Oscheius sp. TEL-2014*. A=0min, B=30min, C=60min, D=90min, E=120min, F=150min and 180min. It is evident that as the length of surface sterilisation time increases, the larva displays less infection symptoms.

3.5 Pathogenicity of Infective juveniles which are not surface sterilised and identification of associated bacterial isolates using Sanger sequencing and Illumina 16S rDNA-based metagenome sequencing

Table 3

Sanger sequencing results which are also support the Illumina 16S rDNA-based sequencing

Bacterial isolate	Species identified
1	*Citrobacter freundii*
2	*Serratia* species
3	Uncultured *Klebsiella*

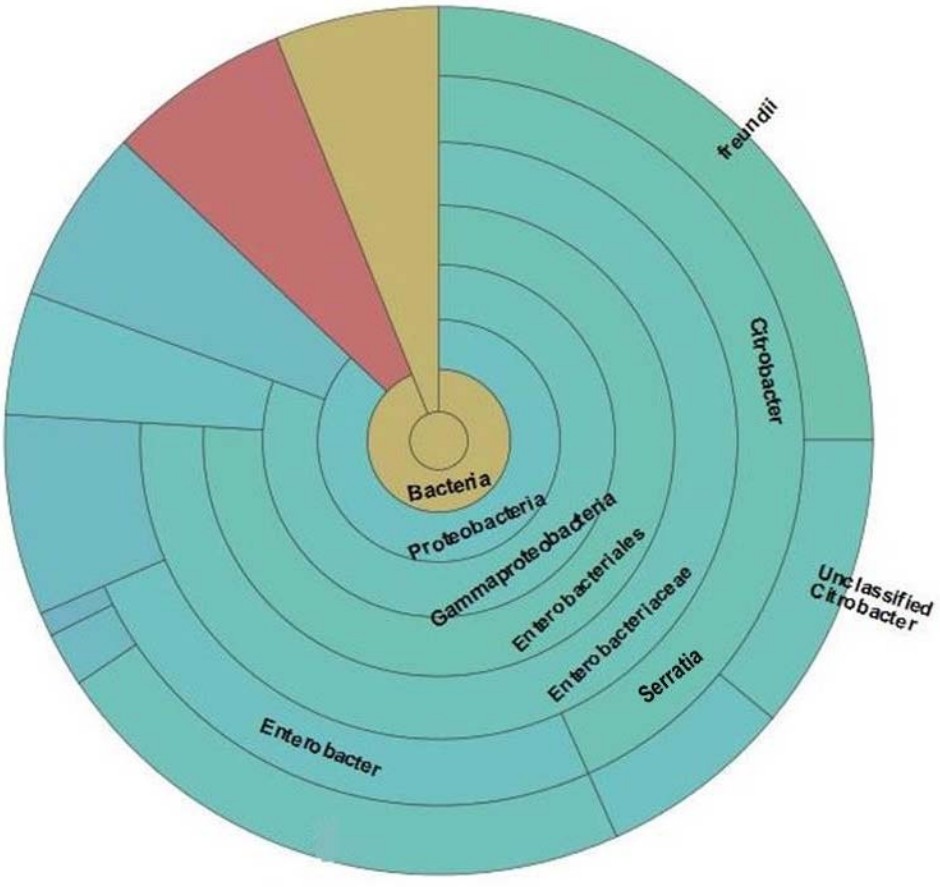


Fig. 6 Sunburst Classification Chart. This sunburst chart shows the relative abundance of the classification results within each taxonomic level. The bacterial species identified using Illumina 16S rDNA-based metagenome analysis were obtained from infective juveniles not surface sterilised.

Surface sterilised nematodes were able to infect and cause mortality of the insect larvae which indicates that the *Serratia* species isolated from the gut of the nematode was pathogenic.

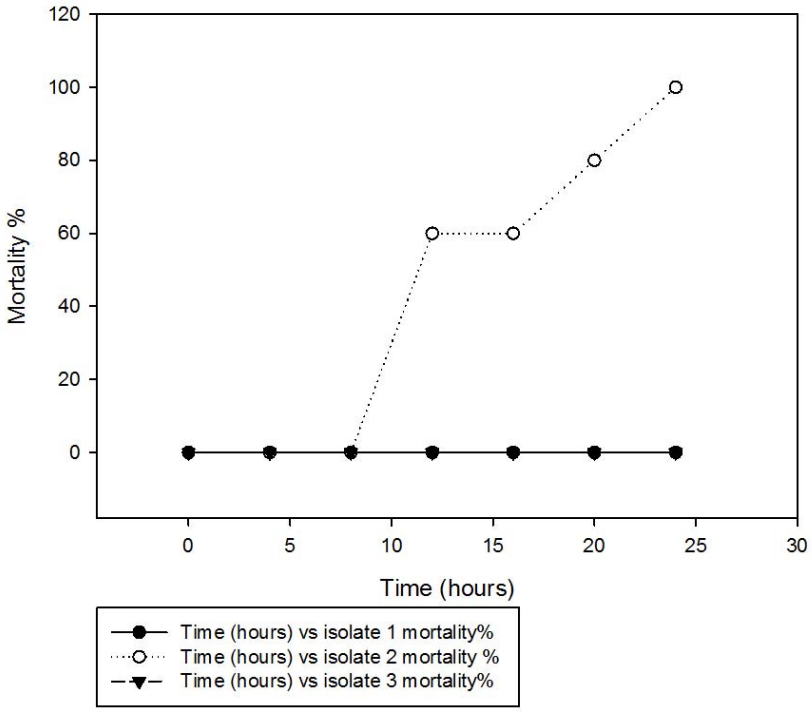


Fig. 7 Mortality observed when 10 insect larvae per bacterial isolate were inoculated with bacterial isolate no. 1, which was identified as *Citrobacter freundii*, isolate no. 2 *Serratia* sp and isolate no. 3 Uncultured *Klebsiella*. Mortality was monitored over 24 hours.

No death or any signs of infection were observed in insect larvae inoculated with *Citrobacter freundii* and Uncultured *Klebsiella* after 24 hours, while all 10 larvae died within 24 hours of infection by the *Serratia* species.

3.6 Phylogenetic analysis

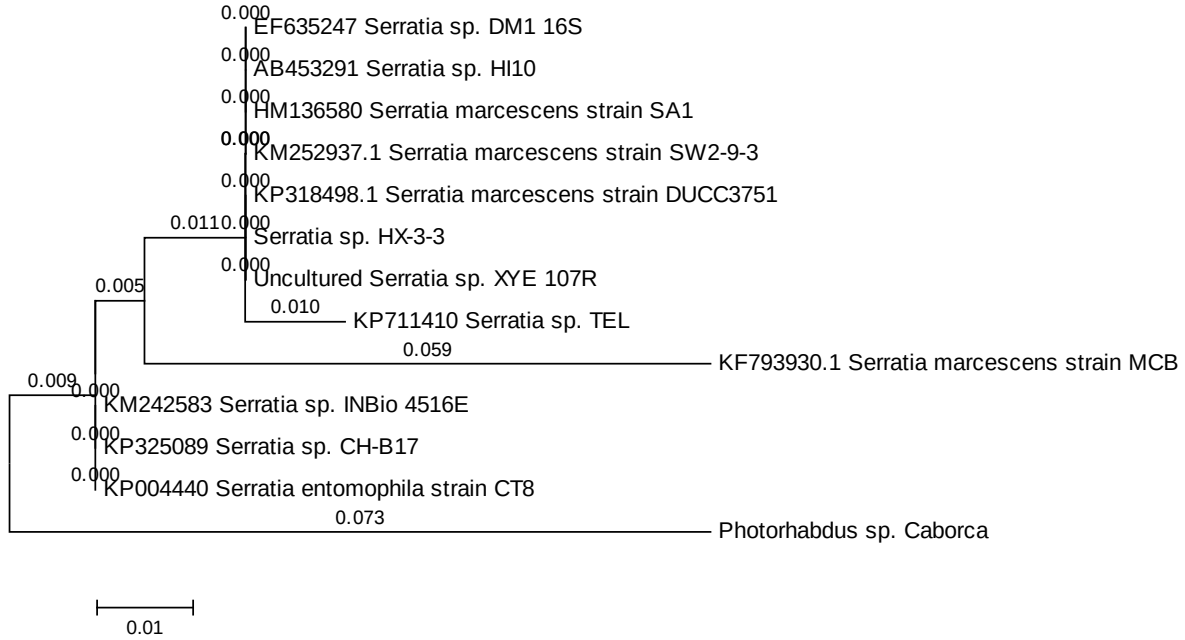


Fig. 8 the evolutionary history of several species of *Serratia* and other selected sequences was centred on the analysis of 16S rDNA ITS region inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The bootstrap consensus tree inferred from 1000 replications and tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches).

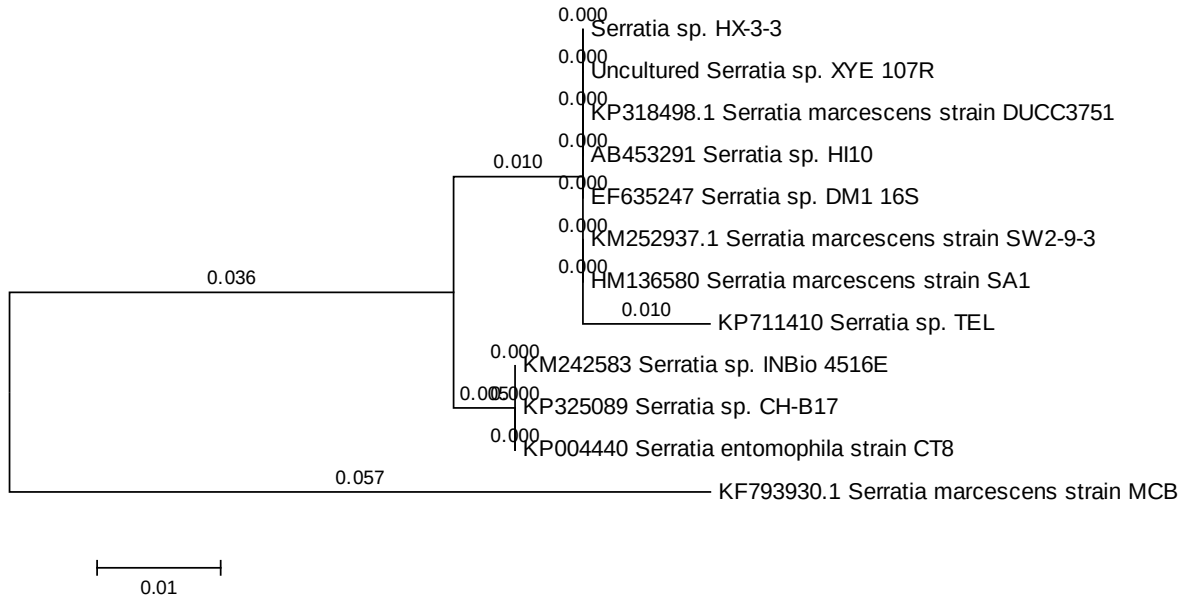


Fig. 9 the evolutionary history of several species of *Serratia* sequences only was centred on the analysis of 16S rDNA ITS region inferred by using the Maximum Likelihood method based on the Tamura-Nei model.

3.7 Bacterial growth curve

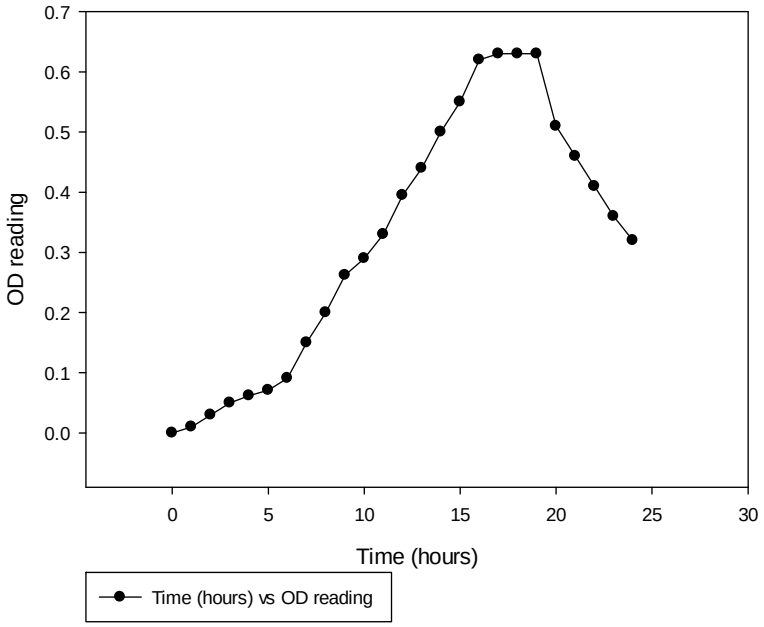


Fig. 10 *Serratia* sp. TEL grown in Luria Broth. Bacterial growth rate was measured for 24 hours and optical density (OD) readings were taken every hour using a spectrophotometer at 600nm wavelength.

Interestingly, unique colour changes occurred in the Luria Broth inoculated with *Serratia sp. TEL* when the growth rate was determined. After 7 hours of culturing the bacteria in Luria Broth, the media changed from yellow to red due to phase I bacteria releasing pigments. Colour changes made on the selective media also show that the *Serratia* bacteria in this study grows into red colonies after 24 hours incubation at 25C on McConkey agar plates and green on NBTA agar plates when they are in Phase I. As they enter Phase II, colonies change from red to some yellow and mostly brown colonies on McConkey agar plates and change from green to red on NBTA agar plates.

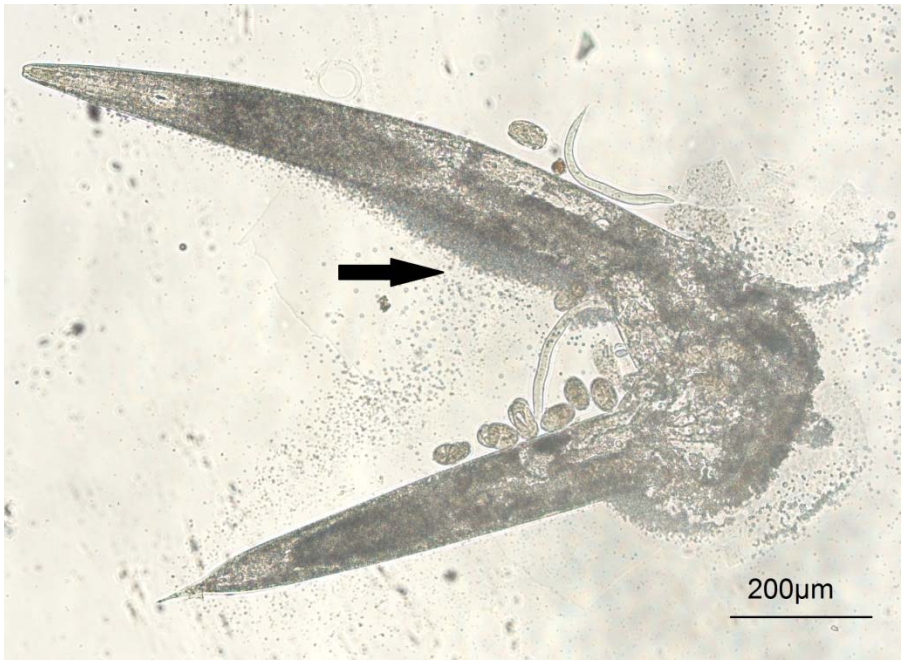


Fig. 11 *Oscheius* sp. TEL-2014 has evolved a mutualistic relationship with *Serratia* sp. TEL (pointed with black arrow) which is responsible for the infection of target insects. This female nematode going though endotokia metricida (Intra-uterine birth instigating maternal death) also released intestinal bacteria and infective juveniles. The length of the bacterial cell = 2.028µm.

3.8 Gram stain

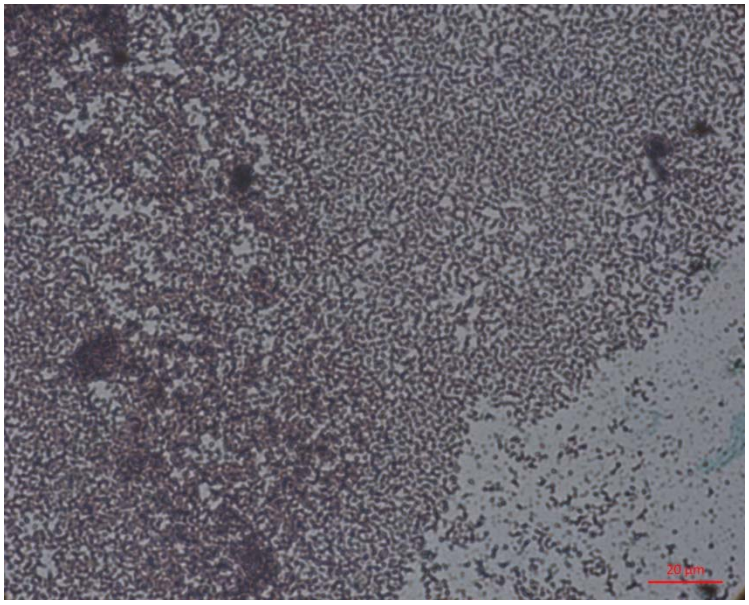


Fig. 12 *Serratia* sp. TEL is gram negative, rod shaped cells.

Serratia sp. *TEL* is gram negative, rod shaped enterobacteria. *Serratia marcescens* is the main representative of the genus *Serratia* and has been reported to possess red pigmentations which are also seen on insect's larvae within 48 hours of infection. The same may be said with *Serratia* sp. TEL, as some insect cadaver were maroon after infection, while some only had dark black spots on their bodies as seen in fig. 5.

3.9 Grams reaction

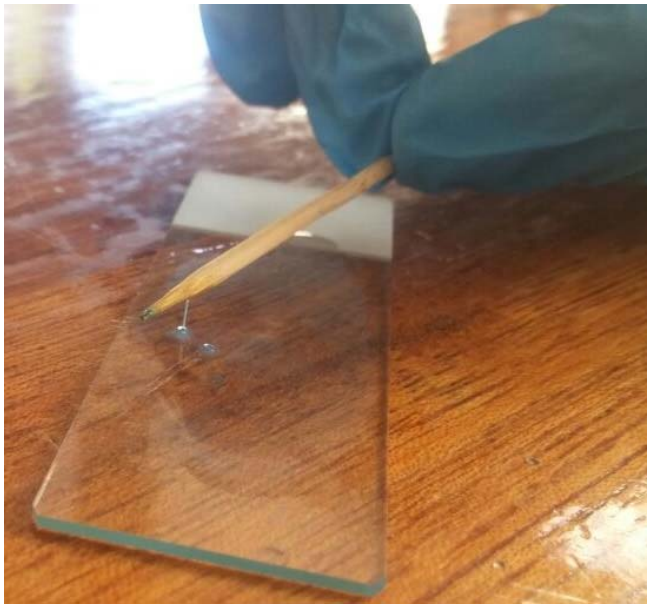


Fig. 13 *Serratia* sp. TEL is gram negative for the gram's reaction. The gram reaction helps to determine if bacteria is gram-positive or gram-negative. Grams reaction method adopted from (Gustav, 2015).

3.10 Spore stain

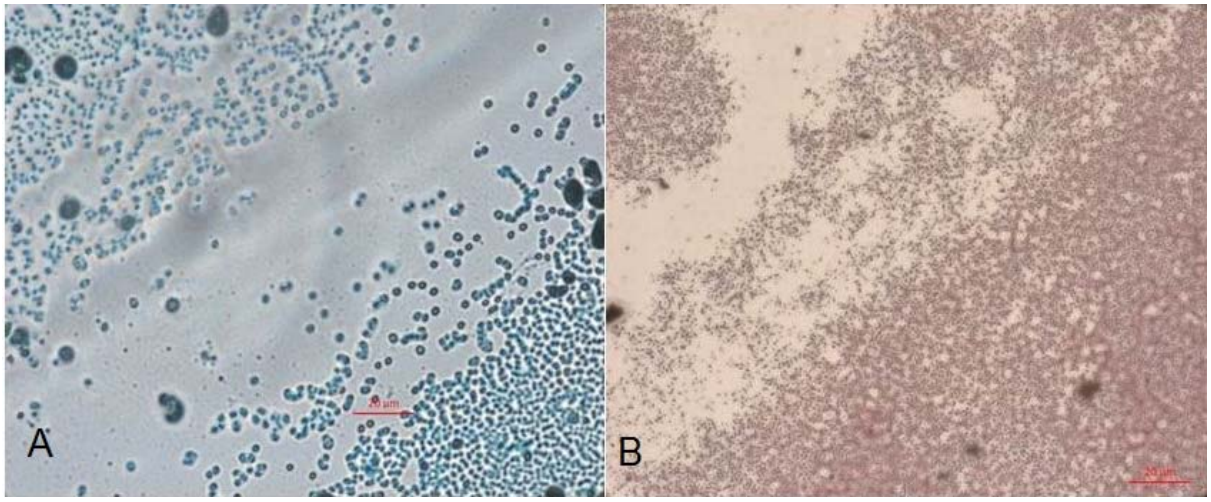


Fig. 14 Spore stain was done to verify if *Serratia* sp. TEL produces spores. A (positive control): *Bacillus thuringiensis* which is a gram-positive bacterium. *B. thurengiensis* forms crystals during sporulation and B: *Serratia* sp. TEL showing no spores.

4. Conclusions

This study provides evidence that a newly isolated and identified entomopathogenic nematode *Oscheius* sp. TEL-2014 has a symbiotic relationship with pathogenic enterobacteria *Serratia* sp. TEL which was isolated from all surface-sterilised IJs and therefore resides inside the nematodes. One previous study had shown that *Serratia* bacteria were found on the surface of the nematode's cuticle and responsible for causing death of insect larvae.

Based on outcomes from this study it would appear that *Oscheius* sp. TEL-2014 relationship with *Serratia* sp TEL bacteria allows the nematode to be categorised as an entomopathogen because of their ability to cause insect mortality.

5. References

Abede E, Jumba M, Bonner K, Gray VM, Morris K, Thomas WK. (2010). An entomopathogenic *Caenorhabditis briggsae*. The Journal of Experimental Biology 213: 3223 – 3229.

Brunel B, Givaudan, A Lanois, A Akhurst, R J and Boemare N, 1997. *Fast and Accurate Identification of Xenorhabdus and Photorhabdus Species by Restriction Analysis of PCR-Amplified 16S rRNA Genes*. Applied and Environmental Microbiology. 63(2):574-580.

Bouwer G, 2015. Advanced bacteriology III practical manual.

Ciche T.A, Darby C, Ehlers R, Forst S, and Goodrich-Blair H, 2006. Dangerous liaisons: The symbiosis of entomopathogenic nematodes and bacteria. Biological Control, 38:22–46.

Ishii K, Adachi T, Hara T, Hamamoto H and Sekimizu K, 2014. Identification of a *Serratia marcescens* virulence factor that promotes hemolymph bleeding in the silkworm, Bombyx mori. Journal of Invertebrate Pathology, 117:61–67.

Ji D, Yi Y, Kang G, Choi Y, Kim P, Baek N and Kim Y, 2004. Identification of an antibacterial compound, benzylideneacetone, from *Xenorhabdus nematophila* against major plant-pathogenic bacteria. FEMS Microbiology Letters, 239:241–248.

Kaya H.K and Stock S.P, 1997. Techniques in insect nematology. In Manual of Techniques in Insect Pathology 281-324.

Kwak Y, Khan A R and Shin J, 2015. Genome sequence of *Serratia nematodiphila* DSM 21420T, a symbiotic bacterium from entomopathogenic nematode. Journal of Biotechnology, 193:1–2.

Lephoto T.E, Featherston J, Gray V.M, 2015. Draft whole-genome sequence of *Serratia* sp. strain TEL, associated with *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae) isolated from a grassland in South Africa. Genome Announc, 3(4):e00747-15.

Lephoto T.E and Gray V.M, 2015. Genome sequencing and annotation of *Serratia* sp. strain TEL. Genomics data 6:54-56.

Malan A.P, Nguyen K.B, and Addison M.F, 2006. Entomopathogenic nematodes *(Steinernematidae* and *Heterorhabditidae*) from the southwestern parts of South Africa. African Entomol, 12:65–69.

Park Y, Kim Y, Tunaz H and Stanley D W, 2004. An entomopathogenic bacterium, *Xenorhabdus nematophila*, inhibits hemocytic phospholipase A2 (PLA2) in tobacco hornworms *Manduca sexta*. Journal of Invertebrate Pathology, 86: 65–71.

Stuart R.J, Barbercheck M.E, Grewal P.S, Taylor R.J and Hoy W.C, 2006. Morphological and functional dimorphism in *Xenorhabdus spp.*, bacteria symbiotically associated with the insect pathogenic nematodes *Neoaplectana and Heterorhabditis.* Microbiololgy, 121: 303-309.

Serwe-Rodriguez J, Sonnenberg K, Appleman B, and Bornstein-Forst S, 2004. Effects of host desiccation on development, survival, and infectivity of entomopathogenic nematode *Steinernema carpocapsae*. Journal of Invertebrate Pathology, 85:175–181.

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.

Zhang C, Liu J, Xu M, Sun J, Yang S, An X, Gao G, Lin M, Lai R, He Z, Wu Y, and Zhang K, 2008. *Heterorhabditidoides chongmingensis* gen. nov., sp. nov. (Rhabditida: Rhabditidae), a novel member of the entomopathogenic nematodes. J. Invertebr. Pathol, 98, 153–168.

CHAPTER 5

Whole genome sequencing, assembly and annotation of *Serratia* sp. TEL

Abstract

Bacteria belonging to the genus *Serratia* are gram-negative and rod shaped. Some species in this genus have evolved a symbiotic association with a number of species of nematode belonging to the genus *Oscheius* nematodes. So far all species of *Oscheius* that have a symbiotic association with bacteria belonging to the genus *Serratia* also display virulence and pathogenicity against a range of different insect hosts. The bacteria exist as an endosymbiont in non-feeding infective juveniles of a number of *Oscheius* species. The infective juveniles acting as vectors for the bacteria can locate, infect and release the bacteria into the hemecoel of the target insect. In this respect they are behaving in manner that is functionally equivalent to the entomopathogenic behaviour that is typical of steinernematids and heterorhabditids. Based on these observations there are grounds for including those species of *Oscheius* which demonstrate similar insect infecting behaviour into the functional group of nematodes that have been characterized as 'EPNs'. These nematodes are therefore referred to as entomopathogenic nematodes. Whole genome sequencing of bacteria *Serratia* sp TEL (GenBank accession number KP711410) associated with *Oscheius* sp. TEL-2014 (GenBank accession number KM492926) was performed and a genome draft was obtained. The sequenced genome was assembled using SPADES assembly programme which generated 19 contigs with an average length of 301,767 bp and an N50 of 200,110 bp. The genome size of the isolate is 5,000,541 bp with a G+C content of 59.1%. The genome was annotated using NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) and the annotation method utilized was best-placed reference protein set; GeneMarkS+. Additionally, the functional annotation was carried out by RAST (Rapid Annotation using Subsystem Technology). The annotation on RAST revealed that this species contains 542 subsystems and 4647 genes. The genome sequence can be accessed at DDBJ/EMBL/GenBank under the accession number LDEG00000000.

Keywords: Entomopathogenic nematode; *Serratia* species, whole-genome sequencing, genome assembly, genome annotation

1. Introduction

Bacteria belonging to *Serratia* are gram-negative and are found in diverse environments which include soil, water, digestive tracts of nematodes and other animals all over the world. Members of the *Serratia* genus are known to secrete various extracellular enzymes and secondary metabolites which assist them to survive and persist in certain environments (Lephoto *et al*, 2015). The *Serratia* genus includes species that are insect-pathogens, plant-pathogens, nematode pathogens and human pathogens. However, it is rare to find a species or strain of bacteria in this genus which has evolved a symbiotic association with any invertebrate which is able to functions as its vector for transmitting it to a susceptible insect host which it can infect. Recent findings have confirmed the occurrence of such symbiotic associations with *Serratia marcescens* in *C. briggsae* and in some *Oscheius* species, confirming the existence of the entomopathogenic nematodes in addition to the *Steinernematids* and *Heterorhabditids* (Abebe *et al*, 2010). The *Serratia-Oscheius*-insect tripartite relationships share many similar symbiotic features with the more familiar bacteria-nematode-insect relationship that we see in the *Photorhabdus-Steinernema*-insect and *Xernorhabdus-Heterorhabditis*-insect tripartite symbiotic associations. *Serratia* produces lytic enzymes which is a common trait seen in this genus (Torres-Barragan *et al*, 2011). These enzymes represent components of a chemical attack and defence system. They play a role in virulence and defence against competitors. The also play a role in virulence against susceptible hosts and for overcoming the hosts immune defence systems. Furthermore, most of the *Serratia* bacteria secrete secondary metabolites and other biomolecules with antibacterial activity necessary to assist with survival and interspecies competition in polymicrobial habitats/environments. Adhesion to various surfaces is also a characteristic of many *Serratia* bacteria (Peterson and Tisa, 2013).

Serratia has the ability to cause disease in various insects and cause lethal infections when they have invaded the hemocoel of the insect host. *Serratia spp* targets the hemocytes by attacking them through the secretion of toxins. Cytotoxicity towards mammalian cells has been attributed to hemolysis activity but the mechanism of hymocytic destruction is less clear. Insect pathogenic bacteria found on the surface of the nematodes or colonised in their gut cause mortality upon entry into the insect's hemocoel. The bacteria carry powerful immune depressors and proteins with insecticidal and proteolytic activity which results in the

death of the insect host. Host death is followed by a proliferation of both the bacteria and nematodes in the cadaver. The nematodes go through several stages of development and generations until the nutrients have been depleted in the host cadaver. On depletion of the nutrients the infective juveniles leave the insect cadaver. These infective juveniles are able to survive and persist in the soil for extended periods through anhydrobiosis, and when conditions become favourable the able to recover and infect new hosts (Pervez *et al*, 2013).

There are some *Serratia* strains that express proteases which contribute to their virulence such as the metaloprotease secreted by *S. marcescens* HR-3 (Pr596) which has been shown to exhibit insecticidal activity and proteolytic activity. This protein has led the bacterial strain to be recommended as a biological control agent against of locust. *S. marcescens* is an effective chitin degrader due to the chitinases they secrete which contributes to the bacteria virulence and pathogenicity towards insects. Some *Serratia* species are externally associated with nematodes and are linked with pathogenicity to insects, for example *S. marcescens Bizio*'s is found on the surface of *Steinernema carbocapsae*. In addition, *Oscheius chongmingensis* was also reported to be a true entomopathogenic nematode in a symbiotic relationship with *Serratia nematodiphila* which is a gram-negative, red pigmented and fluorescent bacterium (Peterson and Tisa, 2013).

Next generation sequencing (NGS) has become more affordable in recent years and this economic fact has encouraged many scientists to make greater use of genomic data to explain certain biological processes. This includes sequencing of the genomes of novel species without reference genomes, for example the Illumina sequencing technology is a good example of next generation sequencing. Illumina sequencing services are now done for as little as 20$ for Miseq and 60$ for Hiseq at some institutions in Australia.

When raw genomic data has been obtained by the researcher, it is important to check for any sequencing mistakes in the data, for example too may duplicates or overrepresented sequences which may result from biases in the chemical reactions during the NGS process of sequencing by synthesis. Trimming of poor or low quality reads is one of the quality control methods vastly used prior to assembly of genomes and annotation (Salzberg, 2012).

The availability of high through-put NGS systems has facilitated the production of the draft whole genome sequence for the bacteria *Serratia* sp. TEL (GenBank accession number KP711410) which is symbiotically associated with *Oscheius* sp. TEL-2014 (GenBank accession number KM492926). Using the genetic information derived from results of genome

annotation of *Serratia* the different groups of genes that possibly play a role in *Serratia-Oscheius* relationship have been reviewed.

2. Materials and Methods

2.1 Isolation of bacteria

The protocol used here has been described by (Lephoto *et al*, 2015). Healthy *Galleria mellonella* larvae were infected with *Oscheius* sp. TEL-2014 infective juveniles (IJs) and after 3 days the insect cadavers were placed on White-traps to recover the IJs under a sterile laminar flow cabinet. These IJs were collected and treated with 0.1% of sodium hypochloride for 4 hours and the solution was removed by rinsing the IJs with sterilized distilled water. Sterilized IJs were finely mashed with a sterile plastic pestle after they had been allowed to sediment in a sterile 50 ml Falcon tube. Inoculants consisting of serially diluted homogenized tissue extracts were streaked on sterile nutrient bromothymol blue agar (NBTA) and McConkey agar selective media using a sterilized inoculation loop flamed for each streaking of dilution series on the plate. The agar plates were sealed with Parafilm to prevent or reduce the risk of contamination and they were incubated at 25ºC for 24 hours according to Kaya and Stock, 1997. Single colonies from each agar plate were sub-cultured for species identity verification steps. These colonies were grown in 1 ml of sterile Luria Broth at 25ºC for 24 hours and 100µl of each culture was spread on sterile lipid agar plates and also grown at 25ºC overnight. The resulting bacterial lawns were inoculated with surface sterilized IJs. All the lipid agar plates with cultures which supported EPNs growth were then used for the molecular 16S rDNA species identity procedure.

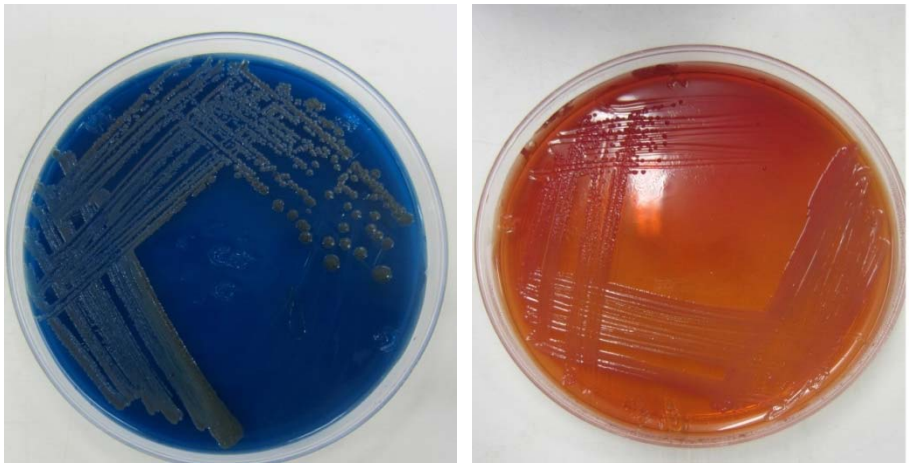


Fig. 1 NBTA and McConkey agar selective media for isolation of symbiotic bacteria from *Oscheius* sp. TEL-2014.

2.2 Bacterial DNA extraction

Genomic DNA was isolated from freshly cultured pure solid bacterial colonies as seen in fig. 1 using the ZR fungal/bacterial DNA MiniPrep kit (Zymo Research, catalog #D3050).

2.3 Agarose gel electrophoresis

To verify the integrity and quality of the genomic DNA, a 1% agarose gel was prepared and ran at 50 volts for 45 minutes in 1X TBE buffer. The ingredients are on appendix 2.

2.4 Polymerase chain reaction

Procedure briefly described on (Lephoto *et al*, 2015). A polymerase chain reaction was used for the amplification of the 16S rDNA gene which is widely conserved in many bacterial species. EUB968 forward primer 5′-ACGGGCGGTGTGTC-3′ Tm (°C) = 62 and UNIV1382 reverse primer 5′-AACGCGAAGAACCTTAC-3′ Tm (°C) = 66 were used for the PCR and short read Sanger sequencing of this DNA region at Inqaba Biotech, South Africa. Bioedit and FinchTV software were used to access the quality of the sequence obtained and also remove low quality regions on both ends of the sequence and unknown bases were also trimmed out. The cleaned sequence was subjected to NCBI BLAST under the default settings for highly similar alignments. The unknown sequenced best aligned with a novel *Serratia* bacterial species which was then registered on NCBI Genbank assigned the name *Serratia* sp. TEL. The Genbank accession number is KP711410. The sequence information may be accessed on: http://www.ncbi.nlm.nih.gov/nuccore/KP711410

2.5 Whole genome sequencing: Library preparation

Genomic DNA paired-end libraries were generated with the Nextera DNA sample preparation kit (Illumina) and indexed using the Nextera index kit (Illumina). Paired-end (2 × 300 bp) sequencing was performed on a MiSeq Illumina using the MiSeq reagent kit v3 at the Agricultural Research Council (ARC) Biotechnology Platform. Miseq sequencing system was used because it is faster and good for small genomes.

2.6 Quality control

Quality and adapter trimming was performed with fastq-mcf toolkit. Low quality reads were cleaned out of the genome and the nextera adapter sequences were removed from the genome as well prior assembly. Another quality control report was generated to access the quality of the data after trimming and to ensure that all adapter sequences have been removed. Brief explanatory guidelines and examples of the quality control report are attached in appendix 2.

