IN SILICO AND MOLECULAR ANALYSIS OF XptA PROTEIN TOXIN GENES OF *Xenorhabdus* sp. AND EFFICACY OF THE BACTERIUM ON *Sitophilus zeamais* AND *Prostephanus truncatus.*

By

NG'ANG'A PETER NJENGA

I576/60295/2013

BSc. Biology – University of Nairobi

Thesis submitted in partial fulfilment of the requirements for the award of the degree of Master of Science in Genetics of the University of Nairobi.

September 2015

DECLARATION

Candidate

This is my original work and has not been submitted for a degree in any other university or any other award.

Signed: ……. Date: ……

Reg. No: I576/60295/2013

Ng'ang'a Peter Njenga

Supervisor

We confirm that the candidate under our supervision carried out the work reported in this thesis

Dr. Amugune Nelson Onzere

Signature: ……. Date: ……..

School of Biological Sciences, University of Nairobi

Dr. Ong'amo George

Signature: ……. Date: ……..

School of Biological Sciences, University of Nairobi

Dr. Masiga Daniel

Signature: ……. Date: ……..

Molecular Biology and Bioinformatics Department,

Duduville campus, ICIPE

DEDICATION

I dedicate this work to my family, friends, supervisors and all those who have played a role in building and moulding me into the person I am today and the scientist that I hope to become in the near future.

ACKNOWLEDGEMENTS

I thank the Almighty God for the gift of life and the opportunity to work towards my dream. I thank my parents for their love and support. I thank the University of Nairobi for my degree, NACOSTI for funding the project, ICIPE for lab space, resources, as well as guidance through the capacity building unit, KALRO for material support and World Federation of Scientists (WFS) for the scholarship.

I thank my University supervisors Dr. Nelson Amugune and Dr. George Ong'amo for believing in me, for the support and mentorship. I thank Dr. Daniel Masiga for his kindness, guidance and the opportunity to work at ICIPE. I have learnt a lot and it has been a worthwhile experience.

I thank all my friends particularly, Lorine Nanjala Nyongesa, Ryan Musumba Awori, Fridah Wambui Kariuki, Rose Mbeya, Waruguru Wanjau, Nyotu Gitau and Murugi Kagotho for their immeasurable kindness, support, and for being a pillar to lean on.

I thank Dr. Vincent Owino, Dr. Bernard Kulohoma, Simon Ngao, Rosaline Macharia, Mr. Mathew Miti, Joyce Njuguna, as well as all colleagues and friends at ICIPE for their guidance, assistance and best wishes. Additionally, I thank all my classmates at the University.

I may not be able to mention everyone but I appreciate all support, assistance and best wishes, may God's grace be forever with all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS

BLAST	Basic local alignment search tool
Bp	Base pairs
Bt	*Bacillus thuringiensis*
CYMMIT	International Maize and Wheat Improvement Centre
°C	Degrees Celsius
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ICIPE	International Centre for Insect Physiology and Ecology
g	Gravitational force
Kb	Kilobases
KDa	Kilodaltons
LB	Luria Bertani
mM	Millimolar
M	Molar
Mb	Megabases
MWCO	Molecular weight cut off
NBTA	Nutrient broth supplemented with triphenyltetrazolium salt and bromothymol blue
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
RH	Relative humidity
SDS	Sodium dodecyl sulphate
TAE	Tris acetate EDTA
TEMED	Tetramethylethylenediamine
µL	Microliter
µM	Micromolar

ABSTRACT

Bacteria of the genus *Xenorhabdus* are entomopathogens that produce insecticidal protein toxins against a wide range of insects. The main proteins involved are the Xenorhabdus protein toxins (Xpts), categorized as class A, B and C. They work best as a complex, though individually, XptA has been found to be more potent against some insect pests such as *Pieris brassicae* and *Heliothis virescens*. The maize weevil (*Sitophilus zeamais)* and the larger grain borer (*Prostephanus truncatus*) are storage insect pests that cause the greatest postharvest losses of maize in Kenya, generally estimated to range between 20-30% annually. Many of the control strategies used against these pests have significant drawbacks. In particular, chemical insecticides are costly and may pose health and environmental hazards. There is thus a need to find an efficient and sustainable alternative. This project sought to identify an isolate of *Xenorhabdus* sp. that is effective against the two storage pests, *Sitophilus zeamais* and *Prostephanus truncatus*. It also sought to isolate and analyse the Xpt proteins that are involved in insecticidal activity. The bacterium was characterized using morphological features typical to the genus, as well as, molecularly via a 16S rDNA sequence homology search. Bacterial cells were lysed to isolate the proteins and the crude lysate purified using a size exclusion column before separation on a PAGE gel. The target bands were identified by comparison to a protein ladder. The insecticidal activity of *Xenorhabdus* sp. against the target pests was then tested by incorporation of live bacteria into an artificial diet before feeding the insects with it. XptA sequences as well as those of their homologs were then retrieved from databases and analysed using bioinformatics techniques and tools. The bacterium was identified as *Xenorhabdus griffinae*. Purified protein bands were found in the expected size range indicating that they were XptA. Whole bacterial cells caused mortality of *S. zeamais* $(\chi_2^2 = 78.32; p < 0.05)$ and reduced consumption of maize pellets by *P. truncatus* $(\chi_4^2 = 117.87; p < 0.05)$. Through multiple sequence alignment, XptA, TcdA and SepA genes were found to be homologous. Similarly, through superimposition of protein tertiary structures on MATRAS and Pymol, XptA, TcdA and SepA were found to be closely related, thus indicating a common mode of action. XptA were found to be homologous to TcdA2, SepA and TccA proteins that have been reported to show toxicity to mammals and plants. All the proteins analysed were found to contain similar conserved domains among which was VRP1, a domain originally found in the virulence plasmid protein SpvA of *Salmonella* spp.. Seven motifs were identified in the VRP1 domain. Of these, motifs number three and four, as well as their inter-motif sequences, were highly conserved across the board while the other motifs were absent in the XptA, TcdA and SepA sequences. A homology search on the HHpred server using both the VRP1 and full class A amino acid sequences, returned TcdA1 (PDB ID: 1VW1A) Tc toxin crystal structure from *P. luminescens* as the PDB top hit. The *Xenorhabdus* sp. used in this study showed potential for use in control of these storage insect pests. Furthermore, through bioinformatic analysis, the XptA proteins also showed a potential for use in pest control. Therefore, overall, *Xenorhabdus* sp. and their class A proteins, could be useful in the development of a cleaner, greener and sustainable crop pest control strategy.

CHAPTER ONE

1.0. INTRODUCTION

Entomopathogenic bacteria, *Xenorhabdus*, belonging to the family enterobacteriaceae, produce insecticidal protein toxins against a wide array of insects (Morgan *et al.,* 2001; Sergeant *et al.,* 2003). There are four active insecticidal proteins from *Xenorhabdus* spp. that have been studied as potential biocontrol agents, these are: - Xpt A1, A2, B1 and C1 categorized as class A, B and C proteins respectively (Morgan *et al.,* 2001; Sergeant *et al.,* 2003). The XptA toxin itself forms a complex of four XptA proteins which then combines with B and C to complete the active insecticidal complex (Lee *et al*., 2007; Sheets *et al*., 2011).

Xenorhabdus live in symbiosis with entomopathogenic nematodes (EPNs) of the genus *Steinernema* which act as vectors through which they infect target insect hosts (Akhurst & Boemare, 1988; Gauler & Boemare, 2002). In this symbiotic relationship, the bacterium kills the host and produces enzymes that help to digest the cadaver, thus providing food and a breeding environment for the nematode (Akhurst, 1982; 1983; Arkhust and Boemare, 1988). *Steinernema* nematodes of various species have been found in Kenya (Waturu *et al.,* 1997; Mwaniki *et al.,* 2008) some of which have been shown to harbour bacteria (Waturu *et al.,* 1997; Tailliez *et al.,* 2006).

Xenorhabdus bacteria do not survive well without their nematode hosts and this has limited their use as biopesticides (Morgan *et al.,* 2001; Sergeant *et al.,* 2003). The toxicity detected in these bacteria can be exploited by genetically transforming other bacteria, microorganisms, or plants with genes coding for the toxins (Morgan *et al.,* 2001; Joo Lee *et al.,* 2004). Alternatively, protein toxins could be harvested, purified, formulated and used

(either as a sprayable compound or powder) as has been done with *Bacillus thuringiensis* (Bt) toxins (Lee *et al*., 2007). A toxin homologous to XptA, TcdA1 from *Photorhabdus luminescens*, another entomopathogenic bacterium closely related to *Xenorhabdus*, has already been expressed in a plant, *Arabidopsis thaliana*, for insect control (Liu *et al*., 2003).

Over the years, there has been considerable progress in cloning and expression of several toxin genes from *Photorhabdus* and *Xenorhabdus* spp. Resulting products have been tested against various hosts where they have proved effective (Morgan *et al*., 2001; Hinchliffe *et al*., 2010). Several insect species from a variety of orders including Lepidoptera, Coleoptera, Hymenoptera and Dictyoptera are reported to be susceptible to *Photorhabdus* and *Xenorhabdus* (Bowen & Ensign, 1998; Morgan *et al*., 2001). However, insecticidal activity of bacteria of the genus *Xenorhabdus*, and/or *Photorhabdus*, against dominant storage pests of maize, *Sitophilus zeamais* and *Prostephanus truncatus*, is yet to be reported. This is despite the potential utility, based on the knowledge that bacteria in these genera produce toxins.

This project explored the possibility of using the *Xenorhabdus* entomopathogenic bacterium, particularly the putative XptA toxin, as a novel means of control of *S. zeamais* and *P. truncatus*. It also sought to further elucidate the mechanism of action of the XptA toxin through structure analysis and comparison to other established class A toxins.

1.1. Statement of the problem

Xenorhabdus have been considered the new frontier in biological control as they produce novel insecticidal toxins. These toxins, particularly the Xpt, have been tested against a wide array of insect species that have been shown to be susceptible. However, the potential for

use of *Xenorhabdus* in biological control of other economically important insect pests such as *Prostephanus truncatus* and *Sitophilus zeamais* is yet to be determined.

Maize (*Zea mays* L) is the most important cereal crop in sub-Saharan Africa and an important staple food (Tefera *et al*., 2010). In Kenya, it plays an integral role in national food security (Short *et al*., 2012). However, only a small proportion of the produce is available for consumption, a significant amount is lost to postharvest pests. Postharvest losses associated with storage insect pests are estimated to range between 20-30% and are a serious concern among small scale farmers (Tefera *et al.*, 2011).

1.2. Justification

Due to the economic importance of these pests, several control strategies have been initiated with limited success (Aktar *et al*., 2009; Clair *et al*., 2012). Some of such strategies, particularly the insecticides, have caused serious health concerns, not only to the end consumers, but also to the environment. The cost of control has also become a burden for the farmers, especially the small holder farmers.

In this regard, there is need for a 'greener', sustainable and cost effective method of plant pest control that is able to augment the already existing ones. *Xenorhabdus* bacteria and their XptA toxins present such an opportunity. As such, it is paramount to study their insecticidal toxins of Xenorhabdus to better understand the mode of action and subsequently, their potential for use. It is also important to assess the potential of using the bacterium itself as a biocontrol agent against these pests. Effective use of these novel potential biopesticides could usher in a new regime in the management of storage pests of maize.

1.3. Broad objective

To analyse putative XptA insecticidal genes of *Xenorhabdus* sp. isolated from a Steinernema nematode found in Kenya and test the efficacy of the bacterium on two storage insect pests, *S. zeamais* and *P. truncatus*.

1.4. Specific objectives

1. To characterize the XptA genes from *Xenorhabdus* sp. obtained from a *Steinernema* entomopathogenic nematode using gene-specific and degenerate primers.
2. To characterize the XptA proteins from *Xenorhabdus* sp. using native-polyacrylamide gel electrophoresis and size exclusion chromatography.
3. To analyse the XptA genes and proteins using bioinformatics techniques.
4. To determine the efficacy of *Xenorhabdus* sp. against the maize storage pests *S. zeamais* and *P. truncatus*.

1.5. Research hypothesis

The *Xenorhabdus* bacterium harbours genes that code for toxins that can be used to control the maize weevil (*Sitophilus zeamais*) and the larger grain borer (*Prostephanus truncatus*).

CHAPTER TWO

2.0. LITERATURE REVIEW

2.1. Entomopathogenic bacteria as sources of insect toxins

Xenorhabdus and *Photorhabdus* are entomopathogenic bacteria that produce insecticidal toxins. In this regard, they are considered possible biocontrol agents. (Akhurst, 1980; Bowen & Ensign, 1998; Guo *et al.*, 1999; Morgan *et al.*, 2001; Sergeant *et al.*, 2003; Lee *et al.*, 2007; Sheets *et al.*, 2011). The two genera are members of the family Enterobacteriaceae encompassing the intestinal bacterial symbionts living as commensals in entomopathogenic nematodes (EPNs) of the genera *Steinernema* and *Heterorhabditis*, respectively (Boemare & Akhurst, 2006; Herbert & Goodrich-Blair, 2007).

The bacterial-nematode relationship is very specific, only one bacterial species inhabits a nematode (Akhurst, 1982; 1983; Boemare & Akhurst, 2006). A *Xenorhabdus* isolate has not been recovered from *Heterorhabditis* or a *Photorhabdus* from *Steinernema* (Boemare & Akhurst, 2006). EPNs have been found in soils worldwide, except in Antarctica (Kawaka *et al*., 2014). Different species of EPNs have been found in Kenya (Waturu *et al*., 1997; Mwaniki *et al.*, 2008), the majority being *Steinernema* spp. (Mwaniki *et al*., 2008; Kawaka *et al.*, 2011). Although some bacteria have been isolated from nematodes found in Kenya (Tailliez *et al*., 2006), there is limited knowledge on the bacterial symbionts that these Kenyan isolates harbour.

These entomopathogenic bacteria are rarely found exclusive of their nematode hosts, they do not survive for long in the external environment (Morgan *et al.*, 2001; Herbert & Goodrich-Blair, 2007). Most of them are pathogenic to insects when injected into the

haemocoele and eventually lead to mortality (Akhurst *et al.*, 1993; Morgan *et al.*, 2001; Boemare & Akhurst, 2006; Herbert & Goodrich-Blair, 2007). In this symbiotic relationship, the nematode delivers the bacteria into the insect host, the insect dies, and both the nematode and bacteria benefit from the resulting insect cadaver (Akhurst, 1982; 1983; Arkhust & Boemare, 1988).

Both *Xenorhabdus* and *Photorhabdus* produce similar insecticidal protein toxins that are effective against a wide range of insects (Forst & Nealson, 1996; Ffrench-Constant & Bowen, 2000; Morgan *et al.*, 2001; Liu *et al.*, 2003). Proteins produced by *Xenorhabdus* spp. are referred to as Xenorhabdus protein toxins (Xpts) (Morgan *et al.*, 2001; Sergeant *et al*., 2003) while those from *Photorhabdus* spp. are referred to as Tc proteins (Bowen and Ensign, 1998; Sheets *et al.*, 2011). There are four active insecticidal proteins from *Xenorhabdus* spp. that have been studied as potential biocontrol agents, these are: - Xpt A1, A2, B1 and C1 categorized as class A, B and C proteins respectively.

Class A proteins have been reported as being toxic on their own. However, they produce a greater effect when in complex with the class B and C proteins (Morgan *et al*., 2001; Lee *et al*., 2007; Sheets *et al.*, 2011). Xpts show both oral toxicity and toxicity when injected directly into the haemolymph (Morgan *et al.*, 2001; Sergeant *et al.*, 2003; 2006; Lee *et al.*, 2007; Hinchliffe *et al.*, 2010; Sheets *et al.*, 2011). Genes similar to the above insecticidal toxin complexes have been found in the bacteria *Pseudomonas syringae* pv. *tomato*, *Fibrobacter succinogenes*, *Treponema denticola*, *Yersinia enterocolitica*, *Yersinia pestis* and *Yersinia pseudotuberculosis* (Lee *et al.*, 2007) as well as *Serratia entomophila* (Sergeant *et al.*, 2003).

2.1.1. General characteristics of *Xenorhabdus* spp.

Xenorhabdus spp. are asporogenous, rod-shaped (0.3–2 μm by 2–10 μm) and occasionally have filaments, which are 15– 50 μm in length. Cells move by means of peritrichous flagella, and swarming may occur on 0.6–1.2% agar. These bacteria are facultative anaerobes, with both respiratory and fermentative types of metabolism. Optimum temperature is usually 28°C or less; a few strains grow at 40°C (Boemare & Akhurst, 2006). The nematode-bacterium association is highly toxic to many insect species, however, in most cases, the bacteria alone are highly virulent when they enter the insect haemocoele (Forst & Nealson, 1996; Herbert & Goodrich-Blair, 2007).

2.1.2. Habitat and life-cycle of *Xenorhabdus* spp.

The *Xenorhabdus* bacteria live in two different habitats during their life cycle: first, the bacteria occur naturally in the intestinal vesicle of non-feeding infective stage entomopathogenic nematodes of the family Steinernematidae (Herbert & Goodrich-Blair, 2007), and second, they are inoculated into and multiply in the body cavity of insects, thus creating a pure monoxenic culture (Boemare & Akhurst, 2006). The infective juveniles search (cruiser species) or wait (ambusher species) for an insect prey (step 1). When the target insect is found, they penetrate by natural openings.

During penetration, the nematodes exsheath the second stage cuticle, which was retained during their period in the soil, before entering into the body cavity of the insects (step 2) where they release the bacteria which proceed to suppress the insect immune system through immunodepressive factors (Step 3) (Boemare & Akhurst, 2006). This immunodepressive factor is presumably a protease that facilitates release of the nematode's bacterial symbionts (step 4) (Boemare & Akhurst, 2006). The bacteria multiply, and at the

final stage, the insect dies due to the septicaemia (step 5). Depending on the insect, nematode and bacterial symbionts, the pathogenic process can be the result of the action of one partner or both. Sometimes a toxemia induced by the symbiont precedes the resulting septicaemia.

At the end of the bacterial multiplication, production of a large variety of antimicrobial compounds prevents microbial contamination, mainly from the insect intestinal microflora (Boemare & Akhurst, 2006). The nematodes feed on the food supplies provided by the bacterial biomass and the metabolized insect tissues (step 6). They then proceed to reproduce in the insect cadaver (step 7) for one, two or three generations. Thus, the bacteria create suitable conditions for the development of their nematode host in the insect cadaver. At the end of the parasitism, the dauer larvae of the nematodes take up the bacteria in to their gut (step 8) before leaving the insect (step 9) and thus maintain the continuity of the symbiosis through the generations (Boemare & Akhurst, 2006).

2.1.3. Phase variation of *Xenorhabdus* spp.

Xenorhabdus spp. exhibit a phenomenon known as phase variation. They exhibit two phases during growth, phase I and II. Phase variation occurs during the stationary period of the growth cycle (Akhurst, 1980) and is a highly significant feature of both *Photorhabdus* and *Xenorhabdus*. Variants can be detected easily by two major properties: dye adsorption and antibiotic production. Only phase I of the symbionts has been detected in nature, but under *in vitro* conditions, a variable proportion undergoes profound change affecting colony and cell morphologies, motility, secondary metabolites, endo- and exo-enzymes (including respiratory enzymes) (Akhurst, 1980; Boemare & Akhurst, 2006).

The timing and extent of the phase change is, largely, unpredictable. However, the rate of change from phase I to phase II is generally greater than the reverse (Boemare & Akhurst, 2006). For every character that can be quantified (e.g. antibiotic production), it is clear that the difference between phases is a matter of magnitude, not presence or absence. It is highly probable that this holds true for all phase-related characters (Volgyi *et al*., 1998; Boemare & Akhurst, 2006).

Phase I colonies are mucoid and stick to the loop when streaked on plates, produce antibiotic molecules (Akhurst, 1982), adsorb dyes when incorporated into agar (e.g., the neutral red in MacConkey agar). Phase I *X. nematophila* appear dark blue on nutrient broth supplemented with bromothymol blue and triphenyltetrazolium chloride (NBTA) (Akhurst, 1980). Phase II appears spontaneously during stationary growth period from *in vitro* culture and during nematode rearing on artificial diets. Phase II colonies are not mucoid and do not produce antibiotics (Boemare & Akhurst, 2006). They also do not adsorb dye but reduce the 2,3,5-triphenyltetrazolium chloride (TTC) to formazoan. Because of this they appear red on NBTA (Akhurst, 1980; Boemare & Akhurst, 2006).

Phase I variants provide and protect essential nutrients for the nematodes by killing and metabolizing the insect host and producing antimicrobial agents. Phase II variants are less effective in providing growth conditions for the nematodes, although they may kill the insect host and are also capable of colonizing the dauer vesicle of *Steinernema* or the anterior part of the intestine of *Heterorhabditis*, (Akhurst, 1980; Akhurst, 1982; Boemare & Akhurst, 2006). However, phase II have never been found associated with naturally occurring nematodes, in fact, some *Photorhabdus* phase II variants may be deleterious to their original nematode (Boemare & Akhurst, 2006).