2.7 Genome assembly

The genome was assembled using SPADES

```
Running SPADES
spade.py —o output (create output folder)
Spades.py - - pe1-1 data/bacteria fy.qz - - pe1-2 data/bacteria z.qz - output
```

Fig. 2 An example of the script used to run a SPADES genome assembly

2.8 Genome annotation

The contigs were annotated using NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) and the annotation method employed was best-placed reference protein set; GeneMarkS+ shown in fig. 3. The functional annotation was carried out by RAST (Rapid Annotation using Subsystem Technology) hosted by Fellowship for Interpretation.

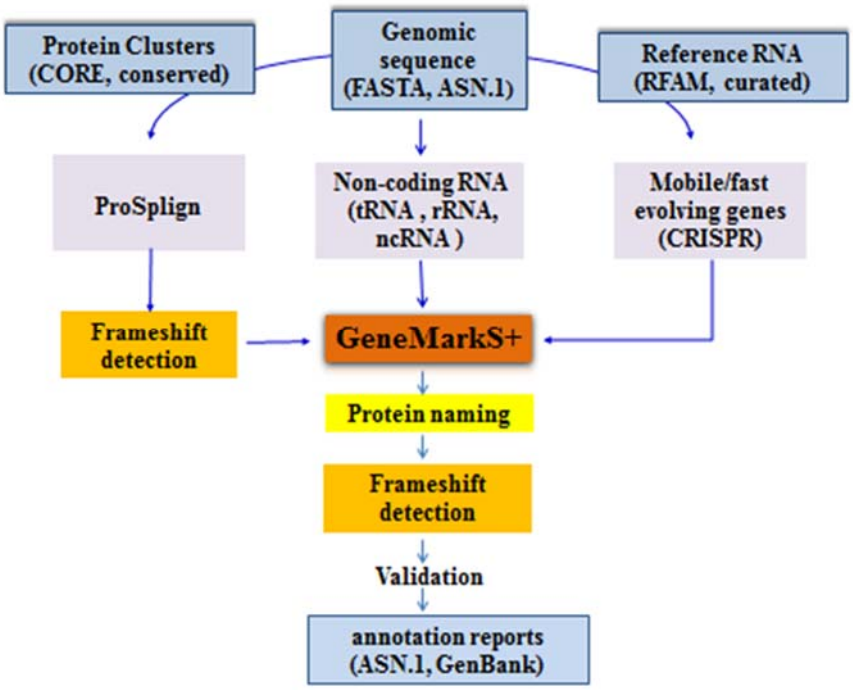


Fig. 3 Annotation pipeline from NCBI which was used to annotate *Serratia* sp TEL. The diagram was adopted from (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/process/)

3. Results and discussion

Table 1

Summary of the *Serratia* sp. TEL genome sequencing, assembly and annotation (Lephoto and Gray, 2015)

Description	
Organism	*Serratia sp.*
Strain	TEL
GenBank accession number	KP711410
Sequencer or array type	Illumina MiSeq
Quality control tool	Fastq-mcf toolkit
Genome assembler	*SPADES de novo*
Number of contigs	19 contigs
Average contig length	301,767 bp
N50	200,110 bp
Genome annotation tools	RAST and PGAAP
Whole genome DDBJ/EMBL/GenBank	LDEG00000000
Size of genome	5,000,541 bp
Number of subsytems	542
G+C content	59.1%
Number of genes	4,647
Entomopathogenic nematode (vector)	*Oscheius* sp. TEL-2014
Genbank accession number	KM492926
Sample source location	Grassland in Suikerbosrand Nature Reserve near Johannesburg in South Africa

Table 2 *Serratia* sp. TEL genome assembly results showing contigs length. 19 contigs were generated from the assembly and the smallest contig length was 1868bp and the longest was 232,434bp.

contigs	**Length of contig**
1	1,868
2	4,506
3	5,552
4	30,700
5	373,610
6	613,447
7	671,070
8	83,051
9	840,889
10	120,220
11	123,757
12	124,218
13	150,719
14	845,737
15	156,082
16	182,591
17	217,529
18	222,561
19	232,434

Table 3

Comparison of physical parameters of *Serratia* sp. TEL genome draft with two *Serratia* genomes

Feature	Organism		
	Serratia sp. TEL	*Serratia* sp. SCBI	*Serratia* Db11
Genome size (MB)	5.05	5.04	5.12
GC (%) content	59.1	59.3	59.1
Number of genes or predicted ORFs	4,647	4599	4736
Pseudogenes	36	N/A	N/A
rRNA genes	13	7	7
tRNA genes	88	84	88
Non-coding RNAs	15	N/A	N/A
Frameshifted genes	9	N/A	N/A

Physical properties for *Serratia* sp. TEL were obtained by the using the NCBI Prokaryotic Genome Automatic Annotation Pipeline.

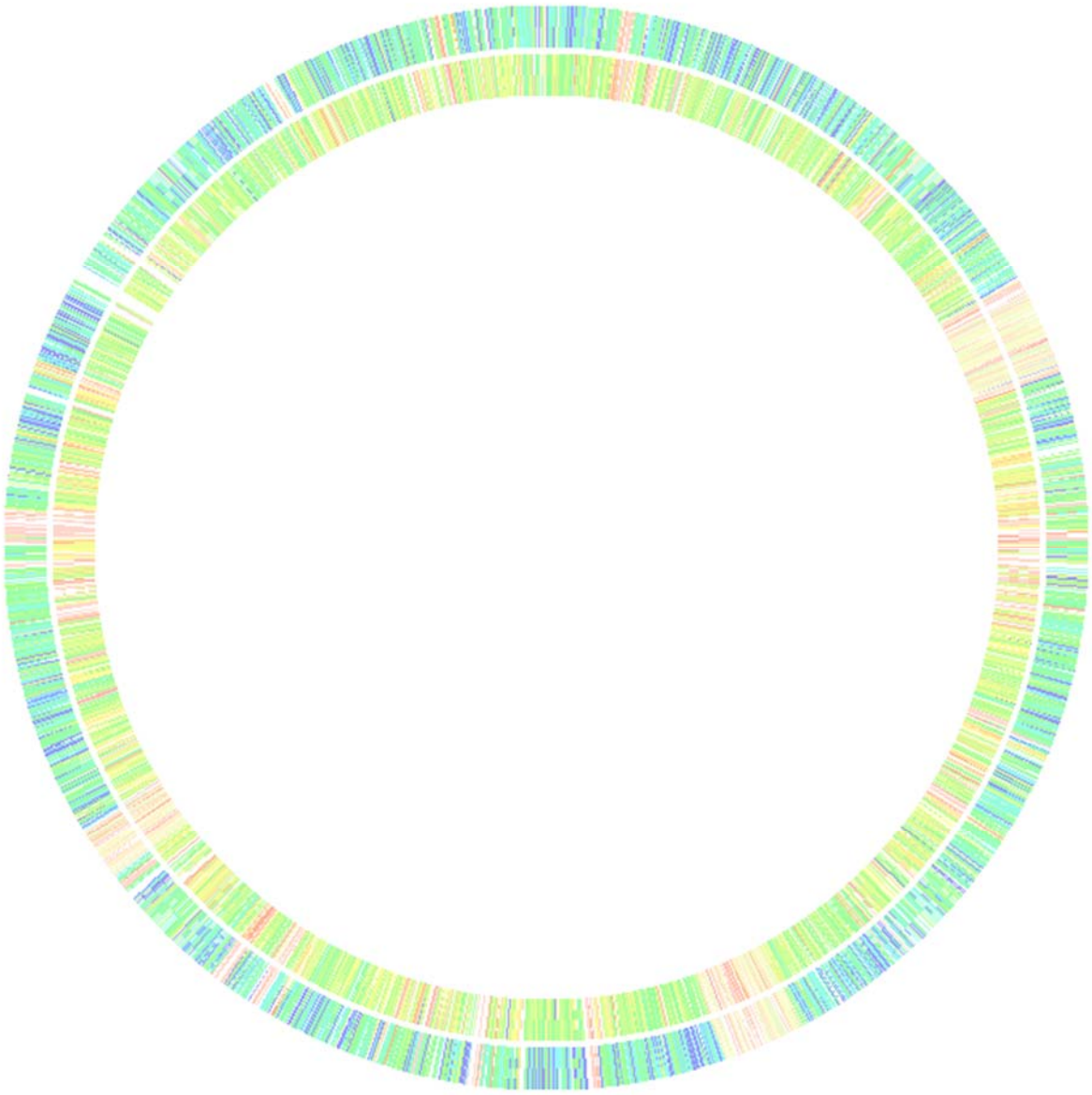


Fig. 4 Circular representation of the aligned genomes comparison of *Serratia* sp. TEL to two references: *Serratia marcescens* Db11 and *Serratia proteamaculans*.

The percentage of sequence identity is represented in colours visible on the circle and shown in fig 6.

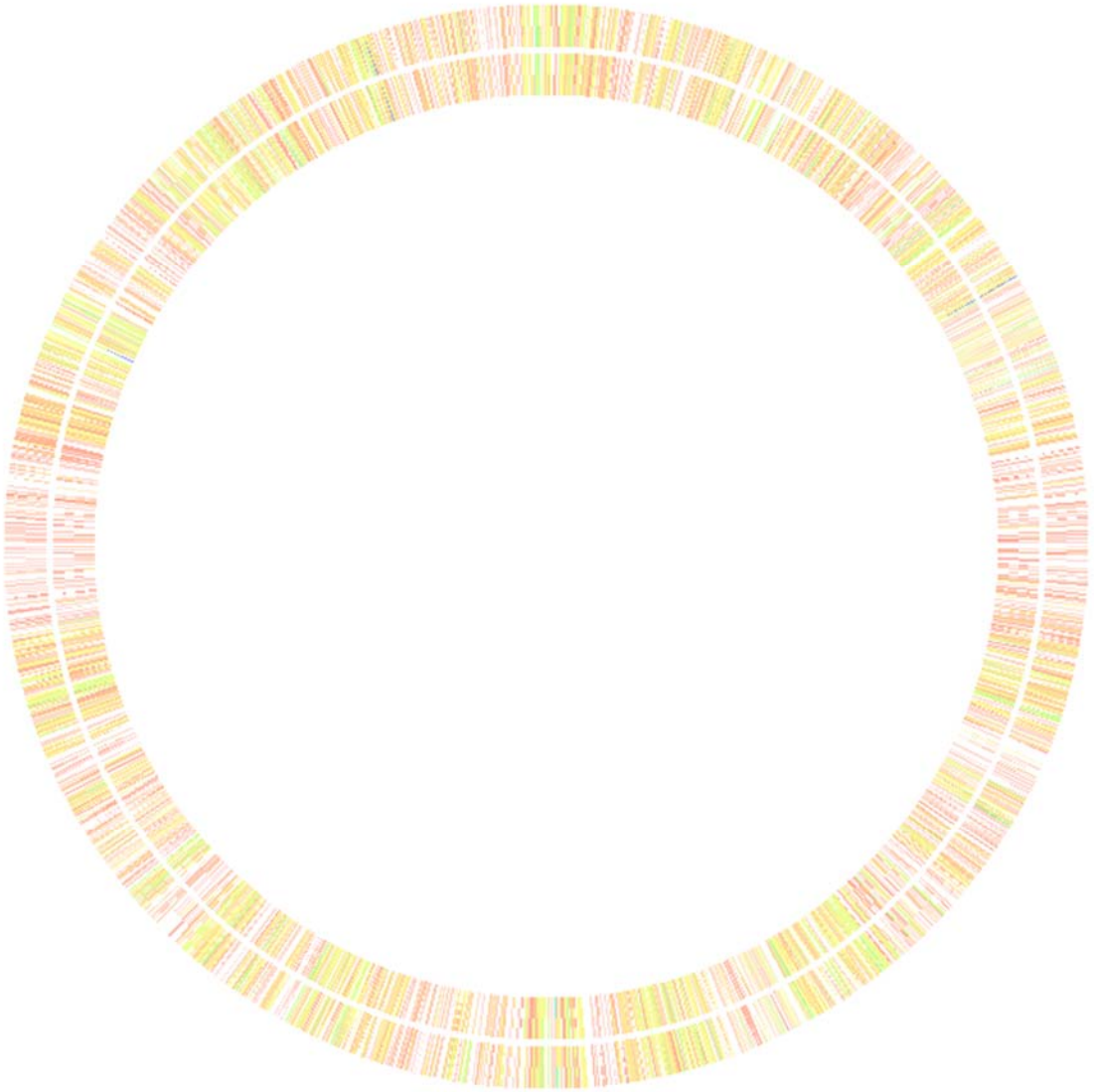


Fig. 5 Circular representation of the aligned genomes comparison of *Serratia* sp. TEL to two references: *Photorhabdus asymbiotica* subsp. *asymbiotica* and *Photorhabdus luminescens* subsp. *laumondii* TT01

.

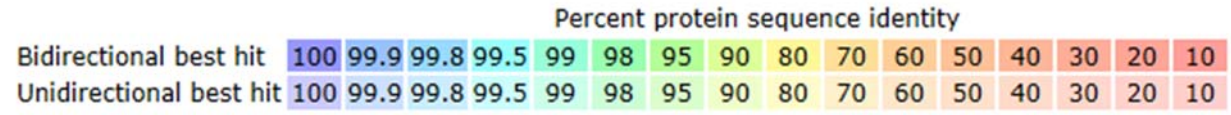


Fig. 6 Percent protein sequence identity

The percent protein sequence identity in fig. 6 was used as a colour coded reference for sequence comparison of the genomes in fig 4 and 5 in the circular representation of the compared genome.

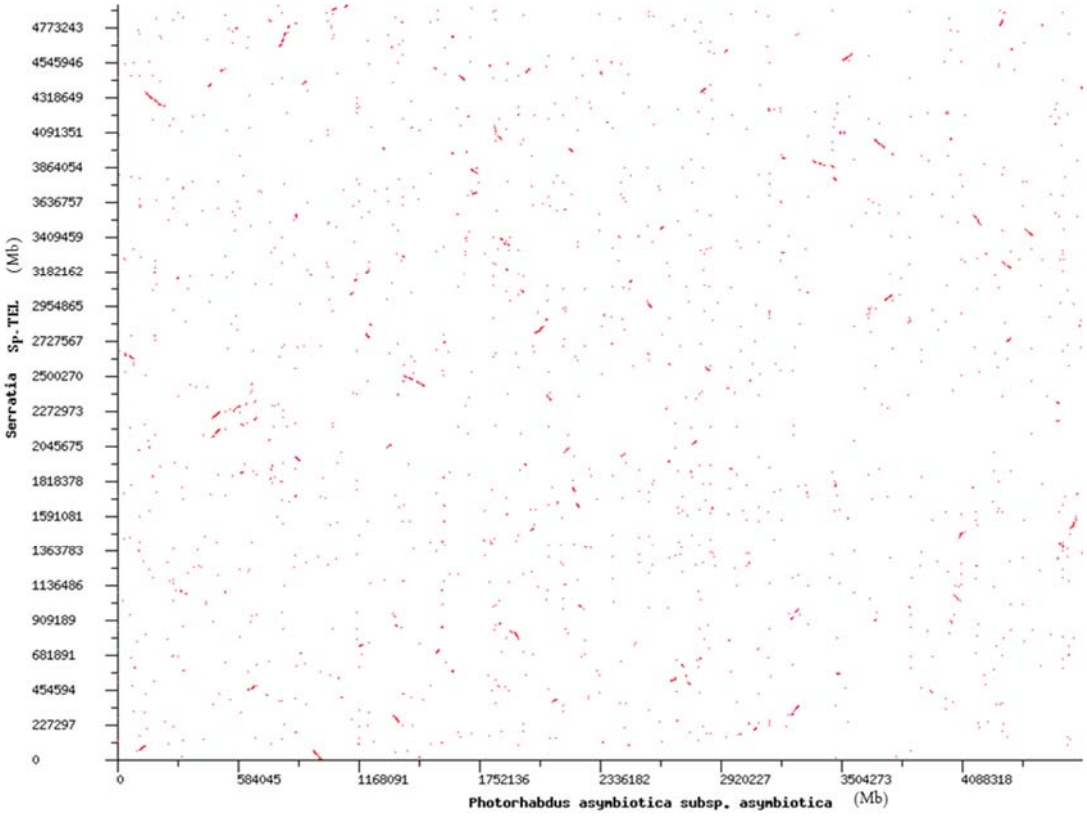


Fig. 7 Dot plot comparing of *Serratia* sp. TEL genome to *Photorhabdus asymbiotica* subsp. *asymbiotica* genome.

Genome sequence similarity is seen where the dots are closely packed and differences are symbolised by scattered dots which do not merge. Bacteria compared in fig. 7 and fig. 8 belongs to two different genera namely *Serratia* and *Photorhabdus.*

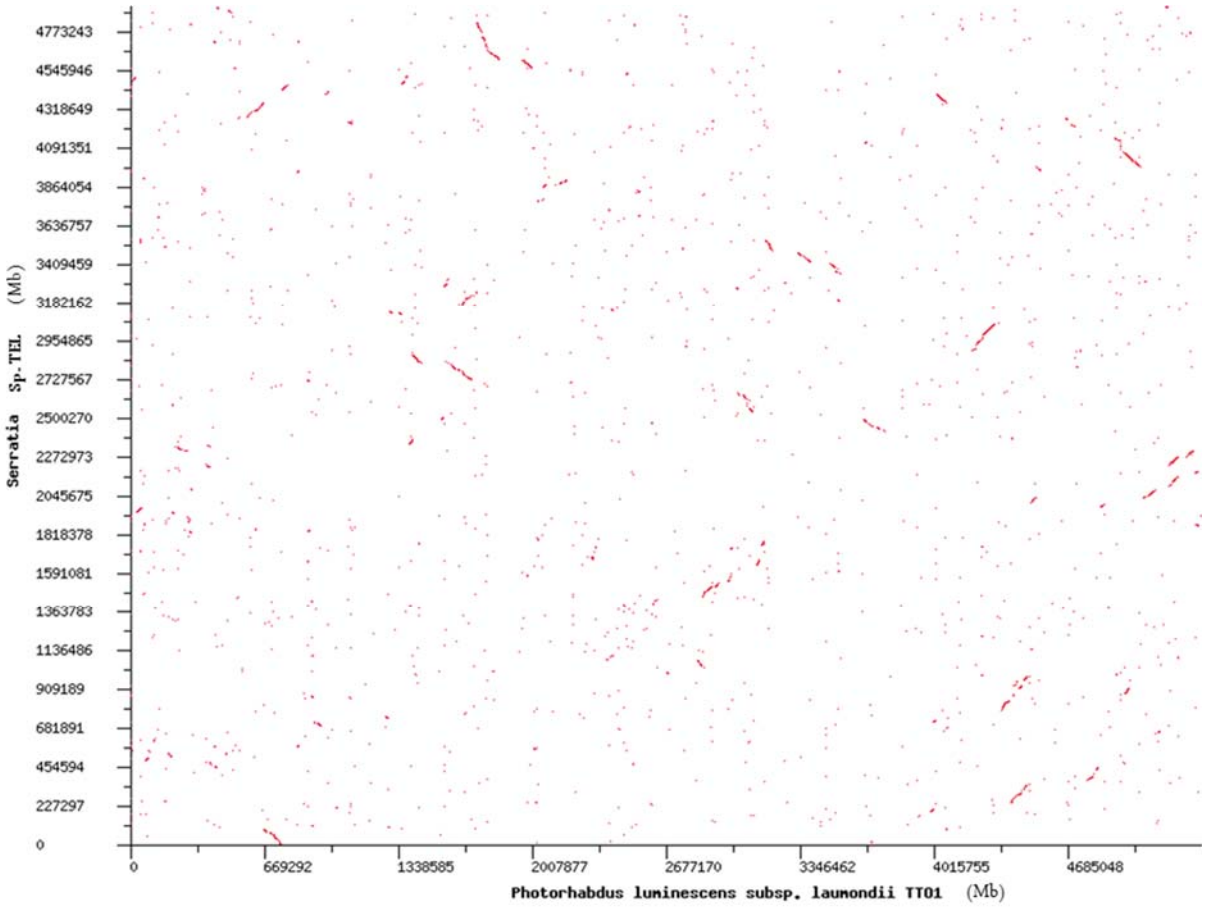


Fig. 8 Dot plot comparing of *Serratia* sp. TEL genome to *Photorhabdus luminescens* subsp. *laumondii* TT01 genome.

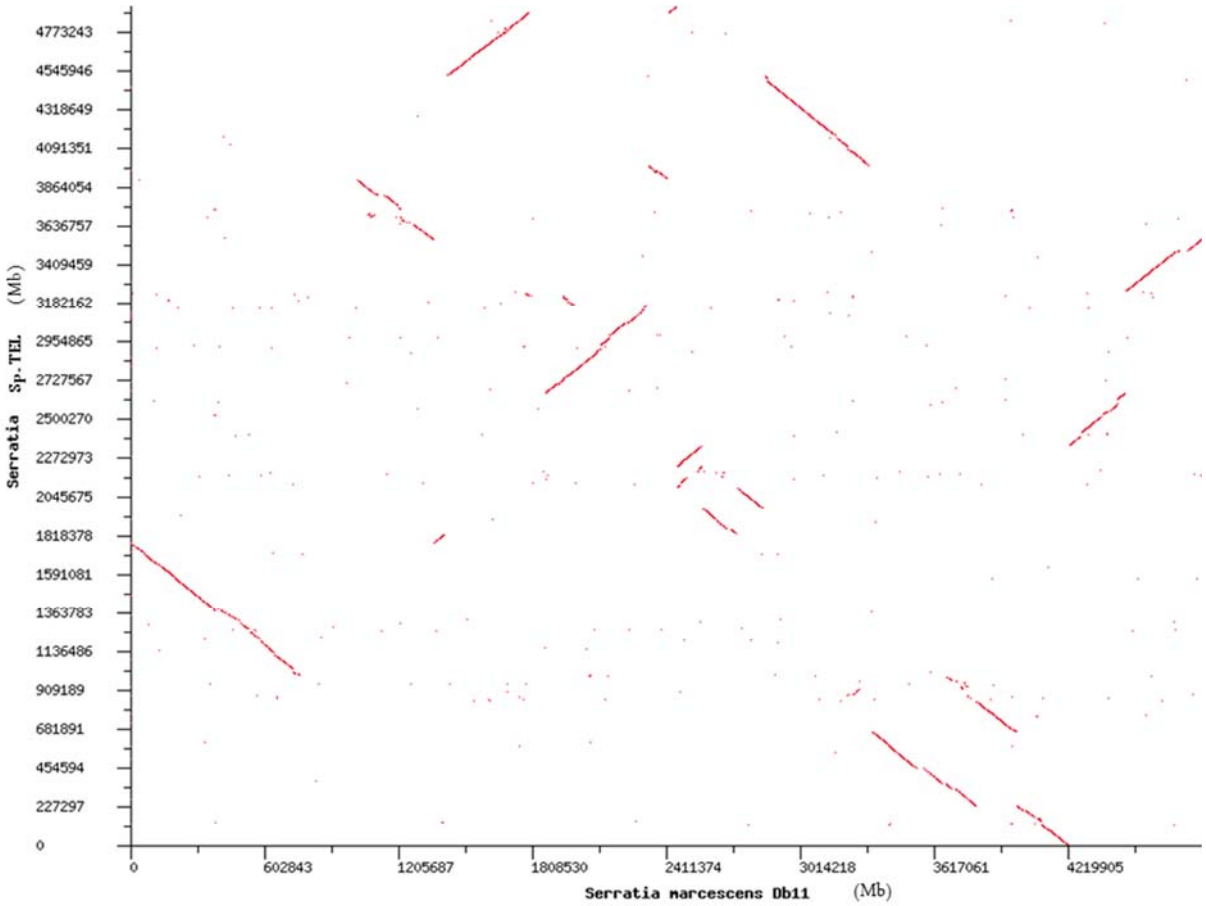


Fig. 9 Dot plot comparing of *Serratia* sp. TEL genome to *Serratia marcescens* Db11 genome. Sequence similarity is observed along the genome sequences of the two bacterial species. This was expected because both bacteria belong to the genus *Serratia*.

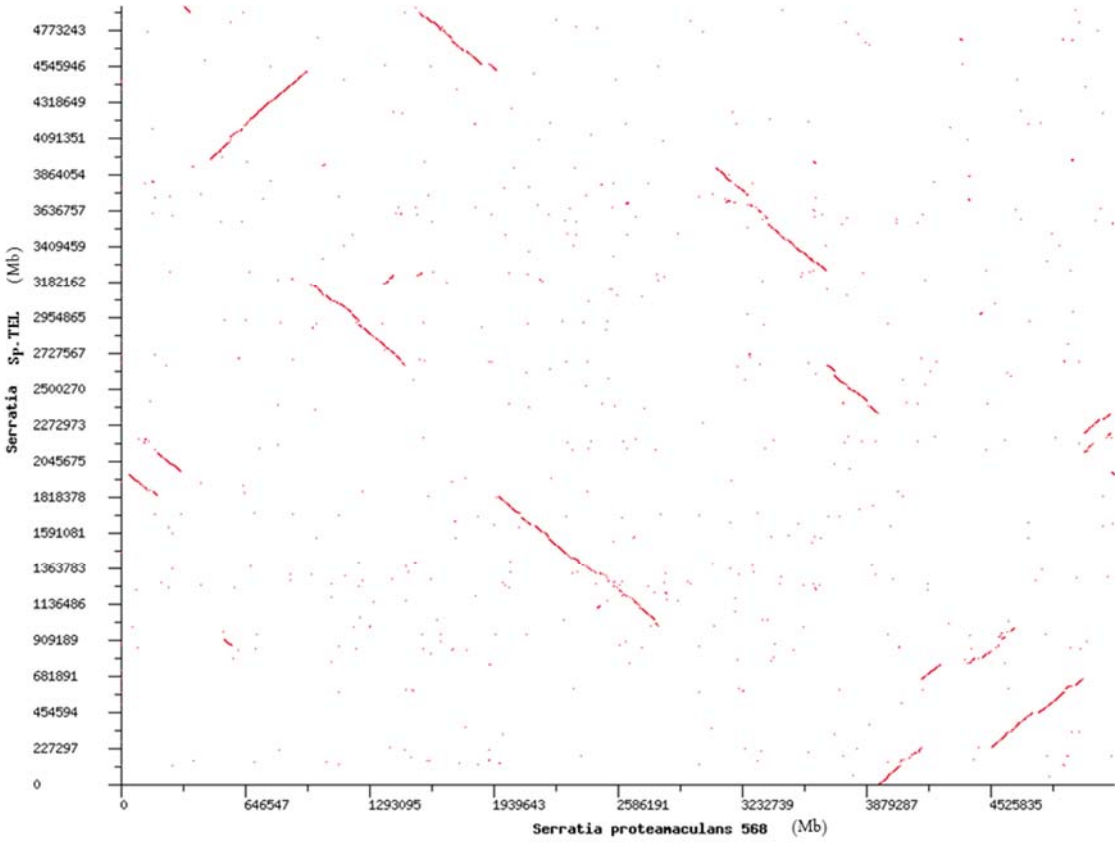


Fig. 10 Dot plot comparing of *Serratia* sp. TEL genome to *Serratia proteamaculans* genome. A great degree of sequence similarity is observed along the genome sequences of the two bacterial species. This was expected because both bacteria belong to the genus *Serratia*.

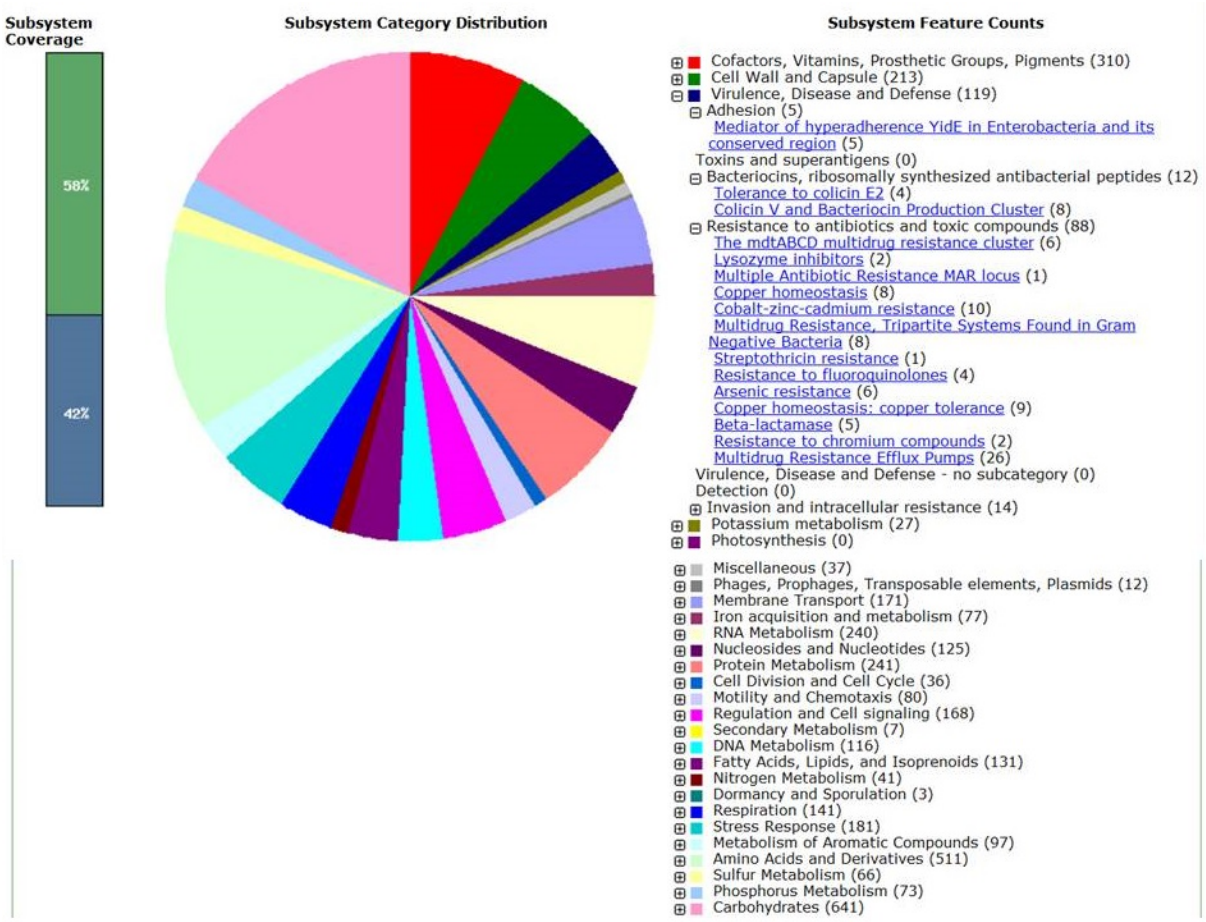


Fig. 11 The distribution of subsystems

What is a subsystem?

A subsystem in genomics is defined as a set of genetic functional roles which cooperate or work together in facilitating or driving a wide diversity of different biological processes. The word subsystem may be assumed as a comprehensive term for a given signal transduction, metabolic or biosynthetic pathway. Most pathways represent functional and structural adaptations or responses to the internal and external environment of the organism. RAST annotation software uses the term subsystems to group genes which work together to bring about a specific role, adaptation or function in an organism.

RAST annotation software tool identified a total of 4618 genes and 103 RNA encoding genes in *Serratia* sp. TEL genome. This annotated genome was found to have 542 subsystems each playing its specific role in this bacterium. This chapter discusses the virulence, defence and disease subsystem which contains genes in involved in cell adhesion to other cells or surfaces. Other subsystems will also be briefly discussed to support some of the physical observations made in the study.

The virulence, defence and disease subsystem

It is interesting to note that the virulence disease and defence subsystem contains 119 genes which may be implicated in the pathogenicity of *Serratia* sp. TEL. Table 4 summarises the role of these genes which are also briefly described in (Lephoto and Gray, 2015).

Table 4

Category: Virulence, disease and defence gene family obtained from RAST.

Subcategory	subsystem	Role
Adhesion	Mediator of hyperadherence YidE in Enterobacteria and its conserved region	16 kDa heat shock protein B
Adhesion	Mediator of hyperadherence YidE in Enterobacteria and its conserved region	Mediator of hyperadherence YidE
Adhesion	Mediator of hyperadherence YidE in Enterobacteria and its conserved region	16 kDa heat shock protein A
Adhesion	Mediator of hyperadherence YidE in Enterobacteria and its conserved region	Uncharacterized protein YidR
Adhesion	Mediator of hyperadherence YidE in Enterobacteria and its conserved region	Outer membrane lipoprotein YidQ
Bacteriocins, ribosomally synthesized antibacterial peptides	Tolerance to colicin E2	Conserved uncharacterized protein CreA
Bacteriocins, ribosomally synthesized antibacterial peptides	Tolerance to colicin E2	Two-component response regulator CreC
Bacteriocins, ribosomally synthesized antibacterial peptides	Tolerance to colicin E2	Inner membrane protein CreD
Bacteriocins, ribosomally synthesized antibacterial peptides	Tolerance to colicin E2	Two-component response regulator CreB
Bacteriocins, ribosomally synthesized antibacterial peptides	Colicin V and Bacteriocin Production Cluster	Folylpolyglutamate synthase (EC 6.3.2.17)
Bacteriocins, ribosomally synthesized antibacterial peptides	Colicin V and Bacteriocin Production Cluster	Amidophosphoribosyltransferase (EC 2.4.2.14)
Bacteriocins, ribosomally synthesized	Colicin V and Bacteriocin Production Cluster	Colicin V production protein

antibacterial peptides		
Bacteriocins, ribosomally synthesized antibacterial peptides	Colicin V and Bacteriocin Production Cluster	Dihydrofolate synthase (EC 6.3.2.12)
Bacteriocins, ribosomally synthesized antibacterial peptides	Colicin V and Bacteriocin Production Cluster	Acetyl-coenzyme A carboxyl transferase beta chain (EC 6.4.1.2)
Bacteriocins, ribosomally synthesized antibacterial peptides	Colicin V and Bacteriocin Production Cluster	DedA protein
Bacteriocins, ribosomally synthesized antibacterial peptides	Colicin V and Bacteriocin Production Cluster	DedD protein
Bacteriocins, ribosomally synthesized antibacterial peptides	Colicin V and Bacteriocin Production Cluster	tRNA pseudouridine synthase A (EC 4.2.1.70)
Resistance to antibiotics and toxic compounds	The mdtABCD multidrug resistance cluster	Multidrug transporter MdtC
Resistance to antibiotics and toxic compounds	The mdtABCD multidrug resistance cluster	Response regulator BaeR
Resistance to antibiotics and toxic compounds	The mdtABCD multidrug resistance cluster	Multidrug transporter MdtB
Resistance to antibiotics and toxic compounds	The mdtABCD multidrug resistance cluster	Probable RND efflux membrane fusion protein
Resistance to antibiotics and toxic compounds	The mdtABCD multidrug resistance cluster	Sensory histidine kinase BaeS
Resistance to antibiotics and toxic compounds	The mdtABCD multidrug resistance cluster	Multidrug transporter MdtD
Resistance to antibiotics and toxic compounds	Lysozyme inhibitors	Membrane-bound lysozyme inhibitor of c-type lysozyme
Resistance to antibiotics and toxic compounds	Lysozyme inhibitors	Inhibitor of vertebrate lysozyme precursor
Resistance to antibiotics and toxic compounds	Multiple Antibiotic Resistance MAR locus	Multiple antibiotic resistance protein MarC
Resistance to antibiotics and toxic compounds	Copper homeostasis	Cytochrome c heme lyase subunit CcmF
Resistance to antibiotics and toxic compounds	Copper homeostasis	Cytochrome c heme lyase subunit CcmH
Resistance to antibiotics and toxic compounds	Copper homeostasis	Copper-translocating P-type ATPase (EC 3.6.3.4)
Resistance to antibiotics and toxic compounds	Copper homeostasis	Blue copper oxidase CueO precursor
Resistance to antibiotics and toxic compounds	Copper homeostasis	Copper resistance protein D
Resistance to antibiotics and toxic compounds	Copper homeostasis	Copper resistance protein C precursor

Resistance to antibiotics and toxic compounds	Cobalt-zinc-cadmium resistance	DNA-binding heavy metal response regulator
Resistance to antibiotics and toxic compounds	Cobalt-zinc-cadmium resistance	Cobalt-zinc-cadmium resistance protein
Resistance to antibiotics and toxic compounds	Cobalt-zinc-cadmium resistance	Cobalt-zinc-cadmium resistance protein CzcD
Resistance to antibiotics and toxic compounds	Cobalt-zinc-cadmium resistance	Cobalt-zinc-cadmium resistance protein CzcA
Resistance to antibiotics and toxic compounds	Cobalt-zinc-cadmium resistance	Zinc transporter ZitB
Resistance to antibiotics and toxic compounds	Cobalt-zinc-cadmium resistance	Transcriptional regulator, MerR family
Resistance to antibiotics and toxic compounds	Cobalt-zinc-cadmium resistance	Cation efflux system protein CusA
Resistance to antibiotics and toxic compounds	Multidrug Resistance, Tripartite Systems Found in Gram Negative Bacteria	Outer membrane component of tripartite multidrug resistance system
Resistance to antibiotics and toxic compounds	Multidrug Resistance, Tripartite Systems Found in Gram Negative Bacteria	Membrane fusion component of tripartite multidrug resistance system
Resistance to antibiotics and toxic compounds	Multidrug Resistance, Tripartite Systems Found in Gram Negative Bacteria	Inner membrane component of tripartite multidrug resistance system
Resistance to antibiotics and toxic compounds	Streptothricin resistance	Streptothricin acetyltransferase, Streptomyces lavendulae type
Resistance to antibiotics and toxic compounds	Resistance to fluoroquinolones	DNA gyrase subunit B (EC 5.99.1.3)
Resistance to antibiotics and toxic compounds	Resistance to fluoroquinolones	DNA gyrase subunit A (EC 5.99.1.3)
Resistance to antibiotics and toxic compounds	Resistance to fluoroquinolones	Topoisomerase IV subunit B (EC 5.99.1.-)
Resistance to antibiotics and toxic compounds	Resistance to fluoroquinolones	Topoisomerase IV subunit A (EC 5.99.1.-)
Resistance to antibiotics and toxic compounds	Arsenic resistance	Arsenical resistance operon repressor
Resistance to antibiotics and toxic compounds	Arsenic resistance	Arsenic efflux pump protein
Resistance to antibiotics and toxic compounds	Arsenic resistance	Arsenate reductase (EC 1.20.4.1)
Resistance to antibiotics and toxic compounds	Copper homeostasis: copper tolerance	Secreted protein, suppressor for copper-sensitivity ScsC

Resistance to antibiotics and toxic compounds	Copper homeostasis: copper tolerance	Suppression of copper sensitivity: putative copper binding protein ScsA
Resistance to antibiotics and toxic compounds	Copper homeostasis: copper tolerance	Magnesium and cobalt efflux protein CorC
Resistance to antibiotics and toxic compounds	Copper homeostasis: copper tolerance	Membrane protein, suppressor for copper-sensitivity ScsB
Resistance to antibiotics and toxic compounds	Copper homeostasis: copper tolerance	Copper homeostasis protein CutE
Resistance to antibiotics and toxic compounds	Copper homeostasis: copper tolerance	Membrane protein, suppressor for copper-sensitivity ScsD
Resistance to antibiotics and toxic compounds	Copper homeostasis: copper tolerance	Copper homeostasis protein CutF precursor
Resistance to antibiotics and toxic compounds	Copper homeostasis: copper tolerance	Periplasmic divalent cation tolerance protein CutA
Resistance to antibiotics and toxic compounds	Beta-lactamase	Beta-lactamase class C and other penicillin binding proteins
Resistance to antibiotics and toxic compounds	Beta-lactamase	Beta-lactamase (EC 3.5.2.6)
Resistance to antibiotics and toxic compounds	Beta-lactamase	Metal-dependent hydrolases of the beta-lactamase superfamily I
Resistance to antibiotics and toxic compounds	Resistance to chromium compounds	Chromate transport protein ChrA
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	RND efflux system, membrane fusion protein CmeA
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	RND efflux system, outer membrane lipoprotein CmeC
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Multi antimicrobial extrusion protein (Na(+)/drug antiporter), MATE family of MDR efflux pumps
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Transcription repressor of multidrug efflux pump acrAB operon, TetR (AcrR) family
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	RND efflux system, inner membrane transporter CmeB
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Multidrug-efflux transporter, major

		facilitator superfamily (MFS) (TC 2.A.1)
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Macrolide export ATP-binding/permease protein MacB (EC 3.6.3.-)
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Acriflavin resistance protein
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Membrane fusion protein of RND family multidrug efflux pump
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Macrolide-specific efflux protein MacA
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Type I secretion outer membrane protein, TolC precursor
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	SSU ribosomal protein S7p (S5e)
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	Translation elongation factor G
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Membrane fusion protein of RND family multidrug efflux pump
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Macrolide-specific efflux protein MacA
Resistance to antibiotics and toxic compounds	Multidrug Resistance Efflux Pumps	Type I secretion outer membrane protein, TolC precursor
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	SSU ribosomal protein S7p (S5e)
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	Translation elongation factor G
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	Translation elongation factor Tu
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	SSU ribosomal protein S12p (S23e)
Invasion and intracellular resistance	Mycobacterium virulence operon involved in DNA transcription	DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)
Invasion and intracellular resistance	Mycobacterium virulence operon involved in DNA transcription	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)

Invasion and intracellular resistance	Mycobacterium virulence operon possibly involved in quinolinate biosynthesis	Quinolinate synthetase (EC 2.5.1.72)
Invasion and intracellular resistance	Mycobacterium virulence operon possibly involved in quinolinate biosynthesis	Quinolinate phosphoribosyltransferase [decarboxylating] (EC 2.4.2.19)
Invasion and intracellular resistance	Mycobacterium virulence operon possibly involved in quinolinate biosynthesis	L-aspartate oxidase (EC 1.4.3.16)
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins)	LSU ribosomal protein L35p
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins)	Translation initiation factor 3
Invasion and intracellular resistance	Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins)	LSU ribosomal protein L20p

Within the virulence, defence and disease subsystem, there are 5 genes involved in adhesion and 12 encoding for bacteriocins, ribosomally synthesised antibacterial peptides. This subsystem also contains 88 genes which are responsible for resistance to antibiotics and production of toxic compounds such as lysozyme inhibitors, beta-lactamase as listed in table 4.

The *Serratia* sp. TEL genome was found to contain genes encoding for pigments; specific genes encoding enzyme for pigments biosynthesis were not present in the annotation.

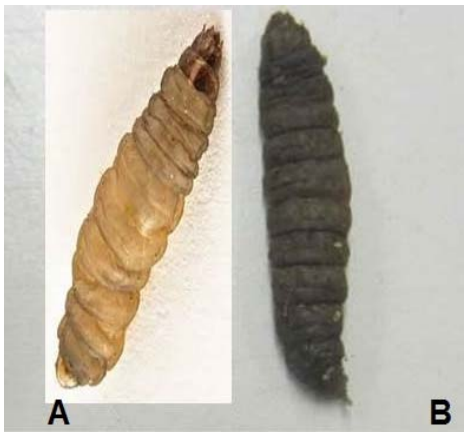


Fig. 12 colour change observed in *Galleria mellonella* insect larva when infected by *Oscheius* sp. TEL-2014 within 48 hours. A: healthy lava and B: brown infected larva.

The brown colour changes observed when *Galleria mellonella* is infected by *Oscheius* sp. TEL-2014 is due to the production pigments. These pigments are characteristic of the presence of *Serratia* sp. TEL as demonstrated in fig. 12.

The cell wall and capsule subsystem

The cell wall and capsule subsystem is encoded for by 213 genes in which 37 of these genes code for capsular an extracellular polysaccharides. *Serratia* sp. TEL was found to be gram-negative based on the gram stain test. RAST also confirms that this species has 75 proteins forming the gram-negative cell wall components and 0 genes encoding for gram-positive cell wall components. These gram-negative components include genes for lipopolysaccharide assembly, lipid A modifications, and inner membrane protein YhiD.

The cell division and cell cycle subsystem

The cell division and cell cycle subsystem has 36 genes with 24 of them responsible for bacterial cytoskeleton, 5 genes for chromosome condensation and some for the macromolecular synthesis operon. There were no coding regions identified for the checkpoint formation. This subsystem is of great importance as it may in also ultimately provide some understanding for factors influencing the proliferation of these bacteria, and also with reference to its growth rate during the EPNs life cycle, and insect host infection. All these aspects need to be further investigated in future studies.

The regulation and cell signalling subsystem

The regulation and cell signalling subsystem are comprised of 168 genes. There are 5 genes that are linked to ppGpp metabolism. It was previously highlighted that most pathogenic bacteria symbiotically associated with EPNs secrete antimicrobial compounds such as xenocoumacins, indoles and dithiolopyrrolones. These antibacterial compounds were postulated to inhibit protein and RNA synthesis by amplifying production of guanosine-30, 50-bispyrophosphatase (ppGpp), a regulatory protein (Ji *et al*, 2004). The metabolites of these insect-killing bacteria have applications in medicine and agriculture.

Stress response subsystem

According to the results generated from RAST, *Serratia* sp. TEL has proteins linked to dormancy. A stress response subsystem was found to contain osmotic stress genes, oxidative

stress genes, acid stress proteins, cold shock, heat shock and detoxification genes but no desiccation tolerance genes.

Additionally, the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) and the annotation method also recognized CDS; rRNA; tRNA; ncRNA and repeat regions features. 4,647 genes were found and 4,495 out of the total were protein-coding sequences (CDS) while RAST found 4618 coding regions. According to PGAAP there are 36 pseudo genes, 2 CRISPR Arrays, 13 rRNA genes with five operons (5S, 16S, 23S), 88 tRNAs, 15 non-coding RNA (ncRNA) and 9 frameshifted genes are present in the genome.

4. Conclusion

The draft genome sequence *Serratia* sp. TEL will allow for the investigation of all the identified genes and subsystems and will be imperative in furthering the understanding of their pathogenicity against insects.

5. References

Ji D, Yi Y, Kang G, Choi Y, Kim P, Baek N and Kim Y, 2004. Identification of an antibacterial compound benzylideneacetone from *Xenorhabdus nematophila* against major plant-pathogenic bacteria. FEMS Microbiology Letters, 239:241–248.

Kaya H.K and Stock S.P, 1997. Techniques in insect nematology. In Manual of Techniques in Insect Pathology 281-324.

Lephoto T.E, Featherston J, Gray V.M, 2015. Draft whole-genome sequence of *Serratia* sp. strain TEL, associated with *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae) isolated from a grassland in South Africa. Genome Announc 3(4):e00747-15.

Lephoto T.E and Gray V.M, 2015. Genome sequencing and annotation of *Serratia* sp. strain TEL. Genomics data, 6:54-56.

Pervez R, Eapen S.J, Devasahayam S and Jacob T.K, 2012. A new species of entomopathogenic nematode *Oscheius gingeri* sp. n. (Nematoda: *Rhabditidae*) from ginger rhizosphere, Archives of Phytopathology And Plant Protection, 46:526-535.

Petersen L.M, Tisa L.S, 2013. Friend or foe? A review of the mechanisms that drive *Serratia* towards diverse lifestyles. Canadian journal of microbiology, 59:627-640.

Salzberg S.L, Phillippy A.M, Zimin A, Puiu D, Magoc T, Koren S, Treangen T.J, Schatz M.C, Delcher A, L, Roberts M, Marcxais G, Pop M and Yorke J.A, 2012. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Cold Spring Harbor Laboratory Press.

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

CHAPTER 6

Analysis of the bacterial biodiversity of hemocoel isolated from *Gallaria Mellonella* infected with *Oscheius* sp. TEL-2014 using 16S rDNA Based Metagenomics.

Abstract

Through continuous natural selection entomopathogenic nematodes have successfully evolved strategies to remain virulent by overcoming the immune defence systems of their insect hosts. In this study we analysed the hemolyph of *Galleria Mellonella* infected by *Oscheius* nematodes using 16S rDNA based metagenomics. We also aimed to confirm the presence of *Serratia* bacteria and possibly other bacterial species present in *Oscheius* nematodes. The 16S rDNA region was sequenced using Illumina miseq sequencer and analysed the results using Miseq Reporter Software by Illumina. Results grouped the total reads obtained from the sequencing according to taxonomic levels and show that *Serratia* bacteria are present in the infected insect hemolymph and absent in uninfected insect larvae. 96% of the identified bacteria belong to the genus *Serratia*; supporting previous studies which have proved that entomopathogenic *Oscheius* nematodes have a relationship with *Serratia* bacteria.

Keywords: Entomopathogenic nematodes, 16S rDNA and metagenomics.

1. Introduction

Entomopathogenic nematodes (EPNs) have evolved unique relationships with virulent insect pathogenic bacteria which they carry in their gut (Alia *et al*, 2014). These infectious bacterial species use EPNs as vectors to enter the target insect host's hemolymph and destroy the hemocytes of the insect. Powerful toxins and immune depressors are secreted by the bacteria which help to facilitate the infection of the host which ultimately results in the death of the host. A representative life cycle of EPNs also highlights the exploitation of the nutrients that are made available in the dead insect host for both the symbiotic bacteria and the infective juveniles thereby making possible their proliferation as shown in fig. 1. Once the nutrients reservoirs become depleted, the non-feeding dauer stage of nematodes are released into the soil where they either persist or search for new hosts to infect (Park *et al*, 2011 and Dillman *et al*, 2012).

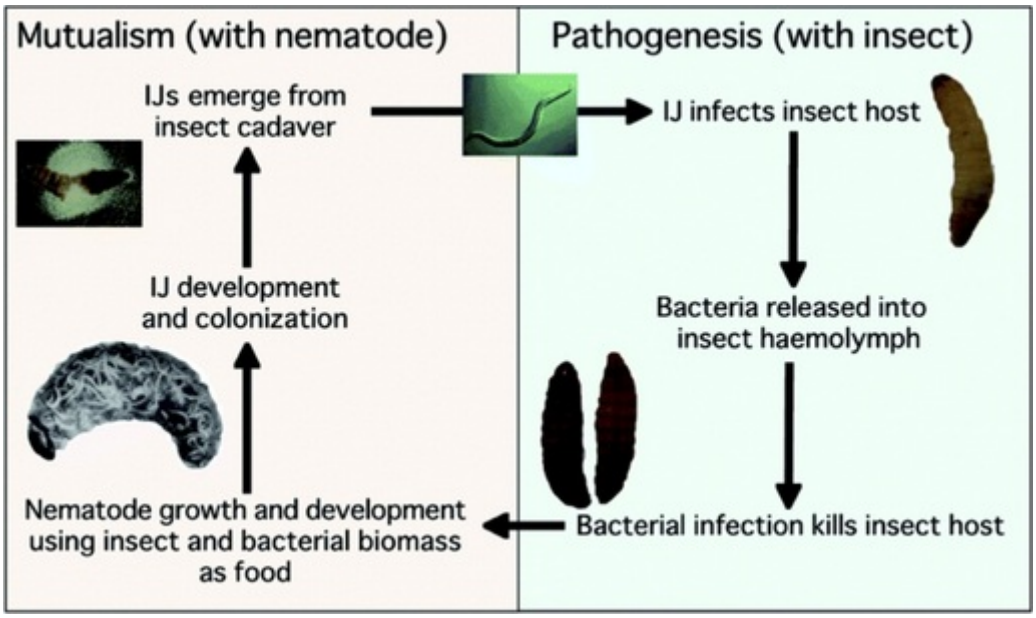


Fig. 1 Life cycle of entomopathogenic nematodes

(Image adopted from: http://www.djpark.org/xenorhabdus-nematophila.html)

Several metagenomic studies have been conducted to investigate the biodiversity of prokaryotic communities from specific environments using 16S rDNA sequences (Zeisel *et al*, 2013). Our study attempts to investigate bacterial communities present in the hemolyph of *Galleria mellonella* infected with EPNs *Oscheius* sp. TEL-2014 isolated from South Africa.

Next generation sequencing techniques were used to analyse sequence reads generated by 16S rDNA polymerase chain reaction which was followed by Illumina miseq sequencing.

Results from our study support the information previously stated about *Oscheius* EPNs, showing that these EPNs have established symbiotic associations with *Serratia* which proliferate in the insect cadaver and thus considered to be necromenic acquaintances of insects (Ye *et al*, 2010).

2. Materials and methods

2.1 Surface sterilisation of nematodes (Kaya *et al*, 1997)

Fresh infective juveniles were collected from White-traps and transferred into sterile falcon tubes and allowed to sediment. The excess water was removed and the EPNs sediment was

re-suspended in 10ml of 0.1% sodium hypochlorite (NaClO) and left for 1 hour. IJs were transferred into 10ml fresh 0.1% NaClO solution and sterilised for a further 3 hours. Lastly, IJs were rinsed twice with sterile Ringer's solution under a laminar flow.

2.2 Description of samples prepared for metagenomic analysis.

Infected *Galleria Mellonella*

Wax moth larvae were inoculated with surface sterilised infective juveniles and the infection was done on sterile coarse river sand with 8% moisture content to facilitate movement of IJs towards the insect larvae. After 72 hours, the hemocoel of the dead larvae was aseptically extracted and added to a sterile Eppendorf tube with 0.5ml of sterile nutrient broth. NBTA and McConkey agar plates were streaked with the fresh inoculum and incubated at 25℃ for 24 hours. Genomic DNA was isolated from freshly cultured pure bright green colonies and pink solid bacterial colonies using the ZR fungal/bacterial DNA MiniPrep kit (Zymo Research), protocol provided in appendix.

Uninfected *Galleria Mellonella*

Uninfected larvae hemocoel was extracted and grown on NBTA and McConkey agar plates and incubated at 25℃ for 24 hours. Genomic DNA was also extracted from solid colonies obtained using similar protocol as above.

2.3 Metagenomic sequencing using illumina Miseq system

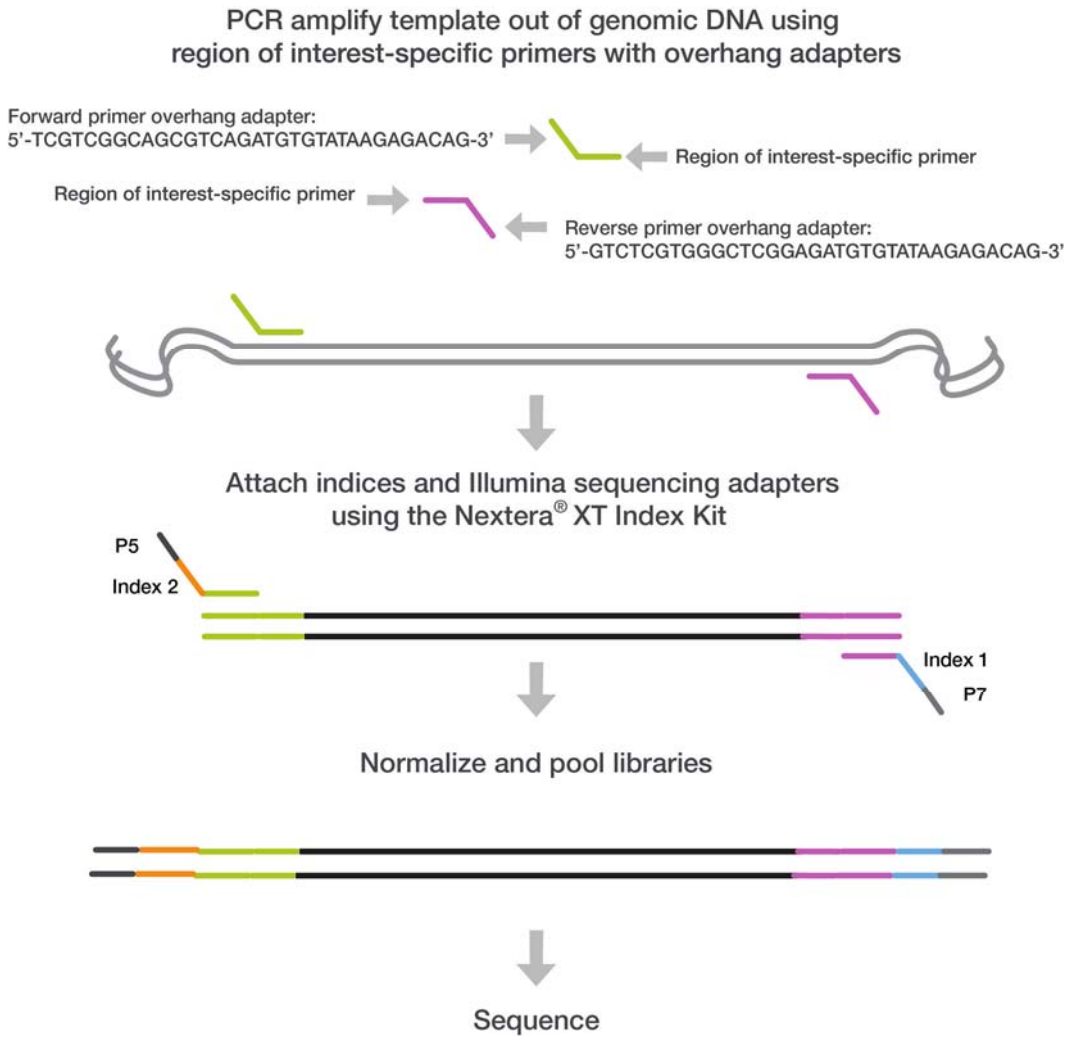


Fig. 2 Illumina workflow for 16S rDNA based metagenomic sequencing.

The part of the 16S rDNA region amplified is about 400bp however, illumina sequencing generates a huge amount of reads is in order obtain deep coverage which makes it possible to find 1 specific bacterial species in 100 or more. The Miseq Reporter Software was used to analyse the metagenomic data and details of the software are found on http://www.illumina.com/systems/miseq/software/miseq-reporter.html

3. Results and discussion

3.1 *Galleria mellonella* infected with *Oscheius* sp. TEL-2014

Sequencing statistics show that a total of 164, 153 reads were generated with only 149, 903 reads passing the quality filtering resulting in 91.3 % of good quality reads. The reads were classified to taxonomic level: kingdom, phylum, class, order, family, genus and species.

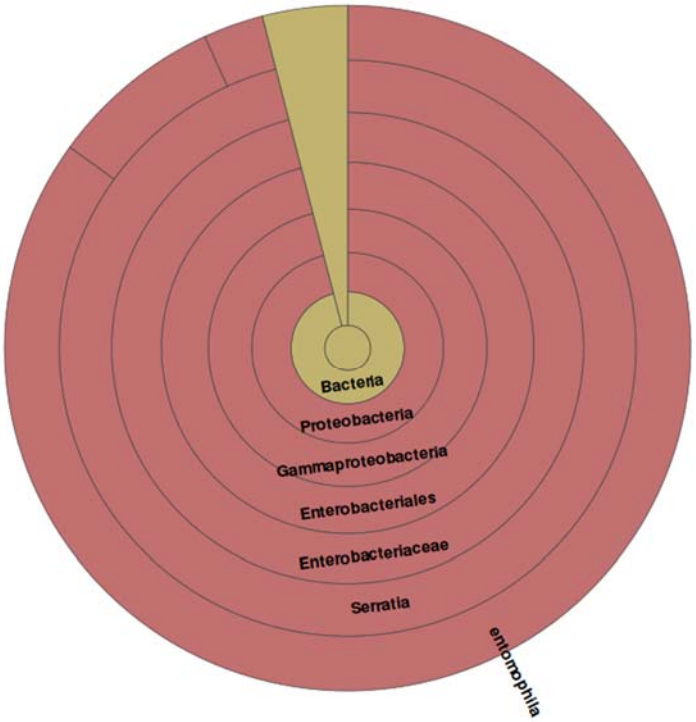


Fig. 3 Sunburst Classification Chart. This sunburst chart shows the relative abundance of the classification results within each taxonomic level.

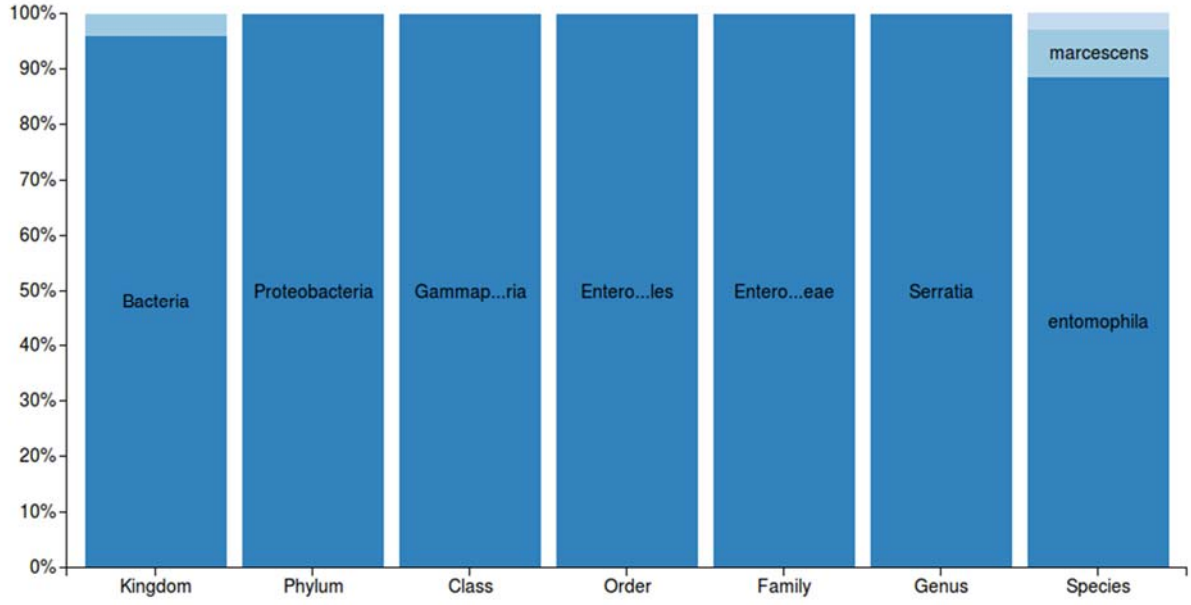


Fig. 4 Top 20 classification results by taxonomic level. This bar chart shows the relative abundance of the top 20 classification results within each taxonomic level.

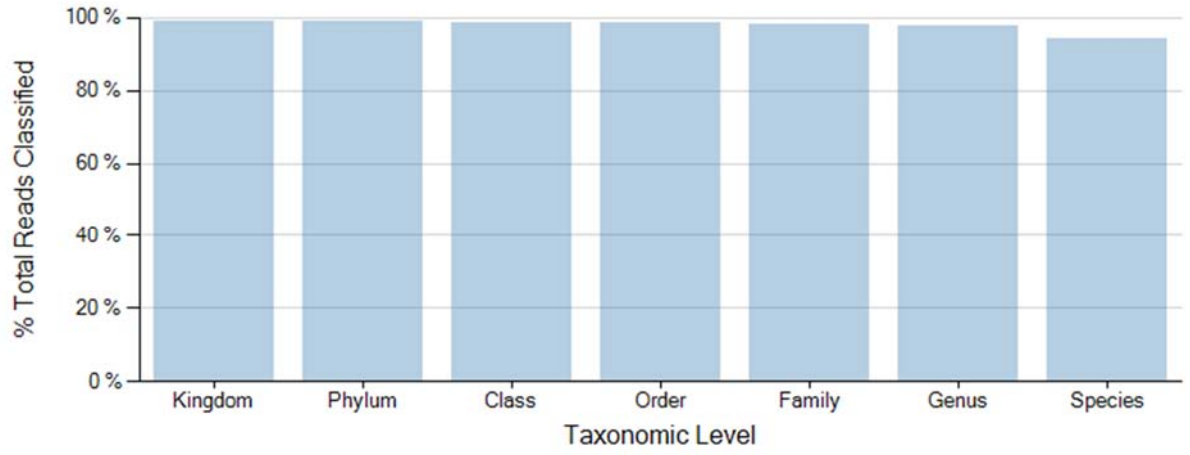
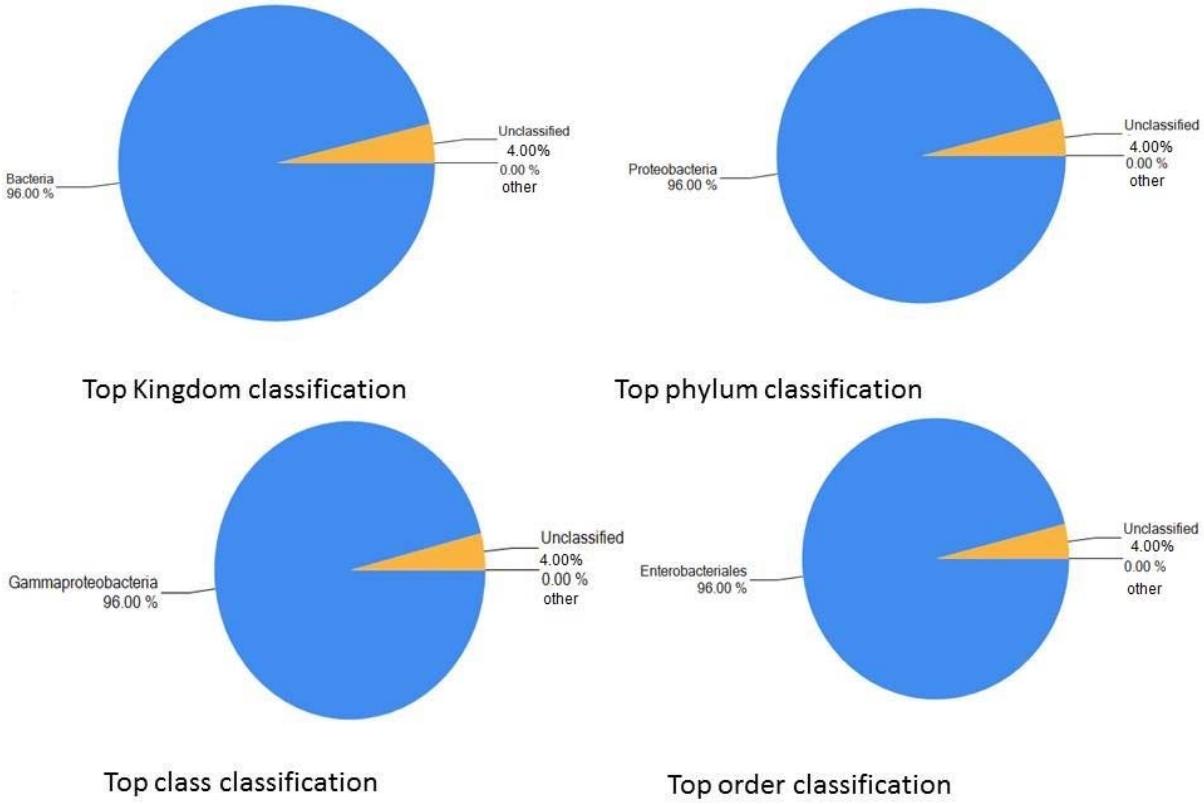


Fig. 5 Classification rate by taxonomic level



A

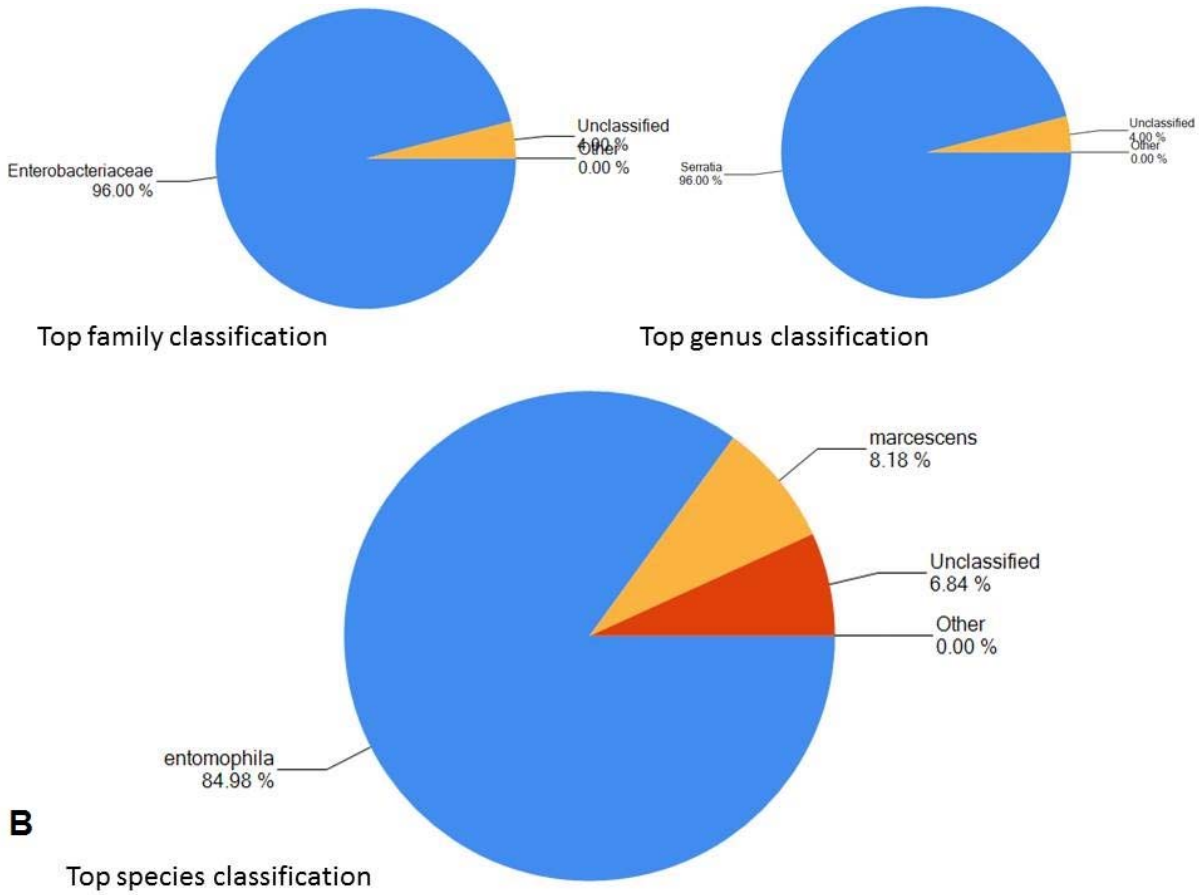


Fig. 6 Pie charts A-B shows all classifications above 3.5% abundance. The classifications are grouped according to taxonomic level i.e. kingdom, phylum, class, order, family, genus and species.

Top kingdom classification results by taxonomic level show that 99.02% of the reads aligned to bacteria, 0.98% aligned to unclassified at kingdom level. Top phylum classification results revealed that 98.65% of the reads are under the Proteobacteria phylum; however 1.09% was unclassified at phylum level. The "other" category in the top phylum pie chart is the sum of all classifications with less than 3.5% abundance. 98.32% of the reads classified the isolated bacterial DNA as Gammaproteobacteria under top class classification results and about 1.23% was unclassified at class level.

Based on the top family classification outputs, 97.75% of the total reads was classified as Enterobacteriaceae and 1.67% was unclassified at family level. The classification by genus level indicates that 96.0% of the reads align to the *Serratia* which significantly maintain findings in chapter 3 whereby *Serratia* sp. TEL was isolated from the gut of surface sterilised

Oscheius sp. TEL-2014 infective juveniles and thus recognised as a pathogenic symbiotic bacteria residing in these EPNs. Interestingly, under species level 84.98% of the total reads were classified as *Serratia entomophila*, 8.18% as *Serratia marcescens* and 5.68% was unclassified at species level. A full summary of the taxonomic levels classification results is included in the supplementary information section.

3.2 *Galleria mellonella* uninfected with *Oscheius* sp. TEL-2014

The sequencing statistics show that a total of 117,596 reads were generated with only 108,236 reads passing the quality filtering resulting in 92.0 % of good quality reads. The reads were classified to taxonomic level: kingdom, phylum, class, order, family, genus and species.

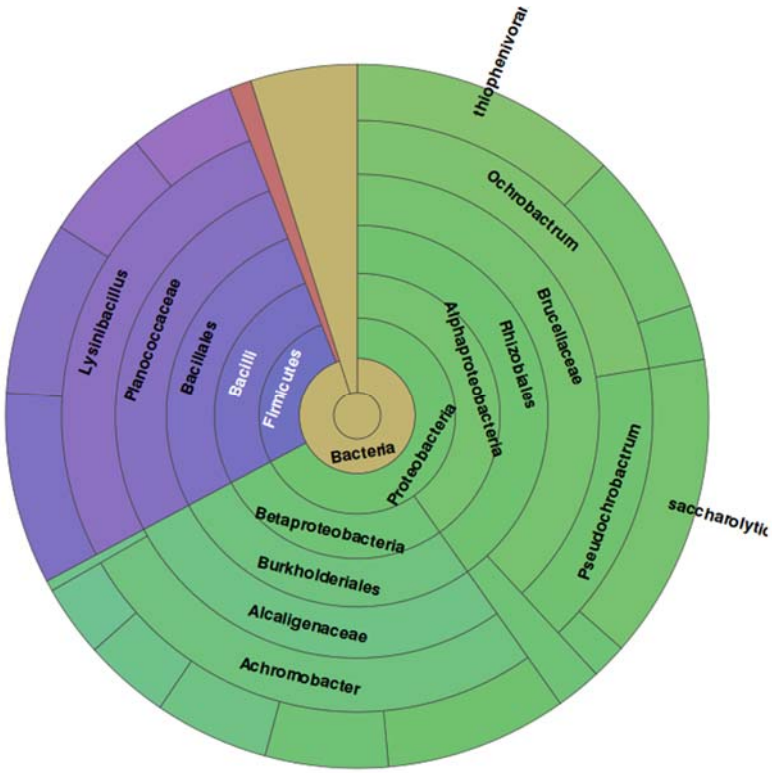


Fig. 7 Sunburst Classification Chart. This sunburst chart shows the relative abundance of the classification results within each taxonomic level.