2.1.4. Swarming motility in Xenorhabdus

Givaudan *et al.* (1995) demonstrated that *Xenorhabdus* phase I variants displayed a swarming motility when they were grown on a suitable solid medium (0.6–1.2% agar). Unlike most phase I variants of different *Xenorhabdus* spp., phase II variants were unable to undergo cycles of rapid and coordinated population migration (swarming and even swimming) over the surface of semisolid agar, particularly those of *X. nematophila*. Their optical and electron microscopic observations showed that the non-motile phase II cells of *X. nematophila* F1 lacked flagella.

2.1.*5*. Production of secondary metabolites

Just like *Photorhabdus*, phase I variants of *Xenorhabdus* produce a variety of secondary metabolites, some of which have antimicrobial properties (Akhurst, 1982). Four chemical groups have been characterized: indole derivatives, xenorhabdins, xenorxides (oxidized xenorhabdins) and xenocoumacins. *Xenorhabdus* antibiotics have been reported to play an important role by preventing microbial contamination in the insect carcass during the development of nematodes (Park *et al*., 2009). The production of antimicrobials provides at least a partial explanation for the "natural monoxeny" during parasitism (Boemare & Akhurst, 2006).

Moreover, when grown in peptone broth, even in the absence of the nematodes, the bacteria produce a range of protein toxins that are lethal to the insect host when they are fed to or injected into the haemolymph of many insect species, such as *Galleria mellonella*, *Manduca sexta* and *Helicoverpa armigera* (Bowen & Ensign, 1998; Ffrench-Constant & Bowen, 2000;Yang *et al.*, 2012).

2.1.6. Phylogeny of *Xenorhabdus* spp.

Analyses of 16S rDNA sequences showed that *Xenorhabdus* is most closely related to *Photorhabdus* followed by *Proteus vulgaris* and *Arsenophonus nasoniae*) (Brunel *et al*., 1997*;* Szállás *et al*., 1997; Saux *et al*., 1999; Boemare & Akhurst, 2006). However, *Xenorhabdus* can be distinguished from its nearest phylogenetic neighbour, *Photorhabdus*, by the sequence TTCG at positions 208–211 (*E. coli* numbering) of the 16S rDNA. *Photorhabdus* has a longer version TGAAAG (Szállás *et al*., 1997; Saux, *et al*., 1999; Boemare & Akhurst, 2006).

2.1.7. Pathogenicity of *Xenorhabdus* spp.

No effects of the bacterium on vertebrates have been demonstrated (Boemare & Akhurst, 2006). *Xenorhabdus* is an insect pathogen only when delivered into the insect haemocoele, either by their nematode symbiont or by injection; they are not pathogenic when applied topically, that is, on the insect cuticle. Most are highly pathogenic for larvae of the greater wax moth, *Galleria mellonella*, with an LD_{50} of less than 20 cells (Boemare & Akhurst, 2006). However, a concentration of 4×10^7 cells mL^{-1} was shown to be effective against several insect larvae(Mahar *et al*., 2004).

Xenorhabdus poinarii has very little pathogenicity for *G. mellonella*, (LD50 = 5,000 cells) when injected alone, however, it is highly pathogenic when co-injected with axenic *Steinernema glaseri*, its natural host (Boemare & Akhurst, 2006). Axenic *S. scapterisci* and its symbiont *Xenorhabdus* strain UY61 alone are also not pathogenic to *G. mellonella*. The combination of both partners re-established the pathogenicity of the complex towards *G. mellonella* (Boemare & Akhurst, 2006). Pathogenicity of *Xenorhabdus* varies between insects, with *X. nematophila* having an LC50 of about 500 for *Hyalophora cecropia*

caterpillars, and no effect on maggots of the genus Chironomus (Gotz *et al*., 1981). The use of a less susceptible host (e.g., *Manduca sexta*) has enabled detection of differences in pathogenicity between the two phases of a strain (Boemare & Akhurst, 2006).

2.1.8. Insecticidal protein toxins

Two insecticidal toxin complexes from *Xenorhabdus* spp. have been described which show activity against a broad-range of insect pests including *Pieris brassicae* (Cabbage butterfly), *Plutella xylostella* (diamondback moth), *Heliothis virescens* (Tobacco budworm) and *Heliothis zea* (corn earworm) (Morgan *et al.*, 2001; Lee *et al.*, 2007). Each complex involves three proteins (their size from primary sequence is indicated in parentheses): XptA1 (287 kDa), XptB1 (110 kDa), XptC1 (158 kDa); and XptA2 (284 kDa) XptB1, XptC1 (Lee *et al.,* 2007) combined in a respective 4:1:1 stoichiometry (Sheets *et al.,* 2011). These toxin complexes are formed in the same manner in *Photorhabdus* spp., in fact, class B and C proteins from *Photorhabdus* (TcdB2 and TccC3) can substitute for the B and C proteins from *Xenorhabdus* to form an active hybrid toxin complex that has greater insecticidal activity than the native toxin complex (Sheets *et al.*, 2011).

The XptA proteins have been found to control host range specificity, XptA1 directs insecticidal activity towards *P. brassicae* (Morgan *et al.,* 2001; Sergeant *et al.*, 2003) while XptA2 directs insecticidal activity towards *H. virescens* (Sheets *et al.*, 2011) and are believed to be the starting point for the assembly of the active complexes. Sergent *et al.* (2003) expressed the four Xpt genes individually in *E. coli*: XptA1, XptA2, XptB1, and XptC1. They also determined the combinations of these genes needed for activity against different insect species. Most of the XptA1 protein expressed in *E. coli* (pLEX-xptA1) by Lee *et al*. (2007) was found to be present in inclusion bodies. However, the Xpt and Tc

proteins are excreted by their native bacteria as soluble proteins (Sheets *et al.,* 2011; Gatsogiannis *et al.*, 2013).

Biophysical characterization of XptA1 suggested a mechanism of action whereby, XptA1 first binds to the cell membrane forming a structure with a central cavity, it then complexes with its partners XptB1 and XptC1 to produce the full insecticidal toxin (Lee *et al.*, 2007; Sheets *et al.*, 2011). Each of the four symmetry-related subunits of XptA1 has three well-defined domains and a longitudinal twist with one end narrower than the other. One third of the residues of XptA1 are α-helical and it is suggested the subunits associate partly via an α-helical coiled-coil interaction. XptA1 itself shows the same secondary structure at neutral pH and in an alkaline environment up to pH 10.5. This pH tolerance indicates that the folded XptA1 can pass through the midgut of lepidopteran insects susceptible to the toxin complex. This implies therefore that its folded structure is important for its biological activity (Lee *et al.*, 2007).

XptA1 is a 1.15 MDa tetramer (287KDa per protomer) with a cage-like structure that can fold, assemble into a quaternary structure, and then interact with both target vesicles and cells without any requirement for the presence of the other two toxin components (Lee *et al*., 2007). Sheets *et al.* (2011) suggested the same structure and showed similar activity in XptA2 which bound to midgut membranes from *Helicoverpa zea* larvae. Similarly, Tc proteins from *Photorhabdus* spp. and their homologs in other bacteria, are excreted as soluble proteins which bind to the cell surface, are endocytosed and perforate the host endosomal membrane by forming channels that translocate toxic enzymes into the host, this damages and ultimately kills the cells (Sheets *et al*., 2011; Gatsogiannis *et al.*, 2013).

Membrane insertion is triggered not only at low pH, but also at high pH, thus explaining Tc action directly through the endosome and midgut of insects respectively (Gatsogiannis *et al.*, 2013; Meusch *et al.*, 2014). The TcdA1's large bell-shaped pentameric structure enters the membrane like a syringe, forming a translocation channel (vuvuzela-shaped) through which the cytotoxic domain (part of the C protein) is transported into the cytoplasm (Figure 1) (Meusch *et al.*, 2014). This is a unique mechanism which differs from that of typical pore-forming toxins (Lesieur *et al.*, 1997) or other toxins that form translocation pores like diphtheria (Murphy, 2011) and anthrax toxin (Young & Collier, 2007). These proteins therefore present a novel mechanism of insecticidal activity that can be useful in insect pest control.

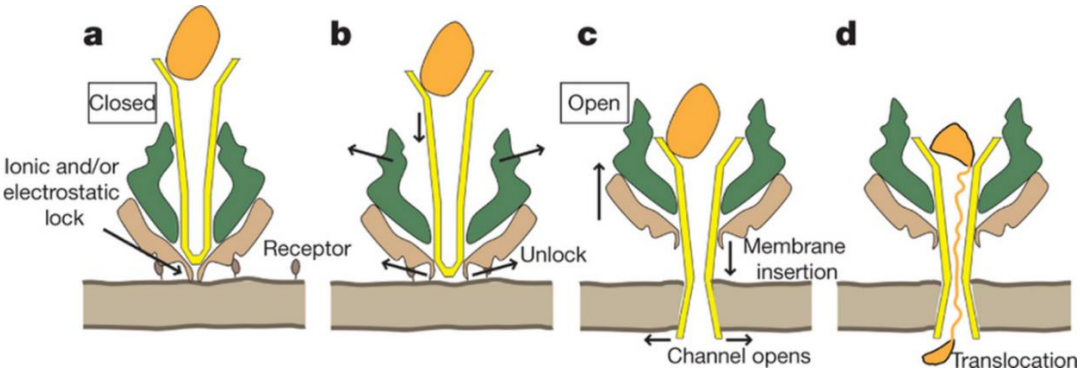


Figure 1: Model for membrane insertion of the Tc protein complex and translocation of the B-C component (Adapted from Gatsogiannis *et al.*, 2013).

a. The complex binds to the cell membrane, **b**. The alpha-helical shell and electrostatic lock open, **c**. The pore domain is injected into the cell membrane and **d**. The enzyme component of the toxin is translocated into the cell.

2.2. Protein structure determination

The 3D structure of a protein is crucial to unveiling its function at the atomic level. Protein structure reveals binding sites, domain interactions, ligand-binding and spatial relations of sub-site residues, these aspects are key to protein function. Advances in high throughput

DNA sequencing technology have led to massive increase in generated sequenced data. The protein sequences currently available in biological databases number in the millions (Pierri *et al.*, 2010). As at 3rd November 2014, the Protein Data Bank held a total of 96649 protein structures (http://www.rcsb.org/pdb/statistics/holdings.do). It is practically impossible to obtain the 3D structures of all novel proteins through experimental methods. This sequence-to-structure gap can be filled through computational biology by predicting the structure and function of uncharacterized proteins (Pierri *et al*., 2010).

The elucidation of the function and biological significance of a protein is heavily reliant on determination of its 3D structure (Nayeem *et al.*, 2006). Structures are conserved to a greater extent than sequences, therefore, protein structure comparison also provides an effective way of finding distant evolutionarily related homologs (Gherardini & Helmer-Citterich, 2008). X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and electron microscopy are experimental techniques used to generate high quality 3D structures of proteins (Di Luccio & Koehl, 2011). These techniques are, however, laborious, time consuming and expensive, hence computational techniques like homology modelling are used to predict the structure of proteins thus subsequently helping to understand their function and relationships.

2.2.1. Homology modelling

The basis of homology modelling is the principle that the 3D structure of a protein can be predicted given its sequence is similar to other proteins of known experimentally-determined structure. Homologs diverge from a common ancestor; over time accumulating non-destabilizing mutations hence their protein structures remain relatively similar. Higher sequence identity increases similarity in the backbone structure of proteins (Bajorath *et al.*,

1993). In the absence of experimentally derived 3D protein structure, homology modelling plays a major role in elucidation of putative protein structure, function and evolutionary relations.

Modelling the 3D structure of a target protein involves a series of steps:

a) Template identification

This involves searching protein structure databases for a homologous protein of known high-resolution which has an experimentally determined 3D structure (a template). The target and template protein sequences are subsequently aligned. The main chain of the target protein is modelled based on the coordinates of the template protein and varied regions unrepresented in the template are modelled as loop regions. The side chain atoms are then built and the resulting model is refined, optimized and verified to assess the quality of the model (Di Luccio & Koehl, 2011). There are several tools available to automate the model building process. For instance, Modeller, an efficient homology modelling program, builds 3D models of protein structures from structural alignments and subsequently refines the obtained models (Sali & Blundell, 1993).

Templates with a sequence identity greater than 30% to the target protein are acceptable as the two proteins are considered to be homologs and are likely to have common 3D structure (Hillisch *et al.*, 2004; Pearson, 2013). Sequence identity between 30-50% between target and template protein results in high quality models that are satisfactory for drug target studies. Sequence identity of 15-30% is considered to be low; hence accurate alignment is required to identify homology between sequences. Models generated at 15-30% identity can be used for protein function studies (Hillisch *et al.*, 2004).

Sequence identity below 15% results in models which may not reliably represent the protein structure (Nayeem *et al.*, 2006). In the case that there are many available possible templates, the one with the highest sequence identity, highest resolution and best coverage of the target protein is selected. If no particular template fully structurally represents the target, a combination of multiple templates can be used to model the various domains of the target protein (Moult, 2005). Templates are searched from the protein structure databases using sequence similarity search tools such as PSI-BLAST, HHpred and PDB sequence search.

b) Template-target alignment

After selecting a suitable template(s), the sequences of the target and template are aligned. Accuracy of this sequence-structure alignment is a prerequisite for high quality resultant models. To achieve optimum alignments, structurally equivalent residues need to align. Errors made in alignment of residues adversely distort the structure of the model (Venclovas, 2003). Target to template alignment takes into consideration both the matching of similar and identical residues and structural correctness of the alignment. It is thus important to use alignment programs that incorporate protein structure prediction such as HHpred and PROMALS-3D (Söding *et al.*, 2005; Pei *et al.*, 2008). Finally, alignments produced by the alignment programs are visually reviewed for any mistakes and the necessary adjustments are made.

c) Model building

With an appropriate target-template alignment, the template can be used as a guide for the main chain atom positions which are then integrated with loop and side-chain building algorithms (Sali & Blundell, 1993). Model building is based on the target-template

structural alignment in regions where identical residues align, the template's main-chain and side-chain atom coordinates are copied to the model. In regions where aligned residues differ, the main-chain atom coordinates are copied and the side chain is later built. In gap regions of the alignment, that have insertions and deletions, loops are modelled (Xiong, 2006). Loops are mostly found on the surface of the protein as flexible structures that connect secondary structures. Loop regions are challenging to model as they represent regions of the target protein that are structurally different from the template. Loops are often biologically significant as they can be involved in the functional roles of the protein like protein-protein interactions (Di Luccio & Koehl, 2011).

d) Model refinement

Finally the complete models are refined and optimized in both geometric and energetic aspects so as to attain a structure of stable conformation (as close to the native protein as possible). Model refinement involves energy minimization and regulating bond lengths and angles of atoms, to attain the appropriate stereochemistry of the model without distorting its structural conformation (Levitt & Lifson, 1969). Refinement requires methods that effectively sample the conformational space and identify near native-like model conformations. Molecular dynamics simulation (MD) techniques are used to identify near native states enhancing accuracy of conformations selected in the refinement process (Lee *et al.*, 2001). These methods are incorporated into model building programs like Modeller ensuring that optimum models are generated.

2.2.2. Model validation

This involves the assessment of the quality of a predicted model. Computational models are typically not as accurate as the true experimentally-derived structures. Evaluation of the accuracy of models is thus important. Validation of proteins involves: assessment of stereochemical properties, examination of protein folding quality using energy calculations, and verification of model compatibility with its amino acid sequence (Kosmoliaptsis *et al.*, 2011). Available model quality assessment programs (MQAPs) include the following:-

I. **MetaMQAPII** (https://genesilico.pl/toolkit/unimod?method=MetaMQAPII)

Unlike most other MQAPs which are based on global evaluation of the protein structure, MetaMQAPII focusses on local areas of inaccuracy that are common in computational models. MetaMQAPII is a meta-predictor that combines results of 8 other validation programs, namely; PROSA, VERIFY3D, ANOLEA, BALA, PROVE, PROQRES, REFINER and TUNE. This combines the strengths of the different MQAPs whilst eliminating individual flaws. Most of the above MQAPs showed bias to trivial features such as residue depth in the structure and residue hydrophobicity. MetaMQAPII thus combines the various MQAPs and uses a multivariate regression model that controls bias to trivial parameters hence providing a better quality assessment tool (Pawlowski *et al.*, 2008).

MetaMQAPII scores predict the absolute deviation of C-α atoms (backbone atoms) in each residue of the model from its corresponding residue in the native protein structure; this is expressed as the root mean square deviation (RMSD) (Moult *et al.*, 2007). Additionally, it gives a Global Distance Test Total score (GDT_TS), a measure of the global structure

similarity between proteins (Zemla, 2003). GDT-TS score is dependent on the length of the protein; hence the optimum score varies with protein size. MetaMQAPII generates a PDB file, in which it colours the structure by quality; colours are incorporated in the B-factor column enabling visualization of errors in the structure. A spectrum of colours ranging from blue to red is used where blue represents highly scored residues and red poorly scores residues. This eases the prediction of regions of lower quality in the model that can then be further refined (Pawlowski *et al.*, 2008).

II. **PROCHECK** (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/)

It is a MQAP that evaluates the stereochemical quality of a protein structure. This includes assessing the bond length, chirality and chiral torsion angles. Its output is Ramachandran plots that analyse the overall and residue-by-residue geometry. A Ramachandran plot has colouring representing different regions where red represents the most favoured region. The ideal is to have 90% of the residues in the most favoured regions (Yadav *et al.*, 2010).

III. **PROSA** (https://prosa.services.came.sbg.ac.at/prosa.php)

PROSA uses knowledge-based force fields to validate the quality of a model (Sippl, 1993). These knowledge-based force fields are energy functions derived from statistical analysis of all experimentally determined protein structures in databases. This gives the statistical average energy function of correctly folded, stable proteins that can be used to assess correctness of modelled structures (Sippl, 1995).The program calculates the surface energy of the protein giving a Z-score that determines global model quality. The Z-score is interpreted by displaying it in a plot containing Z-scores of all experimentally-derived protein structures. The protein Z-score can be observed from the plot to see if it is in the desired range of native conformations (Wiederstein & Sippl, 2007). PROSA also gives a

local model quality energy plot, this plots energy scores against the amino acid positions over a 10 and 40 residue window. Residues with positive energy values show erroneous parts of the structure.

2.3. Maize: an important crop for Africa

Maize is one of the most important staple food crops for the people of Sub-Saharan Africa, it provides food and income to over 300 million resource-poor smallholder farmers in eastern and southern Africa (CIMMYT, 2015). It is estimated that about 90% of rural households in Kenya grow maize. The annual maize production is between 2.2-2.7 million tons and forms the bulk of stored grain (Mutambuki *et al*., 2010). However, the national maize supply by these small scale farmers declines annually due to a combination of crop failures in the predominantly short rains-dependent regions, coupled with pre- and post-harvest losses which range from 20-30% (Kimatu *et al.*, 2012). In addition, farmers in Africa use traditional granaries to store their grains. These are not effective against storage pests (Tefera *et al.*, 2011). The main insect storage pests of maize are the Larger Grain borer (LGB) (Boxall, 2002; Tefera *et al.*, 2010) and the Maize Weevil (MW) (Bosque-Perez, 1995; Demissie *et al.*, 2008; Tefera *et al.*, 2011).

2.4. Maize weevils (*Sitophilus spp*.)

2.4.1. Biology

Two weevil species are implicated in grain losses in the store, the maize weevil (*Sitophilus zeamais*) and the lesser grain weevil/rice weevil (*Sitophilus oryzae*) (Bosque-Perez, 1995). The two belong to the order Coleoptera and family Curculionidae. *S. zeamais* is a small weevil about 2.4−4.5 mm in length with its head protruded into a snout or a distinct beak

or proboscis. At the end of this structure, there is a pair of mandibles or jaws. It is generally reddish brown in colour, sometimes dark brown or almost black. The newly emerged weevil is light brown to reddish brown (Bosque-Pérez, 1995; Tefera *et al.*, 2010). It has a long and narrow snout, with clubbed and eight-segmented elbowed antennae. It is further identified by the presence of four light reddish-brown or yellowish pale oval spots on the elytra (Tefera *et al.*, 2010). Separation of the two species (*S. zeamais* and *S. oryzae*) requires examination of the genitalia (Bosque-Pérez, 1995).

The maize weevils use their elongated snouts for boring into the grain. The females, however, also use them for digging a shallow hole into which they lay eggs. Due to the higher fecundity of females, their numbers increase at a very high rate if the weevils are not controlled (Tefera *et al.*, 2011). The male weevils produce a hormone that attracts and aggregates young weevils thus increasing the infestation (Larrain *et al.*, 1995).

Mating in the maize weevil does not take place before the adults are 3 days old (Tefera *et al.*, 2010). The females remain fecund throughout their life-time, but their fecundity and feeding is highest when they are in the ages of 0-3 weeks after which there is a steady decline (Larrain *et al.*, 1995; Siwale *et al.*, 2009; Tefera *et al.*, 2010). Total development periods range from 35-110 days depending on the humidity, temperature, host, diet and varietal differences within cereals (Bosque-Pérez, 1995; Tefera *et al.*, 2010). The weevils breed well at temperatures between 20°C and 32°C, and in food with a moisture content of between 11 and 13 %. Median development time is reported to be 42 days but adults may live up to a year (Kranz *et al.*, 1977; Bosque-Pérez, 1995; Tefera *et al.*, 2010).

The maize weevil female lays up to 4 eggs in a single maize kernel and in total produces 300 to 400 eggs over a period of 4 to 5 weeks. The egg is white in colour and oval in shape.

A fully-grown larva is white in colour and is about 4 mm in length (Hill, 1983). Eggs are deposited singly in narrow cavities chewed into kernels (Bosque-Pérez, 1995) and covered with a gelatinous material that quickly hardens to form a protective plug (Tefera *et al.*, 2010). However, Danho *et al.* (2002) showed that *S. zeamais* tended to cluster eggs on the grains at lower grain quantities as well as when there is an increase in maize weevil density and the duration of oviposition/feeding.

Upon hatching, after six days at 25°C, the larvae burrow further into the grain and form a winding tunnel that increases in size as it grows. The larva moults four times in a period of approximately 25 days (Hill, 1983; Bosque-Pérez, 1995) before pupation takes place. Larval development and pupation take place inside a single grain and larval damage is thus hidden from visual inspection. Generally, weevil larvae are not able to migrate between grains and larval competition is very high in a host grain with multiple larvae (Danho *et al*., 2002). The new adult chews its way out of the kernel leaving a characteristic emergence hole. The adult may remain inside the kernel for some time after eclosion but eventually emerges. The sex ratio of the newly emerged maize weevil is 1:1 and female weevils live longer than male weevils (Tefera *et al.*, 2010).

2.4.2. Economic importance

The maize weevil is an important pest of stored maize in the tropics, especially unprotected maize. Grain weight loss of 12 to 20 % caused by the weevil has been reported. Losses of up to 80 % may occur for untreated maize grain stored in traditional structures in tropical countries (Boxall, 2002). Weevil damage results in lost food (reduced grain weight) and reduced maize production for farmers who plant saved grain as seed. Consumption of weevil-infested grain also poses a health risk, such grain has been reported to have higher

levels of *Aspergillus flavus* fungal contamination than non-infested maize kernels (Tefera *et al.*, 2010).

2.5. *Prostephanus truncatus* (Larger Grain Borer)

2.5.1. Biology

The larger grain borer, *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae) is part of a family whose members are known as powder post or false powder post beetles (Tefera *et al.*, 2010). Bostrichidae are widely recognized as pests of timber and mainly live on felled and/or dried wood, but some species also attack green timber. (Tooke & Scott, 1994). The deflexed head, strong mandibles, and cylindrical body shape of *P. truncatus* are typical features of xylophagous insects (wood feeders/borers).

The larger pronotum protects the head during tunnelling and provides strong support for the mandibular muscles (Li, 1988). *P. truncatus* has a remarkable tunnelling ability through hard materials, adults have been found to penetrate 35 mm thick plastic (Li, 1988). Such mandibular strength can, however, only be applied if the beetle is able to get sufficient leverage (such as between two kernels on maize cob), since it has difficulties when attacking a smooth surface (Tefera *et al.*, 2010).

The body length of an adult ranges from 2−3.5 mm and 1−1.5 mm in width, with a sex ratio of 1:1 (Tefera *et al.*, 2010). *P. truncatus* is capable of flying and is estimated that it can fly 25 km in 45 hours (Tefera *et al.*, 2010). Flight activity is initiated by a reduction in food quality and seasonality in tree growth. Dispersing *P. truncatus* can be captured in flight traps baited with the insect-male-produced aggregation pheromone (Tefera *et al.*, 2010). Host finding is done through pheromone and volatile compounds emitted from the

maize grain. Males produce pheromones at the highest rate when they are on a suitable substrate and are not in the presence of females (Scholz *et al.*, 1998).

Prostephanus truncatus reproduces on maize grain and ears, dry cassava and other stored produce. The eggs are laid in small clutches in tunnels, and an egg clutch is usually protected by tightly packed frass when reared on loose maize grain. The female lays on average 5-8 eggs in each oviposition chamber. This chamber is half as long as the insect's body and slightly wider than its abdomen. It has a lifetime fecundity of 300 eggs when reared on yellow maize. Fecundity and survival reduce when very hard maize varieties are used (Tefera *et al*., 2010).

Prostephanus truncatus undergoes 3 larval instar stages and the average larval period is 16 days. The larvae of *P. truncatus* can be differentiated from other stored product insect larvae by their C-shaped body and head retracted into the prothorax (Tefera *et al.*, 2010). The last larval instar makes a pupal case from frass bound with secretions from the larvae within the grain or surrounding flour. The developmental time from egg to adult at 70% relative humidity (RH) ranged from 25 days at 32°C to 165 days at 18°C (Tefera *et al.*, 2010).

2.5.2. History of *P. truncatus* existence in Africa

The larger grain borer is a serious pest of maize of introduction to Africa (Bosque-Pérez, 1995; Tefera *et al.*, 2011). It is native to meso-America where it has long been recognized as a destructive pest of maize stored 'on the cob' (Tefera *et al.*, 2010). Its entry and establishment in Kenya for almost 20 years introduced a new dimension in the levels of storage losses for maize (Mutambuki, 2012). It was first found in Tanzania from where it spread to other East African countries (Tefera *et al.*, 2011). In West Africa it was first

found in Togo in the early 1980s. It has now spread to at least 18 African countries becoming the most invasive destructive pest of stored maize in eastern, central, western and southern Africa (Tefera *et al.*, 2010).

2.5.3. Economic importance

The larger grain borer is a major pest of stored maize and will attack maize on the cob, both before and after harvest. Adults bore into the maize husks, cobs or grain, making neat round holes and tunnelling extensively. This causes considerable losses in stored maize; weight losses as high as 35% have been observed after only three to six months of storage in East Africa (Tefera *et al*., 2010). Additionally, they produce large quantities of grain dust (flour) as they tunnel. Flour production by *P. truncatus* is higher than that of *S. zeamais* due to extensive tunnelling in the grain by the former. The flour produced during the insects feeding consists of the insect eggs, excreta and exuviae; hence, neither fit for animal nor human consumption due to its unattractive taste (Tefera *et al.*, 2011).

The larger grain borer is spread over longer distances almost entirely through the import and export of infested grain. Local dispersal is through the movement of infested maize from surplus to deficit areas, and by flight activity. Although the larger grain borer develops best at high temperatures and relatively high humidity, it tolerates dry conditions, and may develop in grain that has as low as 1 % moisture content (Haines, 1991). This is in contrast to many other storage pests, which are unable to increase in number under low moisture conditions. (Tefera *et al.*, 2010).

2.6. Pest control strategies

2.6.1. Pesticides and fumigants

Control of these pests has proven to be an arduous task. Control has mainly been through insecticides, supplemented by cultural control methods such as use of ash, botanicals and store hygiene. The use of insecticides is most prominent but not always most effective. Synthetic insecticides like Actellic Super® (Pirimiphos-methyl (1.6%) + Permethrin (0.3%)) and Sofagrain® (Pirimiphos-methyl (1.5%) + Deltamethrin (0.5%)) are most commonly used for grain protection, but, their effectiveness is limited. Even with appropriate application, long term storage may still result in considerable grain damage and dry weight loss (Mutambuki & Ngatia 2010; Njoroge *et al*., 2014). The use of less effective adulterated pesticides has also been observed (Mutambuki & Ngatia, 2006).

2.6.2. Resistant plant varieties

Another strategy for control of these storage pests has been development of resistant plant varieties. There is a wide genetic diversity of maize in relation to susceptibility to weevil attack, and it is possible to develop varieties with some degree of resistance to weevils (Bosque-Pérez, 1995; Mwololo *et al.*, 2012) and other pests**.** Experiments by Abebe *et al.* (2009) found considerable variation in susceptibility to weevil attack among the maize varieties with respect to F1 progeny, median developmental time, seed damage, seed weight loss and the susceptibility index.

2.6.3. Biological control

One of the most promising alternatives to pesticides and fumigants for postharvest pest management is biological control (Chaisaeng *et al.*, 2010). Biological control of *P. truncatus* has been tried through the release of a histerid beetle, *Teretrius nigrescens* in

selected areas with the aim of suppressing the pest population in the wild vegetation (Bosque-Pérez, 1995; Mutambuki & Ngatia, 2006). Among natural enemies that could act as biological control agents of the maize weevil, the wasp *Anisopteromalus calandrae* (Howard) (Hymenoptera: Pteromalidae) is a dominant parasitoid naturally found in granaries (Chaisaeng *et al.*, 2010). However these methods face challenges and have thus been slow to be applied in the field. For instance, the inherent genetic limitations of *T. nigrescens* populations introduced into Africa, particularly Kenya, have made their use in the control of LGB challenging. These populations have faced challenges in adapting to the environment, thus limiting their effectiveness (Omondi *et al*., 2011; 2014).

A popular biocontrol method that has been used is the bacterium *Bacillus thuringiensis* (Bt), particularly transgenic crops having Bt toxins (Estruch *et al.*, 1996). Bt toxins have been used for the control of insect pests in commercial crops for more than 40 years, initially as a sprayable application, and in more recent years in transgenic crops (Lee *et al.*, 2007). It has been the most successful protein toxin applied to develop transgenic crops (Joo Lee *et al.*, 2004). It currently comprises 75% of the biopesticide market in the USA, Canada, and Mexico (Cory & Franklin, 2012). However, its wide use on a large scale as well as the use of transgenic plants expressing these toxins has enhanced the development of resistant insect populations (Shikano & Cory, 2014).

The first generation of Bt crops was dominated by plants producing single toxins to kill key caterpillar pests, for example, corn producing Bt toxin Cry1Ab and cotton producing the closely related toxin Cry1Ac. Although Bt corn and Bt cotton still dominate, varieties of these crops currently registered in the United States collectively produce 18 different combinations of 11 Bt toxins. Each variety produces one to six Bt toxins that kill

caterpillars, beetles, or both. However, the major threat to the continued success of Bt crops is evolution of resistance by pests (Tabashnik *et al.*, 2009).

Resistance to Bt products was not anticipated (Cory & Franklin, 2012) but is a reality. Numerous selection experiments performed in the laboratory have demonstrated the ability of many insects to evolve resistance, including those belonging to the Orders Lepidoptera, Coleoptera, and Diptera. Cases of field-evolved resistance have also been documented such as in *Plutella xylostella, Trichoplusia ni* and *Culex pipiens* (Cory & Franklin, 2012). New protein toxins are therefore required to provide a greater diversity of genes for use in pest control (Estruch *et al.*, 1996; 1997; Sergeant *et al.*, 2003; Lee *et al.*, 2004).

CHAPTER THREE

3.0. MATERIALS AND METHODS

3.1. Materials

The *Steinernema* entomopathogenic nematode (EPN) (*Steinernema carpocapsae*) and *Galleria mellonella* larvae were obtained from Dr. Charles Waturu of the Kenya Agricultural and Livestock Research Organization (KALRO), Horticultural Research Institute, Thika, Kenya. The maize weevil (Figure 2) were obtained from an infested farm in Nairobi County, Kenya and the larger grain borer (Figure 3) were obtained from an established colony at the International Centre of Insect Physiology and Ecology (ICIPE), Duduville campus, Kasarani, Kenya. The weevils were identified through morphological features typical for the species: the shape of the head, shape and length of the snout and the general body shape and size (Figure 2). The untreated maize grain (H614D variety Kenya Seed Company Ltd) used to rear the insects was procured from farmers in Kitale, Trans-Nzoia County, Kenya. Its storage history was obtained by enquiry from the farmers to ascertain that it had not been treated.

Figure 2: The maize weevil (S. zeamais) (X27 Magnification). **A.** Larval stage. **B.** Immature stage. **C.** Adult stage.

Figure 3: Larger grain borer (*P. truncatus*) (X23 Magnification). **A.** Larva. **B.** Adult (Adapted from Dr. Robert Copeland, ICIPE, 2015).

3.2. Experimental site

The research was done at the Molecular Biology and Bioinformatics Unit (MBBU) in ICIPE. Bacteria (*Xenorhabdus* sp. and *Escherichia coli*) and *Steinernema carpocapsae* nematodes were reared in the microbiology laboratory of MBBU. Molecular and protein analysis was done in the molecular biology laboratory of the same unit. The insect colonies of *P. truncatus* and *S. zemais* were reared in the ICIPE insectary. The insects selected for the bioassays were kept in a special rearing room attached to the microbiology laboratory.

3.3. Isolation, proliferation and preservation of *Xenorhabdus* sp.

3.3.1. Isolation of *Xenorhabdus* sp.

Xenorhabdus sp. was obtained by the indirect isolation method using *Galleria mellonella* larvae (5^{th} instar) infected with *Steinernema* entomopathogenic nematodes (Akhurst, 1983; Boemare & Akhurst, 2006). A total of 15 larvae were enclosed in petri dishes lined with pieces of cotton cloths inoculated with the nematodes in distilled water. Larvae killed by the nematodes were identified based on colour, absence of dark spots, mushy/spongy cadavers and absence of the smell of putrefaction.

Twenty four hours after larval death, the cadavers were surface sterilized by dipping in 70% isopropanol followed by 90% isopropanol each for 30 seconds and flame sterilization before a final dip in sterile water. The sterile cadavers were then dissected under aseptic conditions and their haemolymph streaked onto NBTA medium (Nutrient Agar supplemented with 0.004 % (w/v) Triphenyltetrazolium chloride (TTC) and 0.025 % (w/v) Bromothymol blue) (Akhurst, 1980) in petri dishes. The NBTA plates were then incubated for 48 hours at 28°C. Blue colonies were observed and sub-cultured to achieve a pure colony.

3.3.2. Proliferation and preservation of *Xenorhabdus* sp.

Pure-coloured blue colonies from NBTA plates were inoculated into Lysogeny Broth/Luria-Bertani broth (LB) (10g Bacto-Tryptone, 5g Bacto-Yeast extract and 10g NaCl in 1L of distilled water) (Hanahan, 1983). This was then incubated for 24–48 hours in an ENVIRON-SHAKER 3597-1 rotary incubator (lab-Line Instruments Inc., Kerala, India) at 28-30°C with vigorous shaking at 150 rpm (Morgan *et al*., 2001; Boemare & Akhurst, 2006). The bacterium was then pelleted by centrifugation at 6,000g for 15 minutes using an Avanti J-25 I centrifuge (Beckman, USA) before cell lysis. The isolated bacterium was used for gene and protein isolation as well as in the bioassay.

Medium-term storage (up to four months) of the bacteria was done through stabs into nutrient agar (without TTC and Bromothymol blue) and subsequent storage at room temperature in the dark. Alternatively, the bacteria were stored on NBTA plates incubated at 28°C and then kept at 4°C. For long term storage, an overnight broth culture of bacterium was mixed with 15% of pure sterile glycerol and stored in a -80°C freezer (Tailliez *et al.*, 2006).

3.4. Characterization of *Xenorhabdus* sp.

3.4.1. Morphological characterization

Isolated *Xenorhabdus* sp. was identified using its morphological characteristics on an NBTA plate and Gram staining. The following morphological characteristics were assessed: convex, circular colonies with slightly irregular margins and a diameter of 1.5–2 mm; rod shaped; swarming motility on NBTA medium; dendritic growth pattern, blue coloured phase I and red phase II colonies (Akhurst, 1980).

3.4.2. Molecular characterization

3.4.2.1. Genomic DNA extraction

Genomic DNA was extracted from freshly grown bacteria cultured as detailed in section 3.3. Genomic DNA was extracted using FastDNA® SPIN Kit for soil (MP Biochemicals, Santa Ana, CA) following the manufacturer's instructions or using the procedure by Chen & Kuo, (1993) where blue colonies of bacteria were scrapped off from an NBTA plate using a sterile pipette tip and resuspended in 200 µL of lysis buffer (40 mM Tris-acetate pH 7.8, 20 mM sodium-acetate, 1 mM EDTA, 1% SDS). Bacterial cell lysis was then achieved by vigorous pipetting.

To remove most proteins and cell debris, 66 µL of 5M NaCl solution was added and mixed well, then the viscous mixture was centrifuged at 12,000g for 10 min at 4°C. After transferring the clear supernatant into a new vial, an equal volume of chloroform was added and mixed (by gentle tube inversion) until a milky solution was formed. The solution was centrifuged at 12,000×g for 3 min and the extracted supernatant transferred to another vial. DNA was then precipitated in absolute ethanol, washed twice with 70% ethanol, air dried and finally re-dissolved in 50 µL 1×TE buffer. RNA was removed by adding RNase in the

lysis step and incubation for 30 min at 37°C. The concentration of the extracted DNA was determined and quantified via a nanodrop 2000/2000c spectrophotometer (ThermoScientific, USA).

3.4.2.2. PCR amplification of the Xenorhabdus 16S rRNA gene

The 16S rRNA gene was amplified using the following PCR mix: 1 × Taq polymerase buffer (Genscript®), 200µM of each deoxynucleotide triphosphate (dNTP) (New England Biolabs®), 0.2µM of each primer, forward (27f-5'AGAGTTTGATCATGGCTCAG 3') and Reverse (1392r-5'ACGGGCGGTGTGTGC 3') (Lane, 1991), 1 unit of Taq polymerase (Genscript®) and template DNA of ≤1µg per 100µl reaction. Cycle conditions used for the PCR were as follows: initial denaturation at 94°C for 5 minutes, 35 cycles of subsequent denaturation at 94°C for 30 seconds, annealing at 48°C for 15 seconds, extension at 72°C for 1 minute, and a final extension at 72°C for 10 minutes in Arktik Thermal Cycler (ThermoScientific, USA).

Amplified PCR products were separated by gel electrophoresis in a 1% (w/v) molecular grade Tris acetate EDTA agarose gel supplemented with ethidium bromide at final concentration of 0.5ug/ml. The gel was run at 70V for 1 hour 12 minutes in 1× TAE buffer (0.04M Tris Acetate, 0.01M EDTA, pH 8). A 1000 base pair ladder was used (New England Biolabs® Inc.) to determine the band sizes. Visualization was done using a UV Polaroid camera linked to visualization software (Kodak, USA). The band of interest was cut out of the gel using a sterile scalpel and purified using a QuickClean II gel extraction kit (GenScript®) following the manufacturer's protocol. The purified PCR products were then sequenced (Macrogen™, Amsterdam, Netherlands) using the Sanger platform (Shendure & Ji, 2008).

3.4.2.3. Sequence analysis

The sequences were analysed using bioinformatics tools as follows: editing of the sequence was done using MEGA 5.2 trace editor (Tamura *et al.*, 2011). The first 25 bases on each end of the sequence were truncated leaving a 1300 nucleotide long sequence. Sequence alignment was then done using MUSCLE on MEGA 5.2 sequence alignment editor and analysis program. A similarity search of the Genbank database was then done using basic local alignment search tool (BLAST) (Altschul *et al*., 1997) to identify the bacteria. The best eight BLAST hits obtained were retrieved and their sequences aligned using the MEGA 5.2 alignment tool. Ten 16S rRNA gene sequences from the bacterium *X. nematophila*, a well-studied bacterium with morphological characteristics similar to the isolated *Xenorhabdus* sp., were used for comparison purposes.

3.4.2.4. Phylogenetic analysis of *Xenorhabdus* spp.

Using all the previously retrieved 16S rRNA gene sequences, a phylogenetic tree was reconstructed using the maximum likelihood (ML) phylogenetic inference method on MEGA 5.2 (Tamura *et al.*, 2011). For maximum-likelihood tree inference, gaps or missing data were deleted, the "nearest-neighbour-interchange" (NNI) option was selected and the initial tree was made automatically. Bootstrap analyses were performed with 100 replications. This tree was used to provide information on the clustering of the sequences in order to augment the information obtained from the BLAST search. A 16S rRNA sequence from *Pseudomonas fluorescens* p10-1 was also included in the tree as an out-group.

3.5. Rearing of post-harvest insect pests of maize, *S. zeamais* and *P. truncatus*

3.5.1. Insect colony

The maize weevil and the larger grain borer were reared on whole maize grain sterilized by heat treatment. A sample of 500g of untreated maize grain was placed in 1L glass jars and sterilized as described in subsection 3.5.2. The colony jars were covered with ventilated lids made by placing a brass screen against the inside of a lid whose top surface had been removed via a cylindrical cut-out of the top surface. The screen provided ventilation while at the same time prevented the movement of insects in and out of the colony jar (Tefera *et al.*, 2011).

One hundred unsexed adult insects of the two species were separately introduced into the glass jars containing maize grain (moisture content, 12-14%) (Abebe *et al*., 2009; Tefera *et al.,* 2010; Tefera *et al*., 2011) and allowed to reproduce. Any adults, larvae and pupae collected from degraded grain were introduced into fresh grain for continuity of the culture. The conditions used in rearing were as follows: 28 ± 1°C, 65 ± 5 RH (Relative Humidity), and 12:12 light: dark regime (Tefera *et al.*, 2011). A humidifier and heater were used to achieve the aforementioned conditions.

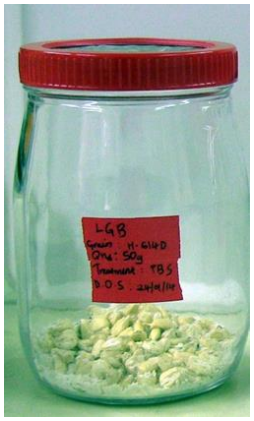


Figure 4: Basic insect colony set-up showing a glass jar containing infested maize and lid fitted with a brass screen.