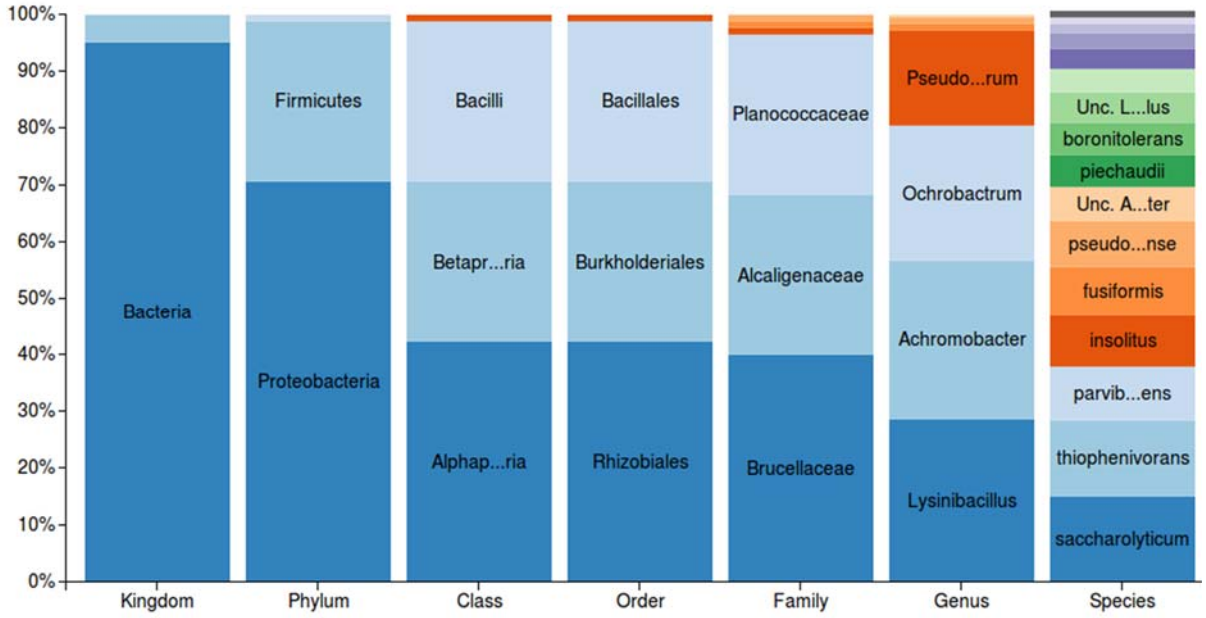


Fig. 8 Top 20 classification results by taxonomic level. This bar chart shows the relative abundance of the top 20 classification results within each taxonomic level.

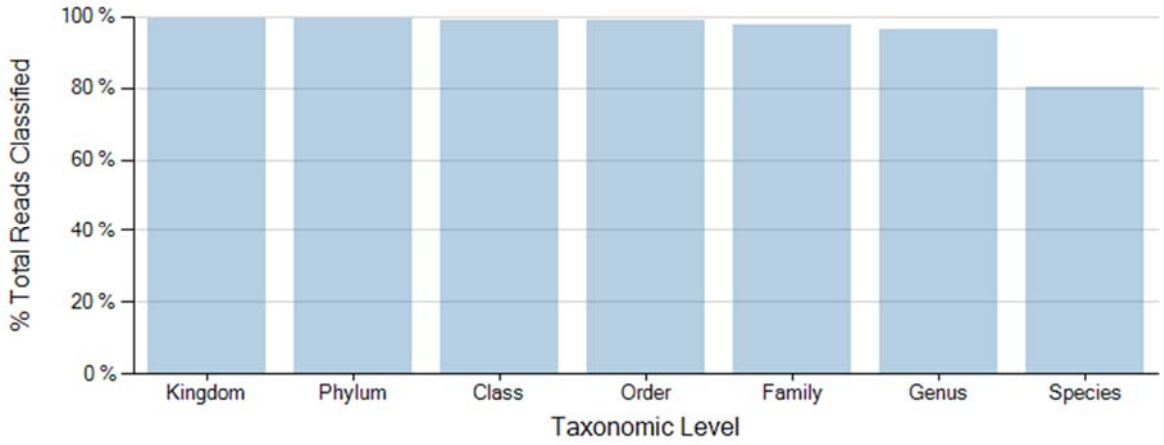
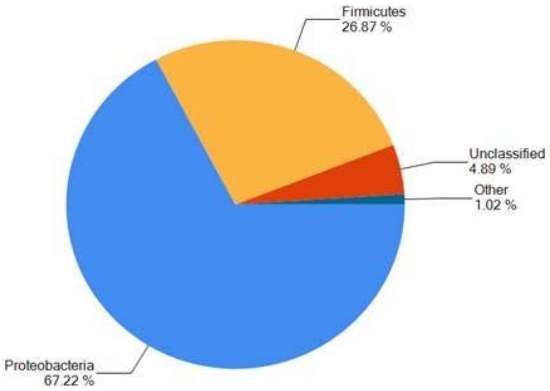
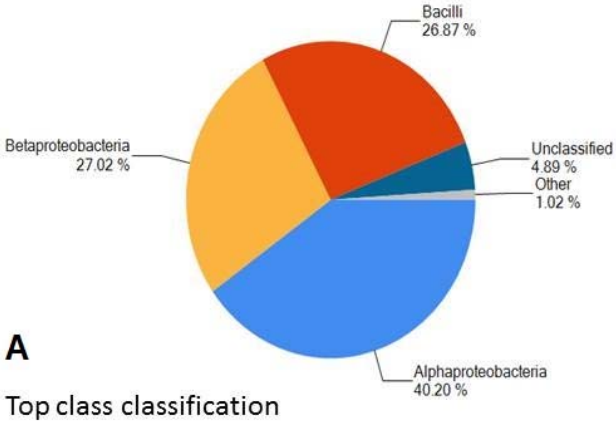


Fig. 9 Classification rate by taxonomic level

Top Kingdom classification

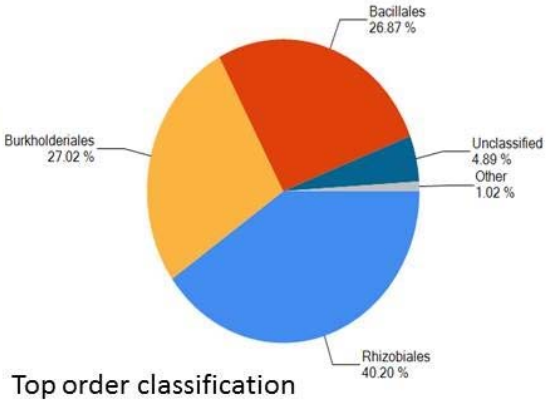


Top phylum classification

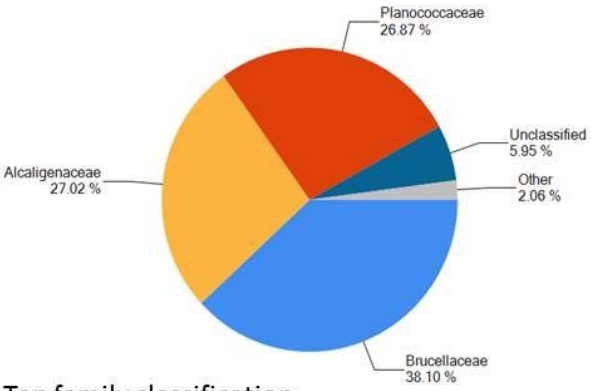


A

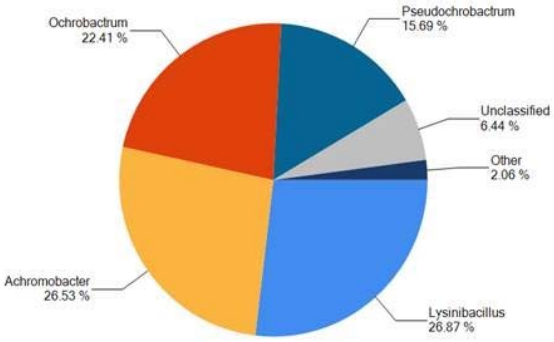
Top class classification



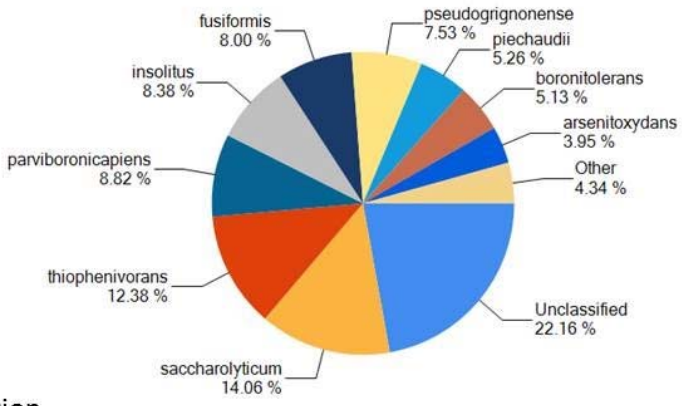
Top order classification



Top family classification



Top genus classification



B

Top species classification

Fig 10. Pie charts A-B show all classifications above 3.5% abundance in the identified taxonomic levels.

At kingdom classification level, 99.64% of the total reads was classified as bacteria and only about 0.35% was unclassified. In the uninfected insect host, 69.37% of the reads was grouped under Proteobacteria, 27.42% Firmicutes and 2.47% was Actinobacteria under top phylum classification level and only 0.57% of the reads was unclassified. Top class classification results indicated that 41.39% was Alphaproteobacteria, 27.46% Betaproteobacteria, 27.38% Bacilli, 2.47% Actinobacteria and 0.82% was unclassified.

Top order classifications have a cocktail of bacteria in which 41.24% was Rhizobiales, 27.42% Burkholderiales, 27.21% Bacillales, 2.46% Actinimycetales, a minor percentage of 0.06% was Enterobacteriales and only 1.06% was unclassified. Brucellaceae 38.79%, Alcaligenaceae 27.05% and 27.03% Planococcaceae were identified under top family classification and top genus classification results showed 27.02% of the total reads categorized as *Lysinibacillus*, 26.54% *Achromobacter* and 22.93% *Ochrobactrum*. Lastly, the species classification showed that 19.52% of the reads was unclassified at species level, 14.06% *Pseudochrobactrum*, 12.38% *Ochrobactrum thiophenivorans*, 8.82% *Lysinibacillus parviboronicapiens*, 8.38% *Achromobacter insolitus*, 8.00% *Lysinibacillus fusiformis*, 7.53% *Ochrobactrum pseudogrignonense* and 5.26% *Achromobacter peichaudii.*

4. Conclusion

Findings support that *Ocheius* nematodes have a symbiotic relation with bacteria belonging to the *Serratia* genus as reported in (Lephoto *et al*, 2015; Petersen *et al*, 2013 and Torres-Barragan *et al*, 2011). 16S rDNA Based Metagenomics sequencing may be considered as a reliable tool for studying the biodiversity and taxonomic levels of hymolyph samples collected from insect larvae infected with *Oscheius sp. TEL-2014*.

5. References

Alia V.I.K, Harma G.A.S, Lan D.A.I.S.H, and Anguly S.U.G, 2014. Biocontrol Potential of *Steinernema thermophilum* and Its Symbiont *Xenorhabdus indica* Against Lepidopteran Pests : Virulence to Egg and Larval Stages. J. Nematol, 46:18–26.

Kaya H.K, Stock S.P, 1997. Techniques in insect Nematology. In: Lacey LA, editor. Techniques in Insect Pathology. London: Academic Press. 822–831.

Lephoto T.E and Gray V.M, 2015. Genome sequencing and annotation of *Serratia* sp. strain TEL. Genomics data, 6:54-56.

Petersen L.M, Tisa L.S, 2013. Friend or foe? A review of the mechanisms that drive *Serratia* towards diverse lifestyles. Canadian journal of microbiology, 59:627-640.

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.

Amir A, Zeisel A, Zuk O, Elgart M, Stern S, Shamir O, Turnbaugh P.J, Soen Y, Shental N, 2013. High-resolution microbial community reconstruction by integrating short reads from multiple 16S rRNA regions. Nucleic Acids Research, 41:e205.

CHAPTER 7

Virulence, disease and defence subsystem, discussing proteins and domains involved in adhesion established from genome annotation of *Serratia* sp. TEL.

Abstract

The virulence, disease and defence subsystem was identified from the annotation of insect pathogenic strain of *Serratia sp. TEL* genome using Rapid Annotation using Subsystem Technology (RAST) hosted by Fellowship for Interpretation. This bacterial strain isolated from *Oscheius sp. TEL-2014, an* entomopathogenic nematode, has a total of 4618 coding regions, 103 RNAs carried and contains 119 genes which may contribute to pathogenicity. Genes encoding the adhesins were also identified. Adhesins are a type of virulence factors and have been shown to be involved in bacterial pathogenicity in several species of bacteria. The ability of a bacterium to attach to specific surfaces and cells is a critical step in its pathogenicity. In this chapter we explore 5 adhesion groups of genes encoding for the mediator the hyperadherence in *Serratia* sp. TEL.

Keywords

Entomopathogenic nematodes, whole genome sequencing, genome annotation and adhesion proteins

1. Introduction

Serratia sp. TEL which belongs to the family of Enterobacteriaceae, is a gram negative, motile and non-spore forming (Lephoto *et al*, 2015). This bacterium was isolated from the gut of surface sterilised infective juveniles of *Oscheius* sp. TEL-2014 which is an entomopathogenic nematode (EPNs). Several strains of *Serratia marcescens* belonging to the *Serratia* genus have been found to be insect pathogens and also pathogens of specific mammals (Abebe *et al*, 2011; Ishii *et al*, 2014). This particular strain of *Serratia marcescens* was observed to produce brown-pigments unlike other pathogenic *Serratia* species which produce red-pigments such as *S. nematodiphila* DSM2140T strain and *S. marcescens* (Ishii *et al*, 2014). Bacteria belonging to the genera *Photorhabdus* and *Xenorhabdus* have already

been confirmed to be insect pathogenic endosymbionts associated with *Heterorhabdistis* and *Steinernema,* respectively. These two species of bacteria have demonstrated pathogenicity to a wide host range of susceptible insects that are crop pests. However, not all bacteria belonging to the genus *Serratia* are pathogenic towards insects (Kwak *et al*, 2015). Several studies have reported that nematode species other than steinernematids and heterorhabditids were capable of acting as vectors of insect pathogenic bacteria belonging to the genus *Serratia* (Alberti *et al*, 1990; Forst *et al*, 1997). Here we report on the genomics of an insect pathogenic strain of *Serratia marcescens*. A new EPNs, *Oscheius* sp. TEL-2014 has been confirmed to act as vector of these bacteria.

Various species of entomopathogenic nematodes (EPNs) express different virulence and infection symptoms against specific insects which have been widely discussed in the field of entomology and nematology (Ansari *et al*, 2008). In this chapter we discuss genes involved in the insect pathogen *Serratia* sp. TEL which include those playing a role in the virulence, defence and disease subsystem, and particularly the adhesion genes encoding for mediator of hyperadherence, as revealed by Rapid Annotation Subsystem Technology) hosted by Fellowship for Interpretation (RAST).

2. Materials and methods

Whole genome shotgun sequencing on *Serratia* sp. TEL was conducted. Genomic DNA paired-end libraries were generated with the Nextera DNA sample preparation kit (Illumina) and indexed using the Nextera index kit (Illumina). Paired-end (2 X 300 bp) sequencing was performed on a MiSeq Illumina using the MiSeq reagent kit version 3 at the Agricultural Research Council Biotechnology Platform (Lephoto *et al*, 2015). The genome annotation was performed using RAST and the mediator of hyperadherence found in the virulence, disease and defence subsystem will be discussed below.

3. Results and discussion

The annotation of *Serratia* sp. TEL on RAST showed that this bacterium has 542 subsystems of which one of them is involved in virulence, disease and defence. This subsystem contains 119 genes which may contribute to pathogenicity of the bacteria towards insects.

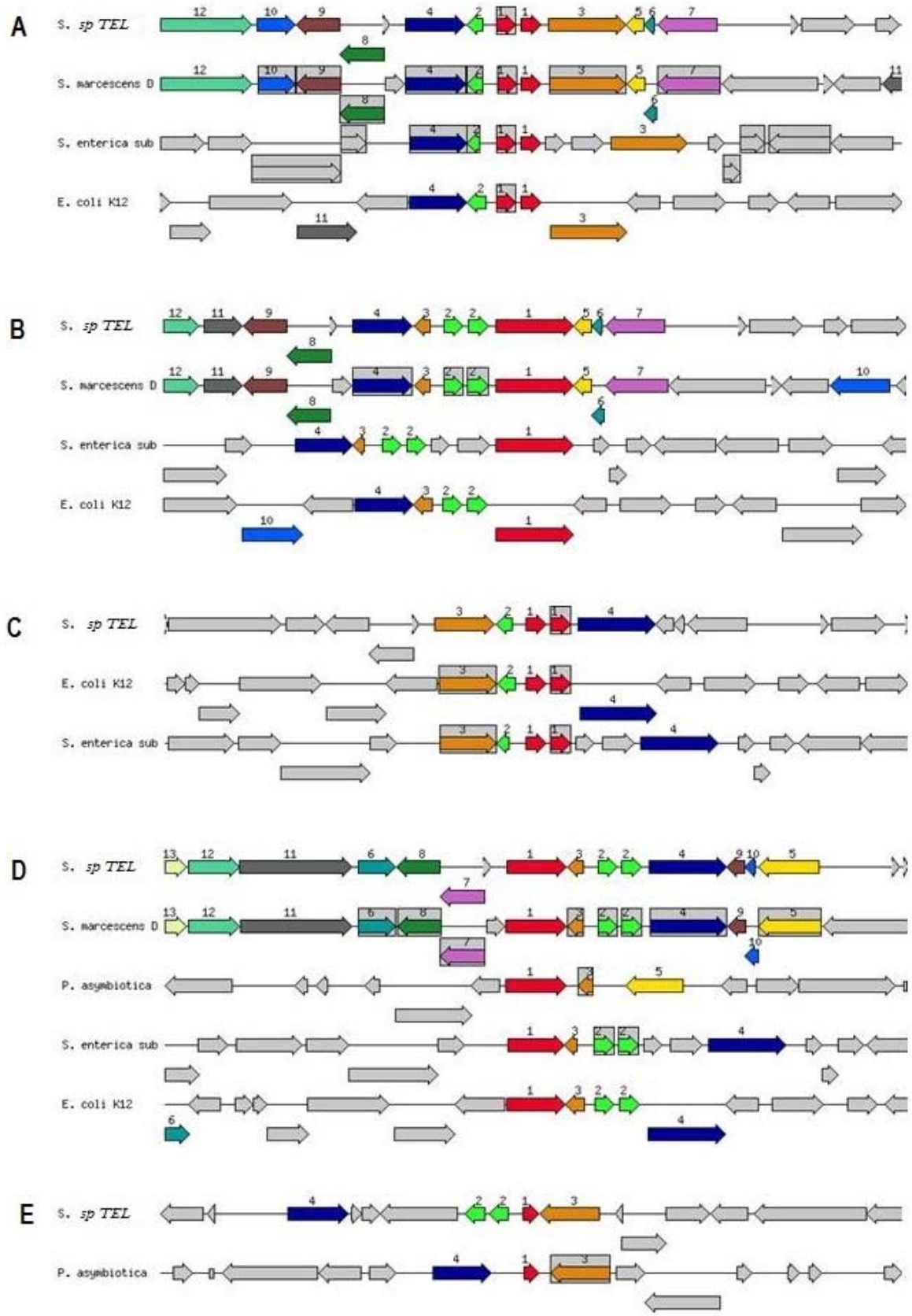

A
S. sp TEL
S. marcescens D
S. enterica sub
E. coli K12
B
S. sp TEL
S. marcescens D
S. enterica sub
E. coli K12
C
S. sp TEL
E. coli K12
S. enterica sub
D
S. sp TEL
S. marcescens D
P. asymbiotica
S. enterica sub
E. coli K12
E
S. sp TEL
P. asymbiotica

Fig. 1 comparative analyses of genomic regions found in various bacteria species. A: 16 kDa heat shock protein B found in *Serratia* sp. TEL to *Serratia marcescens*, *Salmonella enterica subsp. enterica* and *Escherichia coli K12*. B: *Mediator of hyperadherence YidE* found in *Serratia* sp. TEL to *Serratia marcescens*, *Salmonella enterica subsp. enterica* and *Escherichia coli K12*. C: *16 kDa heat shock protein A* found in *Serratia* sp. TEL to *Escherichia coli K12* and *Salmonella enterica.* D: *Uncharacterized protein YidR* found in *Serratia sp. TEL. Serratia marcescens, Photorhabdus asymbiotica, Salmonella enterica subsp. Enterica and Escherichia coli K12.* E: *Mediator of hyperadherence YidQ* in *Serratia* sp. TEL and *Photorhabdus asymbiotica*.

The chromosomal region of the focus gene (top) is compared with four similar organisms. The graphic is centered on the focus gene, which is red and numbered 1. Sets of genes with similar sequence are grouped with the same number and colour. Genes whose relative position is conserved in at least four other species are functionally coupled and share grey background boxes.

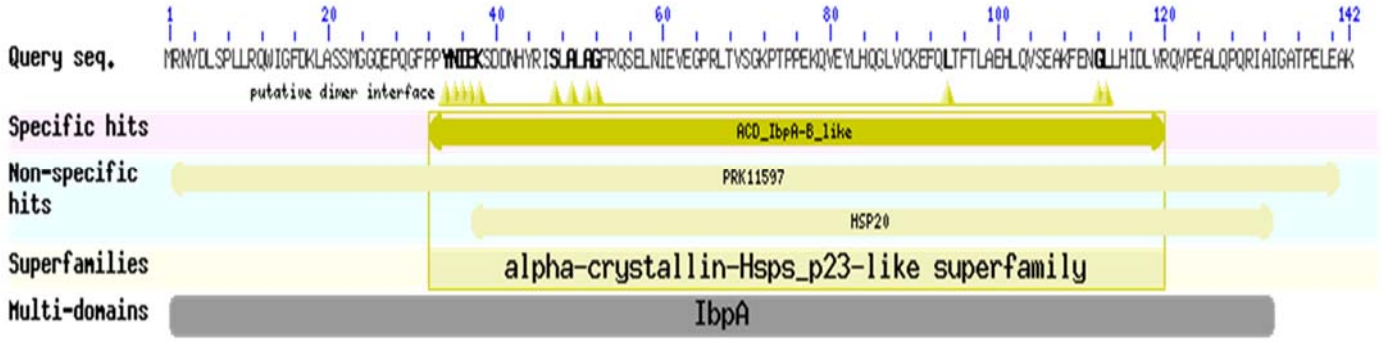


Fig. 5 Conserved domains in *heat shock protein A* and *heat shock protein B* of *Serratia* sp TEL showing specific hits, non-specific hits, super-families and multi-domains

The specific roles and description of each domain was found on the RAST linked NCBI database for conserved domains. One of the specific hits that came up was for alpha-crystallin domain (ACD) found in the *Escherichia coli* inclusion body-associated proteins IbpA and IbpB, and other similar proteins. IbpA and IbpB are 16 kDa small heat shock proteins (sHsps). sHsps are molecular chaperones that suppress protein aggregation and protect against cell stress, and are generally active as large oligomers consisting of multiple subunits. IbpA and IbpB are produced during high-level production of various heterologous proteins, specifically human prorenin, renin and bovine insulin-like growth factor 2 (bIGF-2), and are strongly associated with inclusion bodies containing these heterologous proteins.

IbpA and IbpB which work as an integrated system to stabilize thermally aggregated proteins in a disaggregation competent state. The chaperone activity of IbpB is also significantly elevated as the temperature increases from normal to heat shock. High temperatures results in the disassociation of 2-3-MDa IbpB oligomers into smaller approximately 600-kDa structures. This elevated activity seen under heat shock conditions is retained for an extended period of time after the temperature is returned to normal. IbpA also forms multimers.

Non-specific hits included PRK11597heat shock chaperone IbpB; Provisional, Hsp20/alpha crystallin family; HSP20 family involved in posttranslational modification, protein turnover and with chaperones. The multi-domain IbpA Molecular chaperone IbpA was also identified.

Fig. 3 Alignment of *Serratia* sp TEL compared to other domains in the Conserved domains database on NCBI

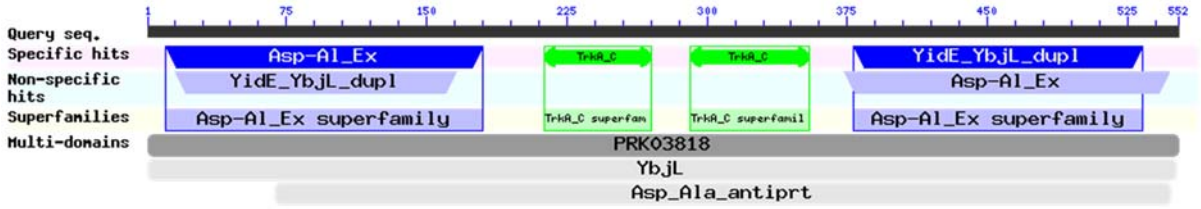


Fig. 4 Conserved domains in *Mediator of hyperadherence YidE* of *Serratia* sp TEL showing specific hits, non-specific hits, super-families and multi-domains.

The specific domain found and shown in fig.4 is aspartate-alanine antiporter AspT/YidE/YbjL antiporter duplication domain. The figure also shows TrkA which is a domain involved in potassium uptake in the cell. Other specific hits include the TrkA-C domain and its exact function is unknown. Some of the multi-domains predicted are PRKO3818 (a putative transporter), YbjL (uncharacterized membrane protein) and aspartate-alanine antiporter. All members of the seed alignment for this model are asparate-alanine anti-transporters (AspT) encoded next to the gene for aspartate 4-decarboxylase (AspD), which converts asparate to alanine, releasing CO2. The exchange of Asp for Ala is electrogenic, so the AspD/AspT system confers a proton-motive force.

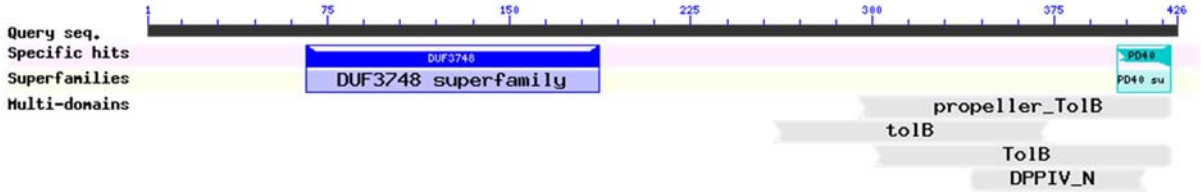


Fig. 6 Conserved domains in *Uncharacterized protein YidR* of *Serratia* sp. TEL showing specific hits, non-specific hits, super-families and multi-domains

The specific alignment hit here is a protein of unknown function (DUF3748) which is approximately 120 amino acids in length. This domain family is originates in bacteria and eukaryotes (Adindla *et al*, 2004). Also shown in fig. 5 is the tol-pal system beta propeller repeat protein TolB. RAST annotation shows that members of this protein family are the periplasmic proteins of Gram-negative bacteria. TolB is part of the Tol-Pal (peptidoglycan-

associated lipoprotein) multiprotein complex, comprising five envelope proteins, TolQ, TolR, TolA, TolB and Pal, which form two complexes. The TolQ, TolR and TolA inner-membrane proteins interact via their transmembrane domains. The beta-propeller domain of the periplasmic protein TolB is responsible for its interaction with Pal. TolB also interacts with the outer-membrane peptidoglycan-associated proteins Lpp and OmpA. TolA undergoes a conformational change in response to changes in the proton-motive force, and interacts with Pal in an energy-dependent manner. The C-terminal periplasmic domain of TolA also interacts with the N-terminal domain of TolB. The Tol-PAL system is required for bacterial outer membrane integrity. E. coli TolB is involved in the tonB-independent uptake of group A colicins (colicins A, E1, E2, E3 and K), and is necessary for the colicins to reach their respective targets after initial binding to the bacteria. It is also involved in uptake of filamentous DNA. Study of its structure suggests that the TolB protein might be involved in the recycling of peptidoglycan or in its covalent linking with lipoproteins. The Tol-Pal system is also implicated in pathogenesis of *E. coli, Haemophilus ducreyi, Salmonella enterica* and *Vibrio cholerae*, but the mechanism(s) is blurred.

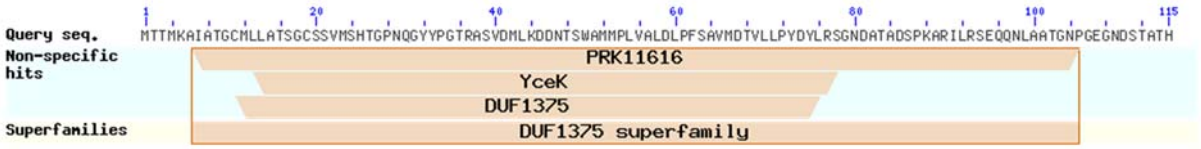


Fig. 7 Conserved domains in *Mediator of hyperadherence YidQ* of *Serratia* sp. TEL showing specific hits, non-specific hits, super-families and multi-domains.

A hypothetical protein, uncharacterized conserved protein with an unknown function was found. Protein of unknown function (DUF1375) is present in *Mediator of hyperadherence YidQ* and belongs to a family consisting of several hypothetical, putative lipoproteins of around 80 residues in length. Members of this family seem to be specific to the Class Gammaproteobacteria. The function of this family is unknown.

The annotation further revealed that the virulence, disease and defense subsystem has 12 genes that encode for bacteriocins, ribosomally synthesized antibacterial peptides and within this group 4 genes are involved in tolerance to colicin E2 and 8 genes encode for colicin V and bacteriocin production cluster. 88 genes involved in resistance to antibiotics and toxic compounds were also identified. The mdtABCD multidrug resistance cluster has 6 genes for which 2 encode for lysozyme inhibitors, 1 gene for multiple antibiotic resistance MAR locus,

8 genes encoding for copper homeostasis and 10 cobal-zinc-cadium resistance genes. 8 genes coding for multidrug resistance, tripartite systems found in gram-negative bacteria, 1 streptotretothricin resistance genes, 4 resistance to fluoroquinolones and 6 arcenic resistance genes. Copper homeostasis: copper tolerance is coded for by 9 genes and beta-lactamase by coded for by 5 genes. 2 genes code for resistance to chromium compounds and 26 genes for multidrug resistance efflux pumps. Moreover, 14 genes are involved in invasion and intracellular resistance, with 5 encoding for mycobacterium virulence operon involved in protein sysnthesis (SSU ribosomal proteins), 2 mycobacterium virulence operon involved in DNA transcription, 4 genes encoding for mycobacterium virulence operon possibly involved in quinolinate biosynthesis and 3 mycobacterium virulence operon involved in protein synthesis (LSU ribosomal protein).

More investigations need to be conducted to understand the role and contribution of these domains in *Serratia* sp TEL pathogenicity, also exploring their functions.

4. Conclusion

Knowing the presence of proteins and domains involved in virulence, disease and defence is crucial for further understanding the infection mechanism of *S.* sp. TEL and its importance or contributions it may have in pest control. The existence of these adhesion proteins as revealed by the RAST genome annotation may give clues about the pathogenicity of this bacterial strain.

5. References

Abebe E, Akele F.A, Morrison J, Cooper V, Thomas W.K, 2011. An insect pathogenic symbiosis between a *Caenorhabditis* and *Serratia*. Virulence, 2(2):158–61.

Alberti L, Harshey R.M, 1990. Differentiation of *Serratia-Marcescens* 274 into Swimmer and Swarmer Cells. J Bacteriol, 172(8):4322–8.

Ansari M.A, Adhikari B.N, Ali F and Moens M, 2008. Susceptibility of *Hoplia philanthus* (Coleoptera: Scarabaeidae) larvae and pupae to entomopathogenic nematodes (Rhabditida: *Steinernematidae, Heterorhabditidae*). Biological Control, 47: 315–321.

Adindla, 2004. Comparative and Functional Genomics 5:2-16

Forst S, Dowds B, Boemare N, Stackebrandt E, 1997. Xenorhabdus and Photorhabdus spp.: Bugs that kill bugs. Annu Rev Microbiol, 51:47–72.

Ishii K, Adachi T, Hara T, Hamamoto T and Sekimizu K, 2014. Identification of a *Serratia* marcescens virulence factor that promotes hemolymph bleeding in the silkworm, Bombyx mori. Journal of Invertebrate Pathology, 117:61–67.

Kwak Y, Khan A R and Shin J, 2015. Genome sequence of *Serratia nematodiphila* DSM 21420T, a symbiotic bacterium from entomopathogenic nematode. Journal of Biotechnology, 193:1–2.

Lephoto T.E, Featherston J, Gray V.M, 2015. Draft whole-genome sequence of *Serratia* sp. strain TEL, associated with *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae) isolated from a grassland in South Africa. Genome Announc, 3(4):e00747-15.

Lephoto T.E and Gray V.M, 2015. Genome sequencing and annotation of *Serratia* sp. strain TEL. Genomics data, 6:54-56.

CHAPTER 8

Whole genome sequencing, assembly and partial annotation of a novel Entomopathogenic nematode *Oscheius* sp. TEL-2014

Abstract

The fact that entomopathogenic nematodes (EPNs) have demonstrated potential as biocontrol agents of insect pests has generated strong incentives to isolate and characterise novel EPNs species. The hazardous environment impacts and the unsustainability of synthetic chemical insecticides due to the evolution of insect pest resistance provides an additional incentive for developing biocontrol agents based on EPNs for insect pest control. These incentives have also played a role in motivating the development of a new wave of genomic research into the molecular genetic basis of the bacterial-EPN-insect tripartite relationship. This research has become increasingly feasible as a result of the advances that have been made in the next-generation sequencing whole genome shotgun sequencing technologies. Whole genome sequencing, sequence assembly and genome annotation have become the necessary steps for increasing our understanding of the molecular genetics underlying the bacterial-EPN-insect tripartite relationships. This chapter also explores genome assembly and annotation of *Oscheius* sp. TEL-2014 EPNs partially wild type population (line 7) and the inbred line 13. Whole genome sequencing was performed on Illumina Hiseq sequencing system and paired end library preparation was used. The quality of the reads was checked using fastq software and trimming of poor quality reads and removal of Nextera adaptors was done using trimmomatic. The genomes were then assembled on Velvet and 75965 contigs (line 7) and 53190 contigs (line 13) were generated. Gene prediction tools revealed the presence of proteins involved in gene expression and DNA replication in nematodes. The draft genome of *Oscheius* nematodes will serve as a foundation for future studies aimed at understanding molecular and metabolic processes in this genus.

Keywords: Entomopathogenic nematode; *Oscheius* species, whole-genome sequencing, genome assembly, genome annotation

1. Introduction

1.1 Entomopathogenic nematodes

Nematodes are the most profuse animals on earth and entomopathogenic nematodes have received a lot of attention in science research because of their ability to be used as biological control agents of insect pests (Zhang *et al*., 2008). The potential of EPNs to be used as effective biological control agents has been investigated globally as more researchers aim to eliminate the use of synthetic chemical pesticides to manage harmful insect pests targeting and damaging various crops (Stuart *et al*, 2006).

Entomopathogenic nematodes are insect killing microscopic worms with the ability to invade insect host, elicit infections and cause death within 24-48 hours depending on the species. (Hazir *et al*, 2003). They carry virulent and pathogenic bacteria which use them as vectors to reach the insect's hemocoel, secrete antimicrobials and immune depressors with the goal to ultimately lead to death of the host (Malan *et al*., 2006). Great steps have been taken to isolate, identify and characterise more EPNs and test them for their entomopathogenecity. In this study we shift gears to genomics and use next-generation sequencing and bioinformatics tools to explore the genome of novel EPNs *Oscheius sp. TEL-2014*.

1.2 What is inbreeding?

Inbreeding is linked with inclined homozygosity because relatives mate with each other. A wild type population is therefore expected to have more heterozygosity as a result of variable organisms mating, which are not closely related. Inbreeding may be defined differently depending on the size of the population and also based on the nature of the reference population usually used in inbreeding and genetic studies to calculate inbreeding resulting in various biological consequences.

The study therefore also attempts to find any differences or similarities in terms of genome assembly and repetitive sequences in wild type population and inbred line genomes.

1.3 Nematode genome

About 1200 whole genomes of organisms were absolutely sequenced by the year 2010, approximately 1,000 bacteria, 80 archaea, and 124 eukaryotes. Whole genome sizes are different from organism to organism. Bacteria genome size ranges from one to six million base pairs and animals such as nematodes have genome sizes which may be greater than 100 million base pairs. This also implies that the number of genes will vary.

The availability of nematode information resources such as Wormbase which contain *Caenorhabditis elegans* complete genomes also containing homologous genes recognized among various species of nematodes has made it possible for more researchers to work on gene annotations and comparative genomic studies (Zhou *et al*, 2015). Genomes of nematodes belonging to the *Caenorhabditis* have been fully sequenced, assembled and annotated. *C. elegans* are an example of nematodes which their genome has been intensively studied and its now about 41 years since Sydney Brenner reported forward genetic screens of these worms (Hu, 2014). *C. elegans* are free-living microscopic worms with a genome size of 100.3Mb. This genome was identified to have over 20 000 protein coding regions and over 3000 RNA genes using specific bioinformatics tools and functional genomics investigation tools. Interestingly, the genome was found to have small number of introns, and there was one gene per 4.9kb contributing to the genome compactness compared to other animals or eukaryotic genomes which have also been fully sequenced. The central core of the autosomes is a compartment for conserved genes and housekeeping genes, while duplicated regions, new genes and repeats were observed to reside on the arms of the autosomes. It was further highlighted that the X (sex chromosome) had a uniform distribution of repeats and genes (Whitton et l, 2004).

Caenorhabditis briggsae, a nematode with a symbiotic relationship with bacteria pathogenic to insects, *Serratia marcescens* (Abebe *et al*, 2011) was also sequenced and its genome contains over 4000 chromosomal aberrations or reorganizations like inversions, translocations and transpositions mostly within the chromosome. These findings were supported by phylogenetic comparison of this group of worms to *C. elegans* and it appeared that *Caenorhabditis briggsae* last had a mutual ancestor 100 million years ago and this accounts for its increased rate of chromosomal evolution. Nematodes belonging to the recently identified *Oscheius* genus also have symbiotic relationships with insect-killing

Serratia bacteria particularly carried in the free-living stage of these worms (Abebe *et al*, 2010). The *C. elegans* genome ws superlative as a reference for gene prediction in *Oscheius* sp. TEL-2014 mainly because it has provided so much knowledge about genetics, developmental biology and cell biology (Wilson, 1999). This was all made possible by understanding the whole genome sequence composition, structure and location of coding and non-coding regions, protein coding regions, the number of genes and their functions.

1.4 Why whole genome sequencing?

Whole genome sequencing and annotation of genomes has become a powerful approach in understanding on several aspects concerning nematodes development, reproduction, and infectivity, genetic and metabolic processes in these microscopic insect-killing worms. This method has also assisted many researchers to find repeats, predict genes, identify open reading frames and perform functional studies using new sequencing and bioinformatics technology. For example, the genome of *Caenorhabditis elegans* has been fully sequenced and assembled, with available information on genome annotation on the Wormbase database. This then allows us to sequence novel EPNs belonging to a recent genus, *Oscheius* and employ genomics and bioinformatics tools to identify genes existing in *Oscheius sp. TEL-2014*.

Sequencing technology has improved over the past decade allowing whole genome sequencing projects to be swiftly completed. These sequencing tools are able to generate millions or even billions of short reads within a short period of time. In some cases, de novo assembly protocols are employed to generate contigs. The quality of the assembly depends on several factors which include the initial quality of the reads, the type of assembler selected to assemble a particular genome (Salzberg, 2012).

In whole genome sequencing related studies, data cleaning is the most imperative step that needs to be done before assembling the genome. Errors that occur during the sequencing process affect how the data is assembled. Quality control software readily available online such as Quake, fastqc and fast-mcf may be used for improving the quality of data prior further analysis. Cleaning of data may improve the N50 contig value and thus contributing to achieving high quality assembly which one may anticipate to use with ease for downstream

analysis. Some assemblers contain error corrections steps which add an extra advantage to achieving good data. N50 value is the size of the minimum contig length "such that 50% of the genome is contained in contigs of size N50 or larger" as defined by Salzberg, 2012.

Once data cleaning has been successfully achieved, the genome is assembled to generate contigs which will be annotated to identify present genes. This is accomplished by using tools to align the de novo genome to reference genomes exiting in specific nematodes genes databases.

1.5 Genome features

Most eukaryotic genomes have been found to contain a huge amount of non-coding DNA which has been previously referred to junk DNA however, a small proportion of the genome encodes for proteins and functional RNA.

The genome of an organism is made up of various sequence features. There are transcribed regions such as mitochondrial RNA, transfer RNA and ribosomal RNA. There are structural regions such as introns, exons, open reading frames (ORF) and untranslated regions (5' UTR and 3'UTR). There are structures in genomes which may be similar across species of certain types of organisms such as orthologs. The genome may also contain transposable elements and repetitive elements.

During the process of genome annotation studies, they have looked at the transcription features of a genome. These features include promoter elements which are critical in gene expression. There mRNA features such as introns and exons and translation features such as the start codon which is the translation start site and UTRs.

1.6 Repeats

Genomes of different organisms have repetitive sequences which occur in diverse types and have variable functions and found of different positions in the genome. Repetitive DNA is present in numerous duplicates in a genome. There are tandem repeats which have been defined as an arrangement of two or more nucleotides which have been recurring where the repetitions are directly adjacent to each other. Repeat sequences may be masked or basically

marked in a genome as opposed to permanently removing them and eliminating them from downstream analysis. Programs have been developed to find tandem repeats and examples include Tandem and Tandem Repeats Finder. Other bioinformatics programs such as Repeat masker and Repeatscout are able to search for different types of repeats in a genome.

Repeats having different features can be divided or classified into groups. There are transposable elements (TEs) which are scattered in a genome, transposons and retrotransposons. Other types of repetitive DNA include simple DNA sequences which are basically composed of numerous copies of tandemly repeated short sequences which are also referred to as short tandem repeat (STR). These STRs may be found in variable locations within the genome and may be different amongst organisms.

Moderately repetitive DNA sequences such as short interspersed elements (SINEs) and long interspersed elements (LINEs) are grouped under the DNA category of sequences with unknown function. Some genomes also contain highly repetitive DNA sequences which include minisatellites, microsatellites and telomeres.

1.7 Gene prediction

Gene prediction methods are used to predict genes, open reading frames (which lack stop codons), and exons, coding and non-coding sequences. An example of gene prediction tools includes AUGUSTUS which was used in this project to predict genes and the output was generated in a ggf format. This program is used for ab initio gene prediction and strives to identify exons, ORFs, coding sequences, start and stop codon sites, splice site consensus sequences.

The genomics approach will allow us to develop some degree of understanding the molecular foundation of EPNs isolated in this study and predict genes using gene prediction tools. The aim of this study is therefore to identify genes and gene families found in the newly isolated EPNs *Oscheius* sp. TEL-2014 wild type population and inbred line obtained through inbreeding.

2. Materials and methods

2.1 Nematodes inbreeding

Nematodes were collected from White-traps petri dished and surface sterilised for 3 hours with 0.1% sodium hypochlorite in sterile 50ml Falcon tubes. They were allowed to sediment and the sterilisation solution was removed. This was followed by rinsing of the nematodes 3 times with sterile Ringer's solution. Pure cultures of symbiotic bacteria *Serratia sp. TEL* grown on NBTA media were used to prepare bacterial lawns to aid EPNs growth and reproduction. Lipid agar plates were inoculated with the 1.0 ml Nutrient Broth culture and incubated at 25°C for 24hours until a bacterial lawn was established. The solution was removed and the nematodes were transferred to sterile lipid agar plates with bacterial lawns and grown at 25°C for 72 days with the plates facing upwards. The plates were viewed under a compound microscope under a laminar flow to monitor EPNs growth.

A wild population of EPNs was isolated from its natural environment with the assumption that the population collected has high heterozygosity. A factor to consider is that when the EPNs were recovered from the soil using the bating technique, mating patterns may have been affected especially as nematodes emerged from the dead insect larvae into the water placed in the White trap. Nematodes were then surface sterilised and transferred to a lawn of *Serratia* bacteria grown on lipid agar plate and allowed to mate over a period of 7 days. This step may have gradually affected the genetic composition of the nematodes. The first inbred line was obtained from taking one pregnant female with evidence of stage 2 nematodes in its body and transferring it into a new lipid agar plate containing a *Serratia* bacteria lawn to support growth of the *Oscheius* nematodes. From each successful line, the same procedure was done obtain the next inbred lines. The highest inbred line was line 13 and on this stage nematodes were unable to continue growing.

1 adult female with IJs inside its body was quickly transferred to a sterile lipid agar plate with the bacterial lawn before endotokia matricida began. The plate was incubated at 25°C for 7 days and the same process was repeated until the 13th inbred line was obtained. After this line, IJs stopped growing and the inbreeding procedure was therefore halted. Wild type population and line 13 were used for further analysis in the study.

2.2 Genomic DNA extraction from nematodes

Entomopathogenic nematodes were collected from freshly prepared White traps and surface sterilised with 0.1% sodium hypochlorite for 3 hours in sterile 1.5 ml Eppendorf tubes.
The nematodes were rinsed three times with sterile distilled water under sterile conditions in a laminar flow hood. Whole genomic DNA was extracted from the sterile nematodes using a protocol adopted from Puregene® DNA Purification Kit, Gentra systems 2003. Nematodes were centrifuged at 13000 rpm for 3 minutes and placed on ice for at least 30 seconds and the supernatant was carefully discarded. 600µl cell lysis solution and 3µl of proteinase K solution were added and the tubes were inverted gently 50 times. Samples were incubated at 55˚C for 3 hours to allow cell lysis. 3µl of RNase A solution was added into the cell lysate, and inverted 25 times then incubated at 37˚C for 30 minutes. Tubes were cooled to room temperature (approximately 25˚C) and 200µl of protein precipitation solution (Recipe available in Puregene® DNA Purification Kit, Gentra systems 2003) was added to the RNase A treated cell lysate and mixed with a vortex at high speed for 30 seconds. The tubes were centrifuged at 13000 rpm for 3 minutes and a tight protein pellet was formed. The supernatant containing the DNA was poured into a sterile 1.5ml centrifuge tube containing 600µl 100% isopropanol and then inverted gently 50 times. Tubes were centrifuged at 13000 rpm for 1 minute; the DNA was visible as a white pellet. The supernatant was poured off and drained the tube on clean absorbent paper. 600µl of 70% cold ethanol was added into the tubes and inverted to wash the pellet gently. Tubes were centrifuged at 13000 rpm for 1 minute and the ethanol was removed carefully. Eppendorf tubes were inverted and drained on an absorbent paper again and allowed to air dry for 10-15 minutes. 100µl of DNA hydration solution (Recipe available in Puregene® DNA Purification Kit, Gentra systems 2003) was added and the DNA was rehydrated by incubating the sample 1 hour at 65˚C and then stored at 4˚C.

2.3 Gel electrophoresis

0.5% agarose gel was prepared in order to confirm the quality and integrity of the extracted DNA. 0.25g of agarose power was dissolved in 50ml of 1XTBE buffer and 1µ of ethidium bromide was added and mixed gently. The gel was left to solidify at room temperature (25˚C) with the well comb inserted. For each 20µl sample, 5µl of loading dye was added and then

samples were run on the gel for 40 minutes at 90 volts immersed in 1XTBE, with constant current.

2.4 Polymerase chain reaction amplification of nematode ITS rDNA region

A polymerase chain reaction was employed to amplify the 18S rDNA region using TW81 Forward Primer 5'-GCGGATCCGTTTCCGTAGGTGAACCTGC -3', Tm (°C) = 71.94 and AB28 Reverse Primer 5'-GCGGATCCATATGCTTAAGTTCAGCGGGT -3', Tm (°C) = 68.87. 1% agarose gel was prepared in order to confirm the presence of the desired amplified region. 0.5g of agarose power was dissolved in 50ml of 1XTBE buffer and 1µ of ethidium bromide was added and mixed gently. The gel was left to solidify with the well comb inserted. For each 10µl sample, 2µl of 6X loading dry was added and then samples were run on the gel for 30minutes at 90 volts immersed in 1XTBE, with constant current. The same primers were used for the sequencing of this gene. The sequence obtained was subjected to NCBI BLAST under the default settings to identify similar sequences. The analysis revealed that among all the matching for the 18S rDNA gene sequences, our sequence had a high similarity (84%) to a novel *Oscheius* species which identified our isolate as a member of the genus *Oscheius.*

2.5 Next generation sequencing

Illumina Nextera protocol was used for paired-end library preparation and sequencing. (Illumina proprietary, catalog # FC-132-9001DOC, Part # 15035209 Rev. C, January 2013)

Genomic DNA paired-end libraries were generated with the Nextera DNA sample preparation kit (Illumina) and indexed using the Nextera index kit (Illumina). Paired-end (2 × 125 bp) sequencing was performed on a HiSeq 2500 using the Illumina SBS v4 chemistry at the Agricultural Research Council (ARC) Biotechnology Platform, Pretoria, South Africa.

Mate-pair and paired ends library preparation

Sample preparation was prepared over 2 days as shown in fig. 1. The tagmentation reaction (DNA sample is simultaneously fragmented and tagged with a biotinylated mate pair junction adapter) followed by strand displacement reaction (uses a polymerase to fill this gap and ensure that all fragments are flush and ready for circularization) and circularisation steps were done on day 1.

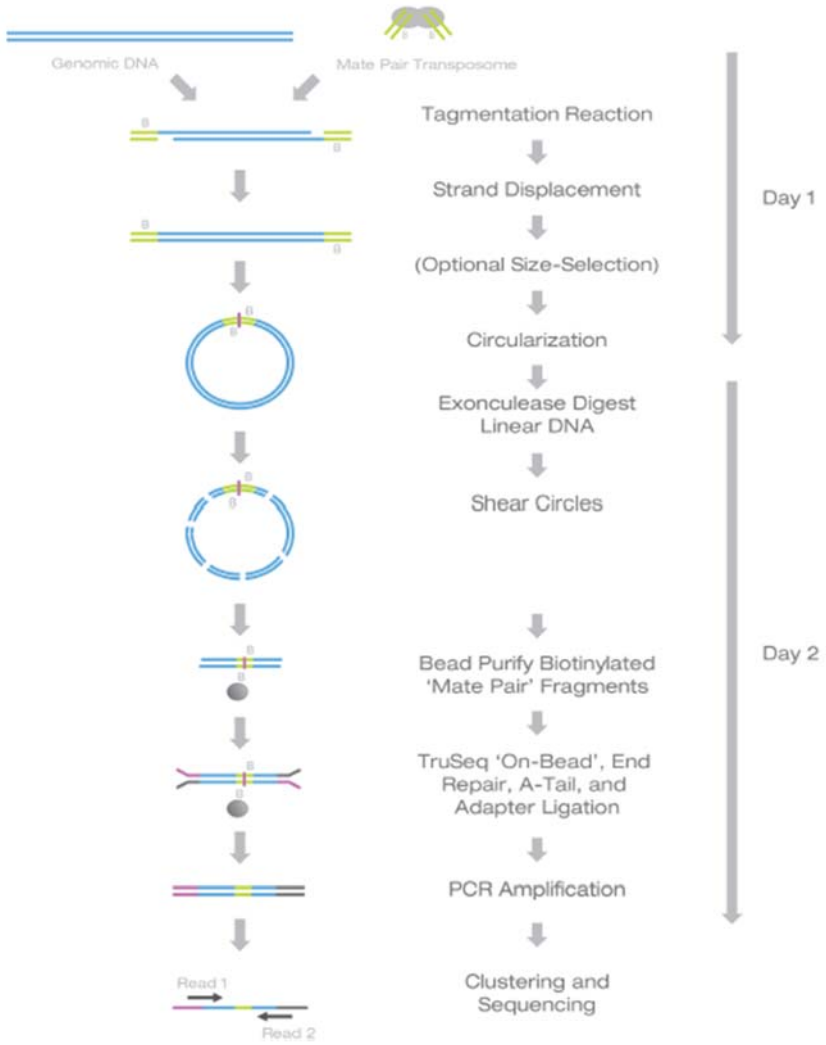


Fig. 1 Nextera Mate Pair sample preparation procedure

On day 2 exonuclease digest was done (any linear molecules still remaining in the circularization reaction are removed by DNA exonuclease treatment, this treatment leaves the desired circular molecules intact), followed by shearing the circularised DNA (approximately 300–1000 bp). Bead purified biotinylated 'mate-pair' fragments were generated. During this step the sheared DNA fragments that contain the biotinylated junction adapter (mate pair fragments) are purified by means of binding to streptavidin magnetic beads, and the unwanted, unbiotinylated molecules are removed through a series of washes. The end repair step which converts the overhangs resulting from the DNA shearing step into blunt ends using an End Repair Mixadaptors was done followed by A-tailing (a single 'A' nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction).

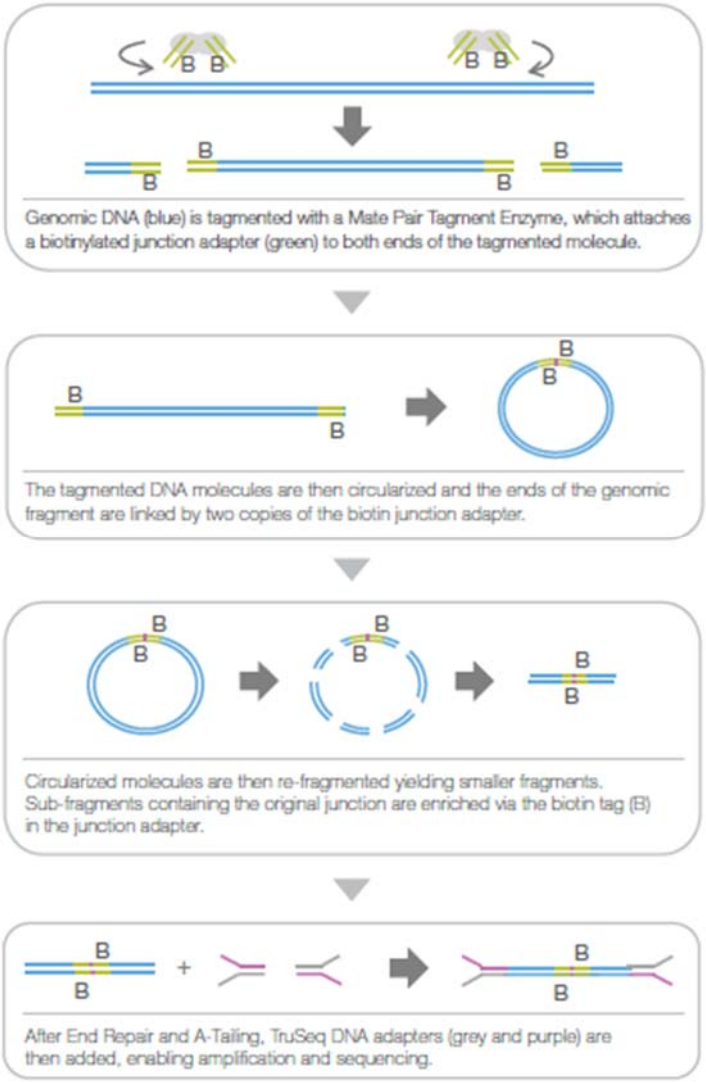


Fig. 2 Nextera Mate Pair workflow adopted from Illumina sequencing technology.

The adaptor ligation step ligates TruSeq indexing adapters to the ends of the DNA fragments, preparing them for PCR amplification and subsequent hybridization onto a flow cell. The adapter ligation reaction was carried out on-bead, and the DNA remains bound to the beads throughout this reaction and subsequent bead wash steps. DNA polymerase chain reaction used to enrich for the mate pair fragments that have TruSeq DNA adapters on both ends. The template material is bound to the streptavidin beads; however the resulting PCR amplified copies are not biotinylated and are not bound to the beads. Lastly the DNA fragments were clustered and sequenced as shown schematically in fig. 2.

Paired-end sample preparation

Libraries of genomic DNA for paired-end sequencing on the Illumina sequencing platform were prepared using the Illumina sample preparation kit (ILLUMINA PROPRIETARY Catalogue # PE-930-1001 Part # 1005063 Rev. E February 2011). The protocol was used to add adapter sequences onto the ends of DNA fragments in order to generate sequencing library format shown in fig. 3.

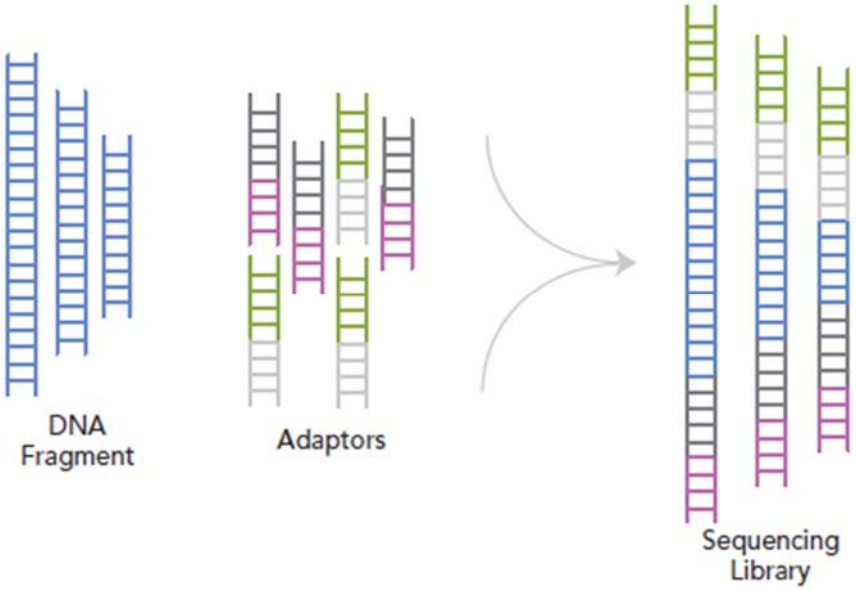


Fig3. Sequencing Library after Paired-End Sample Preparation

Libraries of DNA fragments were prepared by using 50 ng of starting DNA which was put into a tagmentation reaction. Purified genomic DNA extracted from entomopathogenic nematode *Oscheius* sp. TEL-2014 and fragmented by hydrodynamic shearing which produced phosphorylated blunt-ended DNA fragments <800bp as summarised in fig. 4. In preparation for ligation to an adaptor containing a single-base 'T' overhang, a single 'A' nucleotide was added to the 3' ends of the blunt-ended DNA fragments.

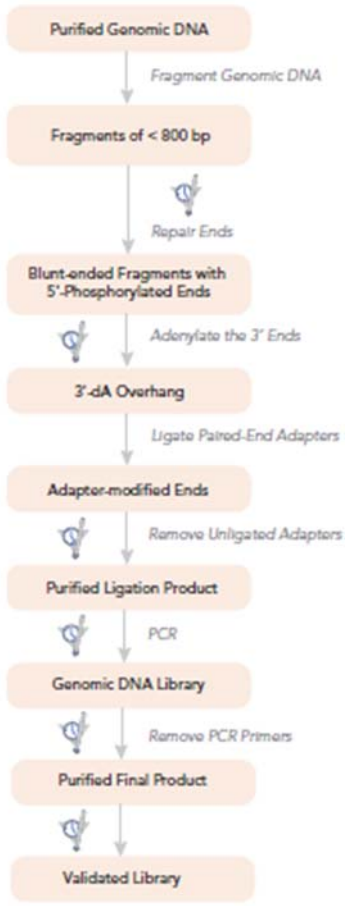


Fig. 4 Paired-End Sample Preparation Workflow

This was followed by ligation of paired-end adaptors which at both ends they put together different sequences at the 5' and 3' end of each strand in the genomic fragment. Ligation products were purified and desirable sizes were selected by agarose gel electrophoresis. The selected DNA fragments are amplified by polymerase chain reaction in order to enrich for fragments that had adaptors on both ends. As a result, a genomic DNA library is obtained and is further purified and the fragments of interest are selected by agarose gel electrophoresis and thus a validated library was produced followed by sequencing.

Illumina genome analyser

The Hiseq sequencing system at the Agricultural research council, Pretoria, South Africa was used for next-generation sequencing of the nematodes samples summarised in the table below. Hiseq was employed because is suitable for sequencing large whole genomes. A description of the system is available at the following link:
http://www.illumina.com/Documents/systems/hiseq/datasheet_hiseq_systems.pdf

Table 1

Summary of EPNs *Oscheius* sp. TEL-2014 GenBank accession (KM492926) samples sequenced, sequencing system and library preparation method used

	Description	Sequencing system	Library preparation method
1	Inbred nematodes (Line 7/partially wild type) Homozygous alleles expected	HiSeq	Paired ends
2	Inbred nematodes (Line 13) Homozygous alleles expected	HiSeq	Paired ends

2.6 Quality control

```
#!/bin/
bash

#PBS -N
FastQC
#PBS -q
WitsLong
#PBS -l
walltime=03:00:00,mem=10gb
#PBS -l
nodes=1:ppn=1
#PBS -o /home/lephotot/TiiQuality/logs/
output.log
#PBS -e /home/lephotot/TiiQuality/logs/
error.log

WORK_DIR=/home/lephotot/TiiQuality

cd $WORK_DIR

fastqc $(ls /home/lephotot/TiiQuality/Sample1/*fastq.gz) -o $WORK_DIR --noextract
```

Fig. 5 Script used for generating a quality control report using FastQC version 0.11.3 (Andrews, 2015). See appendix for reports and interpretation of the reports.

Trimming of poor quality regions on the reads and removing adaptors

```
#PBS -N
Trimomatic
#PBS -q
WitsLong
#PBS -l
walltime=03:00:00,mem=10gb
#PBS -l
nodes=1:ppn=1
#PBS -o /home/lephotot/TiiQualityTrimming/logs/
output.log
#PBS -e /home/lephotot/TiiQualityTrimming/logs/
error.log

WORK_DIR=/home/lephotot/TiiQualityTrimming/

cd $WORK_DIR
OUT_DIR=/home/lephotot/TiiQualityTrimming/Sample13TTrimmingOutput

count=0
for file in $(ls *R1*fastq.gz)
do

(( ++count ))
base_in=$file
base_out="test_trim_"$count".fastq.gz"

java -jar /opt/exp_soft/bioinf/trinity/trinity-plugins/Trimmomatic/trimmomatic.jar PE -phred33 \
-trimlog $OUT_DIR/trimmomatic.log \
-basein $base_in \
-baseout $OUT_DIR/$base_out \
ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:1:true \
MAXINFO:50:0.75 \
MINLEN:50

done
```

Fig. 6 Script used for trimming of adaptor sequences from the raw sequence reads using Trimmomatic version 0.32 (Bolger *et al*, 2014). The reports are in the appendix.

2.7 Genome assembly

An assembly may be tested for its precision by aligning it to a suitable complete reference genome (Salzberg, 2012). In this study, de novo assembly was used to assemble the genomes of the newly isolated species.

The sequences of each samples described in Table 1 were assembled using Velvet version 1.2.10 (Zerbino and Birney, 2013). The workflow in fig. 7 summarises steps taken to assemble the genomes and generate contigs.

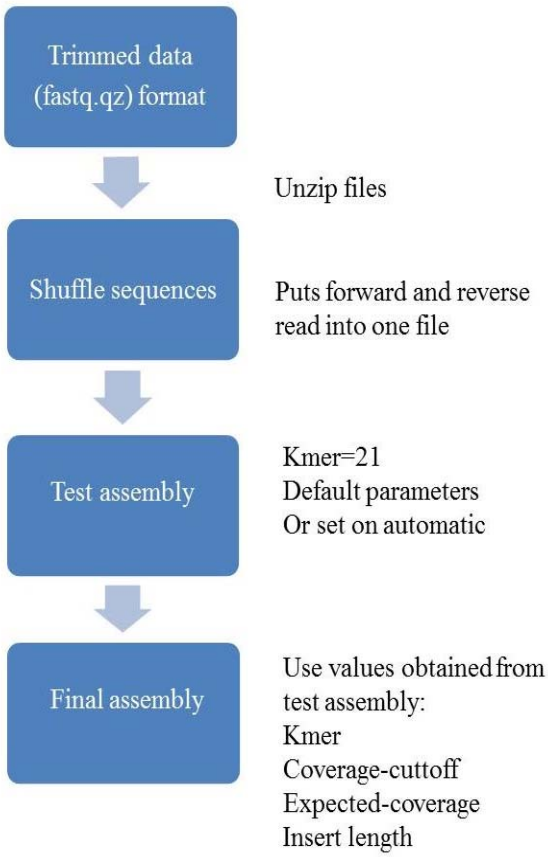


Fig. 7 Assembly workflow used to for assembly of *Oscheius* sp. TEL-2014 genome sequences.

Trimmed reads saved in a zipped format (fastq.gz) were unzipped on the command line. The reads were then shuffled using a perl script on Velvet. Shuffling basically puts the forward read and the reverse read into one file. This is followed by a test assembly where Kmer of 21 was used and all other parameters i.e. coverage cut-off, expected coverage and insert length were set to automatic or default parameters. This test assembly is done to determine suitable parameters for the final assembly.

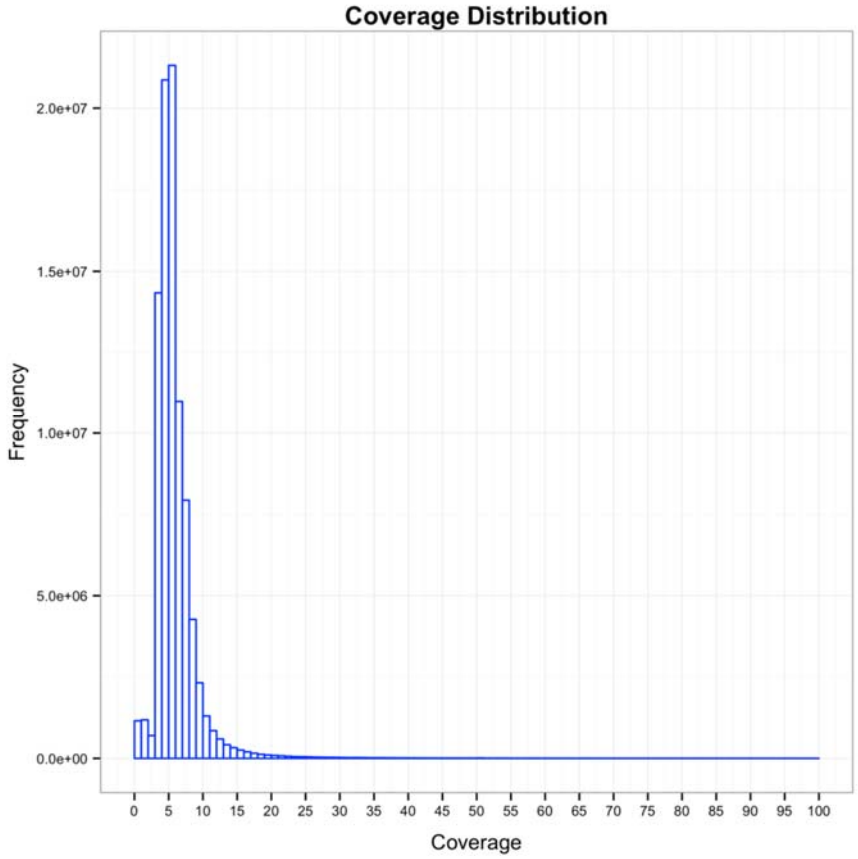


Fig. 8 Plot used to determine the expected coverage which is used in the final assembly step in Velvet assembler

The test assembly includes a graph generated using a velvet script to determine the assembly as shown in Fig 8 and the highest peak on the graph represented the expected coverage. Another script is used to predict values for the other parameters. Velvet h was run to prepare the reads for final assembly and then velvet h was finally employed to achieve the concluding assembly. It is important to note that the k-mers values ranging from 21-59 were used for the test assembly and after running several tests, it was found that K-mer value of 49 was best suited for the data. From observations made during this process, the higher the k-mer, the better the assembly and also hypothetically the better the N50 value.

2.8 Assembly statistics

QUAST version 3.1 (Gurevich *et al*, 2013) was used for the assessment of the assembly

Assembly quality check

BUSCO version 1.1 (Simão *et al*, 2005) was used to assess the completeness of the assembly by aligning the contigs to the reference genome. BUSCO coverts DNA sequences into protein sequences and compared them with proteins of other nematodes in databases. This program will give a brief report to categorise the query gene into these sections: C= complete, PC= partially complete, D= duplicated, F= fragmented, M= missingness.

Further cleaning of contigs

An in-house bash-script was used to extract all the contigs less than 200bp which were removed to meet the requirements of depositing whole-genome shortgun project at Genbank.

```
line_number=0
start=0
begin_seq=''
end_seq=''
sequence=''

declare -a delete=("NODE_124162_length_934_cov_11.359743" "NODE_8612_length_5386_cov_135.856476" "NODE_52687_length_222_cov_18.135136" "NODE_44
\
797_length_382_cov_11.060209" "NODE_70647_length_496_cov_10.592742")

while read line;
do
    let line_number++

    if [ $start -eq "1" ] && [[ ! "$line" =~ ">" ]]
    then
        sequence=$sequence$line
    elif [ $start -eq "1" ] && [[ "$line" =~ ">" ]]
    then
        let start++
        end_seq=$(expr $line_number - 1)
        let start=0

        sequence=$(sed 's/\(N\)\1\+$//' <<< "$sequence")

        if [ "${#sequence}" -ge "200" ] && [[ ! " ${delete[@]} " =~ " ${name:1} " ]]
        then
            echo $name
            sed -r 's/(.{60})/\1\n/g' <<< "$sequence"
        else
            :
        fi
        sequence=''
    fi
 #This is the begining of a sequence (i.e. sequence starts with '>')
    if [[ "$line" =~ ">" ]]
    then
        name=$line
        let start++
        begin_seq=$line_number
    fi
done < $file

#Last
sequence!
sequence=$(sed 's/\(N\)\1\+$//' <<< "$sequence")
if [ "${#sequence}" -ge "200" ] && [[ ! " ${delete[@]} " =~ " ${name:1} " ]]
then
    echo $name
    sed -r 's/(.{60})/\1\n/g' <<< "$sequence"
else
    :
fi
```

Fig. 9 Bash script used to remove all contigs less than 200bp from the nematodes genome.

2.9 Genome annotation

Fig. 10 A typical genome annotation pipeline: building a pipeline for eukaryotes
Diagram adopted from (http://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/)

The genome was sequenced and the quality of the reads was checked and corrected using FASTQC and Trimmomatic, repetitive DNA sequences were masked and identified using Repeat Masker version 3.3.0 (Smit *et al*, 2006). Augustus (Stanke *et al*, 2004) and BLASTx

were used for gene prediction and protein coding DNA identification followed by protein identification using SwissProt.

Repeat Masking

```
#!/bin/bash

#PBS -N RepeatMasker
#PBS -q WitsLong
#PBS -l walltime=20:00:00,mem=20gb
#PBS -l nodes=1:ppn=7
#PBS -o <path>/logs/output_repeatmasker.log
#PBS -e <path>/logs/error_repeatmasker.log

WORK_DIR=<path>
DATA_DIR=<path>

cd $WORK_DIR

start=$(date +"%s")

RepeatMasker \
    -species elegans \
    -pa 6 \
    -dir $WORK_DIR/repeat_masker_out \
    $DATA_DIR/contigs.fa

end=$(date +"%s")
diff=$(($end-$start))

echo -e "\n##########\nAnalysis took $(($diff / 3600)) hours, $(($diff % 3600 / 60)) minutes and $(($diff % 60)) seconds to finish!\n##########
\n"
```

Fig. 11 a bash script used for masking repetitive DNA sequences using Reapeat Masker

Gene prediction

NCBI BlastX was used to search a protein database by means of a translated nucleotide query.

Augustus was used also to predict genes (Open reading frames) which the output was obtained in a gff file.

GFF format (http://www.ensembl.org/info/website/upload/gff.html)

This format is comprised of 9 columns. Column 1=source, 2=software used, 3=feature annotated, 4=start site, 5=end site, 6=score, 7=DNA strand, 8=phase, 9 attribute

```
#!/bin/bash

#PBS -N Gifbaar_Trans_vs_SwissProt
#PBS -q WitsLong
#PBS -l nodes=1:ppn=13,walltime=400:00:00,mem=40gb
#PBS -e <path>/logs/blastx_error.log
#PBS -o <path>/logs/blastx_output.log

WORK_DIR=<path>
DBS=<path>

cd $WORK_DIR

START=$(date '+%A %M-%B-%Y at %X.')
date1=$(date +"%s")
echo ""
echo "#####----- Job started at: $START -----#####"
echo ""

blastx -query <your_contigs> \
    -db $DBS/c_elegans_pep.fasta \
    -out blastx_results.outfmt6 \
    -num_threads 13 \
    -evalue 1e-5 \
    -max_target_seqs 1 \
    -outfmt 6

END=$(date '+%A %M-%B-%Y at %X.')
date2=$(date +"%s")
diff=$(($date2-$date1))
echo ""
echo "#####----- Job ended at: $END -----#####"
echo ""
echo "$(($diff / 60)) minutes and $(($diff % 60)) seconds elapsed."
```

Fig. 12 A blastx bash script used to predict translated proteins. Results were generated with protein IDs which were mapped to protein names on Uniprot

AUGUSTUS

```
#!/bin/bash

#PBS -N AUGUSTUS_Job
#PBS -q WitsLong
#PBS -l walltime=400:00:00,mem=2gb
#PBS -l nodes=1:ppn=1
#PBS -o /home/lephotot/for_augustus/logs/output.log
#PBS -e /home/lephotot/for_augustus/logs/error.log

WORK_DIR=/home/lephotot/for_augustus

cd $WORK_DIR

augustus --species=caenorhabditis --outfile=augustus_output.gff contig_sequences_from_5h_paired_data.txt
```

Fig. 13 an Augustus bash script used for gene prediction.

3. Results and discussion

3.1 **Assembly and sequence analysis of the *Oscheius* sp. TEL-2014**

The *Osheius basothoensis* genome drafts in this study show 44% completeness when using kmer 29 for line 13 and 25%completeness for line 7 when using kmer 25 according to BASCO software. The genome might contain some amount of novel genes and genomic regions which have not been identified in the available reference genomes which BASCO could not align the novel sequences to *C. elegans*.