3.5.2. Colony diet preparation

When the grain was first received, it was cleaned and dried before storage or usage. Cleaning was done by sieving to remove any excess dirt, dust, fine materials, mouldy and broken or shrivelled kernels. Once the grain was clean and dry, it was sprayed with approximately 2mL distilled water and the glass jar covered with a polythene piece of paper before replacing the perforated brass-covered lid. This setup was to achieve a partial vacuum. The covered jars were then incubated at 75°C for 4 hours in an oven to allow sterilization by heat. They were then retrieved from the oven, the polythene cover removed, the normal lid replaced and the seeds kept for two weeks in the rearing room for acclimatization before introduction of the insects (Abebe *et al.*, 2009).

3.6. Bioassay of *Xenorhabdus* sp. against *S. zeamais* and *P. truncatus*

3.6.1. Bacterial colony preparation

A pure colony of *Xenorhabdus* bacterium from an NBTA plate culture was inoculated into 500mL of LB broth and proliferated as detailed in section 3.3. Bacterial cells were pelleted

by centrifugation at 6000 x g for 15 minutes in an Avanti J-25I centrifuge (Beckman, USA). A culture of *E. coli* was also prepared on nutrient agar and incubated at 37°C. Pure colonies were inoculated into LB and incubated at 37°C with shaking at 150 rpm in a rotary incubator for 48 hours after which cells were also pelleted. The cell pellets were then re-suspended in 17mL of phosphate buffered saline (PBS) pH 7.2. Optical density (OD) readings at 600nm wavelength were taken to ascertain the bacterial concentration. The concentration was subsequently obtained using the formula 0.1 OD (optical density) $\approx 10^8$ cells/mL (Ausubel *et al*., 1987). The OD readings obtained were each multiplied by 10^8 to get the cell count.

3.6.2. Preparation of artificial diet inoculated with *Xenorhabdus* sp.

In a sterile hood, 17mL bacterial suspensions of *Xenorhabdus* sp. and *E. coli* were each mixed with 25g of untreated maize flour to make thick pastes. A bacterial concentration of 3.56×10^8 cells/mL, was used in preparing the artificial diet. A third paste was made using PBS only to serve as the negative control. The paste was then spread on a flat plastic-container lid, sterilized by UV, to approximately 0.7cm thick and pellets (approximately 0.7cm in diameter) were made from it. The pellets were kept at 28°C in an incubator for 12 hours to dry and then 72 hours at room temperature (23-26°C) to acclimatize to the culture room conditions before introduction of the insects.

3.6.3. Experimental design

Adult insects of both species (*P. truncatus* and *S. zeamais*) to be used in the bioassay were randomly selected from their rearing containers in the insectary and starved for three hours. This was to ensure that they started feeding immediately after introduction to the maize pellets, within the duration that the bacterium was still viable in the pellets. Each

experiment consisted of three treatments; maize pellets with Xenorhabdus, maize pellets with *E.coli* and maize pellets with PBS (negative control). Each treatment consisted of a total of 80 insects, randomly selected from the starved group, in four replicates with 20 insects per replicate. The insects were introduced into glass jars containing 25 treated pellets each. The experimental set up was kept in a rearing room at room temperature and humidity and monitored daily for 11 days. Within this period, the viability of the Xenorhabdus bacterium was also assessed.

3.6.4. Test for the viability of *Xenorhabdus* sp. in artificial diet

To determine the viability of the bacterium, a pellet was retrieved from the *X. nematophila* and PBS treatments at intervals of 6, 12, 24, 46 hours, 4days and 6 days. These pellets were each re-suspended in 1mL of PBS in a sterile Eppendorf tube, under aseptic conditions and 200µL of this suspension spread on an NBTA plate. This plate was incubated at 28°C for 48 hours after which the bacterium was sub-cultured onto another NBTA plate to achieve a pure colony. Colony morphology was then assessed and compared to that of the starting bacterium.

3.6.5. Data collection and analysis

On the 11^{th} day, all the insects were taken out of the jars. An assessment was then made on the mortality and feeding behaviour of the test insects. Counts were made of the dead and live insects to assess mortality. On the other hand, the undamaged pellets (no emergence holes) and the bored pellets (those with emergence holes but were still intact) were all counted to determine the feeding behaviour of the insects. These values were then added and their sum subtracted from the initial number of pellets in order to determine the number of completely destroyed pellets. The same process was repeated for *E. coli* and the negative

control. A Chi square test of independence at P = 0.01 was carried out on both data sets of mortality and feeding to assess the effect due to the treatments.

3.7. Bioinformatic analyses of class A genes and proteins from *Xenorhabdus* bacterium

3.7.1. Search for XptA genes/protein sequences and their homologs

Two class A genes and proteins given identical names were identified in *Xenorhabdus* spp., XptA1 and XptA2 genes/proteins (Morgan *et al*., 2001). A specific search of the KEGG (Kyoto Encyclopaedia of Genes and Genomes) database was conducted for Class A genes of the *Xenorhabdus* bacterium and their homologs. The KEGG database provided links to other nucleotide and protein databases for the same genes, which included GenBank and UniProtKB/TrEMBL. Both the nucleotide and amino acid sequences for these genes were retrieved as well as details of their lengths and accession numbers.

3.7.2. Analysis of class A protein families, domains and motifs

A domain analysis of the XptA genes was conducted on Interpro (http://www.ebi.ac.uk/interpro/) and Pfam (http://pfam.xfam.org/), using XptA1 and A2 from *X*. *nematophila* as query sequences. This provided information on gene/protein families, conserved domains and also any other organisms that possess similar domains. One domain known as VRP1 was identified. All sequences containing this identified domain were retrieved and archived. In order to identify motifs, the domain sequence was retrieved from Pfam (http://pfam.xfam.org/) and NCBI's conserved domain database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) and then submitted to Motif scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan) for prediction of the conserved

motifs. This sequence showed motifs and their respective positions within it. It also gave statistical scores to enable the interpretation of the significance of these motifs, these scores included the E-value, normalized score and the raw score.

The E-value, which is the number of matches with a score equal to or greater than the observed score that are expected to occur by chance given the size of the database, provided an estimation of the number of false positives likely to be obtained by chance. The normalized score on the other hand was the base ten logarithm of the size (in residues) of the database in which one false positive match was expected to occur by chance. Both the E-value and normalized score enabled the interpretation of the raw score and gave an indication of the likelihood of the motif matches being false positives, that is, the strength of the matches.

All of the VRP1 domain sequences retrieved from the databases were analysed using the conserved domain search on NCBI. An assessment was done on the completeness of the domain (full domain or truncated), the domain length, the region of the query in which the domain occurred, as well as the E-value and bit score which gave an indication of how good an alignment was. A multiple sequence alignment of all the sequences was also performed using MUSCLE on SeaView's graphical user interphase (GUI) version 4 (Gouy *et al.,* 2010) in order to show the conserved residues. These were later compared with the motifs previously predicted through Motif scan.

A Maximum Likelihood (ML) tree was re-constructed using PhyML 3.0 (Guindon & Gascuel, 2003). One hundred bootstrap replicates were performed using the LG amino acid model on PhyML 3.0. The starting tree was generated using the BioNJ algorithm in PhyML. The tree was rooted using the dapE (N-Succinyl diaminopimelate deacylase)

(KEGG entry: b2472) amino acid sequence from *Escherichia coli* K-12 MG1655. The tree was visualized and edited using FigTree program version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). This tree enabled the assessment of the relationships between the sequences obtained from the different organisms.

3.7.3. Prediction of the 3D structure of Class A insecticidal proteins

3.7.3.1. Template identification and verification

The amino acid sequences of XptA protein and XptA1 VRP1 domain amino acid sequence from *X. nematophila* were submitted to HHpred, an online protein structure and function prediction platform (Soding *et al.*, 2005). Other A toxin amino acid sequences were also submitted and their results compared to those obtained from XptA proteins. HHpred first generated a multiple sequence alignment using HHblits and subsequently a profile hidden Markov model (HMM). The HMM was then used to query against a database of protein sequences with known structure (for example PDB, SCOP) or annotated protein families (for example PFAM, SMART, CDD, COGs, KOGs). The output was a list of the closest homologs.

The structure with the highest sequence identity, highest resolution and best coverage of the target protein was selected as the template. Homology models of all retrieved class A sequences were then built using the chosen template from the protein data bank (PDB) database (http://toolkit.tuebingen.mpg.de/). This template was then retrieved from the PDB and its quality assessed using PROCHECK for proper stoichiometry, PROSA for the energy state and MetaMQAPII for overall quality.

3.7.3.2. Template-target alignment

Upon selection of the appropriate template for modelling, alignments of the target and template sequences was done using the HHpred server which created sequence alignments in the *.pir* format that is essential for modelling. The HHpred alignments were then compared to others generated using the PROMALS-3D program (Pei *et al.*, 2008). These enabled adjustments of the HHpred alignments in order to achieve the optimum alignments which is crucial to obtaining good models.

3.7.4. Modelling of the full insecticidal proteins and VRP1 domains

Having attained the appropriate template-target alignments, the adjusted alignments were then used with the Modeller program on the HHpred server in *.pir* format. Modeller was used to calculate the homology models which it generated automatically, once given an input of the alignment in *.pir* format. The eventual output was PDB files containing the coordinates of the generated models. These files were then viewed and compared using PyMOL v1.5.0.4 (Schrödinger, LLC), an open-source, user-sponsored, molecular visualization system (http://www.pymol.org/) and MATRAS (Markovian transition of structure evolution), an online program for protein 3D structure comparison (Kawabata, 2003).

3.7.4.1. Model validation

To validate the structures of the obtained models, a number of model quality assessment programs (MQAPs) were used including MetaMQAPII, PROCHECK and PROSA. The MQAPs evaluated various properties of the structures as previously described in 2.2.2 thus detecting errors arising due to misalignment of residues, regions that were inappropriately modelled and erroneous main chain conformations. The programs were accessed via web

servers through which the PDB files holding model atom coordinates were uploaded. These programs then validated the models automatically and enabled refinement of those models that were found to be below par.

3.7.5. Structure comparisons and location of the position of VRP1 domain on the full sequence

Using PyMol and MATRAS, all models were viewed and superimposed in order to identify regions of similarity and/or difference. Amino acid sequence alignments accompanied these structure comparisons and thus structure similarities/differences could be traced back to the amino acid sequences. The tools were also used to map the VRP1 domain structure on to the full class A protein structure through superimposition of the VRP1 models onto the full protein models. This would help in elucidating the putative function or importance of the VRP1 domain in the full protein.

3.8. Phylogenetic relationship of organisms with class A proteins

Representative 16S rRNA gene sequences for all bacteria whose class A proteins have been analysed were retrieved from Genbank and archived. These sequences were subsequently processed as described in sub-section 3.7.2 and a phylogenetic tree reconstructed. The tree was rooted using a 16S rRNA sequence from *Candida albicans* as an out-group and visualized using the FigTree program version 1.4.2.

3.9. Protein toxin isolation and purification from *Xenorhabdus* sp.

The *Xenorhabdus* bacterium was isolated from its host nematode and cultured as previously described in section 3.3. All protein purification steps were carried out at 4°C (Yang *et al.*, 2012). The protein profiles were visualized on native-PAGE gels stained using

either coomassie brilliant blue (0.02% (w/v) coomassie brilliant blue; 40% methanol; 10% acetic acid and 49.98% dH_2O) or silver staining (Mortz *et al.*, 2001).

3.9.1. Determination of the appropriate lysis procedure

The bacterium was pelleted from broth by centrifugation at 6000 × g for 15 minutes using an Avanti J-25I centrifuge (Beckman, USA) before cell lysis. Three different lysis procedures were compared to determine which one gave the highest protein yields and high molecular weight proteins.

a) A modification of the procedure by Sheets *et al.* (2011) (50mM Tris-HCl pH 8; 100mM NaCl; 1mM DTT; 10% Glycerol; 0.6 mg/mL Lysozyme and sonication at 10-14µ for one 30 second burst).
b) A modification of the procedure by Morgan *et al.* (2001) (10mM PBS, sonication at 10-14µ peak-to-peak for 30s).
c) A combination of enzymatic (98.99% TE buffer, 0.005% Lysozyme and 1% $CaCl_2$) and mechanical lysis (pBAD manual, Thermoscientific, USA).

Each lysed sample was then centrifuged at 13000 × g for 15 minutes in a centrifuge 5417R (Eppendorf, Germany) to pellet cell debris (Morgan *et al.*, 2001). The lysate supernatant was mixed with native-PAGE sample buffer (55.5% distilled water; 0.0625M Tris-HCl, pH 6.8; 30% Glycerol and 0.5% (w/v) Bromophenol blue) (Biorad, USA) and 8µL of each sample loaded into a mini native-PAGE gel composed of a 4% stacking gel (0.189M Tris-HCl, pH 6.8; 6.03% monomer; 0.075% (w/v) APS; 0.015mL TEMED; topped up to 10mL with dH_2O) (Biorad, USA) and a 7% resolving gel (1.125M Tris-HCl pH 8.8; 20.7% monomer; 0.15% APS; 0.02mL TEMED, all topped up to 10mL with dH_2O) (Biorad, USA). The gel was then run at 200V and 30mA for 40 minutes in a Mini-PROTEAN® tetra

cell electrophoresis kit (Biorad, USA). The procedure giving the best results was chosen for subsequent experiments.

3.9.2. Determination of protein yield from different lysis volumes

Using the modified lysis procedure by Morgan *et al.* (2001), crude lysate samples and a subsequent native-PAGE gel were prepared in the same procedure as in subsection 3.9.1 to compare protein yield from different lysis volumes. Three different cell pellets of the same size were obtained by centrifugation from a 200mL LB broth culture grown at 28°C with agitation at 150 rpm for 72 hours in a rotary incubator. The pellets were re-suspended in 5, 10 and 15mL of PBS respectively and then sonicated. Cell debris was then pelleted by centrifugation at 13000 × g for 15 minutes. The lysate supernatant and pellet were both mixed with native-PAGE sample buffer and 8µL of each sample loaded on to a mini native-PAGE gel. The gel was then run at 80V and 12mA for 45 minutes. Cell free supernatant and a high molecular weight protein ladder (Amersham Biosciences, UK) were also included in this gel. The lysis volume giving the best results was chosen for subsequent experiments.

3.9.3. Size exclusion chromatography

A 60cm long, 1.5cm diameter column was set up and packed with a 125 mL bed volume of Sepharose CL-6B, cross-linked 6% beaded agarose gel matrix (Sigma-Aldrich, USA), with a dry bead diameter of 45-165µm and fraction range of 10 – 4000 KDa for globular proteins. The column was mounted on a vertical stage. The matrix was first cleaned by repeated suspension in PBS pH 7.2 (137mM NaCl; 2.7mM KCl; 10mM Na_2HPO_4; 1.8mM KH_2PO_4 and 0.2g NaN_3 in 1L of dH_2O) and aspiration. The gel was finally resuspended in fresh buffer three times the volume of the resin and degassed in a vacuum for one hour.

PBS buffer used in running the column was also degassed. The degassed gel matrix was subsequently poured into the column and 3 column volumes of buffer passed through to pack it by gravity. The packed column was then moved to a cold room (4°C) and 2-3 column volumes of precooled buffer passed in order to equilibrate the gel bed. The column was left at this temperature overnight in order to acclimatize before being connected to a fraction collector model 2128 (Biorad, USA).

3.9.3.1. Calibration of the size exclusion chromatography column

Calibration of the column was done using protein standards as follows: 2.5mg (2% of the total gel bed volume) of Blue Dextran 2000 (Amersham Biosciences, UK) was first run through in order to determine the column void volume (V_o). The Blue Dextran 2000 was reconstituted in 2.5mL of PBS before being loaded into the column. The protein was then eluted by gravity with a buffer reservoir volume of approximately 400mL. Each fraction collected was approximately 1.5mL. Optical density measurements at a wavelength of 280nm were taken for each fraction and the values used to plot a graph of optical density versus tube number. The elution volume of the Blue Dextran 2000 was then determined by measuring the volume of the eluent from the point of application to the centre of the elution peak.

Various protein standards including: bovine serum albumin (67 KDa), Aldolase (158 KDa), Catalase (232 KDa) and Thyroglobulin (669 KDa) (Amersham Biosciences, UK) were run through the column in order to determine the protein elution profile. Each standard was reconstituted at the rate of 5 mg/mL in PBS and 2mL of each pooled and loaded into the column. Elution was by gravity with a buffer reservoir volume of approximately 400mL and each fraction collected was approximately 1.5mL. Optical density measurements at a

wavelength of 280nm were then taken for each fraction and the values used to plot a graph of optical density versus tube number. The elution volumes (V_e) of the calibration proteins were determined by measuring the volume of the eluent from the point of application to the centre of the elution peak of each protein standard. The column was washed by running three column volumes of fresh PBS after complete elution of all standards.

The eluted fractions were pooled according to their peaks and then concentrated by precipitation with 80% of supersaturated ammonium sulphate. This mixture was left to stand on a rocking shaker at 10 rpm overnight at 4°C before centrifugation at 7,000 ×g for 30 minutes to pellet the precipitated salt-protein complex. This pellet was then resuspended in 1mL of PBS and transferred into a Spectrapore dialysis tube of a molecular weight cut-off (MWCO) of 12-14 KDa (Spectrum labs Inc., USA). Dialysis was done overnight at 4°C against PBS in a volume two hundred times the total sample volume. After dialysis, the pooled samples were retrieved and run on a native-PAGE gel as in subsection 3.9.1 above so as to validate the elution profile of the protein standards.

3.9.3.2. Purification of *Xenorhabdus* crude lysate samples

Xenorhabdus bacterium was proliferated as in section 3.3 and lysed using the modified procedure of Morgan *et al.* (2001), as detailed in subsection 3.9.1. Cell debris was then pelleted as previously described (subsection 3.9.1) and 5mL of the supernatant (crude lysate) loaded into the column. The sample was fractionated and eluted by gravity with a buffer reservoir volume of approximately 400mL. Each fraction collected was approximately 1.3mL. Optical density measurements at a wavelength of 280nm were then taken for each fraction and the values used to plot a graph of optical density versus tube number. The elution volumes (V_e) of the eluted proteins were then determined by

measuring the volume of the eluent from the point of application to the centre of the elution peak. Ultimately, a native PAGE gel (subsection 3.9.1) was run in order to validate the profile of the eluted proteins. The samples were then pooled, concentrated, dialyzed as stated before and run on a similar native-PAGE gel.

3.10. PCR amplification of VRP1 domain from XptA gene

3.10.1. Design of gene-specific primers

Gene-specific primers for the VRP1 domain were designed using the XptA1 gene nucleotide sequence from *X. nematophila* ATCC 19061 (Accession: YP_003712778). A nucleotide sequence of approximately 800 bases was identified as the VRP1 domain and used as a template to manually pick the primers. NEBcutter v2.0 (http://nc2.neb.com/NEBcutter2/), an online sequence analysis tool, was used to determine all enzyme restriction sites within the chosen target sequence. This enabled the choice of appropriate restriction enzymes to be included in the primers for downstream applications like cloning and expression. Another tool, ExPASy translate (http://web.expasy.org/translate/), an online tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence, was used to translate the template sequence to ensure maintenance of the open reading frame.

The chosen primer set and the XptA1 template sequence were both submitted to the Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/), a web-based collection of sequence analysis programs. The suite was used to perform an *in silico* PCR in order to predict the amplified sequence, this would then be compared with the original template sequence to check for any discrepancies. The suite was also used to analyse the primer sequences for any potential problems in the primer sequences such as, self-

annealing and hairpin formation, as well as to calculate their annealing temperature. The following primer pair was eventually chosen:

I. Forward $^{5'}$ GCGGATCCATGATAAAAGTTAATGAACTG $^{3'}$

II. Reverse $^{5'}$ GCGAATTCCTAGGAGAGATTGACAAATAAACTG $^{3'}$

3.10.2. Design of degenerate primers

Several other class A sequences from different bacteria phylogenetically close to *Xenorhabdus* spp. were retrieved from the KEGG and Genbank databases. These included sequences from *Xenorhabdus bovienii* SS-2004 (Chaston *et al*., 2011), *Photorhabdus luminescens* (Ghazal *et al*., 2014) and *Photorhabdus asymbiotica* (Wilkinson *et al*., 2009) whose whole genome sequences are available. The chosen sequence homologs are shown in Table 1. The amino acid sequences from the aforementioned bacteria were aligned using CLUSTALW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) in two different sets for both A1 and A2 proteins. The set 1 alignments were composed of sequences from only *Xenorhabdus* spp. while the set 2 on the other hand were composed of sequences from both *Xenorhabdus* and *Photorhabdus* spp.. This in turn yielded set 1 and 2 primer pairs for each protein group (A1 and A2).

From each multiple sequence alignment, the aligned amino acids on each end of the targeted region of the gene were picked for primer design. Amino acids with a high degeneracy were omitted from the choice. The chosen amino acids on the other hand, had their prerequisite nucleotide sequences obtained as shown in Table 1 with the degenerate codons having special letters filled in as shown in appendix 15.

Table 1: Sequence homologs of XptA1 chosen for use in designing degenerate primers.