Two paired-ends libraries were sequenced and assembled where 75 965 contigs (line 7) and 53 190 (line 13) were generated. The total sequencing length or genome size was found was 122 785 442 bp line 7 and 123 752 189 bp line 13. Table 3 shows that the N50 of line 7 is 1 397 and table 4 shows that the N50 contig length, N50 value for line 13 was 3 019. The GC content of the nematodes was slightly different 43.56% (line 7) and 42.24% (line 3).

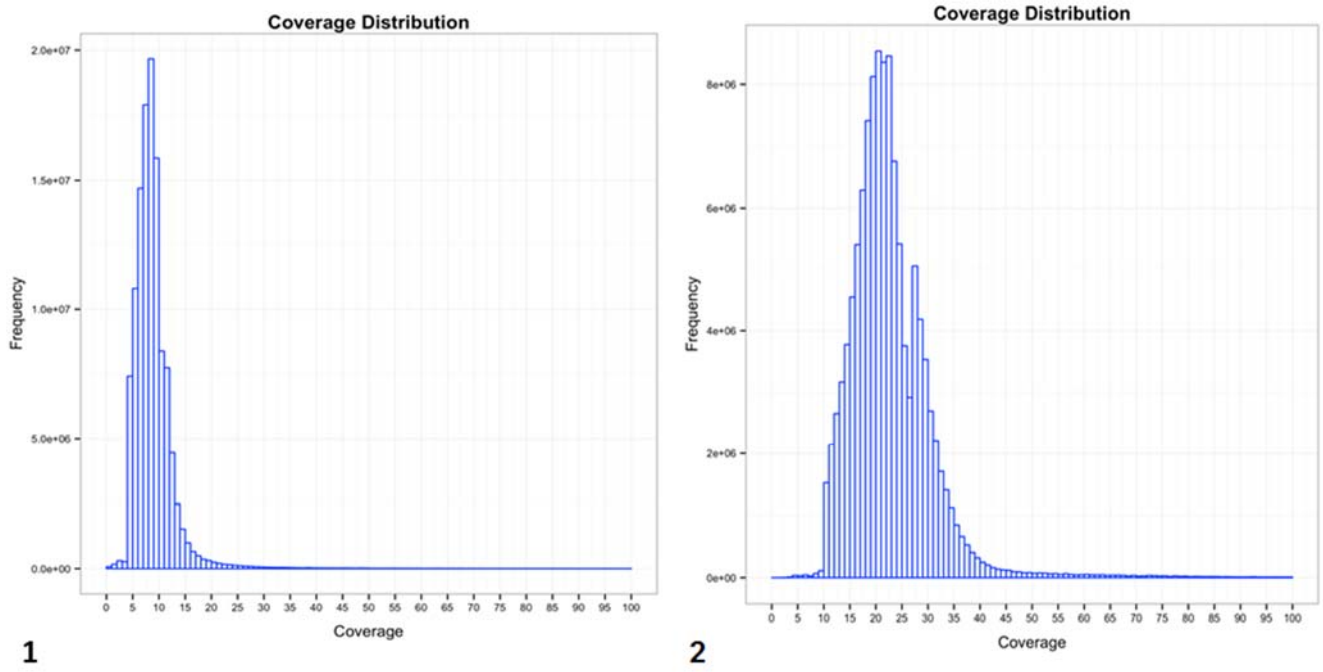


Fig. 14 Plot used to determine the expected coverage which is used in the final assembly step in Velvet assembler. Plot 1: partial wild type nematodes, 2: inbred line 13 nematodes. The heighted peak was used as the coverage value for genome assembly.

Table 2

Summary of the *Oscheius* sp. TEL-2014 genome sequencing, assembly and annotation specifications

Description	Line 7 inbred	Line 13 inbred
Organism	*Oscheius sp. TEL-2014*	*Oscheius sp. TEL-2014*
GenBank accession number	KM492926	KM492926
Sequencer or array type	Illumina HiSeq	Illumina HiSeq
Quality control tool	Fastq-mcf toolkit	Fastq-mcf toolkit
Genome assembler	Velvet	Velvet
Genome annotation tools	BLASTn, AUGUSTUS	BLASTn, AUGUSTUS
WGSsubmission ID	SUB1181822	SUB1158461
BioProject number	PRJNA301782	PRJNA300865
BioSample	SAMN04259800	SAMN0423.445
Aceesion Number	SUB1181822	LNBV00000000
Number of genes predicted using all six reading frames	49947	49947
Sample source location	Grassland in Suikerbosrand Nature Reserve near Johannesburg in South Africa	

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LNBV00000000. The version described in this paper is version LNBV01000000.

Quast reports for assembly statistics

More information is deposited in the link below:

file:///C:/Users/Tiisetso/AppData/Local/Microsoft/Windows/Temporary%20Internet%20Files/Content.IE5/MXXOJHO1/report.html

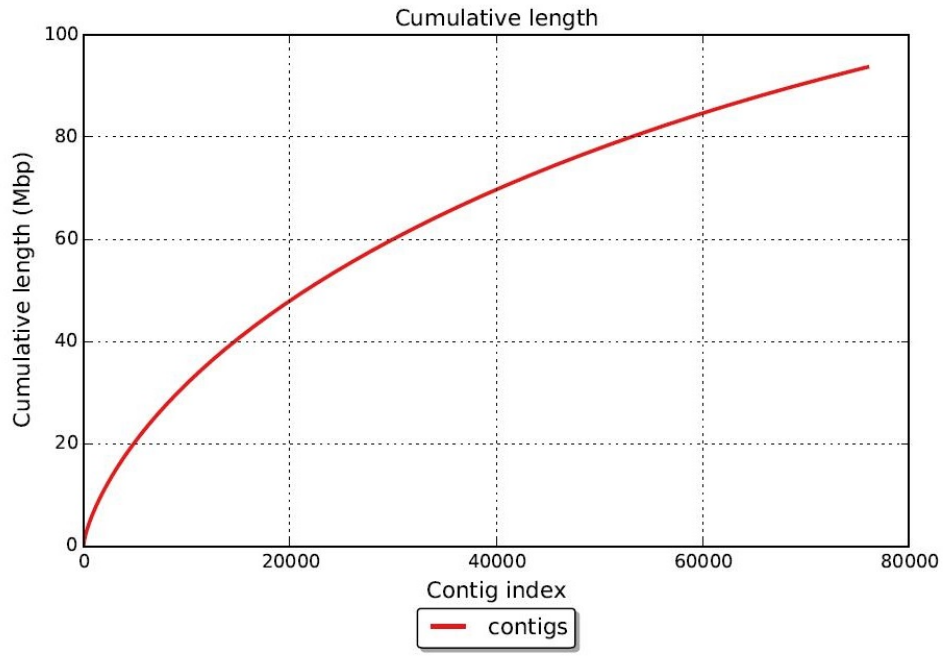


Fig. 15 Cumulative length plot representing the number of contigs generated for line 7

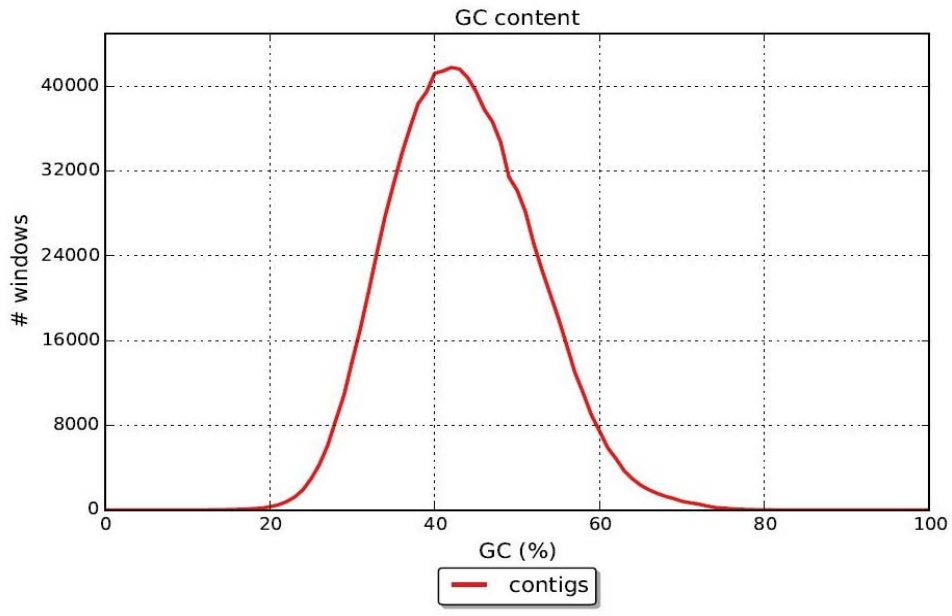


Fig. 16 GC content plot for line 7

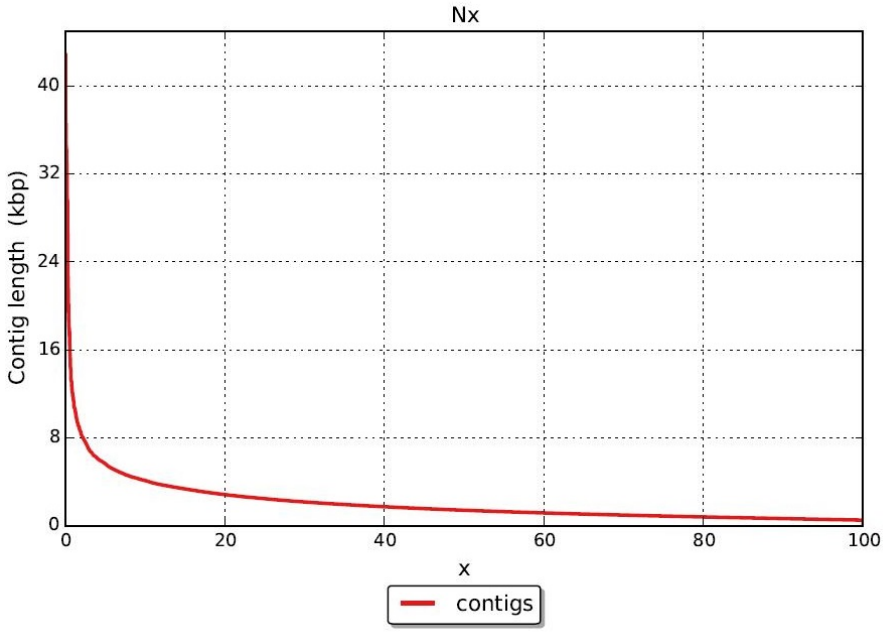


Fig. 17 Plot showing contig length generated by Quast program in line 7

Table 3

Assembly statistics for line 7

Description	
# contigs (>=0 bp)	24 7028
# contigs (>= 1000bp)	33309
Total length (>= 0bp)	122785442
Total length (>=1000bp)	63426717
#contigs	75965
Largest contig	42897
Total length	93641262
GC (%)	43.56
N50	1397
N75	871
L50	19167
L75	40610
# N's per 100kbp	78.82

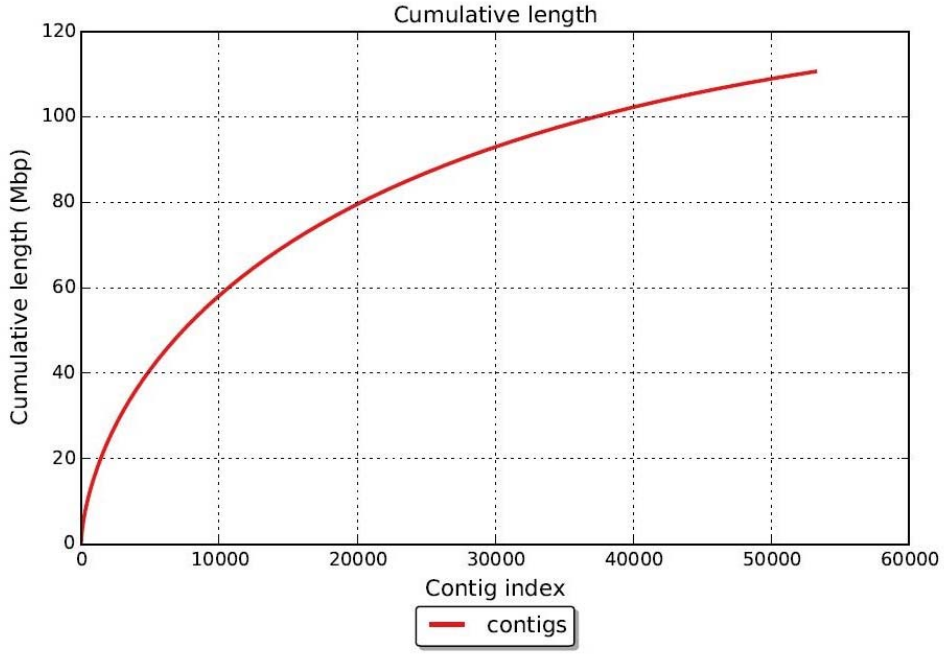


Fig. 18 Cumulative length plot representing the number of contigs generated for line 13

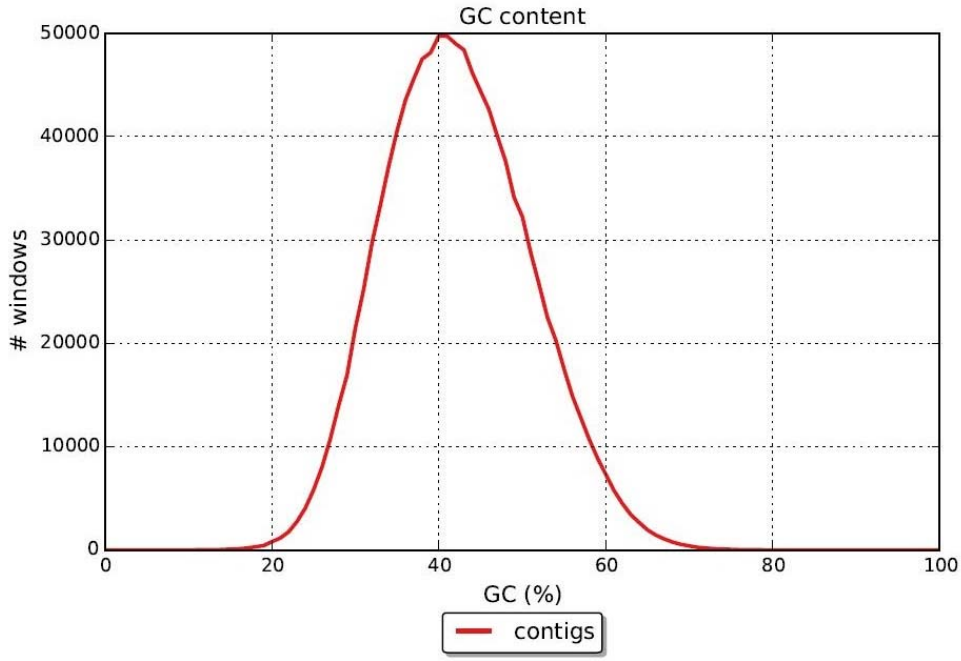


Fig. 19 GC content plot for line 13

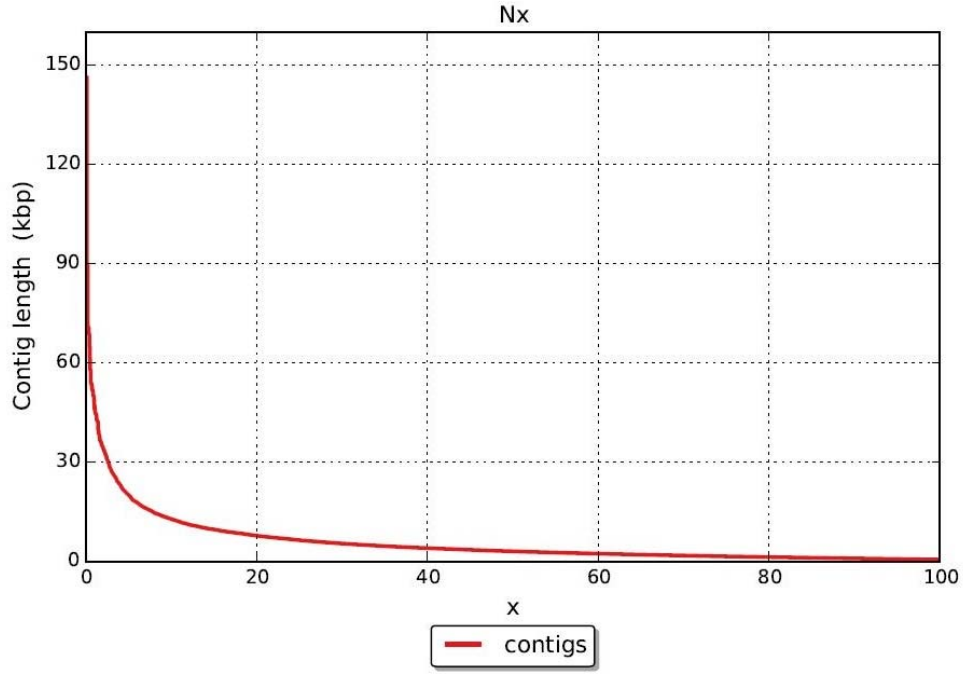


Fig. 20 Plot showing contig length generated by Quast program in Line 13

Table 4

Assembly statistics for line 13

Description	
# contigs (>=0 bp)	145402
# contigs (>= 1000bp)	32787
Total length (>= 0bp)	123752189
Total length (>=1000bp)	95866266
#contigs	53190
Largest contig	146289
Total length	110599558
GC (%)	42.24
N50	3019
N75	1486
L50	9029
L75	22193
# N's per 100kbp	163.41

3.2 Repetitive DNA

Transposable Elements (TEs) receive a lot of attention in genome studies because of their potential to multiply within the genome using variable self-copying mechanisms. This potential allows them cause mutations which may cause structural changes in DNA and hence they able to function as agents of genetic variation and gene evolution in most eukaryotic genomes. Transposable elements are found in genomes of various organisms with varying genome sizes (Kapitonov *et al*, 2009). For example, the pufferfish has a genome size of about 300Mb and 2.7% of its genome is made up of TEs (Gau *et al*, 2014). In our study the nematode's genome size is around 120Mb; 3.21% in line 13 and 2.71% in line 7 is made up of TEs. Interestingly, 45% of the large human genome of about 3000Mb is composed of TEs. Transposable elements therefore affect genome gene density, genome sizes and genome structure (Metcalfe *et al*, 2012).

There are two major classes of TEs: retrotransposons (class I) and DNA transposons (class II). The classification of these TEs is centred on their replication mechanisms. Retrotransposons (class I) and DNA transposons (class II) have dissimilar transposition mechanisms. Class I uses an RNA intermediate and class II uses DNA intermediate for transpositions. Class I TEs are present in greater quantities compared to class II TEs because their replication mechanism is different (de la Chaux *et al*, 2011).

DNA transposons have inverted repeats at the 5' and 3' end and are able to insert at recipient sites within the host genomic that are demarcated by direct repeats of nucleotides. DNA transposons move about from site to site within the host genome by means of a mechanism referred to as cut-and-paste. This cut-and-paste mechanism results in the DNA transposon being literally removed from one region on the genome and assimilated into another (Muñoz-López *et al*, 2010).

In the case of retrotransposons the transposition mechanism involves the transcription of the retrotransposon into an RNA intermediate which is then reversed transcribed into double stranded DNA which is then inserted into sites within the host genome. This process of transposition involving only retrotransposons has been referred to as the copy and paste

mechanism. In this manner multiple copies of retrotransposons are replicated in the host genome.

Mobile transposons have facilitated gene duplication, exon duplication and exon shuffling; and in this manner they have contributed to the evolution of genome through the generation new genes and gene duplication.

DNA transposons (Class II) have been used in other studies to detect genes and pathways involved in pathogenesis of pathogens. DNA transposons may thus provide clues on the pathogenesis of entomopathogenic nematodes in future studies confirm this with detailed investigation strategies and suitable tools for studying TEs (Metcalfe *et al*, 2014).

In summary, transposable elements are described as mobile DNA sequences or mobile genetic elements that possess the genetic capability to move from one location to another in a genome (Havecker *et al*, 2004). They make up a huge percentage of various organisms genomes but the value may vary from one genus or species to another. They have been observed to make up about 12% of *C. elegans* genome and in our study they make up 2.71% to 3.12 % of the *Oscheius* nematodes' genome. TEs mount up overtime and continue to rearrange genomes through their transposition or mobilisation.

A total of 2878544 bp (line 7) and 3772068bp (line 13) sequences were identified as repetitive sequences, which accounted for 2.71% and 3.21%, of the total genome of *Oscheius* sp. TEL-2014 respectively. Repeat Masker analysis showed 1.14 % (line 7) and 1.78% (line 13) of the genome matched simple repeats as shown in table 6 and table 8. The most abundant repeats were the retroelements compared to DNA transposons.

The results showed retroelements and simple repeats were the major repetitive sequences in both *Oscheius* genomes. The high diversity of retrotransposons has been observed in genomes of *C. elegans* (Gau *et al*, 2014). Repeat Masker results before and after removing contigs less than 200bp are shown as evidence that repeats are present in the genomes even after short sequences have been removed to fulfil the requirements of submitting WGS project on NCBI.

Masking repeats BEFORE removing contigs less than 200 bp

Table 5

Partial wild population (line 7) of *Oscheius* nematodes contigs file containing 247028 sequences. The total size of the genome found by Repeat Masker is 122785442bp (122780769bp excluding N/X-runs). The GC level is 43.22 %. The number of bases masked is 3330582bp (2.55 %)

```
==================================================
                number of      length   percentage
                elements*    occupied  of sequence
--------------------------------------------------
Retroelements          935        98758 bp    0.08 %
   SINEs:                3          176 bp    0.00 %
   Penelope            834        84155 bp    0.07 %
   LINEs:              850        86990 bp    0.07 %
    CRE/SLACS            0            0 bp    0.00 %
     L2/CR1/Rex         11         1949 bp    0.00 %
     R1/LOA/Jockey       0            0 bp    0.00 %
     R2/R4/NeSL          2          617 bp    0.00 %
     RTE/Bov-B           3          269 bp    0.00 %
     L1/CIN4             0            0 bp    0.00 %
   LTR elements:        82        11592 bp    0.01 %
     BEL/Pao            41         2767 bp    0.00 %
     Ty1/Copia           0            0 bp    0.00 %
     Gypsy/DIRS1        41         8825 bp    0.01 %
       Retroviral        0            0 bp    0.00 %

DNA transposons        195        15864 bp    0.01 %
   hobo-Activator        6          356 bp    0.00 %
   Tc1-IS630-Pogo       78         5888 bp    0.00 %
   En-Spm                0            0 bp    0.00 %
   MuDR-IS905            0            0 bp    0.00 %
   PiggyBac              9          606 bp    0.00 %
   Tourist/Harbinger     2          231 bp    0.00 %
   Other (Mirage,        0            0 bp    0.00 %
    P-element, Transib)

Rolling-circles          0            0 bp    0.00 %

Unclassified:           61         5328 bp    0.00 %

Total interspersed repeats:      119950 bp    0.10 %


Small RNA:             861        60141 bp    0.05 %

Satellites:             27         3267 bp    0.00 %
```

```
Simple repeats:      16874      1496386 bp    1.22 %
Low complexity:      17997      1665697 bp    1.36 %
================================================
```

*Most repeats fragmented by insertions or deletions have been counted as one element.
The query species was assumed to be caenorhabditis. RepeatMasker version open-3.3.0 default mode runs with rmblastn version 2.2.23+. RepBase Update 20150807, RM database version 20150807.

LINEs which are interspaced nuclear elements have been classified into CR1 and L2 clades. (Havecker *et al*, 2004).

Protein-coding non-TE sequences are present in small proportions in the majority of eukaryotic genomes and have been observed to be the most stable (Gau *et al*, 2014). Repetitive DNA sequences are present in large proportions and often TEs, and thus considered as highly dynamic (Metcalfe *et al*, 2012)

Penelope

Penelope transposable elements belong to the retrotrasposon class and were found to be responsible for causing mutations in *Drosophila virilis* during hybrid dysgenesis leading to hybrid dysgenesis syndrome reported in 1997 and since then they were identified as the only mutagenic elements within the retrotrasposon class (Evgen'ev *et al*, 2005). Penelope elements are found in eukaryotes and are also referred to as Penelope-like elements and they PLE have been established as a unique and new class of eukaryotic retroelements (Arkhipova *et al*, 2006). Penelope elements are present in both lines of the *Oscheius* nematodes.

Masking repeats AFTER removing contigs less than 200 bp

Table 6

Partial wild population (line 7) of *Oscheius* nematodes contigs file containing 133173 sequences. The total size of the genome found by Repeat Masker is 112697346bp (112692673 bp excluding N/X-runs). The GC level is 43.20 %. The number of bases masked is 2878544 bp (2.71 %)

	number of elements*	length occupied	percentage of sequence
Retroelements	906	96380 bp	0.09 %
SINEs:	3	176 bp	0.00 %
Penelope	829	83665 bp	0.07 %
LINEs:	844	86438 bp	0.08 %
CRE/SLACS	0	0 bp	0.00 %
L2/CR1/Rex	11	1949 bp	0.00 %
R1/LOA/Jockey	0	0 bp	0.00 %
R2/R4/NeSL	2	617 bp	0.00 %
RTE/Bov-B	2	207 bp	0.00 %
L1/CIN4	0	0 bp	0.00 %
LTR elements:	59	9766 bp	0.01 %
BEL/Pao	36	2522 bp	0.00 %
Ty1/Copia	0	0 bp	0.00 %
Gypsy/DIRS1	23	7244 bp	0.01 %
Retroviral	0	0 bp	0.00 %
DNA transposons	172	14218 bp	0.01 %
hobo-Activator	5	313 bp	0.00 %
Tc1-IS630-Pogo	64	4725 bp	0.00 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	0	0 bp	0.00 %
PiggyBac	9	606 bp	0.00 %
Tourist/Harbinger	2	231 bp	0.00 %
Other (Mirage, P-element, Transib)	0	0 bp	0.00 %
Rolling-circles	0	0 bp	0.00 %
Unclassified:	58	5106 bp	0.00 %
Total interspersed repeats:		115704 bp	0.10 %
Small RNA:	716	50838 bp	0.05 %
Satellites:	26	3170 bp	0.00 %

```
Simple repeats:      13050      1288631 bp    1.14 %
Low complexity:      14750      1434899 bp    1.27 %
===========================================================
```

Retrotransposons have been further subdivided into LTR retrotransposons and non-LTR retrotransposons based on the presence of long terminal repeat sequences (Piednoël *et al*, 2011). Non-LTR retrotransposons were not identified in *Oscheius* sp. TEL-2014 in both the wild type and inbred line according to results obtained from using Repeat Masker. The program also found a number of unclassified repeats in both nematodes lines which may be hypothesised to account for non-LTR retrotransposons.

There are several kinds of TEs and one of them includes Long Terminal Repeats (LTR) retrotransposons which are found in eukaryotic genomes. LTR retrotransposons (LTRs) are a group of copy and paste transposons, transcribing the TE first into RNA which is then reversed transcribed into DNA. LTRs have the potential to cause and generate mutations on the genome and further lead to rearrangement of the genome. This ability allows LTRs to contribute to genetic variation and evolution (de la Chaux *et al*, 2011).

LTRs have been observed to resemble retroviruses based on their sequence structure and composition (Havecker *et al*, 2004). There are some LTRs elements which contain the *eve* gene which is also present in retroviruses and said to be crucial for infectivity of the virus (Metcalfe *et al*, 2014).

As shown in table 5,6,7,8, LTRs are further divided into 4 subclasses. The division is centred on sequence resemblance and structural features and sequence composition. These subclasses are BEL/Pao, Ty1/Copia, Gypsy/DIRS1 and Retroviral as seen on table. In other papers they referred to the subgroups as Ty1/Copia, Ty3/Gypsy, BEL/Pao and DIRS. Results generated by Repeat Masker in our study show that BEL/Pao elements are present in larger numbers, followed by Gypsy/DIRS1 while the Ty1/Copia and Retroviral elements are not present at all. These findings differ from what other studies have conveyed because they had found that the majority of LTRs elements belong to Ty1/Copia and Ty3/Gypsy subclasses which were found to be present in almost all eukaryotic genomes. Previous studies also indicated that the other subclasses of LTRs occurred recently and not present in most genomes. Ty1/Copia, Ty3/Gypsy and DIRS are the LTRs subclasses which have been reported to exist in almost all eukaryotic genomes (Piednoël *et al*, 2011).

The BEL/Pao subclass is only present in metazoan genomes and existing as second most profuse class of LTR retrotransposons in those genomes. BEL/Pao elements are not found in mammals although they are present in a lot phylum including basal metazoans and thus it was recommended that BEL/Pao elements arose early in animal evolution (de la Chaux *et al*, 2011). It is captivating to find that the *Oscheius* nematodes in our study have the BEL/Pao elements. The BEL/Pao elements are generally 4.2 to 10kb long. Other studies further analysed genome sequences of different organisms and compared their BEL/Pao. Their results showed that *Caenorhabditis brenneri*, *Caenorhabditis briggsae*, *Caenorhabditis elegans*, *Caenorhabditis japonica* and *Caenorhabditis remanei* have 48, 6, 12, 2 and 46 BEL/Pao elements, respectively (de la Chaux *et al*, 2011).

Masking repeats BEFORE removing contigs less than 200 bp

Table 7

Inbred (line 13) of *Oscheius* nematodes contigs file containing 145402 sequences. The total size of the genome found by Repeat Masker is 123752189bp (123736215bp excluding N/X-runs). The GC level is 42.42 %. The number of bases masked is 4253249bp (3.44 %)

	number of elements*	length occupied	percentage of sequence
Retroelements	1402	123995 bp	0.10 %
SINEs:	3	176 bp	0.00 %
Penelope	1245	106529 bp	0.09 %
LINEs:	1263	109396 bp	0.09 %
CRE/SLACS	0	0 bp	0.00 %
L2/CR1/Rex	11	1933 bp	0.00 %
R1/LOA/Jockey	0	0 bp	0.00 %
R2/R4/NeSL	3	617 bp	0.00 %
RTE/Bov-B	4	317 bp	0.00 %
L1/CIN4	0	0 bp	0.00 %
LTR elements:	136	14423 bp	0.01 %
BEL/Pao	94	5554 bp	0.00 %
Ty1/Copia	0	0 bp	0.00 %
Gypsy/DIRS1	42	8869 bp	0.01 %
Retroviral	0	0 bp	0.00 %
DNA transposons	262	20447 bp	0.02 %
hobo-Activator	9	593 bp	0.00 %
Tc1-IS630-Pogo	94	7005 bp	0.01 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	0	0 bp	0.00 %
PiggyBac	12	782 bp	0.00 %
Tourist/Harbinger	2	231 bp	0.00 %
Other (Mirage, P-element, Transib)	0	0 bp	0.00 %
Rolling-circles	0	0 bp	0.00 %
Unclassified:	62	5556 bp	0.00 %
Total interspersed repeats:		149998 bp	0.12 %
Small RNA:	1312	88416 bp	0.07 %
Satellites:	30	3677 bp	0.00 %
Simple repeats:	48235	2418003 bp	1.95 %

```
Low complexity:      20309      1604720 bp    1.30 %
===================================================
```

DNA transposons are categorized into two various families based on *Terminal Inverted Repeats* (TIRs) and *Target Site Duplications* (TSDs) presence on the sequence. TIRs are found on the transposase gene present in DNA transposons and this gene is flanked by 2 TIRs. These TIRs are involved in the excision step of DNA transposons while TSDs are implicated during insertion of the element where the target site DNA is duplicated during the migration of the transposable element from one position to the next on a genome (Muñoz-López *et al*, 2010).

As seen in table 5,6,7,8, the subclasses of DNA transposons are the hobo-Activator, Tc1-IS630-Pogo, En-Spm, MuDR-IS905, *PiggyBac*, and Tourist/*Harbinger.*

Superfamily Tc1

The length of a Tc1 element is between 1 and 5 kb and codes for a transposase flanked by 2 TIRs which their length is between 17 to 1100 bp. One of the sub-classes of the Tc1 elements called *ma*T elements (DD37D) was found in *C. elegans* and *C. briggsae* (Muñoz-López *et al*, 2010).

Masking repeats AFTER removing contigs less than 200 bp

Table 8

Inbred (line 13) of *Oscheius* nematodes contigs file containing 73791 sequences. The total size of the genome found by Repeat Masker is 117481808bp (117465881bp excluding N/X-runs). The GC level is 42.26 %. The number of bases masked is 3772068bp (3.21 %)

	number of elements*	length occupied	percentage of sequence
Retroelements	1353	120039 bp	0.10 %
SINEs:	3	176 bp	0.00 %
Penelope	1225	104801 bp	0.09 %
LINEs:	1241	107558 bp	0.09 %
CRE/SLACS	0	0 bp	0.00 %
L2/CR1/Rex	11	1933 bp	0.00 %
R1/LOA/Jockey	0	0 bp	0.00 %
R2/R4/NeSL	3	617 bp	0.00 %
RTE/Bov-B	2	207 bp	0.00 %
L1/CIN4	0	0 bp	0.00 %
LTR elements:	109	12305 bp	0.01 %
BEL/Pao	84	5163 bp	0.00 %
Ty1/Copia	0	0 bp	0.00 %
Gypsy/DIRS1	25	7142 bp	0.01 %
Retroviral	0	0 bp	0.00 %
DNA transposons	231	18338 bp	0.02 %
hobo-Activator	8	530 bp	0.00 %
Tc1-IS630-Pogo	78	5845 bp	0.00 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	0	0 bp	0.00 %
PiggyBac	12	782 bp	0.00 %
Tourist/Harbinger	2	231 bp	0.00 %
Other (Mirage, P-element, Transib)	0	0 bp	0.00 %
Rolling-circles	0	0 bp	0.00 %
Unclassified:	62	5556 bp	0.00 %
Total interspersed repeats:		143933 bp	0.12 %
Small RNA:	1187	81062 bp	0.07 %
Satellites:	30	3677 bp	0.00 %
Simple repeats:	40958	2086080 bp	1.78 %
Low complexity:	18504	1468441 bp	1.25 %

==========

Superfamily hobo-Activator **(hAT)**

hobo-Activator is found in eukaryotes and its size in length ranges from 2.5 to 5 kb. hAT encodes for a transposase flanked by TIRs which their length is between 5 to 27 bp (Muñoz-López *et al*, 2010).

Superfamily *piggybac*

piggyBac is a DNA transposon isolated from Cabbage Looper moth genome and its transposition mechanism is similar to that of Tc1 elements. *piggyBac* length is 2.4kb and unlike other described elements they contain 13bp TIRs and 19bp internal inverted repeats These elements are also present in human, animals, plants and fungi (Muñoz-López *et al*, 2010).

Simple repeats

Simple repeats are also known as Simple Sequence Repeats (SSR), Single Tandem Repeats or Microsatellites. They typically consist of di or tri-nucleotide repeats and are found in eukaryotes and can originate anywhere in the genome implying that they are interspersed repeats (Meng *et al*, 2015). These repeats were present in larger percentage in both lines of *Oscheius* nematodes in the study.

Satellites

Satellite DNA is the main constituents of functional centromeres and main structural components of heterochromatin. They consist of large quantities of tandem repeats and are found in most eukaryotic genomes (Meštrovic´ *et al*, 2009).

Low complexity DNA repeats were also identified by Repeat Masker and they are found in eukaryotic genomes.

Repetitive DNA serves various roles in different organisms with varying genome sizes.

3.3 Protein coding genes

Prediction of protein-coding genes was performed using Augustus and Blastx. Augustus predicted 49947 in both nematodes lines, among them most of them were complete models

with both the start and stop codon, whereas some of the genes where incomplete Open reading frames as seen in the ggf files obtained. Augustus and Blastx findings revealed some important proteins and conserved domains crucial for gene transcription and translation; some proteins also have a role in DNA replication. Other proteins have been hypothesised to be involved in desiccation tolerance. Protein names and functions were obtained by using BLASTp to align protein sequences generated by Augustus to the protein databases on NCBI. The domain diagrams were obtained from courtesy of NCBI Conserved Protein Domain.

A hypothetical protein CAEBREN_17421 found in *Caenorhabditis brenneri* was also present in the inbred line 13 of *Oscheius sp.* TEL-2014. This protein was isolated from position 1 (start) to 1249 (end) according to Augustus findings.