No.	Gene/ Protein name	Organism	Genbank Accession number	UniProtKB/ TrEMBL
1.	TcdA1	*Photorhabdus luminescens*	NP_928296	Q7N7Y9
2.	TcdA1	*Pseudomonas syringae pv. Syringae* B728a	YP_237273	Q4ZNN7
3.	TcdA2	*Photorhabdus luminescens*	NP_928304	Q7N7Y1
4.	TcdA2	*Photorhabdus asymbiotica*	YP_003039736	C7BNH9
5.	XptA2	*Xenorhabdus nematophila* ATCC 19061	YP_003712778	D3VHH9
6.	XptA2	*Xenorhabdus bovienii* SS-2004	YP_003712781	D3VHI2

3.10.3. PCR amplification

A 25µL gradient PCR reaction was set up as follows: 1× PCR buffer (Genscript), 200µM of each dNTP, 0.2µM of each primer, 1unit of Taq polymerase enzyme (Genscript) and 21.1ng/µL of template genomic DNA. The amplification conditions were: Initial denaturation at 94°C for 3 minutes and 35 cycles of denaturation at 94°C for 30 seconds, annealing at 35-55°C for 1 minute, extension at 74°C for 1 minute as well as a final extension at 74°C for 7 minutes.

For the degenerate primers, a gradient PCR reaction was also set up. Three different primer sets were used as follows; A1VRPSet2, A2VRP1Set1 (described in Appendix 15) and a combination of a VRP1 gene-specific forward primer with A2VRP1Set1 reverse primer (A2S1VF). The amplification conditions were also as previously described but with the

annealing temperatures set as follows; 40-50°C for both A1VRPSet2 and A2VRP1Set1; and 50-60°C for A2S1VF.

All PCR products were run in a 1% TAE agarose gel, visualized and the targeted bands purified from the gel as previously described (subsection 3.4.2.2). These purified products, except the degenerate primers, were then used as template in another reaction set up in the same manner and run at an annealing temperature of 35°C for 1 minute. The resulting PCR products were processed similarly and then sequenced (Macrogen™, Amsterdam, Netherlands) via the Sanger platform as in subsection 3.4.2.2. The resulting sequences were analysed using MEGA 5.2 as previously described (subsection 3.4.2.3) and finally a similarity search carried out on the NCBI database using BLAST to identify the amplified gene fragments.

CHAPTER FOUR

4.0. RESULTS

4.1. Morphological and molecular characterization of *Xenorhabdus* sp.

4.1.1. Morphological characterization

The isolated *Xenorhabdus* bacterium from dead *Galleria mellonella* haemolymph showed the following morphological characteristics on 1% NBTA medium:- phase variation (I and II), blue colour, dendritic growth pattern (Figure 5A), raised convex circular colonies with slightly irregular margins and swarming motility on 0.8 to 1 percent NBTA media (Figure 5A). Phase II bacteria on the other hand were red in colour, had distinct circular colonies and did not show swarming motility on 0.8 to 1 percent media (Figure 5B). Gram staining revealed gram negative rod-shaped bacterial cells.

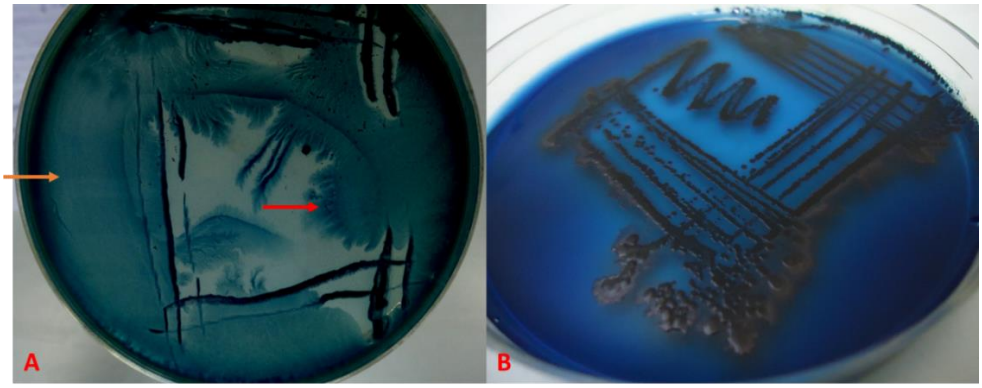


Figure 5: Cultural characteristics of *Xenorhabdus* bacterium grown on 1 % NBTA medium. **A**: Phase I bacteria showing swarming motility (orange arrow) and dendritic growth (red arrow). **B**. Phase II bacteria showing red colonies.

4.1.2. Molecular characterization

Extracted genomic DNA quantified using nanodrop yielded concentrations of between 50 and 70 ng/mL. As expected, PCR amplification of *Xenorhabdus* genomic DNA using 16S

rRNA gene primers gave a single band of ~ approximately 1500bp at 48°C annealing temperature (Figure 6). The resultant sequence queried against the NCBI non-redundant nucleotide database yielded *Xenorhabdus griffinae* as the best hit. It gave an identity of 98% from a query cover of 98% (Appendix 1 and 2). Subsequent phylogenetic analysis of the top eleven BLAST hits and other additional *Xenorhabdus* sequences via the maximum likelihood (ML) tree inference method in MEGA 5.2 revealed that the query sequence clustered with *X. griffinae* (Figure 7 (red label)).

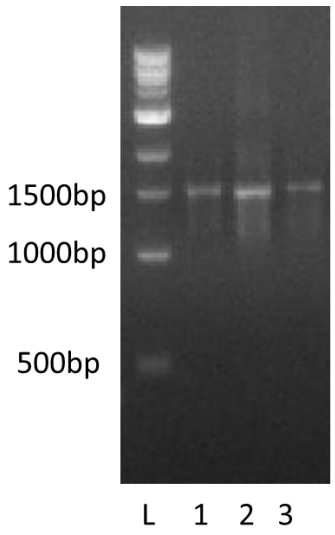


Figure 6: Gel electrophoresis of PCR amplicons of 16S rRNA gene of *Xenorhabdus* sp.: Lane L, 1Kb ladder (New England Biolabs); lane 1, 2 and 3, replicates of 16S rRNA bands.

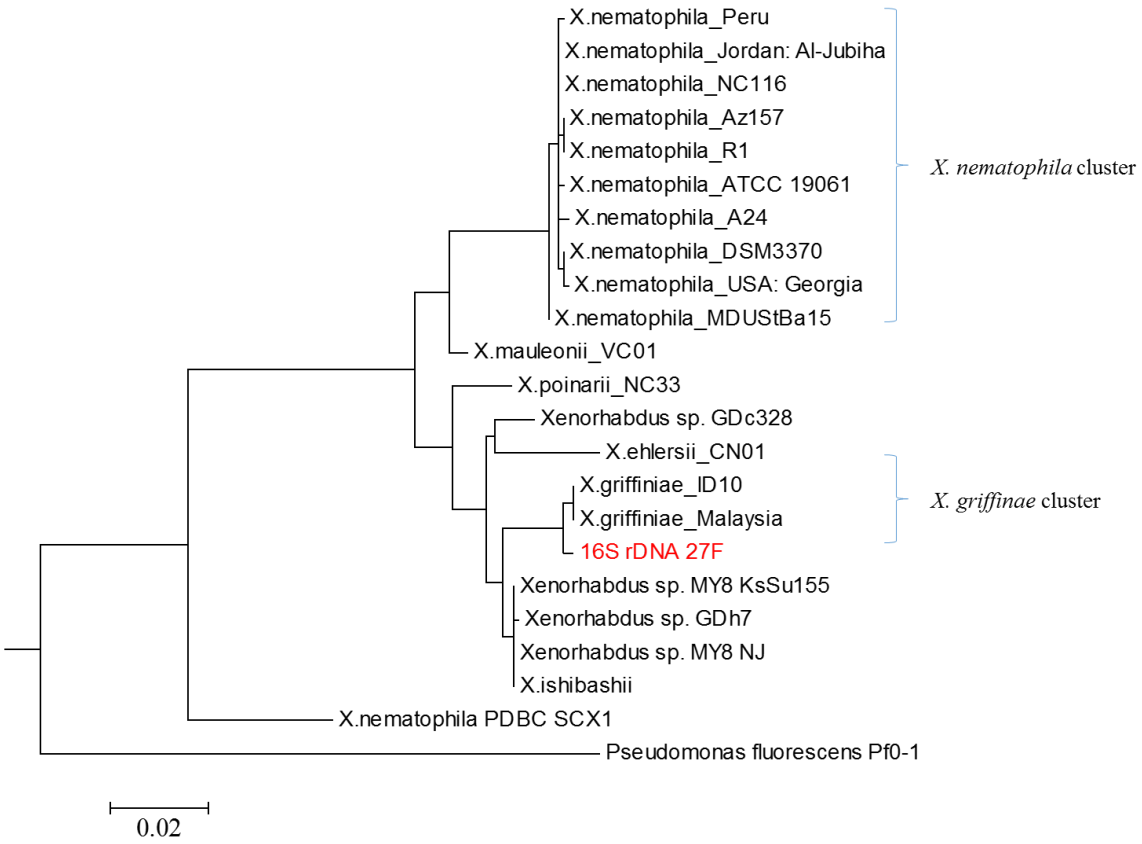


Figure 7: Phylogenetic tree of *Xenorhabdus* 16S rRNA gene sequences. The scale bar indicates nucleotide substitutions per site.

4.2. Bioassay of *Xenorhabdus* sp. against *S. zeamais* and *P. truncatus*

4.2.1. Re-isolation of bacteria from maize pellets

Bacteria recovered from the maize pellets at 6, 12, 24 and 46 hour intervals (Figure 8A to D respectively) showed characteristics similar to the Phase I bacteria of the starting culture (Figure 5). However, bacteria recovered on the 6th day had red distinct circular colonies with reduced elevation and irregular margins, did not show a dendritic growth pattern and had no swarming motility (Figure 8E). These were characteristics reminiscent of Phase II bacterial observed in Figure 5B.

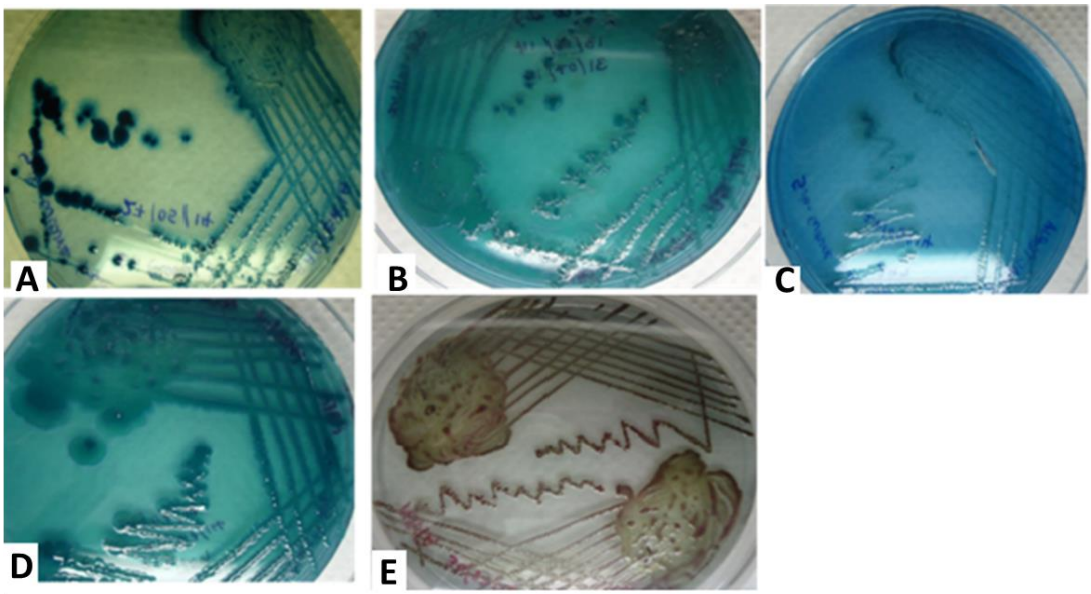


Figure 8: Cultural characteristics of *Xenorhabdus* bacterium, phase I (A to D) and phase II (E), recovered from maize pellets at different time intervals. **A.** 6 hours. **B.** 12 hours. **C.** 24 hours. **D.** 46 hours. **E.** 6 days.

4.2.2. Effect of *Xenorhabdus* sp. against the maize weevil (MW)

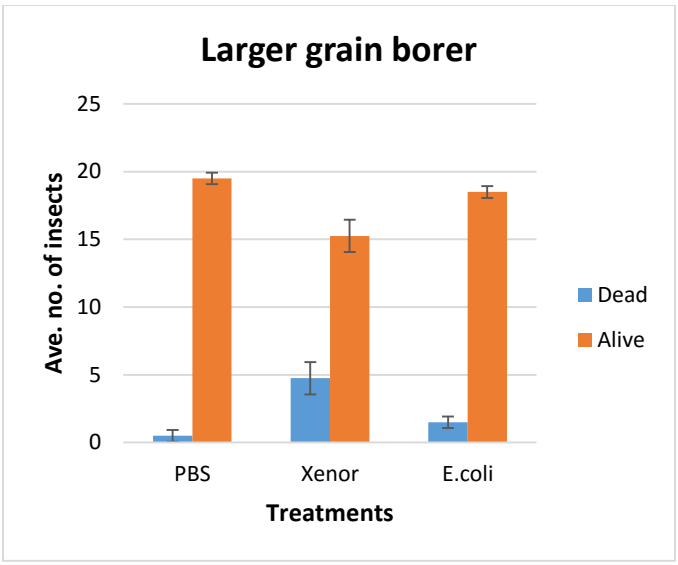
Statistics from the maize weevil bioassay indicated that both pellet destruction and mortality were dependent on the treatment $(\chi_4^2 = 35.28; p < 0.05)$ and $(\chi_2^2 = 78.32; p < 0.05)$ respectively. The data from the treatments showed higher mortality rates for the maize weevil (MW) (average of 35.8%) than for the larger grain borer (LGB) (average of 11.25%) $(\chi_2^2 = 11.82; p < 0.05)$ (Figure 9). The highest rate of mortality was observed in the *Xenorhabdus* treatment (14.75±1.34). Feeding results showed generally more damage caused by the LGB (average of 31%) than the MW (average of 24.3%) $(\chi_2^2 = 15.55; p < 0.05)$ (Figure 10).

Pellet destruction was highest in the PBS treatment (88%). However, in this treatment, there was a higher percentage of pellets having entrance holes (58%), than those that were

completely destroyed (30%). On the other hand, in the bacteria treatments, the completely destroyed pellets formed the greatest proportion (40%) of damaged pellets. Comparatively, the *E. coli* treatment had more pellets with entrance holes at 29% (compared to 22% in *Xenorhabdus* treatment) and fewer undamaged pellets at 31% (compared to 38% in *Xenorhabdus* treatment). Overall, there were more pellets with entrance holes in the control treatments than in the *Xenorhabdus* treatment.

4.2.3. Effect of *Xenorhabdus* sp. against the larger grain borer (LGB)

Data from the LGB assay indicated that both pellet destruction and mortality were dependent on the treatment $(\chi_4^2 = 117.87; p < 0.05)$ and $(\chi_2^2 = 19.78; p < 0.05)$ respectively. The highest rates of mortality were observed in the *Xenorhabdus* treatment (4.75±1.19), followed by *E. coli* (1.5±0.43) then the PBS treatment had the least (0.5±0.43) (Figure 9). Feeding results showed complete destruction of the pellets in the PBS treatment. The *Xenorhabdus* treatment had 21% undamaged, 37% bored (with entrance holes) and 42% completely damaged pellets. The *E. coli* treatment on the other hand had 8% bored pellets and 92% completely destroyed pellets. Pellet destruction by LGB was thus observed to be greatly reduced in the *Xenorhabdus* treatment. Overall, feeding inhibition was greater in the LGB than in the weevil (Figure 10).



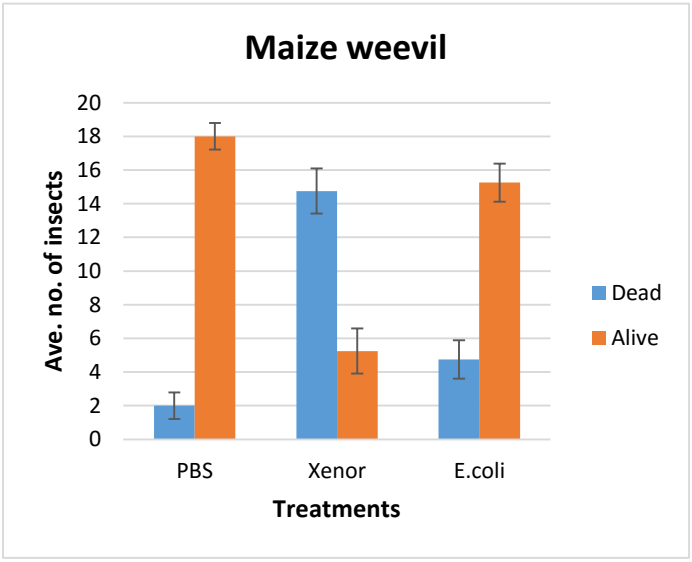
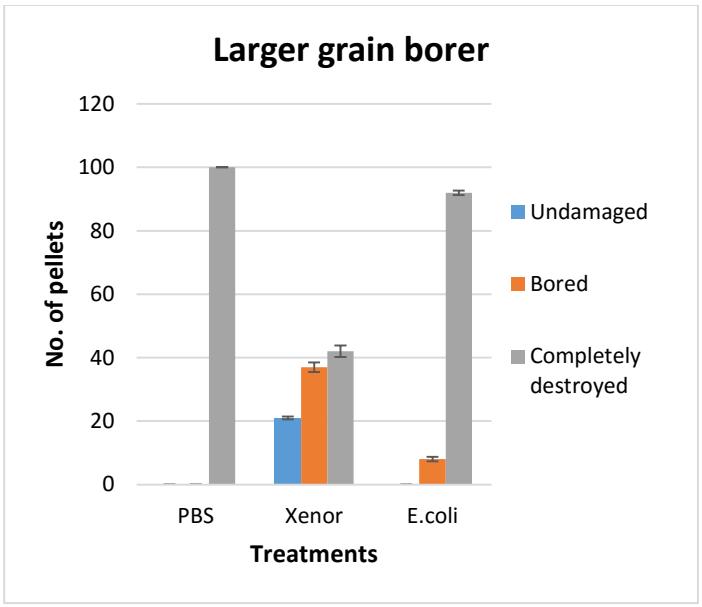


Figure 9: Bioassay results showing the mortality rates of the larger grain borer (LGB) and maize weevil (MW) treated with Xenorhabdus, n=80.

Key: PBS = Negative control

Xenor = Xenorhabdus

E. coli = *E. coli* DH5α



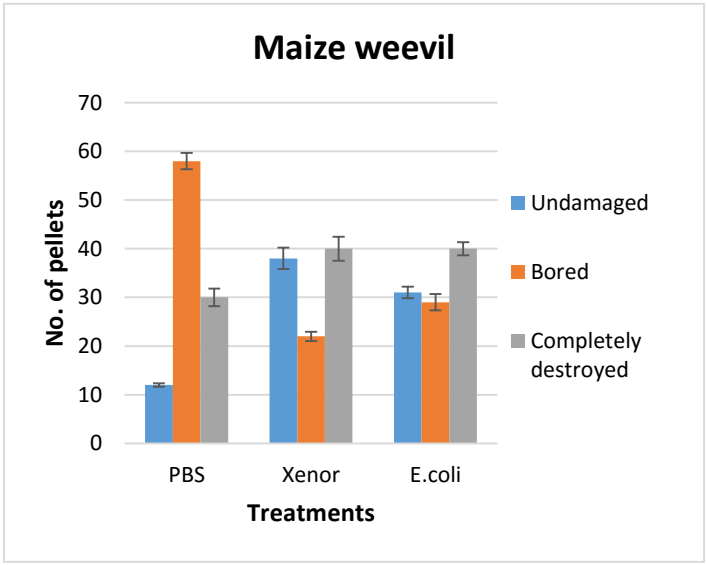


Figure 10: Bioassay results showing the rates of feeding of the larger grain borer (LGB) and maize weevil (MW) treated with Xenorhabdus and *E. coli*. n=100.

Key: PBS = Negative control

Xenor = Xenorhabdus

E. coli = *E. coli* DH5α

4.3. Bioinformatic analyses of class A genes and proteins

Genes and proteins used in the analyses (appendices 2 and 3) were retrieved from public databases using the XptA genes of *X. nematophila* as query sequences. All these class A sequences were homologous but had varying nucleotide and amino acid sequence lengths. They were identified in a wide array in bacteria covering, insect, mammalian (including human) and plant pathogens.

4.3.1. Protein families, domains and motifs of class A proteins

Analysis of class A proteins and genes in the different databases confirmed their identity as insecticidal toxin complex proteins. All the toxins studied here compared to Tc toxins of *Photorhabdus* spp. either in naming and/or description. They were also categorized in line with the classification of the Tc toxins. Four classes of Tc protein complexes were identified by Bowen *et al*. (1998) Tca, Tcb, Tcc and Tcd. The Tc complexes appeared to be the most well studied and were the front runners in this field. Therefore, all other toxins homologous to them were similarly classified. The class A proteins were also named in line with their respective toxin complex (TcaA, TcbA, TccA and TcdA).

Conserved domains were observed in all the proteins (Figure 11). These domains were identified and labelled according to work done by Meusch *et al*. (2014). The genes appeared to be divided into two main parts, the α-helical and the pore-forming. These two parts were then further subdivided as shown in Figure 11. Receptor-binding domains A, B, C and D were found to be absent in some of the proteins. TcdA2 from *P. luminescens* lacked receptor binding domain A, SepA from *Serratia entomophila* lacked domain D while TcA from *Y. pestis* lacked domains B, C and D. Of particular interest was VRP1

domain, originally identified in the *Salmonella* virulence plasmid SpvA family (Interpro ID: IPR003518), which was conserved in all the proteins.

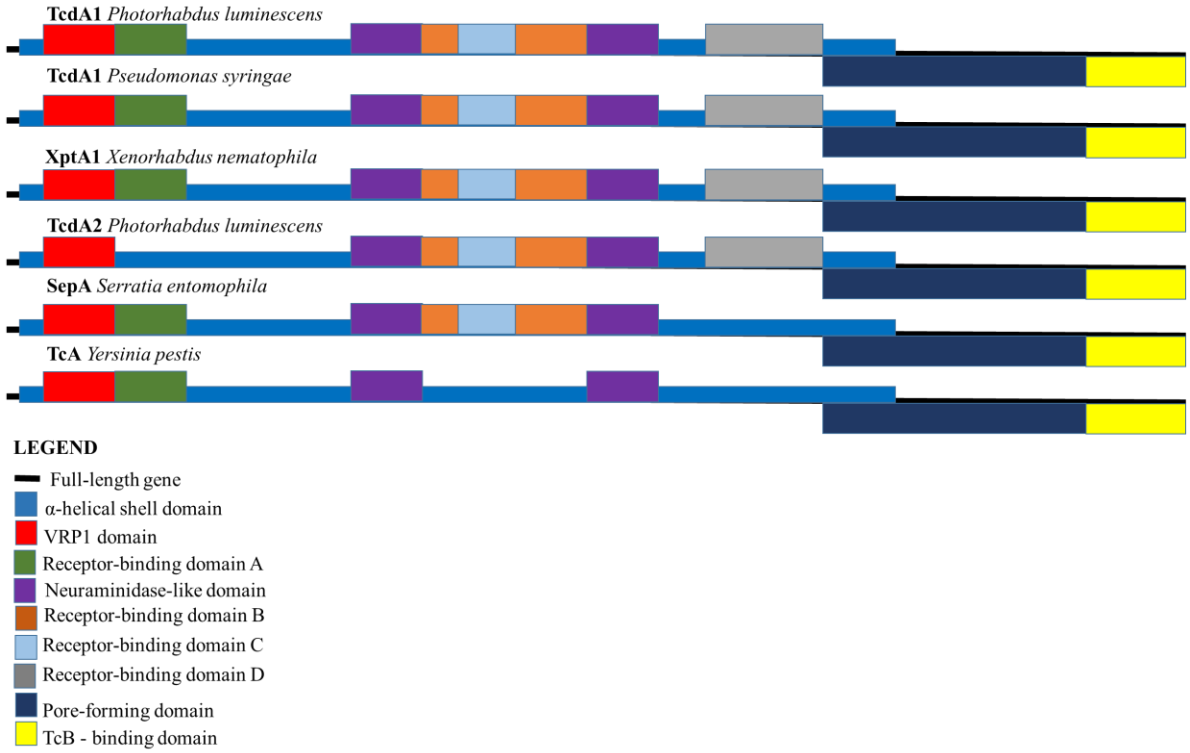


Figure 11: Domain organization of the class A toxin proteins. Each colour represents a particular domain as shown in the legend.

Analysis of VRP1 on Pfam confirmed its presence in all the class A proteins from *Xenorhabdus* spp. including their homologs from the other organisms thus suggesting a wide distribution. BLAST searches of the retrieved domain sequences all confirmed the VRP1 domain from the *Salmonella* virulence plasmid. Its presence in all the searched sequences gave further evidence of its wide distribution in class A proteins. The VRP1 domain was found to vary in sequence length in different class A sequences and appeared to have lost parts of its sequence in some of these proteins. Its location on the genes however, was relatively constant and was generally found toward the 5' end in all class A

sequences. Appendices 4 and 5 show some of the retrieved domains and the lengths of their sequences.

4.3.2. Analysis of VRP1 domain sequences

Two reference domain sequences of VRP1; pfam03538 (325 residues) retrieved from the Pfam database and PRK15212 (255 residues), obtained from NCBI's conserved domain database (CDD), were used in the identification of VRP1 domains and conserved motifs present in different bacteria. The two sequences showed differences, pfam03538 had insertions which were absent in PRK15212, this explained its longer sequence. However, the predicted motifs were largely similar in both sequences save for the ones in the inserted sections (Figure 12).

An analysis of all the retrieved domain sequences on CDD (NCBI) showed that, in general, all of the VRP1 sequences from *Salmonella* spp. aligned against the PRK15212 reference sequence while the rest aligned against pfam03538. Domain searches on CDD (NCBI) revealed complete domains (spanning the full length of the reference sequence) in all sequences from *Salmonella* spp., as well as TccA sequences from *Photorhabdus* spp., *Pseudomonas fluorescens* and *Xenorhabdus bovienii*. The other sequences were incomplete, covering only a section of the reference sequence rather than the full sequence. They appeared to be truncated at either one or both terminal ends of the sequences.

All of the class A domain sequences, except TcdA1 and TccA1 from *P. luminescens*, and the TccA2 sequences, showed truncation at both the N and C-terminal ends. They all lacked the peripheral residues of the reference sequence. TcdA1 from *P. luminescens* had only the C-terminal end truncated while TccA1 was complete. The TcdA and XptA sequences gave

some of the lowest bit scores whilst the highest scores were obtained from the *Salmonella* sequences.

A comparison of all the individual VRP1 sequences with the reference sequences revealed an average identity of 27.9%. Generally, the XptA and TcdA sequences showed lower sequence similarity to the PRK15212 sequence (average of 11.6%), compared to an average of 29.7% similarity to the pfam03538 sequence. Similarly, the *Yersinia, Erwinia* and *Burkholderia* sequences showed higher overall similarity to the pfam03538 sequence.

Prediction of conserved motifs by Motif scan (ExPASy) showed that sequence pfam03538 had motifs in the region between residues 1-325 (raw score = 767.4, E-value = 2.3e-229, normalized score = 235.964), while sequence PRK15212 had motifs in the region between residues 1-255 (raw score = 408.8, E-value = 2.1e-121, normalized score = 128.004). These predicted VRP1 motifs of the XptA, Tc and Tc-like proteins from Motif scan were similar to VRP1 motifs identified in *Salmonella* sequences through a search on FingerPRINTScan, a data bank of protein family fingerprints. Seven conserved motifs were identified in *Salmonella,* belonging to the SALSPVAPROT fingerprint family (PRINTS accession number: PR01340) with a profscore of 7202 and E-value of 1.8e-86 (Figure 12).

Motifs number three and four had the largest number of conserved amino acids in all species (Figure 12B). The sequence between the two motifs also had highly conserved amino acids which were not included in the SALSPVAPROT fingerprint (Figure 12A and B (red-coloured motifs)). Interspecific amino acid changes were noticed that did not alter the motif. The high conservation of the third and fourth motifs coincided with the observed truncation. The truncated sequences had the terminal fragments missing but with the middle intact. This conserved middle portion contained the third and fourth motifs.

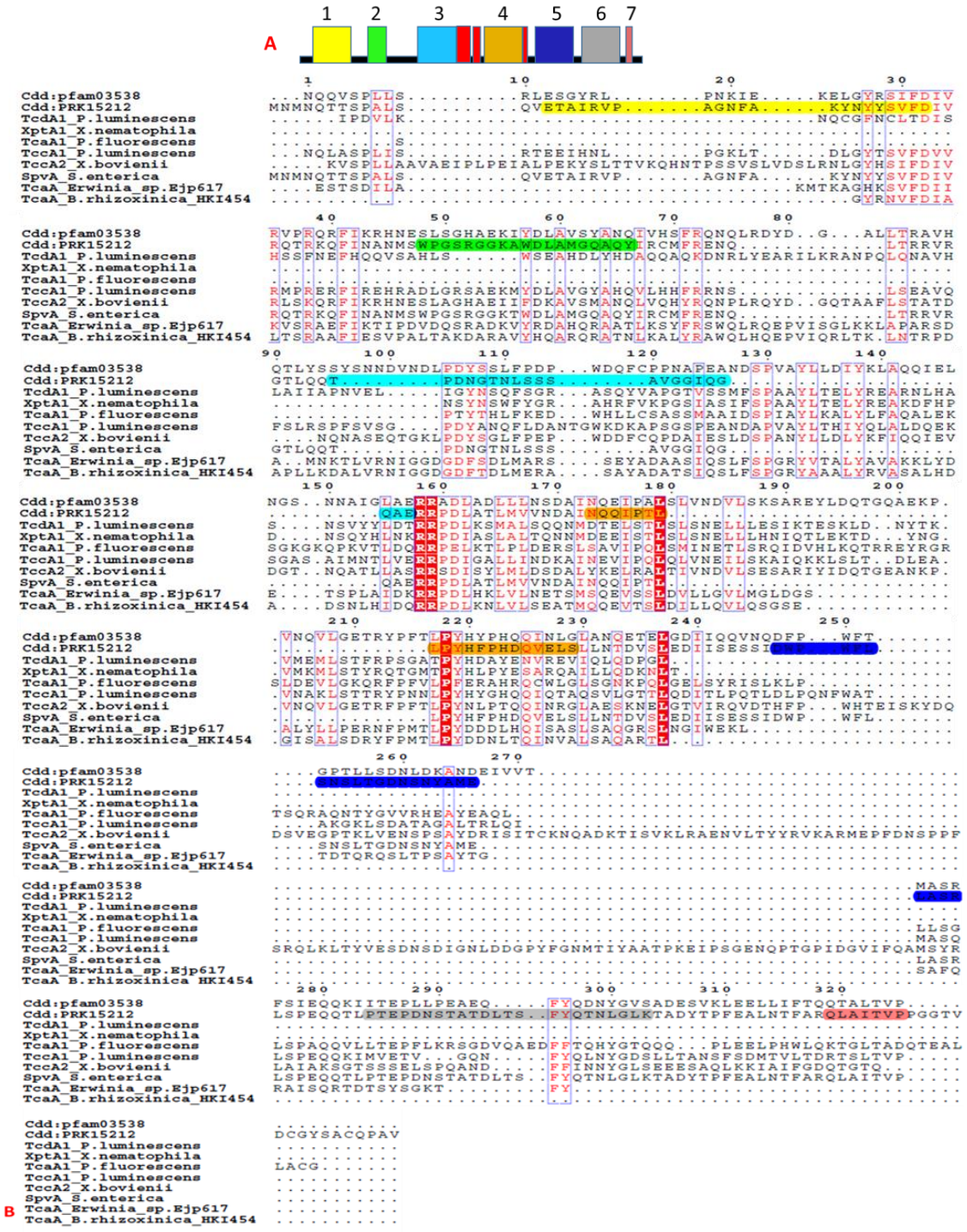


Figure 12: Organization of the VRP1 domain motifs (**A**) in the SALSPVAPROT fingerprint family and (**B**) their conservation in different proteins. Key: The coloured highlights represent the different motifs: Yellow = motif 1, green = motif 2, blue = motif 3, orange = motif 4, blue = motif 5, grey = motif 6 and pink = motif 7. The red highlights represent other conserved inter-motif amino acids not captured in the fingerprint family.

A maximum likelihood tree of all the retrieved VRP1 domain sequences was reconstructed to assess the relationship of the VRP1 domains from the different bacteria. XptA, TcdA and SepA domains clustered together. Similarly, *Salmonella* sequences also clustered together (Figure 13). The TccA2 formed a clade with TccA1 from *P. luminescens*, these had shown a complete domain and high similarities to both PRK15212 and pfam03538 reference sequences from CDD and Pfam. There appeared to be an increase in sequence variation moving up the tree from the root.

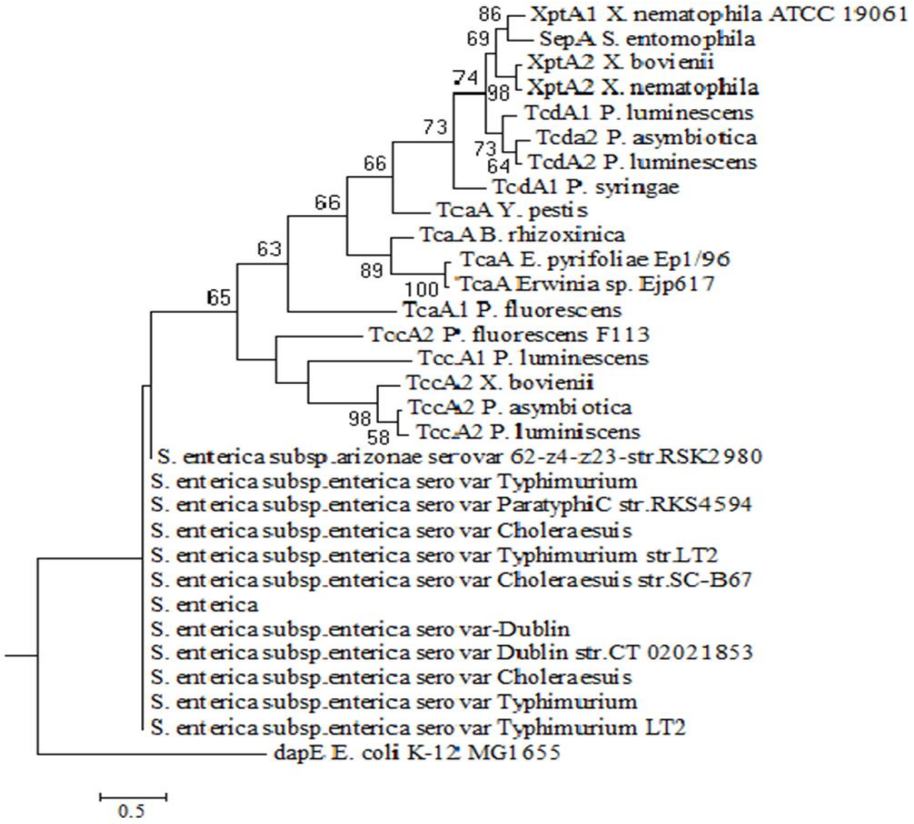


Figure 13: Maximum likelihood tree of all VRP1 domain sequences retrieved from Pfam, and CDD databases. The scale bar represents the number of amino acid substitutions per site.

4.3.3. Prediction of the 3D structure of Class A insecticidal proteins

4.3.3.1. Template selection and verification

A homology search on the HHpred server using both the VRP1 and full XptA1 amino acid sequences, returned TcdA1 (PDB ID: 1VW1A) Tc toxin crystal structure from *P. luminescens* as the PDB top hit with a score of 5689.9, probability of 100 and a resolution of 3.5Å. This hit had a good coverage of the target protein, however, it was incomplete in the C-terminal portion of the XptA1 protein. It covered 2516 residues out of the 2523 residues of XptA1. Searches of the other class A sequences and their corresponding VRP1 domain sequences also yielded 1VW1A as the top hit. The crystal structure of TcdA1 as found in PDB is as shown in appendix 13.

The 1VW1 structure is a complex of five identical TcdA1 proteins joined together. To enable comparison of this structure with the target A models, chain A (1VW1A) representing one of the TcdA1 proteins served as the template (Appendix 13). Validation using model assessment tools confirmed the template to be of good quality (Appendix 8). The VRP1 region of the 1VW1A full protein structure was identified through a sequence search in Pfam to be at the N-terminal portion of the sequence spanning 211 residues. It was also modelled using 1VW1A as the template. Subsequent validation also confirmed its good quality as a template (Appendix 8).

4.3.3.2. Model building

The resultant models showed similarity in line with the clusters formed in the maximum-likelihood tree (Figure 13). VRP1 models from XptA, TcdA and SepA proteins showed greater homology with each other. These models showed the greatest similarity to the template both in their full and VRP1 structures, they are thus referred to as template-like.

The rest which deviated from the template are henceforth collectively referred to as non-template-like. Similarity in the template-like models was found at the middle and C terminal regions of the structures (Figure 14). However, their similarity with the non-template-like models was observed in the middle region only.

The TcdA models were larger than the XptA due to their longer amino acid sequences. Of the template-like, the TcdA1 domain had a greatest number of α-helices, the rest differed by a single α- helix. The non-template-like models had longer loop regions than the template-like (Appendix 9 - 12). The A2 VRP1 models had the longest loop regions, indicating their divergence from the template. The VRP1 secondary structures of revealed α-helices and no β-sheets. This was in line with the placement of this domain in the larger α-helical domain of the class A toxins. The most conserved secondary structures (Figure 14) coincided with the highly conserved motifs 3 and 4, as well as their highly conserved inter-motif amino acids (Figure 12). The secondary structures from the A2 VRP1 models were similar to those of the A1 VRP1 models. They also showed similar conserved structures (Appendix 10 and 11).

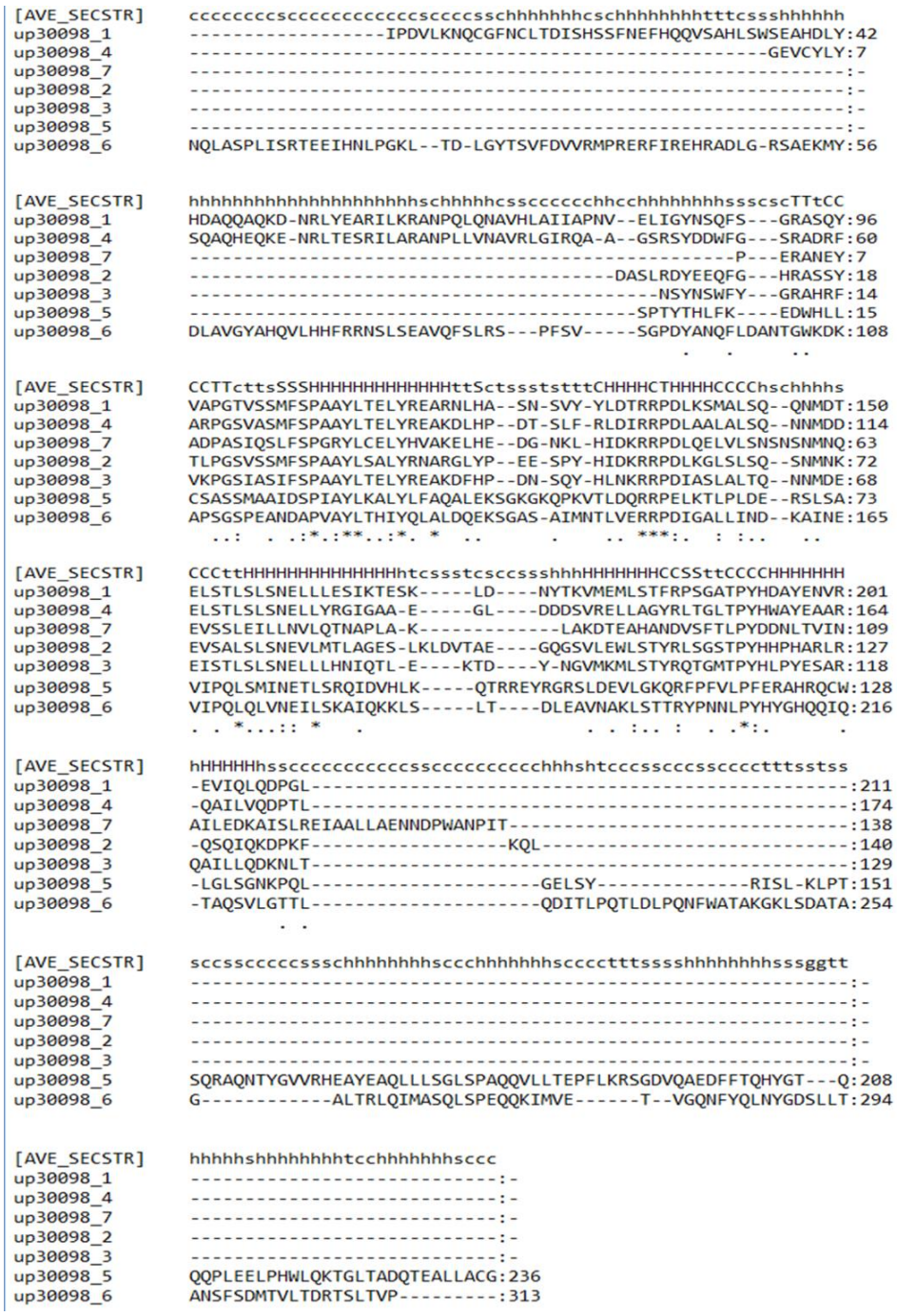


Figure 14: Multiple sequence alignment of representative A1 VRP1 model sequences retrieved from the databases. The bottom line shows the sequence identity (*), conserved residues (**:**) and semi-conserved residues (**.**).