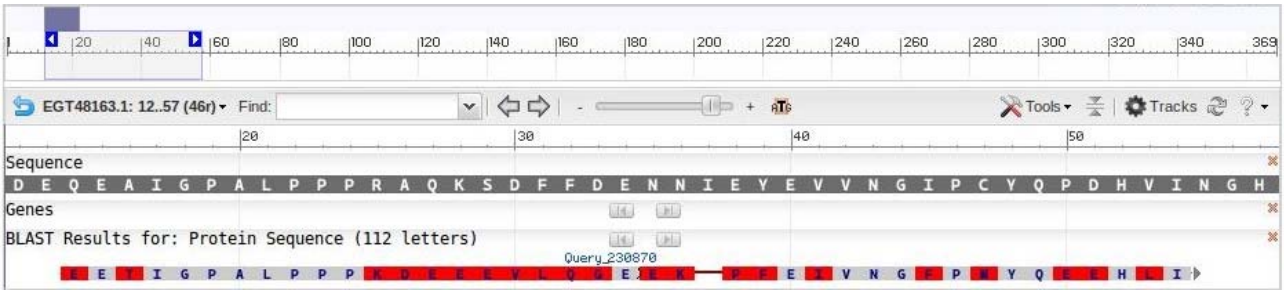


Fig.21 a 112 amino acid long hypothetical protein CAEBREN_17421 predicted by augustus and blastx. Image taken from NCBI Blasx.

WD40 repeat domain was predicted and found on position 5208 (start) to 5385 (end). This protein was identified from *Haemonchus contortus* also known as Barber pole worm which also belongs to the phylum Nematoda, Chromadorea, Rhabditida.

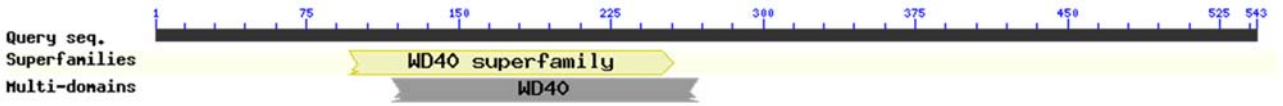


Fig. 22 WD40 repeat domain in line 13 of *Oscheius sp. TEL-2014*. Information from NCBI Conserved Protein Domain

A hypothetical protein Y032_0569g73 found in *Ancylostoma ceylanicum* (Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida)

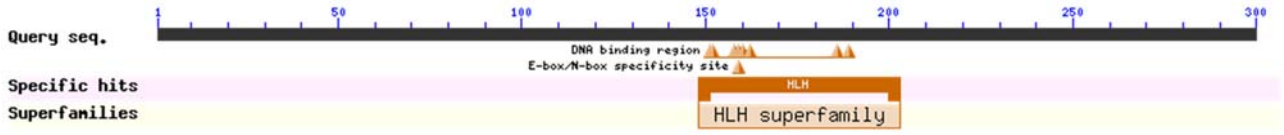


Fig. 23 Helix-loop-helix domain highlighting the specific hits which the Query from line13 of *Oscheius* sp. TEL-2014 contigs had high affinity to.

The Helix-loop-helix domain is found in specific DNA- binding proteins that act as transcription factors. This domain is 60-100 amino acids long and is present in most eukaryotes. It was predicted in position 1072 (start) to 1145 (end) on the genome.

Topoisomerase II large subunit originally found in Escherichia phage PBECO 4 was predicted to be present I the *Oscheius* nematodes genome. Histidine kinase-like ATPases is one of the domains present in this protein.

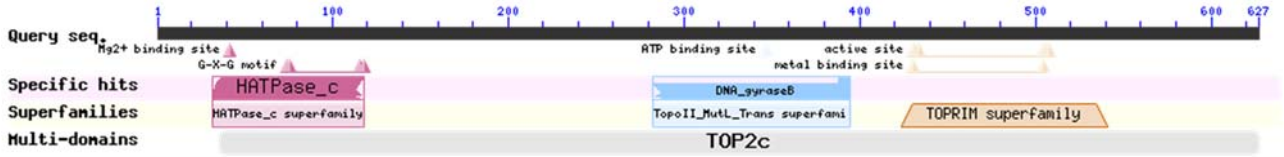


Fig. 24 TOPRIM superfamily found in line 13 of *Oscheius* sp. TEL-2014

A Histidine kinase-like ATPases was predicted to be present in *Oscheius* nematodes. The TOPRIM superfamily also comprises of numerous ATP-binding proteins such as histidine kinase, DNA gyrase B, topoisomerases, heat shock protein HSP90, phytochrome-like ATPases and DNA mismatch repair proteins. The heat shock protein HSP90 may be hypothesised to be involved in desiccation tolerance of these entomopathogenic nematodes.

An uncharacterized protein CELE_C24H12.4 originally found in *Caenorhabditis elegans* was predicted to be present in *Oscheius* sp. TEL-2014 on position 2795 (start) to 2969 (end). A P-loop_NTPase domain superfamily was identified to be associated with this protein.

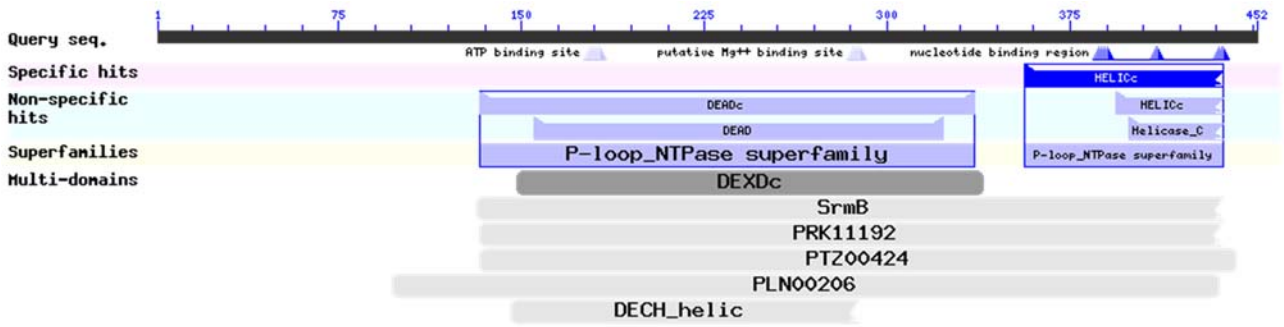


Fig. 25 A P-loop_NTPase domain superfamily observed in line 13 of *Oscheius sp. TEL-2014*

One of the specific domain hits on this protein was the Helicase superfamily c-terminal domain which is linked with DEXDc, DEAD and DEAH box proteins. The DEAD-box helicases is defined as assorted family of proteins involved in ATP-dependent RNA unwinding. Members of this family are the DEAD and DEAH box helicases.

Non-specific hits include the DEXDc which belongs to the DEAD-like helicases superfamily and Helicase_C which is a helicase conserved C-terminal domain.

Superfamilies and multi-domains identified are the SrmB which contain the II DNA and RNA helicase involved in replication, repair and recombination.

Provisional domains isolated are PRK11192 domain recommended as an ATP-dependent RNA helicase SrmB, PTZ00424 which was reported to be part of the helicase and PLN00206 classified as a DEAD-box ATP-dependent RNA helicase. The DECH_helic helicase had no specific function or association mentioned.

A hypothetical protein Y032_0043g788 present in *Ancylostoma ceylanicum* genome, was seen to be present on position 4335(start) to 4335 (end) in the *Oscheius* nematodes most inbred line genome.

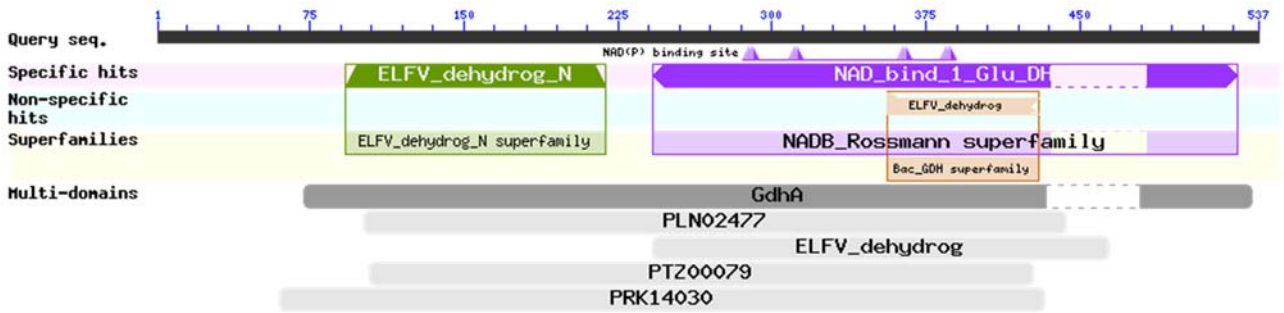


Fig. 26 NAD(P) binding domain of glutamate dehydrogenase, subgroup 1 linked with a hypothetical protein Y032_0043g788 seen in 13 of *Oscheius* sp. TEL-2014

NCBI Conserved Protein Domain has provided information about the amino acid dehydrogenase (DH) is an extensively dispersed family of enzymes that play a role in the catalysis the oxidative deamination of an amino acid to its keto acid and ammonia with concomitant reduction of NADP+.

ELFV_dehydrog_N is one of the specific hits simply described as a dimerization domain and also known as Glutamate/Leucine/Phenylalanine/Valine dehydrogenase. The NAD(P) binding site was found.

The ELFV_dehydrog Glu/Leu/Phe/Val dehydrogenase categorised under superfamilies along with NADB_Rossmann.

Multidomains such as Glutamate dehydrogenase/leucine dehydrogenase (GdhA) have been implicated in amino acid transport and metabolism. PLN02477 which is glutamate dehydrogenase was also identified and ELFV_dehydrog, a Glutamate/Leucine/Phenylalanine/Valine dehydrogenase.

Provisional multidoamins such as PTZ00079, a NADP-specific glutamate dehydrogenase and PRK14030, glutamate dehydrogenase were found.

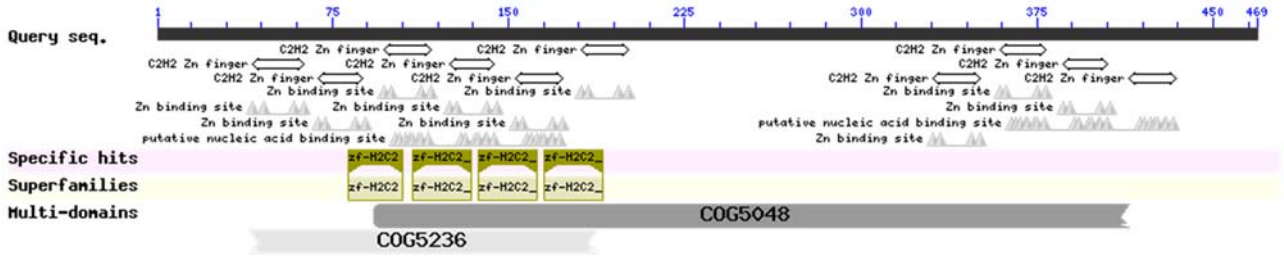


Fig. 27 Zinc-finger domains in 13 of *Oscheius sp. TEL-2014*

Zinc-finger domains were identified and are involved in gene expression, cell-adhesion and protein folding. The predicted zinc finger protein 271-like is present in *Acyrthosiphon pisum* genome. This protein is found on position 2196 (start) to 2196 (end) based on the ggf file report provided by Augustus.

A hypothetical protein Y032_0004g2073 found in *Ancylostoma ceylanicum* (Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida) was also identified in line 7 of *Oscheius* sp. TEL-2014

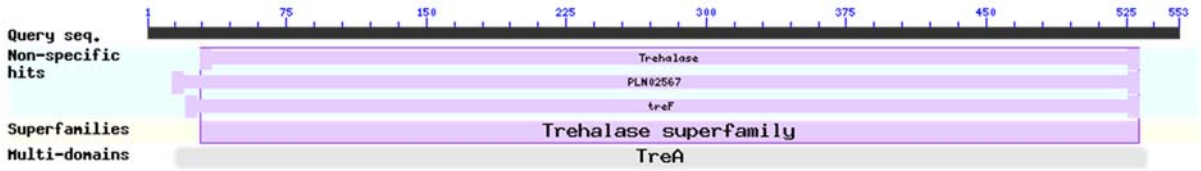


Fig. 28 Trehalase (EC: 3.2.1.28) is recognized as an enzyme that recycles trehalose to glucose. Trehalose is a biological trademark of heat-shock response in yeast and thus and protects it against a variety of stresses.

The eLRR (extracellular Leucine-Rich Repeat) isolated from *Caenorhabditis elegans* was identified and found in line 7 of *Oscheius* sp. TEL-2014

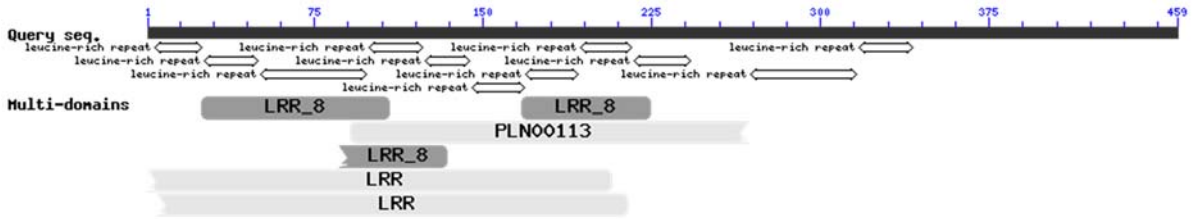


Fig. 29 Leucine rich repeat domains are often seen to have a function in stabilisation protein structures

A hypothetical protein CAEBREN_28360 found in *Caenorhabditis brenneri* genome was also identified. This protein may be further hypothesised to be involved in nematodes chemotaxis and behaviour.

Fig. 30 Neurotransmitter-gated ion-channel ligand binding domain in line 7 of *Oscheius* sp. TEL-2014. It has been reported that members of the LIC family of ionotropic neurotransmitter receptors are found only in vertebrate and invertebrate animals.

4. Conclusion and suggestions

The process of assembling and annotation of genome sequence data is faced with several formidable computational challenges. The whole process is limited by the computation power and reliability of the various genome assembly tools to generate a finished genome. Draft genomes have been on the rise with the hope to complete and improve the assemblies. Sequencing technology continues to undergo rapid development and it is advancing with the development of new processes and technologies which will decrease the time taken for raw genome sequence data to be generated, assembled and annotated. The read length in Illumina has increased from 25bp to 125-300bp over the past few years. There are several reasons which may differ or be the same amongst genome assemblers and they include read length differences, N50 value, read counts, error profile and steps taken by the software to try reduce errors. Furthermore, it has been observed that NGS data "sometimes exist as a mixture of reads produced by different technologies" (Bradnamy *et al*, 2013).

Resequencing whole genome DNA of this organism and assessing the quality of the sequence reads obtained followed by a reasonable assembly may improve gene prediction. Different tools may be used to assemble the genome and choose tools which may improve the assembly. There are a number of assembly validation tools which are used especially for de novo assembly for example Quast. Assembly validation is needed when one is dealing with highly repetitive data.

Comparative genomics is one of the reasons sequencing of entomopathogenic nematodes is gradually increasing. Another reason is based on the availability of reference genomes which

may be used during genome annotation using gene prediction tools. These tools are linked to databases containing established or manually curated gene models. The *C. elegans* genome has been fully annotated and may be used as a suitable reference for newly isolated species genome annotation.

The main objective of gene prediction tools is to predict all genes correctly and making an effort to include all alternative isoforms. Nonetheless, these isoforms are not always efficiently predicted. In our study, a great number of genes were predicted. Further studies need to be conducted to study functions and homology of genes predicted.

Studies on inbreeding may further educate us on the important types of this mating phenomenon which may contribute to how we define inbreeding. The first type previously reported is pedigree inbreeding which states that an individual organism is considered inbred if its parents have a common ancestor. This is imperative to note because in our study we know that both parents are from similar ancestral origin because both parents belong to the same genus and species. The second type is inbreeding as non-random mating which states that this type of inbreeding is "relative" to mating occurring randomly within a population. Other types force inbreeding to occur because of the limited size of the population and thus leading to genetic consequences such as genetic drifting. These types of inbreeding briefly described have given us an idea of the expected outcomes in terms of the genetic composition of the most inbred line 13 mainly on just seeing if the genome becomes homozygous than the genome of partially wild type nematodes population (line 7) which is hypothesised to be heterozygous.

The genome evidence described in the present study offers a valuable platform for future studies of *Oscheius* nematodes and possesses momentous importance in the Agricultural industries and scientific research.

5. References

Abebe E, Jumba M., Bonner K, Gray V, Morris K, and Thomas W.K, 2010. An entomopathogenic *Caenorhabditis briggsae*. J. Exp. Biol, *213*:3223–3229.

Abebe E, Bonner K, Gray V, Thomas W.K, 2011. Response to "Bugs don't make worms kill". J Exp Biol, 214:1053–1054.

Andrews S, 2015. FastQC: a quality-control tool for high-throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arkhipova I. R, 2006. Distribution and Phylogeny of *Penelope-Like* Elements in Eukaryotes. Sysl *Biol.* 55(6):875-885.

Bolger A. M, Lohse M, and Usadel B, 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, 170.

Bradnam K.R, Fass1 J.N, Alexandrov A *et al*, 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience, 2:10.

De la Chaux N and Wagner A, 2011. BEL/Pao retrotransposons in metazoan genomes. BMC Evolutionary Biology, 11:154.

Evgen'ev M.B and I.R. Arkhipov, 2005. Penelope-like elements-a new class of retroelements: distribution, function and possible evolutionary significance Cytogenet Genome Res, 110:510–521.

Gao Y, Gao Q, Zhang H, Wang L, Zhang F, Yang C, and Song L, 2014. Draft Sequencing and Analysis of the Genome of Pufferfish Takifugu flavidus. DNA research, 21:627–637.

Gurevich A, Saveliev V, Vyahhi N, Tesler G, 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics, 29:1072–1075.

Havecker E.R, Gao X and Voytas D.F, 2004. The diversity of LTR retrotransposons. Genome Biology, 5:225.

Hazir S, Kaya H.K, Stock P, and Keskun N, 2003. Entomopathogenic Nematodes (Steinernematidae and Heterorhabditidae) for Biological Control of Soil Pests. Turk J Biology, 181-202.

Hu P.J, 2014. Whole genome sequencing and the transformation of *C. elegans* forward genetics. Methods, 68:437–440.

Kapitonov V.V, Tempel S, and Jurka J, 2009. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. Gene 448, 207–213.

Keller L.F and Walle D.M, 2002. Inbreeding effects in wild populations. TRENDS in Ecology and Evolution, 17(5):230-241.

Malan A.P, Nguyen K.B, and Addison M.F, 2006. Entomopathogenic nematodes (Steinernematidae and Heterorhabditidae) from the southwestern parts of South Africa. African Entomol, 12:65–69.

Meng F, Liu L, Peng M, Wang Z, Wang C and Zhao Y, 2015. Genetic diversity and population structure analysis in wild strawberry (Fragaria nubicola L.) from Motuo in Tibet Plateau based on simple sequence repeats (SSRs). Biochemical Systematics and Ecology, 63: 113-118.

Metcalfe C.J, File´e J, Germon I, Joss J, and Casane D, 2012. Evolution of the Australian Lungfish (Neoceratodus forsteri) Genome: A Major Role for CR1 and L2 LINE Elements. Molecular Biology and Evolution. 29(11):3529–3539.

Meštrovic´ N, Plohl M and Castagnone-Sereno P, 2009. Relevance of satellite DNA genomic distribution in phylogenetic analysis: A case study with root-knot nematodes of the genus Meloidogyne. Molecular Phylogenetics and Evolution, 50:204–208.

Metcalfe C.J and Casane D, 2014. Modular organization and reticulate evolution of the ORF1 of Jockey superfamily transposable elements. Mobile DNA, 5:19.

Muñoz-López M and García-Pérez J.L, 2010. DNA Transposons: Nature and Applications in Genomics. Current Genomics, 11:115-128.

Piednoël M, Isabelle R Gonçalves, Dominique Higuet and Eric BonnivardEukaryote DIRS1-like retrotransposons: an Overview. BMC Genomics, 12:621.

Salzberg S.L, Phillippy A.M, Zimin A, Puiu D, Magoc T, Koren S, Treangen T.J, Schatz M.C, Delcher A, L, Roberts M, Marcxais G, Pop M and Yorke J.A, 2012. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Cold Spring Harbor Laboratory Press.

Simão F.A, Waterhouse R.M, Ioannidis P, Kriventseva E.V, and Zdobnov E.M, 2005. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. 1-6

Smit, A.F.A, Hubley, R and Green, P, 2006. RepeatMasker Open-3.0. <http://www.repeatmasker.org>.

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

Stuart R.J, Barbercheck M.E, Grewal P.S, Taylor R.A.S, and Hoy C.W, 2006. Population biology of entomopathogenic nematodes: Concepts, issues, and models. Biological Control 38: 80–102.

Whitton C, Daub J, Quail M, Hall N, Foster J, Ware J, Ganatra M, Slatko B, Barrell B and Blaxter M, 2004. A genome sequence survey of the filarial nematode *Brugia malayi*: repeats, gene discovery, and comparative genomics. Molecular & Biochemical Parasitology, 137: 215–227.

Wilson R.K, 1999. How the worm was won, the *C. elegans* genome sequencing project. TIG 15(2):51-58.

Zerbino D.R and Birney E, 2013. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Research, 18:821-829.

Zhang C, Liu J, Xu M, Sun J, Yang S, An X, Gao G, Lin M, Lai M, He Z, Wu Y and Zhang K, 2008. Heterorhabditidoides chongmingensis gen. nov., sp. nov. (Rhabditida: Rhabditidae), a novel member of the entomopathogenic nematodes. Journal of Invertebrate Pathology 98:153–168.

Zhou K, Huang B, Zouc M, Lua D, He S and Wang G, (2015). Genome-wide identification of lineage-specific genes within Caenorhabditis elegans. Genomics 106:242–248

CHAPTER 9

Research Conclusions

Oscheius sp. TEL-2014 is described as a novel entomopathogenic nematode species based on its morphometrics and 18S rRNA gene sequence originality.

The whole genome draft of *Serratia* sp. TEL will allow for the investigation of identified genes and will be critical in furthering the understanding of the insect pathogenicity of *Serratia* sp. TEL carried by entomopathogenic nematode *Oscheius* sp. TEL-2014.

Genetic subsystems discussed were highlighted as important because their presence contributed to the insect pathogenicity of the bacterial species for which the nematode acted as the vector for the location and infection of potential insect hosts.

Further studies need to be conducted to investigate entomopathogenecity mechanisms and efficacy in greater detail so that the nematode may be assigned or recommended as biopesticide for controlling specific problematic insects especially in Agricultural industries.

The draft genome of *Oscheius* sp. TEL-2014 described creates a foundation for future studies of *Oscheius* nematodes which will benefit Agricultural industries and scientific research.

Appendix 1

Solutions

TAF

8 ml 35% formaldehyde

2.28 ml triethanolamine

104 ml distilled water

Double-strength TAF

8 ml 35% formaldehyde

2.28 ml triethanolamine

52 ml distilled water

Solution I

20 ml 95% ethanol

1 ml glycerine

79 ml distilled water

Solution II

5 ml glycerine

95 ml 95% ethanol

Lipid agar medium

Lipid agar media
Method adapted from Kaya *et al* 1997.

Corn syrup was replaced with honey

10g honey

5g yeast extract

25g nutrient agar

2.5ml cod liver oil

2g $MgCl_2.6h_20$

1 litre of distilled water

Mix the ingredients well and autoclave at 121^oC and 15psi for 20 minutes. Pour the media into sterile Petri dishes aseptically under a laminar flow system.

Appendix 2

Selective media EPNs symbiotic bacteria

NBTA selective media

1 litre nutrient agar

0.04g triphenyltetrazolium chloride (TTC)

0.025g bromothymol blue (BTB)

Mix nutrient agar and BTB and autoclave the media at 121˚C and 15 psi for 15 minutes. Add TTC, just before pouring into Petri dishes; ensuring that the autoclaved media is less than 50˚C. TTC will break down if added when medium is too hot. Swirl to mix and dispense into sterile Petri dishes and leave to solidify.

McConkey agar selective media

Composition (g/l) 92

20.0g Peptone
10.0g Lactose
1.5g Bile Salts
5.0g Sodium Chloride
0.03g Neutral Red
0.0001g Crystal Violet
13.5g Agar

Weigh out McConkey agar powder and suspend in 1000ml distilled water. Boil whilst stirring until completely dissolved and autoclave at 121ºC and 15 psi for 15 min. Cool to 45 – 50ºC. Mix properly and dispense into sterile Petri dishes and leave to solidify.
0.1% jik solution for infective juvenile sterilization
34ml distilled water
1ml 3.5% jik
Autoclave distilled water. Mix jik and distilled water in bottles and autoclave at 121˚C and 15 psi for 15 min.

Recipe for *in vitro* culture media

Nutrient Broth
4.0% (W/V) Canola oil
25mg/ml glucose
Weigh out nutrient broth powder and suspend in desired volume of distilled water. Add glucose. Mix well and dispense adequate amounts into volumetric flasks. Add 4.0% (W/V) Canola oil to each volumetric flask containing nutrient broth and autoclave at 121˚C and 15 psi for 15 minutes
5X TBE
54g Tris base
27.5g Boric acid
20ml 0.5M EDTA pH 8.0
Mix with 1L distilled water and autoclave at 121˚C at 15psi for 20 minutes

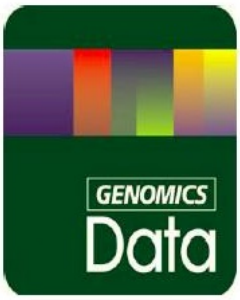
Genomic DNA extraction protocol

Bacteria were grown on NBTA and Mc Conkey plates and an isolated bacterial colony was picked and suspend in a ZR BashingBeadTM Lysis Tube and vortex at maximum speed for 5minutes, this aided in gently breaking down the bacterial wall in order to release nucleic material. The ZR BashingBeadTM Lysis tube was centrifuged at 10 000 rpm for 1 minute. Up to 400µl of the supernatant was transferred into a Zymo-Spin TM IV Spin Filter in a Collection Tube and centrifuged at 7000 rpm for 1 minute. 1200µl of Fungal/ Bacterial DNA binding buffer was added to the filtrate in the Collection Tube and transferred 800µl of the mixture to a Zymo-SpinTM II Column in a Collection Tube and centrifuged at 10000rpm for 1 minute. The flow was discarded from the Collection Tube and 200µl DNA Pre-Wash Buffer was added to the Zymo-SpinTM II Column in a new Collection Tube and centrifuged at 10000rpm for 1 minute. 500µl of Bacterial DNA Wash Buffer was added to the Zymo-SpinTM II Column and centrifuged at 10000rpm for 1 minute. The Zymo-SpinTM II Column was transferred to a clean 1.5 ml microcentrifuge tube and added 100µl DNA Elution Buffer directly to the column matrix. The tube was centrifuged at 100000rpm for 30 seconds to release the DNA from the matrix. The DNA was stored at 4°C to be used for further analysis.

Publications

http://ac.els-cdn.com/S2213596016300174/1-s2.0-S2213596016300174-main.pdf?_tid=46c21ac6-c8b6-11e5-9123-00000aacb362&acdnat=1454312270_0222613fed6355d4746bb1a87e4413fb

Accepted Manuscript



Whole genome sequence of *Oscheius* sp. TEL-2014 entomopathogenic nematodes isolated from South Africa

Tiisetso E. Lephoto, Phelelani Mpangase, Shaun Aron, Vincent M. Gray

PII: S2213-5960(16)30017-4
DOI: doi: 10.1016/j.gdata.2016.01.017
Reference: GDATA 472

To appear in: *Genomics Data*

Received date: 21 January 2016
Accepted date: 29 January 2016

Please cite this article as: Tiisetso E. Lephoto, Phelelani Mpangase, Shaun Aron, Vincent M. Gray, Whole genome sequence of *Oscheius* sp. TEL-2014 entomopathogenic nematodes isolated from South Africa, *Genomics Data* (2016), doi: 10.1016/j.gdata.2016.01.017

http://genomea.asm.org/content/3/4/e00747-15.full.pdf

genomeA
Journals.ASM.org

Draft Whole-Genome Sequence of *Serratia* sp. Strain TEL, Associated with *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae) Isolated from a Grassland in South Africa

Tlisetso E. Lephoto,[a] Jonathan Featherston,[b] Vincent M. Gray[c]

School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, South Africa[a]; Agricultural Research Council, Biotechnology Platform, Pretoria, South Africa[b]; School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, South Africa[c]

Here, we report on the draft genome sequence of *Serratia* sp. strain TEL, associated with *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae, KM492926) isolated from a grassland in Suikerbosrand Nature Reserve near Johannesburg in South Africa. *Serratia* sp. strain TEL has a genome size of 5,000,541 bp with 4,647 genes and a G+C content of 59.1%.

Received 2 June 2015 **Accepted** 4 June 2015 **Published** 9 July 2015

Citation Lephoto TE, Featherston J, Gray VM. 2015. Draft whole-genome sequence of *Serratia* sp. strain TEL, associated with *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae) isolated from a grassland in South Africa. Genome Announc 3(4):e00747-15. doi:10.1128/genomeA.00747-15.

Address correspondence to Tlisetso E. Lephoto, tlisetso.lephoto@students.wits.ac.za.

The majority of entomopathogenic nematode (EPN) species that have been isolated belong to heterorhabditids and steinernematids (1), which act as vectors for insect pathogenic bacteria species belonging to the genera *Photorhabdus* and *Xenorhabdus*, respectively (2). However, other species of nematodes, one of which includes a strain of *Caenorhabditis briggsae*, have also been shown to act as vectors for insect pathogenic strains of *Serratia marcescens* (3).

Oscheius nematodes now also act as vectors for insect pathogenic bacteria belonging to the genus *Serratia* (4). Like heterorhabditids and steinernematids, entomopathogenic *Oscheius* species also have a dauer stage in their life cycle (5). During this stage, *Serratia* spp. persist internally within the infective juvenile (IJ) as endosymbionts (6). On infection of an insect host, entomopathogenic *Serratia* is released by the *Oscheius* IJ into the insect's hemocoel, in an infection process that is very similar to the pathogenic behavior of the heterorhabditids and steinernematids (5). All EPNs complete their cycle by feeding on the bacteria growing inside the host's hemocoel (7). Once the nutrient reserves have been depleted, nonfeeding *Oscheius* sp. IJs, carrying a colony of endosymbiotic *Serratia* sp. bacteria (8), migrate from the cadaver and survive in a state of anhydrobiosis for extended periods in the soil (9).

In this study, the novel insect pathogenic *Serratia* sp. strain TEL (GenBank accession number KP711410) was isolated from the gut of an IJ of *Oscheius* sp. TEL-2014 (KM492926).

Methods described in reference 10 were employed to isolate *Serratia* sp. strain TEL from *Oscheius* sp. TEL-2014. Whole DNA extraction from solid bacterial colony cultures was done using the ZR bacterial DNA miniprep kit (Zymo Research). Genomic DNA paired-end libraries were generated with the Nextera DNA sample preparation kit (Illumina) and indexed using the Nextera index kit (Illumina). Paired-end (2 × 300 bp) sequencing was performed on a MiSeq Illumina using the MiSeq reagent kit version 3 at the Agricultural Research Council Biotechnology Platform. Quality and adapter trimming was performed with the fastq-mcf toolkit. The genome was assembled using SPADES, and 19 contigs were generated with an average length of 301,767 bp and an N_{50} of 200,110 bp. The genome of *Serratia* sp. strain TEL was found to be 5,000,541 bp in size, with a G+C content of 59.1%, which was similar to that of other *Serratia* species (11–13). Furthermore, the contigs were annotated using the NCBI Prokaryotic Genome Automatic Annotation Pipeline.

The annotated features are as follows: 4,647 genes were found, and 4,495 were protein-coding sequences (CDS). The genome contains 36 pseudogenes, 2 CRISPR arrays, 13 rRNA genes with five operons (5S, 16S, 23S), 88 tRNAs, 15 noncoding RNAs, and 9 frameshifted genes.

Several genes involved in virulence, disease, defense, stress response, cell division, motility, and chemotaxis were identified. This draft genome sequence will allow for the investigation of identified genes and will be critical in furthering the understanding of the insect pathogenicity of *Serratia* sp. strain TEL.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number LDEG00000000, which is the first version.

ACKNOWLEDGMENTS

We thank the Gauteng Department of Agriculture and Rural Development (GDARD) for funding the research project. T.E.L. received an Innovation Doctoral Scholarship from the National Research Foundation (NRF) (grant number SFH1208147793) and a Wits Postgraduate Merit Award (PMA) from the University of the Witwatersrand.

We thank the Agricultural Research Council (ARC) Biotechnology platform for Illumina technology sequencing services.

REFERENCES

1. **Malan AP, Nguyen KB, Addison MF.** 2006. Entomopathogenic nematodes (Steinernematidae and Heterorhabditidae) from the southwestern parts of South Africa. Afr Entomol 12:65–69.
2. **Zhang C, Liu J, Xu M, Sun J, Yang S, An X, Gao G, Lin M, Lai R, He Z, Wu Y, Zhang K.** 2008. *Heterorhabditidoides chongmingensis* gen. nov.,

sp. nov. (Rhabditida: Rhabditidae), a novel member of the entomopathogenic nematodes. J Invertebr Pathol **98**:153–168. http://dx.doi.org/10.1016/j.jip.2008.02.011.

3. **Rae R, Sommer RJ.** 2011. Bugs don't make worms kill. J Exp Biol **214**: 1053. http://dx.doi.org/10.1242/jeb.052480.
4. **Abebe E, Bonner K, Gray V, Thomas WK.** 2011. Response to "Bugs don't make worms kill." J Exp Biol **214**:1053–1054. http://dx.doi.org/10.1242/jeb.053827.
5. **Torres-Barragan A, Suazo A, Buhler WG, Cardoza YJ.** 2011. Studies on the entomopathogenicity and bacterial associates of the nematode *Oscheius carolinensis*. Biol Contr **59**:123–129. http://dx.doi.org/10.1016/j.biocontrol.2011.05.020.
6. **Abebe E, Jumba M, Bonner K, Gray V, Morris K, Thomas WK.** 2010. An entomopathogenic *Caenorhabditis briggsae*. J Exp Biol **213**:3223–3229. http://dx.doi.org/10.1242/jeb.043109.
7. **Stuart RJ, Barbercheck ME, Grewal PS, Taylor RJ, Hoy WC.** 2006. Morphological and functional dimorphism in *Xenorhabdus* spp., bacteria symbiotically associated with the insect pathogenic nematodes *Neoaplectana* and *Heterorhabditis*. Microbiol **121**:303–309. http://dx.doi.org/10.1099/00221287-121-2-303.
8. **Ishii K, Adachi T, Hara T, Hamamoto H, Sekimizu K.** 2014. Identification of a *Serratia marcescens* virulence factor that promotes hemolymph bleeding in the silkworm, *Bombyx mori*. J Invertebr Pathol **117**:61–67. http://dx.doi.org/10.1016/j.jip.2014.02.001.
9. **Grewal PS.** 2000. Enhanced ambient storage stability of an entomopathogenic nematode through anhydrobiosis. Pest Manag Sci **56**:401–406. http://dx.doi.org/10.1002/(SICI)1526-4998(200005)56:5@003c401::AID-PS137>3.0.CO;2-4.
10. **Kaya HK, Stock SP.** 1997. Techniques in insect nematology, p 281–324. *In* Lacey LA (ed), Manual of techniques in insect pathology. Elsevier, Amsterdam, Netherlands.
11. **Kwak Y, Khan AR, Shin JH.** 2015. Genome sequence of *Serratia nematodiphila* DSM 21420T, a symbiotic bacterium from entomopathogenic nematode. J Biotechnol **193**:1–2. http://dx.doi.org/10.1016/ j.jbiotec.2014.11.002.
12. **Serepa MH, Gray VM.** 2014. Draft whole-genome sequence of *Serratia marcescens* strain MCB, associated with *Oscheius* sp. MCB (Nematoda: Rhabditidae) isolated from South Africa. Genome Announc 2(5):e00911-14. http://dx.doi.org/10.1128/genomeA.00911-14.
13. **Wang Y, Yuan Y, Zhou L, Su Q, Fang X, Li T, Wang J, Chang D, Su L, Xu G, Guo Y, Yang R, Liu C.** 2012. Draft genome sequence of *Serratia marcescens* strain LCT-SM213. J Bacteriol **194**:4477–4478. http://dx.doi.org/10.1128/JB.00933-12.

Genomics Data 6 (2015) 54–56

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Genome sequencing and annotation of *Serratia* sp. strain TEL

Tiisetso E. Lephoto *, Vincent M. Gray

Microbiology and Biotechnology Department, School of Molecular and Cell Biology, University of the Witwatersrand, Braamfontein 2050, Johannesburg, South Africa

ARTICLE INFO

Article history:
Received 7 August 2015
Accepted 7 August 2015
Available online 14 August 2015

Keywords:
Serratia
Entomopathogenic nematodes
Whole-genome sequencing
Annotation

ABSTRACT

We present the annotation of the draft genome sequence of *Serratia* sp. strain TEL (GenBank accession number KP711410). This organism was isolated from entomopathogenic nematode *Oscheius* sp. strain TEL (GenBank accession number KM492926) collected from grassland soil and has a genome size of 5,000,541 bp and 542 subsystems. The genome sequence can be accessed at DDBJ/EMBL/GenBank under the accession number LDEG00000000.