The sequences up30098_1, _2, _3, _4, _5 and _6 in Figure 14 represent XptA1, TcdA1 from *P. luminescens*, TcdA1 from *P. syringae*, TcaA1 from *P. fluorescens*, TccA1 from *P.*

luminescens, SepA from *S. entomophila* and TcaA1 from *Y. pestis* VRP1 models respectively. The row [AVE_SESTR] is the average secondary structure of each site. The letters H, S, T and C in this row represent alpha helix, bend, hydrogen-bonded turn and coil respectively. Fully and partially conserved sites are shown in upper and lower case letters, respectively.

As observed in the domain models, the full protein models also resembled each other in line with the VRP1 clustering in the maximum-likelihood tree. XptA and TcdA models closely resembled each other (Figure 15 and 16). The SepA full protein model showed a high level of similarity with the XptA and TcdA structures. The TccA2 models also showed a degree of resemblance. The non-template-like A structures showed more structural differences in terms of topology (Figures 15 and 16). They had long loop regions and deviated from the structure described by Gatsogiannis *et al*. (2013) (Figure 15A). Furthermore, they appeared to be incomplete forms of the template. This was clear at the onset due to their much shorter sequences and gaps in sequence alignments.

The template-like models displayed the general three domain (α-helical, β-sheet and pore-forming domains) structure described by Gatsogiannis *et al*, (2013). Furthermore, they showed the more detailed structure described by (Meusch et al. 2014) (Figure 17). This latter structure consisted of all the identified domains. However, missing domains in the TcdA2 models were noted, but, these were not clear for the SepA model. Multiple sequence alignments showed these missing domains as gapped regions of the alignment. These missing domains coincided with those previously described (Figure 11).

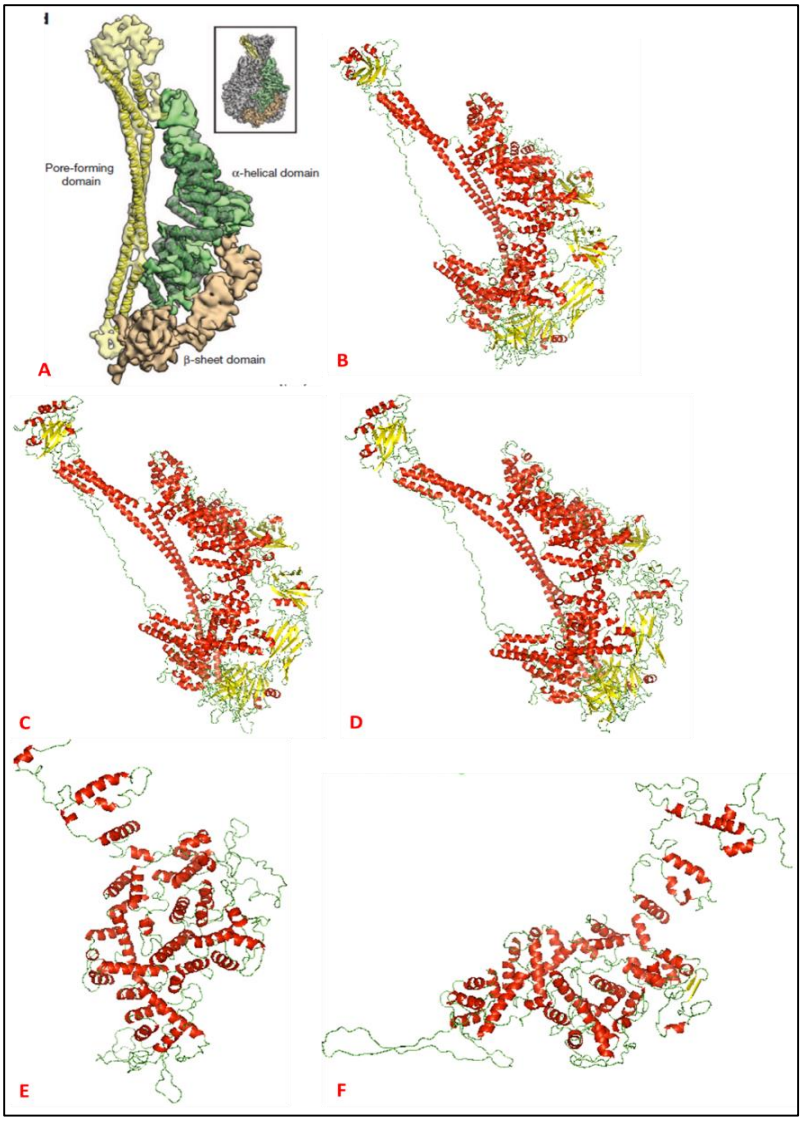


Figure 15: A1 full-length predicted protein models.

A. TcdA1 from *P. luminescens* (adapted from Gatsogiannis et al., 2013). **B.** XptA1 from *X. nematophila.* **C.** TcdA1 from *P. luminescens.* **D.** TcdA1 from *P. syringae.* **E.** TcaA1 from *P. fluorescens.* **F.** TccA1 from *P. luminescens.* The models are coloured to show the different secondary structures corresponding to the different classifications shown in Figure A, red for the α-helix, yellow for the β-sheet and green for the loops.

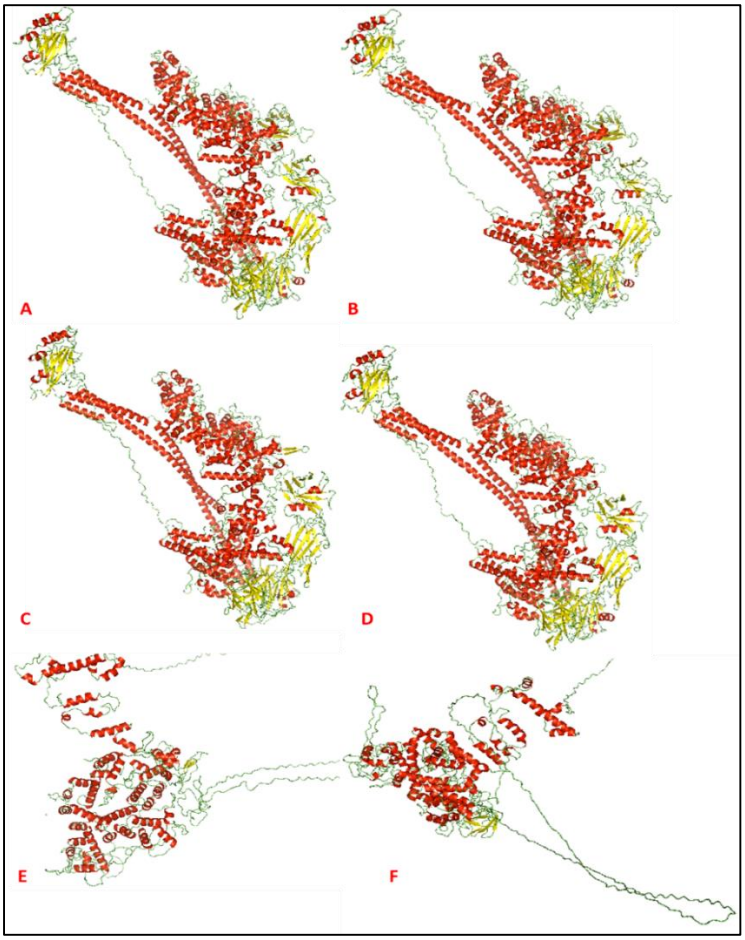


Figure 16: A2 full-length predicted protein models.

A. XptA2 from *X. nematophila*. **B**. XptA2 from *X. bovienii*. **C**. TcdA2 from *P. luminescens*. **D**. TcdA2 from *P. asymbiotica*. **E**. TccA2 from *X. bovienii*. **F**. TccA2 from *P. fluorescens*. The models are coloured to show the different secondary structures corresponding to the different classifications shown in figure 15A, red for the α-helix, yellow for the β-sheet and green for the loops.

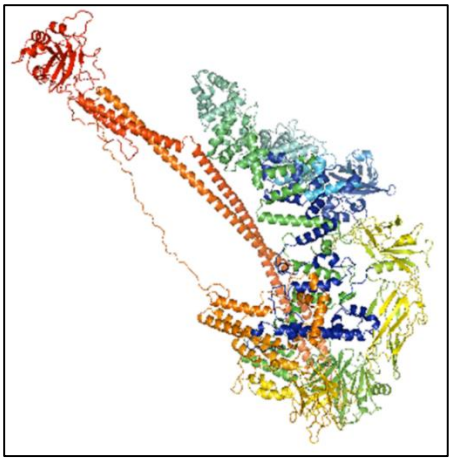


Figure 17: The general 3-Dimensional structure of the full A protein coloured in a spectrum of colours ranging from blue in the N-Terminal region to red in the C-terminal.

4.3.3.3. Model validation

Tables 2 and 3 contain a detailed summary of the validated models, showing the scores of model validation. PROCHECK validation of the structures showed that the models with the most residues in the most favoured region were the TcdA and XptA, this was seen for both the full proteins and their corresponding VRP1 models. These structures also had some of the lowest RMSD scores in the whole group. Furthermore, they had some of the highest GDT_TS scores. These scores from the different tools indicated that these were relatively good quality models and confirmed that they were the most closely related to the template. The TcdA2 models on the other hand showed slightly lower scores than the A1, indicating their slight difference from the template (the missing domains).

The TccA2 and TccA1 models gave very high RMSD scores indicating their distant relationship to the template. The TcaA1 VRP1 model gave the lowest RMSD implying that it was very close in homology to the template, however, its full protein gave a relatively high RMSD score, although not as high as the TccA2 and TcaA1 models. This suggested

that it was more closely related to the template than the TccA2 and TcaA1 structures. The overall poor scores of the non-template-like models were evidence of the extent of their differences with the template.

Table 2: A summary of the scores of model validation for all the full class A proteins.

MODEL	MetaMQAPII GDT_TS	RMSD	PROCHECK Residues in most favoured region
TcaA1 P. fluorescens.pdb	28.616	4.014	89.5%
TccA1 P. luminescens.pdb	28.753	11.837	84.6%
TcdA1 P. luminescens.pdb		0.940	91.8%
TcdA1 P. syringae.pdb	51.039	3.867	91.8%
XptA1X. nematophila.pdb	51.397	1.330	90.4%
TccA2 P. asymbiotica.pdb	50.866	18.237	83.1%
TccA2 P. fluorescens.pdb	50.941	21.941	80.4%
TccA2 P. luminescens.pdb	51.002	19.558	82.6%
TccA2 X. bovienii.pdb	50.971	12.849	85.9%
TcdA2 P. asymbiotica.pdb	50.897	0.943	91.4%
TcdA2 P. luminescens.pdb	50.757	1.732	90.7%
XptA2 X. bovienii.pdb	53.456	0.986	89.9%
XptA2 X.nematophila.pdb		0.951	89.7%
SepA S. entomophila.pdb	50.758	3.289	89.1%
TcaA1 Y. pestis 91001.pdb	38.628	3.945	87.7%

Table 3: A summary of the scores of model validation for all the class A VRP1 domain structures.

MODEL	MetaMQAPII GDT_TS	RMSD	PROCHECK Residues in most favoured region
TcaA1 P. fluorescens.pdb	28.616	0.472	88.0%
TccA1 P. luminescens.pdb	17.492	7.682	89.7%
TcdA1 P. luminescens.pdb	26.066	1.965	92.8%
TcdA1 P. syringae.pdb	17.321	1.356	92.6%
XptA1X. nematophila.pdb	16.085	1.582	94.0%
TccA2 P. asymbiotica.pdb	26.282	4.581	88.3%
TccA2 P. fluorescens.pdb	13.507	2.427	90.2%
TccA2 P. luminescens.pdb	27.500	6.916	89.9%
TccA2 X. bovienii.pdb	25.690	2.558	92.1%
TcdA2 P. asymbiotica.pdb	16.373	3.664	92.9%
TcdA2 P. luminescens.pdb	19.930	4.449	93.8%
XptA2 X. bovienii.pdb	13.346	2.410	90.9%
XptA2 X.nematophila.pdb	17.829	2.320	93.9%
SepA S.entomophila.pdb	17.816	2.851	94.9%
TcaA1 Y.pestis 91001.pdb	12.500	4.269	89.8%

4.3.4. Comparison of protein models by superimposition

4.3.4.1. Comparison of VRP1 models

All the template-like VRP1 models aligned with each other (Figure 18A and 19A). However, they also aligned with the TcaA1 and TccA1 structures from *P. fluorescens*, *Y.*

pestis 91001 and *P. luminescens* respectively (Figure 18B and 19B). The alignment, shown as the coloured regions in the images, confirmed homology in the middle portion of the structures where there was also sequence similarity. This also coincided with the region where the most conserved motifs had been identified. Point mutations in these regions did not seem to have a significant effect on the overall tertiary structure, all the A VRP1 models seemed to retain a common structure in that middle portion (Figure 18C and 19C). Comparison of A1 and A2 VRP1 structures and sequences further elucidated the regions of conservation (Figure 20).

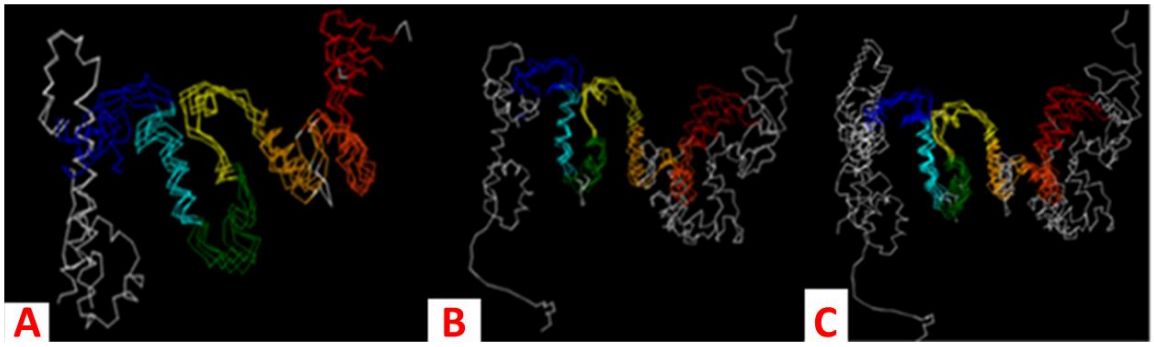


Figure 18: Comparison of A1 VRP1 predicted 3D structures. The coloured regions represent the regions of homology.

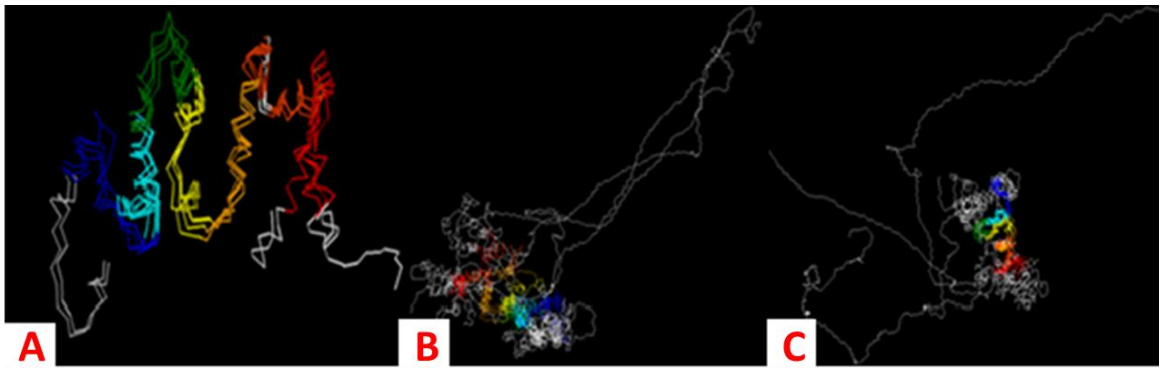


Figure 19: Comparison of A2 VRP1 predicted 3D structures. The coloured regions represent the regions of homology.

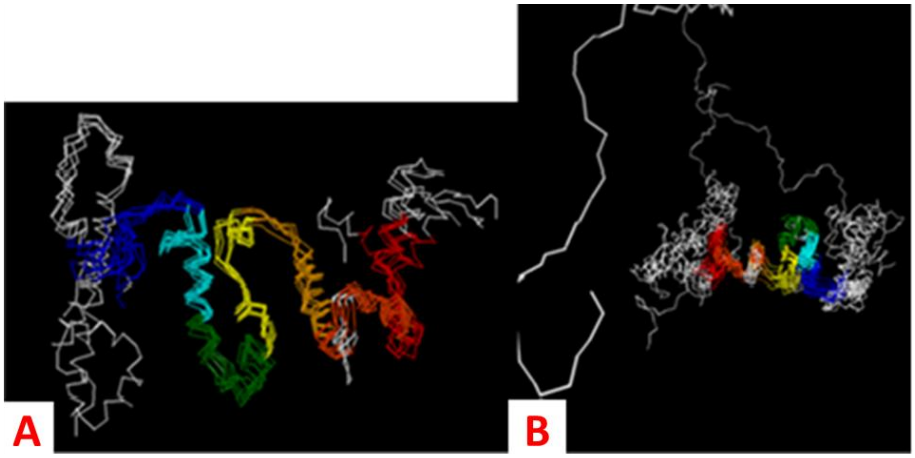


Figure 20: Comparison of the A1 and A2 VRP1 3D models. The coloured regions represent the aligned parts of the structures while the white regions represent the misaligned parts.

4.3.4.2. Comparison of full protein models

A comparison of the full protein models yielded similar results to those from the VRP1 domain. The template-like models aligned very well save for some few areas. As previously observed, the TcdA2 and SepA proteins appeared to have missing parts in the β-sheet region (the white parts in Figure 21A). A confirmation with their respective sequences in the multiple sequence alignment revealed gaps in the regions corresponding with the missing parts. Only the XptA appeared to be completely identical to the template (TcdA1).

A comparison of the template-like and non-template-like revealed a smaller region of homology in the α-helical part of the structures (Figure 21B). The TcaA1 models from *P. fluorescens* and *Y. pestis* showed significant similarity with each other in structure but with relatively poor homology at an RMSD score of 25.8 from PyMol (Figure 21C). The TccA2 models on the other hand (Figure 21D) showed greater homology with each other than with

other models. All the non-template-like models showed homology with each other in the α-helical region (Figure 21E).

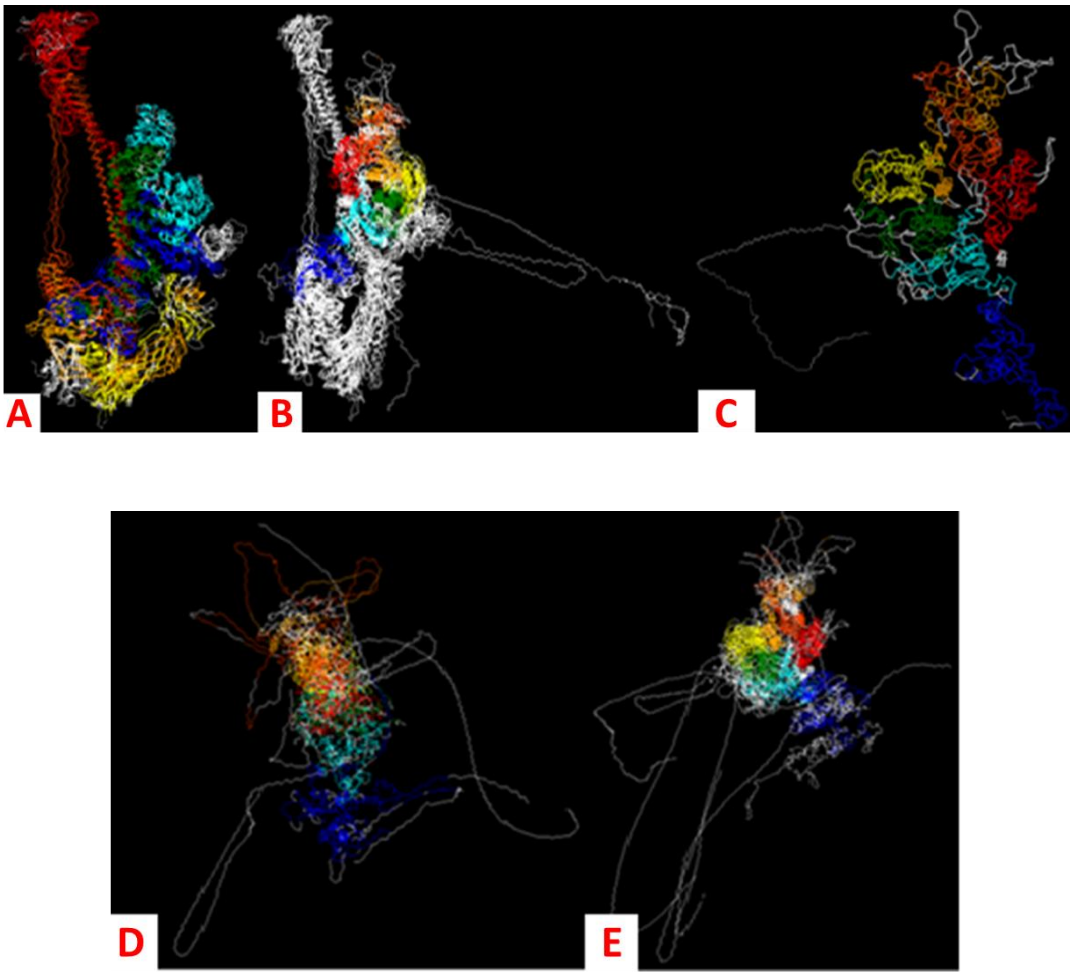


Figure 21: Comparison of the A1 and A2 full protein models, template-like (**A** and **B**) and non-template-like (**C**, **D** and **E**). The coloured regions represent the aligned parts of the structures while the white regions represent the misaligned parts.