Specifications	
Organism	*Serratia* sp.
Strain	TEL
Sequencer or array type	Sequencer; Illumina MiSeq
Data format	Processed
Experimental factors	Microbial strains
Experimental features	Draft genome sequence of *Serratia* sp. TEL, assembly and annotation
Consent	N/A
Sample source location	Grassland in Suikerbosrand Nature Reserve near Johannesburg in South Africa

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/nuccore/LDEG00000000.

2. Experimental design, materials and methods

Entomopathogenic nematodes (EPNs) are soil dwelling insect-killing microscopic worms which have been studied as potential insect pest biological control agents across the world [1]. The non-feeding infective juvenile (IJ) stage functions as vectors for insect pathogenic symbiotic enterobacteria. After the IJs have infected the insect host the bacteria are regurgitated from the alimentary into the insect's haemolymph. The bacteria release antimicrobials that suppress opportunistic colonialization by microbial saprophytes and lethal insecticidal toxins which kill the insect within 48–72 h after IJ ingress into the haemolymph [2]. EPNs belonging to *Heterorhabditis*, *Steinernema*, and *Oscheius* genera have evolved symbiotic associations with *Photorhabdus*, *Xenorhabdus*, and an insect pathogenic species of *Serratia*, respectively [3,1,4]. Not all species or strains of bacteria in the genus *Serratia* are pathogens of insects [5]. An insect pathogenic strain of *Serratia marcescens* has also been found to be associated with *Caenorhabditis briggsae* and the IJs of this nematode are able to infect larvae of *Galleria mellonella* [6]. Similarly, in other studies, insect pathogenic bacteria belonging to the genus *Serratia* have been found to be symbiotically associated with *Oscheius carolinensis* and this nematode is also capable of acting as a vector for the bacteria by infecting insect hosts [1].

In this paper we discuss *Serratia* sp. TEL (GenBank accession KP711410) and several functional features identified from the annotation results. *Serratia* sp. TEL (Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Serratia) is a Gram-negative, non-spore forming bacterium. This bacterium is symbiotically associated as an insect pathogen with a novel entomopathogenic nematode species *Oscheius* species TEL-2014 (GenBank accession number KM492926).

To isolate the bacterial pathogen, non-feeding infective juveniles were surface sterilized with 0.1% sodium hypochlorite for 4 h and rinsed with autoclaved distilled water. Sterile IJs were carefully suspended in 1 ml of sterile nutrient broth in a microtube and homogenized with a small sterile plastic pestle. The homogenate was streaked onto NBTA and McConkey agar plates and incubated at 25 °C for 24 h in the dark. DNA was extracted from the bright green colonies and pink colonies, pure cultures obtained from the selective media and genomic DNA was extracted using the ZR bacterial DNA miniprep kit (Zymo Research). A polymerase chain reaction was employed to amplify the 16S rRNA gene sequence of the isolated bacterial DNA using EUB968

* Corresponding author.

http://dx.doi.org/10.1016/j.gdata.2015.08.010

forward primer 5′-ACGGGCGGTGTGTC-3′ Tm (°C) = 62 and UNIV1382 reverse primer 5′-AACGCGAAGAACCTTAC-3′ Tm (°C) = 66. The same primers were used for the sequencing of this gene. The sequence obtained was subjected to NCBI BLAST under the default settings for highly similar alignments. The analysis revealed that among all the matches for the 16S rRNA gene sequences, the unknown sequence had a high affinity to a novel *Serratia* species which was then assigned the name *Serratia* sp. strain TEL. Other high scoring bacteria 16S rRNA sequences were downloaded from GenBank to be used for phylogenetic analysis. The evolutionary relatedness of *Serratia* sp. strain TEL with several *Serratia* species was based on the 16S rRNA ITS regions using the Maximum Likelihood method based on the Tamura–Nei model using MEGA6. The bootstrap consensus tree was inferred from 1000 replications and the tree was drawn to scale, with branch lengths measuring the number of substitutions per site (next to the branches) as shown in Fig. 1. Bacteria belonging to the well-known genera *Photorhabdus* and *Xenorhabdus* were used as possible out-groups to root the tree and see if there is any significant relationship they might have with the identified *Serratia* species.

The whole genome was shotgun sequenced on the Illumina Miseq, following sequence quality control using Fastq-mcf toolkit and de novo assembly using SPADES were performed as described in [7]. A draft annotated genome of this organism was constructed using the best-placed reference protein set; GeneMarkS+ annotation method set on the NCBI Prokaryotic Genome Automatic Annotation Pipeline and has been reported in [7]. Furthermore, the functional annotation was carried out by RAST (Rapid Annotation using Subsystem Technology) hosted by Fellowship for Interpretation.

RAST revealed the closest neighbours of *Serratia* sp. TEL as *S. marcescens* Db11 (reference genome id 615.1), followed by *Serratia proteamaculans* 568 (399741.3) and *Serratia plymuthica* A30 (1206776.3).

RAST identified a total of 4618 coding regions and 103 RNAs. Fig. 2 shows that the annotated genome has 119 genes potentially involved in virulence, disease and defence including 88 genes for resistance to antibiotics and toxic compounds such lysozyme inhibitors, multiple antibiotic resistance (MAR) locus, beta lactamase, copper homeostasis genes, streptothricin resistance genes, multidrug resistance efflux pumps, multidrug resistance tripartite systems found in Gram negative bacteria and resistance to chromium compounds. An example of a gene involved in stress response, virulence and defence is 16 kDa heat shock protein A and will be briefly discussed.

The focus gene coding for a 16 kDa heat shock protein A is involved in stress response, virulence and defence was compared with two similar organisms. The graphic is centred on the focus gene, which is red and numbered 1. Sets of genes with similar sequence are grouped with the same number and colour. Genes whose relative position is conserved in at least two other species are functionally coupled and share grey background boxes. The focus gene always points to the right, even if it is located on the minus strand.

It was found that some of the genes involved in virulence, disease and defence as shown in Fig. 3 such as the 16 kDa heat shock protein B and the outer membrane lipoprotein YidQ which were closely associated with genomic regions containing 16 kDa heat shock protein A and were also similarly arranged together in *Serratia* sp. TEL, *Escherichia coli* K12 and *S. enterica* subsp. *enterica* serovar Choleraesuis strain SC-B67. Additionally, the annotated genome has 80 genes involved in motility and chemotaxis in which 24 of these genes encode for bacterial chemotaxis and 56 of them encode for flagellum motility.

Genetic subsystems discussed were highlighted as important because their presence contributed to the insect pathogenicity of the bacterial species for which the nematode acted as the vector for the location and infection of potential insect hosts.

3. Nucleotide sequence accession numbers

This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LDEG00000000.

Conflict of interest

The authors declare that there is no conflict of interest on any work published in this paper.

Acknowledgements

Thanks to Gauteng Department of Agriculture and Rural Development (GDARD) (GDARO 12 TS) for funding the research project. Thanks also for the Innovation Doctoral Scholarship from the NRF National Research Foundation (NRF) with the grant number [SFH1208147793] and the Wits Postgraduate Merit Award (PMA) from the University of the Witwatersrand which supported the first author. Thanks to the Agricultural Research Council (ARC) Biotechnology platform for Illumina technology sequencing services.

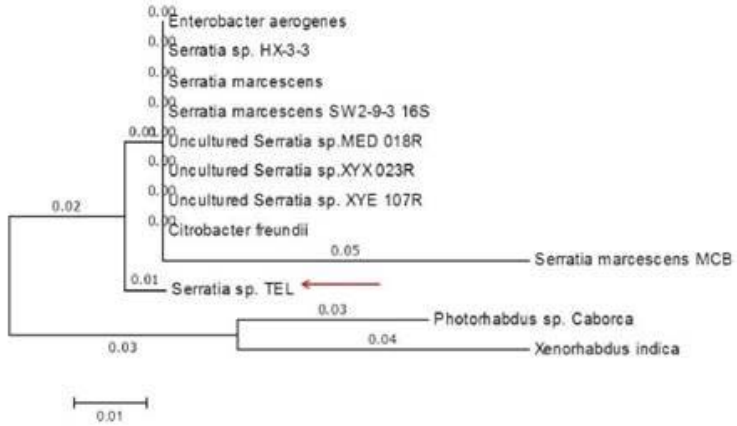


Fig. 1. The evolutionary history of *Serratia* sp. TEL (pointed with a red arrow) inferred by using the Maximum Likelihood method based on the Tamura–Nei model on MEGA6 software. The bootstrap consensus tree inferred from 1000 replications.

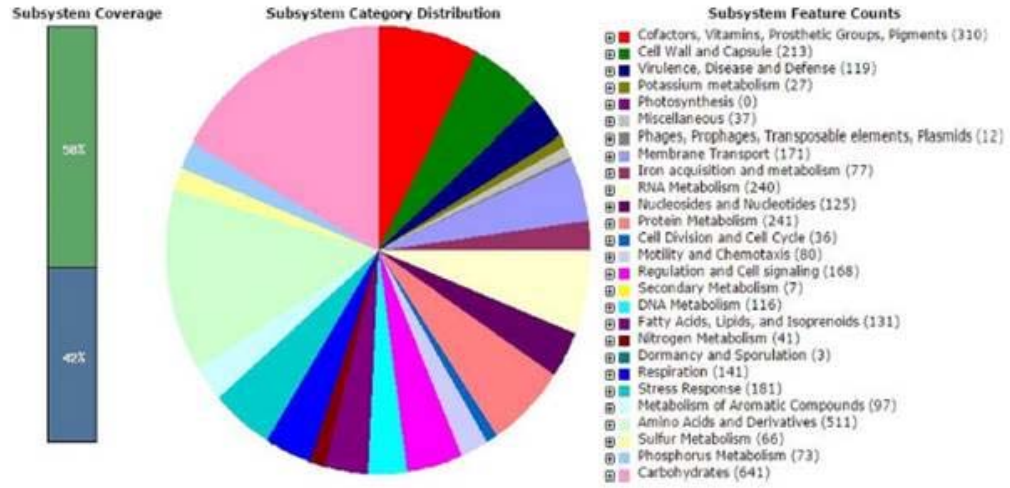


Fig. 2. The subsystem distribution of *Serratia* sp. TEL based on RAST annotation.

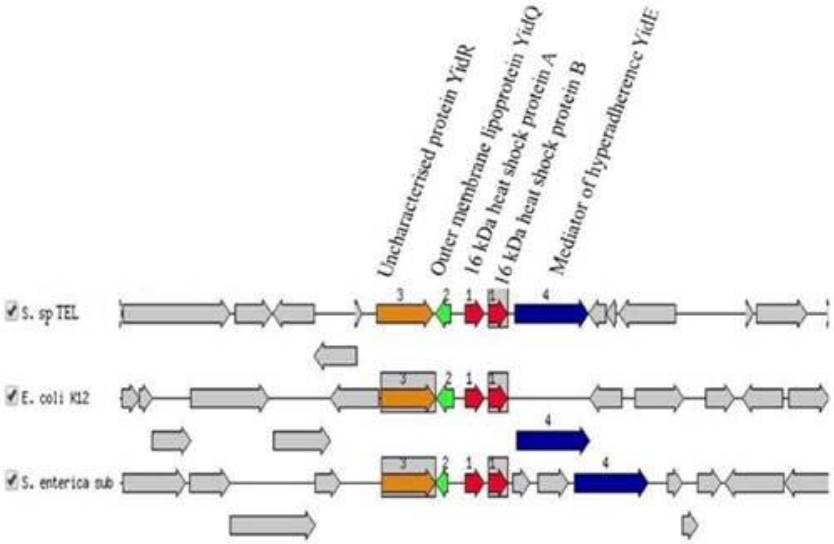


Fig. 3. Comparative analysis of genomic regions containing 16 kDa heat shock protein A found in *Serratia* sp. TEL to *Escherichia coli* K12 and *Salmonella enterica* subsp. *enterica* serovar Choleraesuis strain SC-B67.

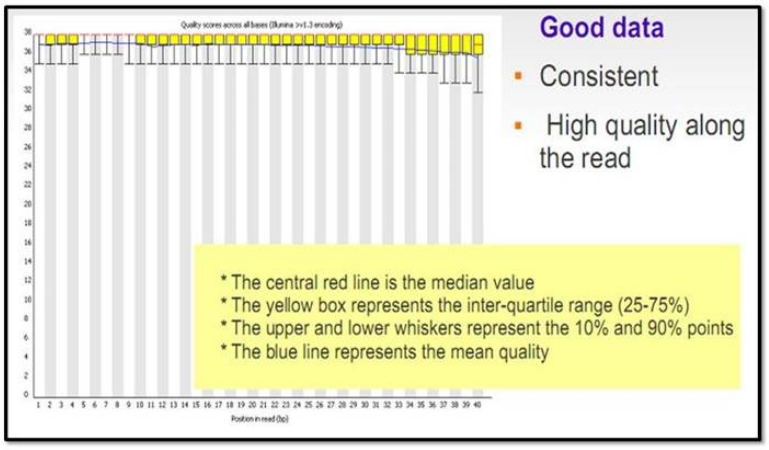
Appendix A. Supplementary data

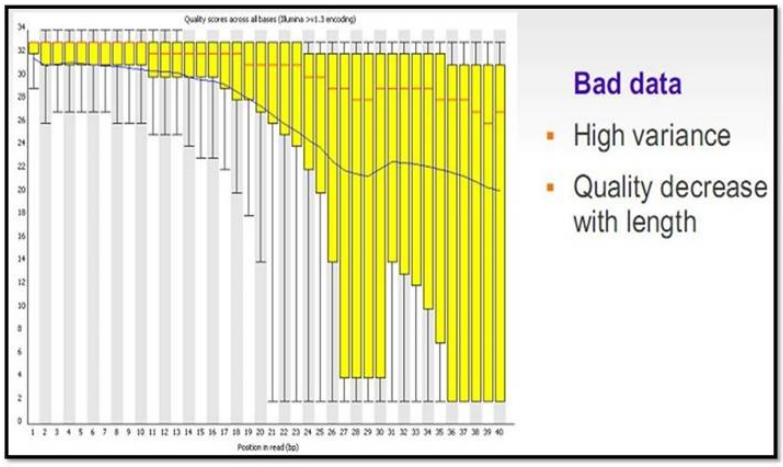
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2015.08.010.

References

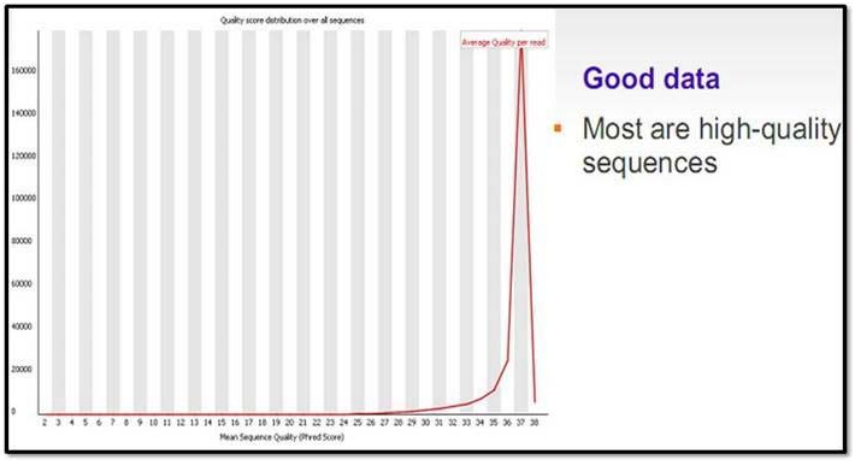
[1] A. Torres-Barragan, A. Suazo, W.G. Buhler, Y.J. Cardoza, Studies on the entomopathogenicity and bacterial associates of the nematode *Oscheius carolinensis*. Biol. Control 59 (2011) 123–129.
[2] A.P. Malan, K.B. Nguyen, M.F. Addison, Entomopathogenic nematodes (Steinernematidae and Heterorhabditidae) from the southwestern parts of South Africa. Afr. Entomol. 12 (2006) 65–69.
[3] R.J. Stuart, M.E. Barbercheck, P.S. Grewal, R.J. Taylor, W.C. Hoy, Morphological and functional dimorphism in *Xenorhabdus* spp., bacteria symbiotically associated with the insect pathogenic nematodes *Neoaplectana* and *Heterorhabditis*. Microbiology 121 (2006) 303–309.
[4] C. Zhang, J. Liu, M. Xu, J. Sun, S. Yang, X. An, G. Gao, M. Lin, R. Lai, Z. He, Y. Wu, K. Zhang, *Heterorhabditidoides chongmingensis* gen. nov., sp. nov. (Rhabditida: Rhabditidae), a novel member of the entomopathogenic nematodes. J. Invertebr. Pathol. 98 (2008) 153–168.
[5] Y. Kwak, A.R. Khan, J. Shin, Genome sequence of *Serratia nematodiphila* DSM 21420T, a symbiotic bacterium from entomopathogenic nematode. J. Biotechnol. 193 (2015) 1–2.
[6] E. Abede, M. Jumba, K. Bonner, V.M. Gray, K. Morris, W.K. Thomas, An entomopathogenic *Caenorhabditis briggsiae*. J. Exp. Biol. 213 (2010) 3223–3229.
[7] T.E. Lephoto, J. Featherston, V.M. Gray, Draft whole-genome sequence of *Serratia* sp. strain TEL, associated with *Oscheius* sp. TEL-2014 (Nematoda: Rhabditidae) isolated from a grassland in South Africa. Genome Announc. 3 (4) (2015) e00747-15 http://dx.doi.org/10.1128/genomeA.00747-15.

Quality- Per base position



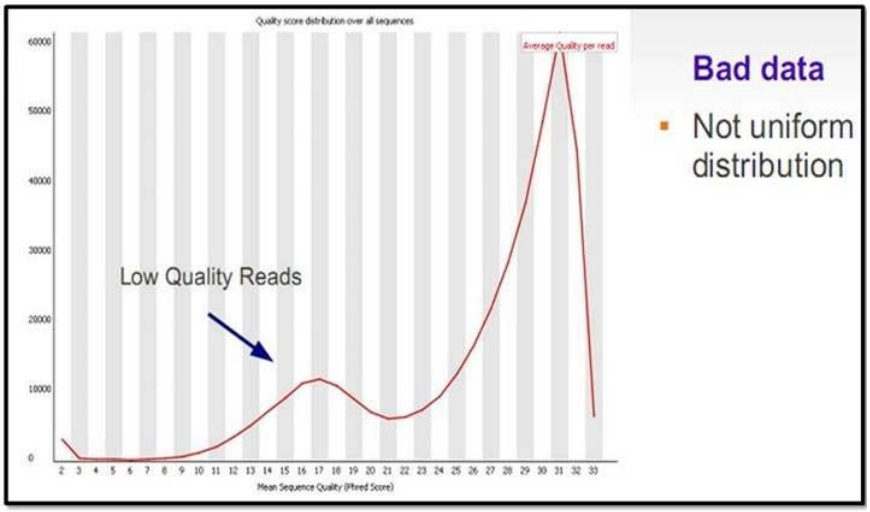


Per Sequence Quality Distribution



Good data

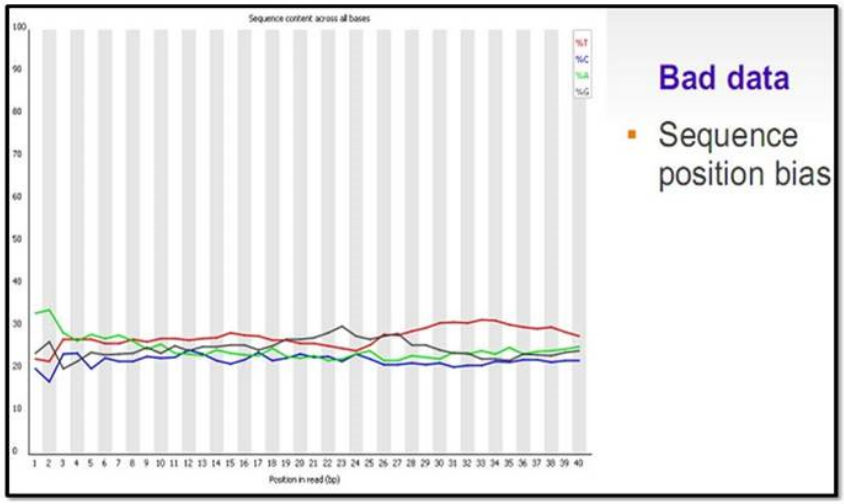
- Most are high-quality sequences

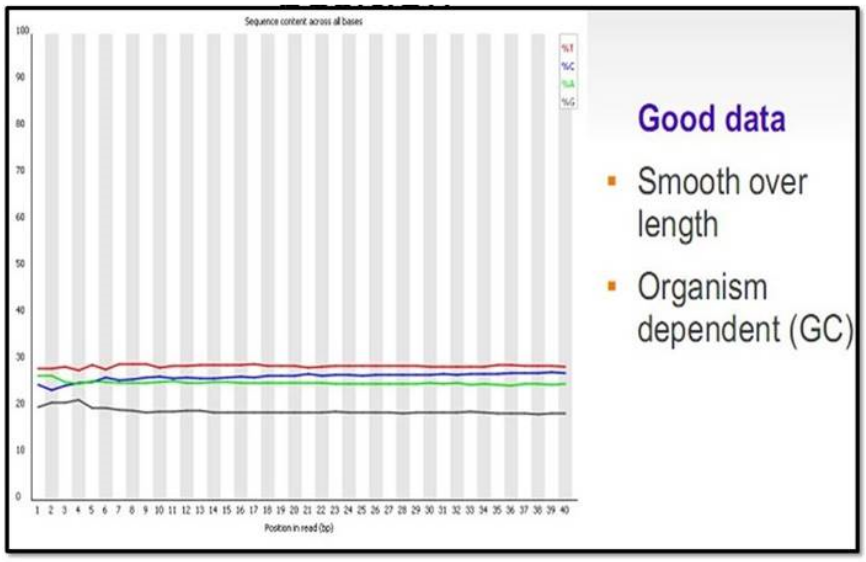


Bad data

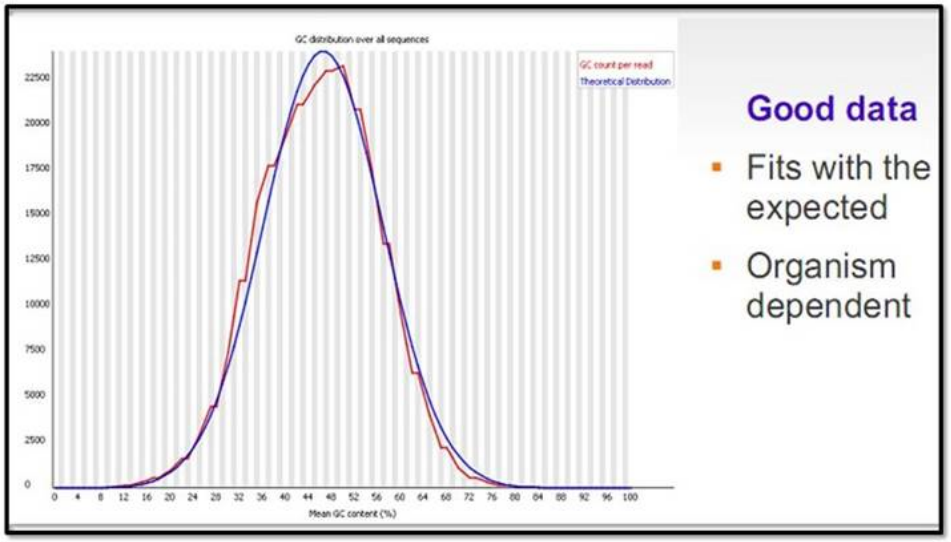
- Not uniform distribution

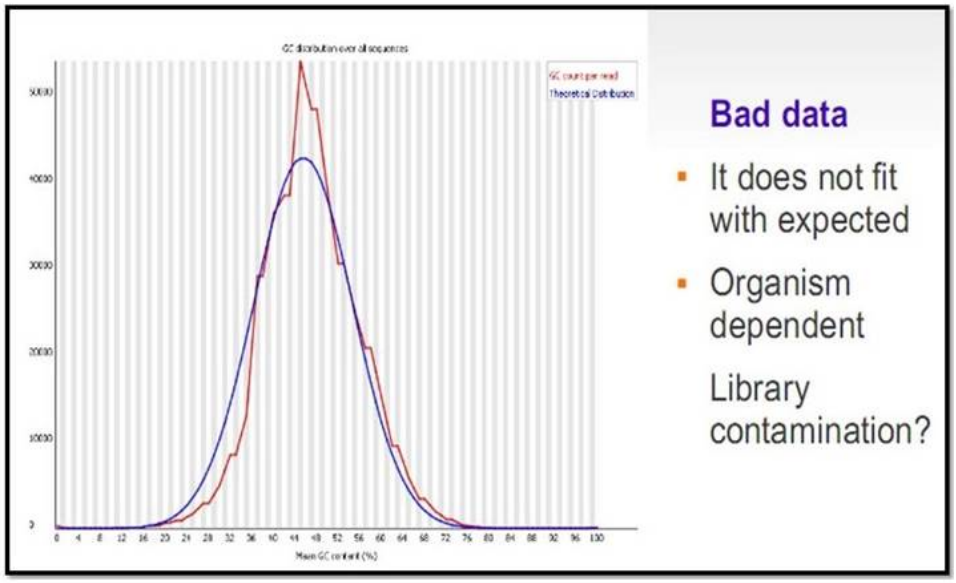
Nucleotide content per position



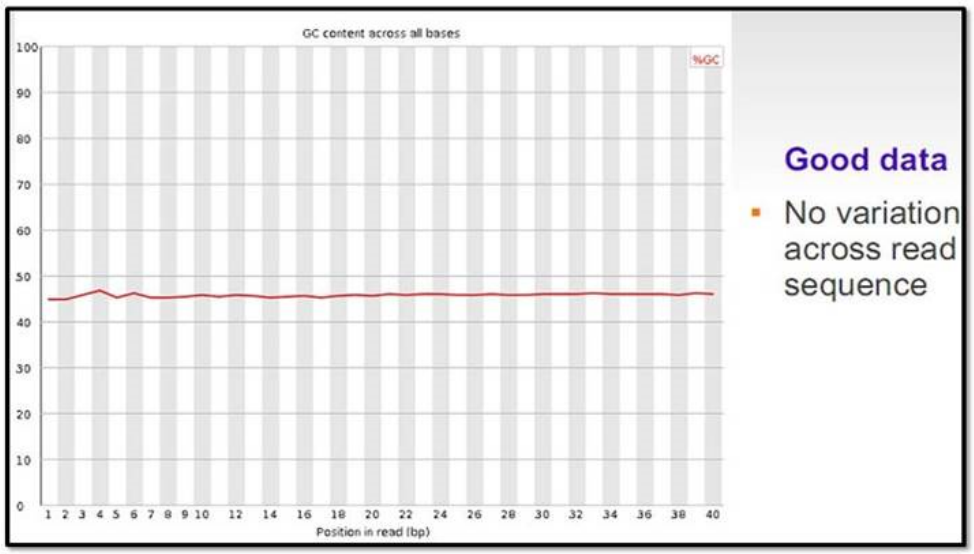


Per sequence GC distribution



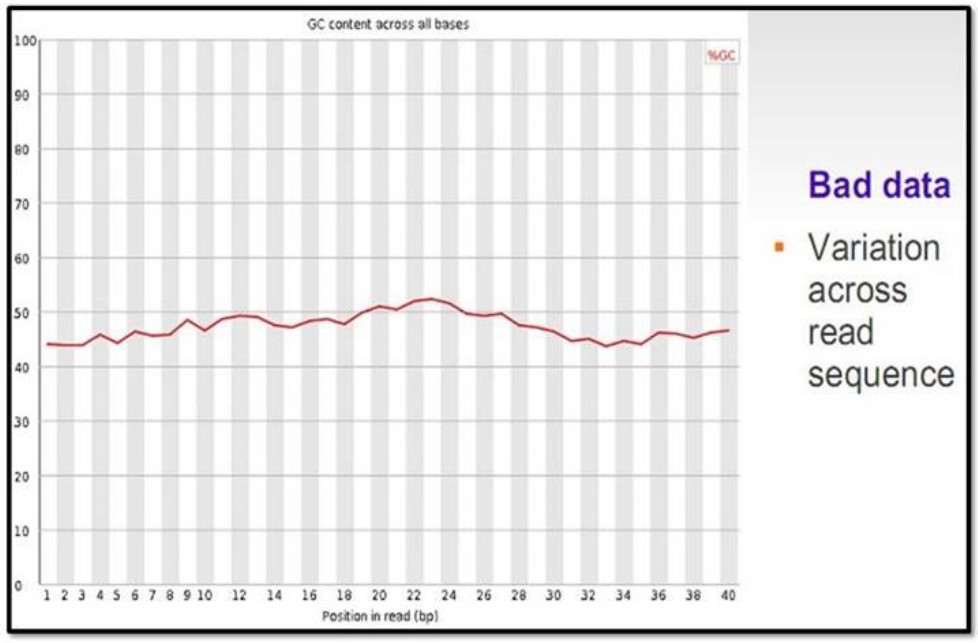


Per base GC distribution



Good data

- No variation across read sequence



Bad data

- Variation across read sequence

Supplementary data

Galleria mellonella infected with *Oscheius sp. TEL-2014*

Sequencing Statistics

Total Reads	Reads Passing Quality Filtering	% Reads Passing Quality Filtering
164,153	149,903	91.3 %

Classification Rate Summary

Taxonomic Level	Reads Classified to Taxonomic Level	% Total Reads Classified to Taxonomic Level
Kingdom	148,429	99.02 %
Phylum	148,274	98.91 %
Class	148,065	98.77 %
Order	147,550	98.43 %
Family	147,406	98.33 %
Genus	146,200	97.53 %
Species	141,391	94.32 %

Classification Results by Taxonomic Level

Tables show the highest 8 taxonomic classifications at each level.
Pie charts show all classifications above 3.5% abundance.

Top Kingdom Classification Results

Classification	Number of Reads	% Total Reads
Bacteria	148,427	99.02 %
Unclassified at Kingdom level	1,474	0.98 %
Viruses	2	0.00 %

Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.

Top Phylum Classification Results

Classification	Number of Reads	% Total Reads
Proteobacteria	147,877	98.65 %
Unclassified at Phylum level	1,629	1.09 %
Cyanobacteria	257	0.17 %
Actinobacteria	71	0.05 %
Firmicutes	43	0.03 %
Bacteroidetes	17	0.01 %
Spirochaetes	4	0.00 %
DNA	2	0.00 %

Total Phylum-level Taxonomic Categories Identified: 11. This table shows the top 8 of 11 classifications.

Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.

Top Class Classification Results

Classification	Number of Reads	% Total Reads
Gammaproteobacteria	147,381	98.32 %
Unclassified at Class level	1,838	1.23 %
Nostocophycideae	233	0.16 %
Alphaproteobacteria	199	0.13 %
Betaproteobacteria	100	0.07 %
Actinobacteria	70	0.05 %
Bacilli	38	0.03 %
Sphingobacteriia	11	0.01 %

Total Class-level Taxonomic Categories Identified: 19. This table shows the top 8 of 19 classifications.

Top Order Classification Results

Classification	Number of Reads	% Total Reads
Enterobacteriales	146,523	97.75 %
Unclassified at Order level	2,353	1.57 %
Stigonematales	225	0.15 %
Alteromonadales	192	0.13 %
Rhizobiales	136	0.09 %
Burkholderiales	93	0.06 %
Actinomycetales	70	0.05 %
Oceanospirillales	52	0.03 %

Total Order-level Taxonomic Categories Identified: 42. This table shows the top 8 of 42 classifications.

Top Family Classification Results

Classification	Number of Reads	% Total Reads
Enterobacteriaceae	146,523	97.75 %
Unclassified at Family level	2,497	1.67 %
Rivulariaceae	225	0.15 %
Alcaligenaceae	86	0.06 %
Brucellaceae	79	0.05 %
Alteromonadaceae	63	0.04 %
Aeromonadaceae	50	0.03 %
Rickettsiaceae	38	0.03 %

Total Family-level Taxonomic Categories Identified: 73. This table shows the top 8 of 73 classifications.

Top Genus Classification Results

Classification	Number of Reads	% Total Reads
Serratia	143,904	96.00 %
Unclassified at Genus level	3,703	2.47 %
Enterobacter	708	0.47 %
Erwinia	432	0.29 %
Calothrix	225	0.15 %
Klebsiella	216	0.14 %
Achromobacter	84	0.06 %
Pseudochrobactrum	65	0.04 %

Total Genus-level Taxonomic Categories Identified: 107. This table shows the top 8 of 107 classifications.

Top Species Classification Results

Classification	Number of Reads	% Total Reads
Serratia entomophila	127,386	84.98 %
Serratia marcescens	12,263	8.18 %
Unclassified at Species level	8,512	5.68 %
Erwinia billingiae	395	0.26 %
Calothrix parietina	225	0.15 %
Enterobacter asburiae	209	0.14 %
Enterobacter amnigenus	173	0.12 %
Klebsiella granulomatis	80	0.05 %

Total Species-level Taxonomic Categories Identified: 96. This table shows the top 8 of 96 classifications.

Galleria mellonella uninfected with *Oscheius sp. TEL-2014*

Sequencing Statistics

Total Reads	Reads Passing Quality Filtering	% Reads Passing Quality Filtering
117,596	108,236	92.0 %

Classification Rate Summary

Taxonomic Level	Reads Classified to Taxonomic Level	% Total Reads Classified to Taxonomic Level
Kingdom	107,853	99.65 %
Phylum	107,621	99.43 %
Class	107,345	99.18 %
Order	107,090	98.94 %
Family	105,602	97.57 %
Genus	104,536	96.58 %
Species	87,107	80.48 %

Classification Results by Taxonomic Level

Tables show the highest 8 taxonomic classifications at each level.
Pie charts show all classifications above 3.5% abundance.

Top Kingdom Classification Results

Classification	Number of Reads	% Total Reads
Bacteria	107,851	99.64 %
Unclassified at Kingdom level	383	0.35 %
Viruses	2	0.00 %

Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.

Top Phylum Classification Results

Classification	Number of Reads	% Total Reads
Proteobacteria	75,083	69.37 %
Firmicutes	29,679	27.42 %
Actinobacteria	2,669	2.47 %
Unclassified at Phylum level	615	0.57 %
Cyanobacteria	143	0.13 %
Tenericutes	23	0.02 %
Bacteroidetes	9	0.01 %
Spirochaetes	4	0.00 %

Total Phylum-level Taxonomic Categories Identified: 14. This table shows the top 8 of 14 classifications.
Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.

Top Class Classification Results

Classification	Number of Reads	% Total Reads
Alphaproteobacteria	44,796	41.39 %
Betaproteobacteria	29,724	27.46 %
Bacilli	29,632	27.38 %
Actinobacteria	2,669	2.47 %
Unclassified at Class level	891	0.82 %
Gammaproteobacteria	317	0.29 %
Nostocophycideae	126	0.12 %
Mollicutes	23	0.02 %

Total Class-level Taxonomic Categories Identified: 22. This table shows the top 8 of 22 classifications.
Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.

Top Order Classification Results

Classification	Number of Reads	% Total Reads
Rhizobiales	44,638	41.24 %
Burkholderiales	29,676	27.42 %
Bacillales	29,455	27.21 %
Actinomycetales	2,668	2.46 %
Unclassified at Order level	1,146	1.06 %
Stigonematales	117	0.11 %
Pasteurellales	98	0.09 %
Enterobacteriales	63	0.06 %

Total Order-level Taxonomic Categories Identified: 46. This table shows the top 8 of 46 classifications.
Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.

Top Family Classification Results

Classification	Number of Reads	% Total Reads
Brucellaceae	41,981	38.79 %
Alcaligenaceae	29,273	27.05 %
Planococcaceae	29,259	27.03 %
Microbacteriaceae	2,642	2.44 %
Unclassified at Family level	2,634	2.43 %
Bartonellaceae	1,137	1.05 %
Burkholderiaceae	198	0.18 %
Rhizobiaceae	131	0.12 %

Total Family-level Taxonomic Categories Identified: 85. This table shows the top 8 of 85 classifications.
Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.

Top Genus Classification Results

Classification	Number of Reads	% Total Reads
Lysinibacillus	29,249	27.02 %
Achromobacter	28,726	26.54 %
Ochrobactrum	24,821	22.93 %
Pseudochrobactrum	16,994	15.70 %
Unclassified at Genus level	3,700	3.42 %
Leucobacter	2,176	2.01 %
Bartonella	1,137	1.05 %
Mycetocola	289	0.27 %

Total Genus-level Taxonomic Categories Identified: 152. This table shows the top 8 of 152 classifications.
Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.

Top Species Classification Results

Classification	Number of Reads	% Total Reads
Unclassified at Species level	21,129	19.52 %
Pseudochrobactrum saccharolyticum	15,218	14.06 %
Ochrobactrum thiophenivorans	13,395	12.38 %
Lysinibacillus parviboronicapiens	9,545	8.82 %
Achromobacter insolitus	9,068	8.38 %
Lysinibacillus fusiformis	8,660	8.00 %
Ochrobactrum pseudogrignonense	8,149	7.53 %
Achromobacter piechaudii	5,689	5.26 %

Total Species-level Taxonomic Categories Identified: 140. This table shows the top 8 of 140 classifications.
Note: The "Other" category in this pie chart is the sum of all classifications with less than 3.5% abundance.