4.3.5. Position of the VRP1 domain within the full A proteins

The VRP1 domain of all the models aligned to the N-terminal end in the α-helical region of the full models (Figure 22). This alignment position coincided with the region in the α-helical domain where all of the full protein models had aligned when superimposed (Figure

21A). Furthermore, it coincided with the region of homology observed in multiple sequence alignments of the full protein sequences.

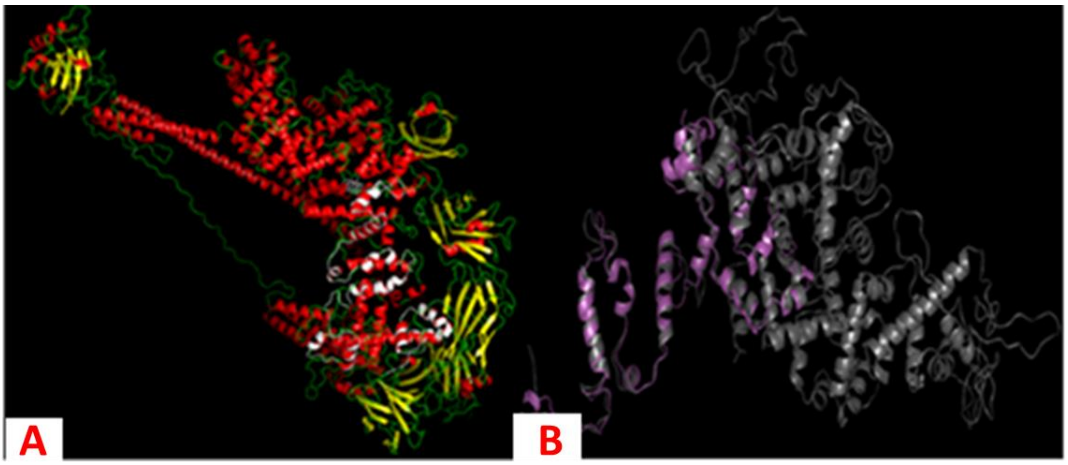


Figure 22: Superimposition of representative models, template-like (**A**) and non-template-like (**B**), to find the position of VRP1 domain within the full protein.

4.3.6. The VRP1 paradox: Related but contradictory domains

A search for the VRP1 domain on the Interpro database yielded two results for that search term, these were Verprolin (IPR028293) and Insecticidal toxin complex/plasmid virulence protein (IPR018003). The latter was the main target. Upon closer inspection, it was revealed that, Verprolin is a budding yeast (*Saccharomyces cerevisiae*) protein with homologs found in vertebrates, for example, Wiskott-Aldrich syndrome protein (WASP)-interacting protein (WIP) (Munn & Thanabalu, 2009). Verprolin is required for normal growth and the correct maintenance of the actin cytoskeleton (Donnelly *et al*., 1993). It therefore appeared that the two protein domains work on the actin cytoskeleton but, Verprolin (VRP1) plays the agonist role, while the insecticidal VRP1 took on the antagonist role. A comparison of the features of the two proteins revealed the following:-

Verprolin (as described by Donnelly *et al*. (1993).	**Insecticidal VRP1 domain**
It has a predicted high content of proline residues (24%). Because of this, the gene was named VRP1 and its product, Verprolin, which is also very rich proline.	All the VRP1 sequences used in this study showed a significantly less proportion of proline residues averaging at~ 5.4%. It is not clear why this domain superfamily was named VRP1.
The high number of proline residues are arranged either in homogeneous stretches of three to nine residues, or in various motifs.	These sequences did not show any stretches of proline residues or motifs.
The presence of these proline-rich motifs gives the protein a repetitive domain structure	There were no repetitive proline-rich domain structures identified on these sequences.
Verprolin has a serine, alanine and threonine content of 15%. 10% and 5.6% respectively.	These sequences showed proportions of serine, alanine and threonine averaging at 9.06%, 8.3% and 7.226% respectively.
It is highly charged, and basic (5.2% lysine) with a pi of 10.68.	Only a few sequences achieved a 5.2% Lysine content; TcaA1 *P. fluorescens*, TccA2 *P. asymbiotica*, TccA2 *P. luminescens*, TcaA *E. pyrifoliae* and TcaA *Erwinia* sp. Ejp617. Many had a predicted pi of~4.81. Many of the domains appeared to be acidic with a combined glutamic acid and aspartic acid content average of~ 10.5%.
Verprolin is required for both normal growth and the correct maintenance of the actin cytoskeleton.	The N-terminal region of TcdA proteins, which contain the VRP1 domain, was reported to cause actin condensation (Hinchliffe *et al.*, 2010).

4.3.7. Phylogenetic relationship of the bacteria containing the VRP1 domain

The phylogenetic tree obtained for 16S rRNA gene is as shown in Figure 23. The *Photorhabdus* and *Xenorhabdus* spp. clusters show close relationship, as did their toxin proteins and VRP1 domains. *Serratia entomophila* and *Pseudomonas syringae* were distantly related to *Photorhabdus* and *Xenorhabdus* in spite of their closely related class A protein toxins. *Yersinia, Erwinia, Salmonella* and *Burkholderia* appeared to be much older than these nematode-symbionts.

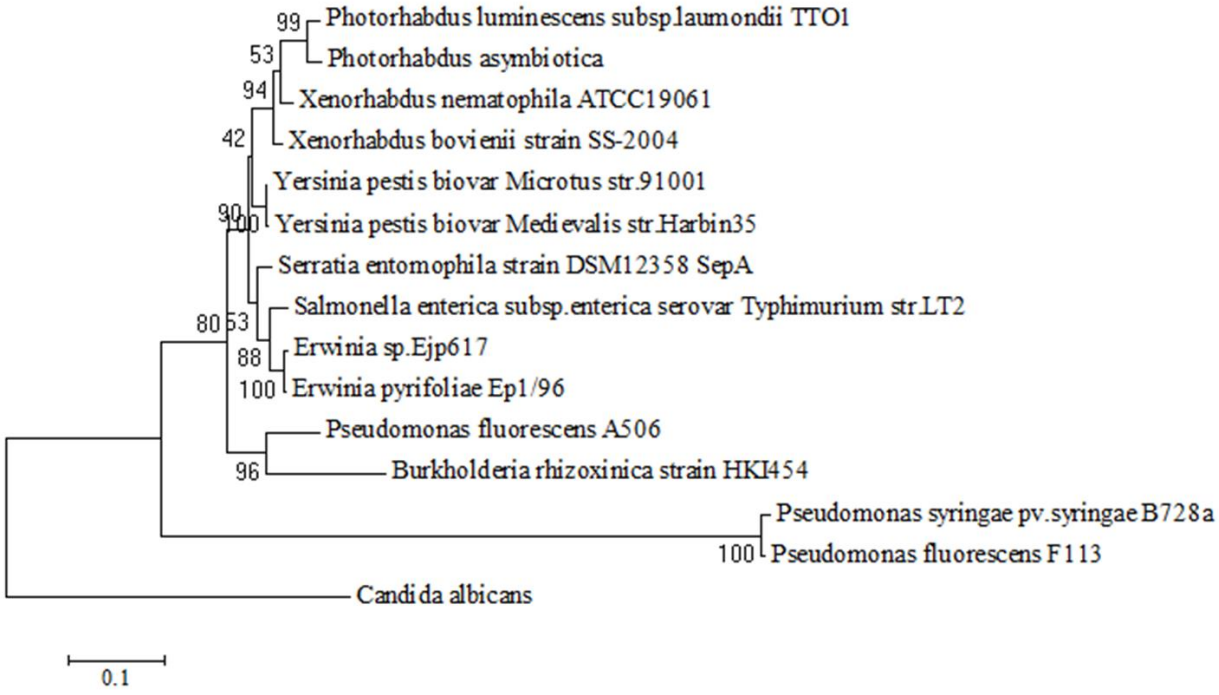


Figure 23: Maximum likelihood phylogenetic tree of bacteria with VRP1 domain, generated from 16S rRNA gene sequences. The scale at the bottom represents nucleotide substitutions per site.

4.4. Protein toxin isolation and purification

4.4.1. Toxin isolation using lysozyme and sonication

The procedure by Morgan *et al*. (2001) (herein referred to as Morgan) and the combination of enzymatic and mechanical lysis (herein referred to as lysozyme) showed some smearing of protein bands along the lanes, while the procedure by Sheets *et al*. (2011) (herein referred to as Sheets) showed clearer, albeit fainter, banding (Figure 24A). Bands greater than 669KDa were observed in the 'Lys' and 'Mor' lanes of Lysozyme and Morgan respectively. Bands of interest (greater than 232KDa) were observed in all lanes. A lot of material was left in the wells after using the three lysis procedures. Some high molecular weight (HMW) bands greater than 440 KDa were lacking in 'Sheets'. The Lysozyme and Morgan procedures produced almost identical results with regards to HMW bands. However, lysozyme also lacked some of the low molecular weight bands below 237 KDa. The procedure by Morgan was therefore chosen as the appropriate procedure due to the better protein yield and ease of use.

To determine the appropriate lysis volume, the sample lysed in 5mL of PBS smeared the most while the sample in 15mL smeared the least (was more dilute) (Figure 24B). The 10mL volume sample showed intermediate smearing and was thus considered to be the better of the two volumes. This volume was thus chosen for use in future lysis procedures. The protein bands from the lysate were distinct in all the lanes. However, the faintest bands were in the lanes containing material from the lysate cell debris. This indicated that this lysis method was relatively good as little material was left in the lysate pellet. Some bands seen in the lanes containing lysate supernatant were missing, between 140 KDa and 440 KDa, in the lanes containing lysate cell debris.

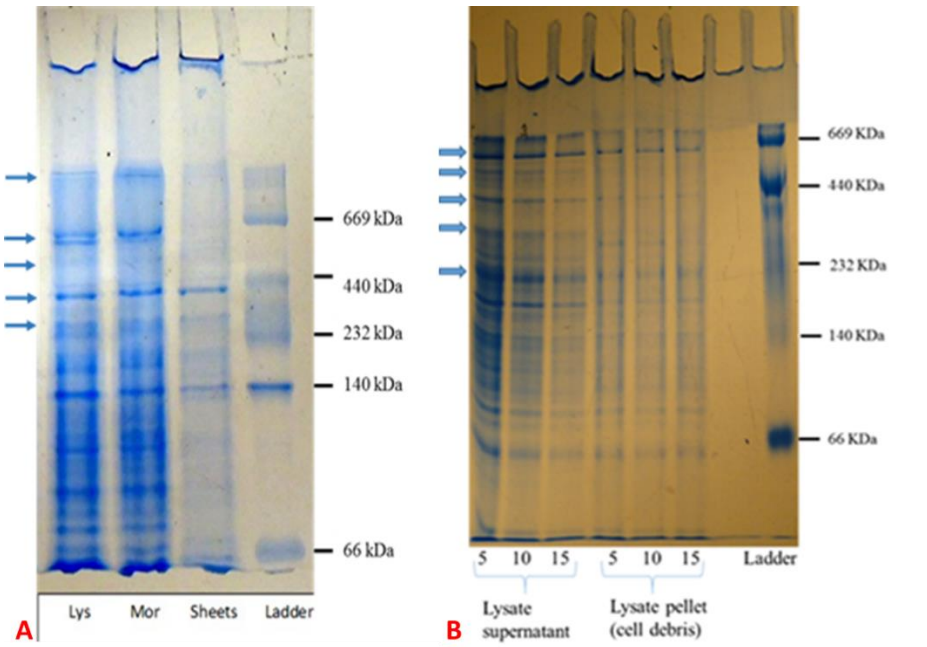


Figure 24A: Native-PAGE gel electrophoresis for comparison of three different bacterial lysis procedures. Key: Lys = Lysozyme, Mor = Morgan, Sheets = Sheets

Figure 24B. Native-PAGE gel electrophoresis for comparison of three different bacterial lysis volumes. The lanes labelled 5, 10 and 15 represent volumes in microliters.

The arrows show the bands in the regions of interest.

4.4.2. Protein purification through size exclusion chromatography

Separation of the *Xenorhabdus* crude lysate yielded results as shown in Figure 25A. The peaks shown in the graph yielded the resultant pooled fractions which were concentrated, dialyzed and run on a native-PAGE gel (Figure 25B). The pooled fractions were run alongside fresh crude lysate which was not subjected to chromatography. The pooled fractions 1 and 5 did not show any bands on the gel. High molecular weight proteins of sizes greater than 232 KDa were seen in pools 2 and 3. These bands generated interest as they were in the regions where the XptA protein bands were expected to be found after separation. Pool 4 contained many of the lower molecular weight bands below 232 KDa. All of the eluted protein bands coincided with bands in the crude lysate in the 'lysate' lane.

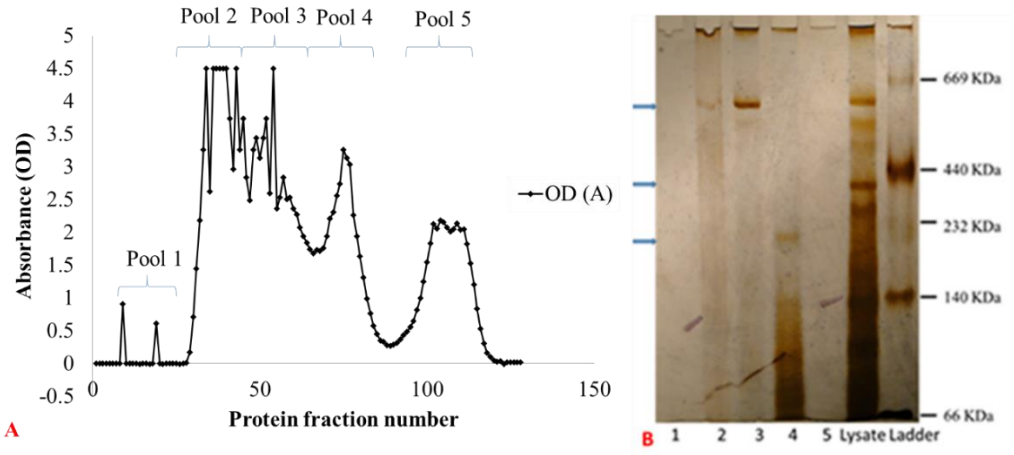


Figure 25: Protein purification by size exclusion chromatography. **A**. Plot of absorbance against protein fraction number. **B**. Native-polyacrylamide gel electrophoresis of pooled protein fractions. Each lane represents a single pool. The arrows indicate the high molecular weight bands of interest.

4.5. Amplification of VRP1 domain DNA fragment

PCR performed using gene-specific primers yielded a band approximately 800 base pairs (bp) long upon re-amplification (Figure 26A). A 'highly similar' BLAST of the obtained sequence (Appendix 16) gave no results and thus a 'somewhat similar' BLAST was performed. A *Rahnella aquatilis* CIP 78.65=ATCC 3301 sequence was obtained as the best and only hit with a maximum score of 198, identity of 79%, E value of 1e-44 and bit score of between 80-200 out of a query cover of 42%.

Amplification using degenerate primers yielded the following bands: one 600bp band for A1VRPSet2 at all temperatures between 40 and 47.2°C (Figure 26B lane 1); two bands, 800bp between 40.7 and 48.3°C, and 1.5kb between 44 and 50°C for A2VRPSet1 (Figure 26B lanes 2 and 3); and two bands of A2S1VF, 600bp and 1.5Kb at temperatures between 50 and 60°C (Figure 26B lanes 4 and 5).

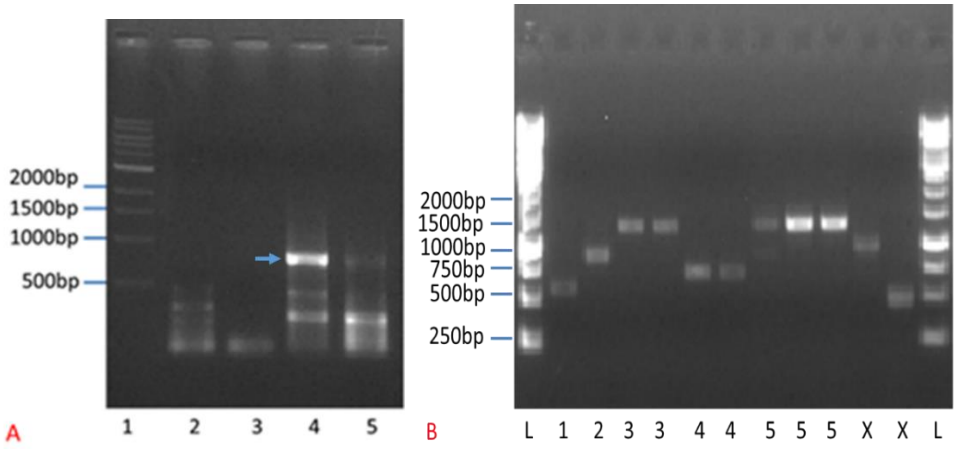


Figure 26: Gel electrophoresis of PCR amplification of VRP1 domain using (**A**) gene-specific and (**B**) degenerate primers. Lane X in figure B is an unrelated PCR.

A. Lane 1. 1Kb ladder (New England Biolabs); lane 2 and 3. First time amplification; lane 3 and 4. Re-amplification using the PCR products of lane 2 and 3 respectively as template. **B.** L. 1Kb ladder (New England Biolabs); 1. A1VRPSet2 600bp; 2. A2VRPSet1 800bp; 3. A2S1VF 1.5Kb; 4. A2S1VF 600bp; 5. A2VRPSet1 1.5Kb; X. Unrelated PCR.

BLAST searches of the edited sequence results (Appendix 16) yielded the following results: The edited 600bp sequence of A1VRPSet2 gave *Xenorhabdus doucetiae* str. FRM16 as the top hit with an identity of 90%, E-value of 2e-162 and bit score of ≥200 from a query cover of 84%. The subsequent best hits were from *X. nematophila*, *X. poinarii* and *X. bovienii* respectively. All the BLAST sequence results from the four different bacteria were for genes responsible for flagella biosynthesis.

BLAST results for the 800bp band of A2VRPSet1 also gave *X. doucetiae* as the top hit with an identity of 78%, E-value of 5e-134 and bit score of ≥200 from a query cover of 99%. However, this hit was of a putative ornithine racemase gene. The 1500bp band on the other hand failed to yield a good useable sequence. Both the 1500 and 600bp bands of the combination primer set A2S1VF gave poor quality sequences and could thus not be used.

CHAPTER FIVE

5.0. DISCUSSION

5.1. Bacterial characterization resulted in *Xenorhabdus griffinae*

The bacterium isolated from the nematodes showed morphological characteristics similar to *Xenorhabdus nematophila* as previously described by Arkhust (1980; 1983). However, molecular identification through PCR amplification and BLAST search of the NCBI nucleotide database suggested this target sequence to be of *Xenorhabdus griffiniae.* Tailliez *et al*. (2006) described *X. griffine* as producing non-pigmented colonies on NBTA similar to *X. nematophila.* This description could not enable morphological distinction of the two species. However, discrimination of the two was achieved through phenotypic characterization and 16S rDNA sequencing.

Following the bacterial taxonomy threshold by Stackebrandt & Goebel (1994), a 16S rRNA gene sequence homology greater than 97% indicates a DNA similarity that is greater than 70% in DNA-DNA hybridization, which is indicative of relatedness at species level. Consequently, a 98% result, as was the case here, makes this target bacterium most likely a strain of the *X. griffiniae* species. However, further testing (biochemical tests and/or amplification of other conserved genes as described by Tailliez *et al.* (2006; 2010) has to be done to confirm the identity of this bacterium at the species level. Identification of the nematode of origin would also give further insight into the identity of the bacterium.

5.2. *Xenorhabdus* sp. affected the feeding behaviour of *P. truncatus* and *S. zeamais*

The larger grain borer showed significantly less damage in the *Xenorhabdus* treatment (42% completely destroyed pellets) compared to the controls (100% and 92% completely

destroyed pellets in the PBS and *E. coli* treatments respectively). This indicated that *Xenorhabdus* reduced feeding of LGB. On the other hand, *E. coli* did not appear to have a major inhibitory impact on the feeding behaviour of the LGB. There was a slightly lower proportion of pellets destroyed in the *E. coli* treatment (92%) when compared to the PBS (100%). This suggested that the presence of *E. coli* in the diet affected the feeding of the insects. Comparatively, LGB caused greater damage to the pellets than the maize weevil. This could be because the LGB have been reported to be more voracious eaters than the MW and are thus more destructive (Tefera *et al*., 2010).

5.3. The bacterium caused significant mortality of the maize weevil

The maize weevil showed notably higher mortality rates in all the treatments compared to larger grain borer. LGB appeared to be more tolerant to the inoculated pellets despite their relatively higher rates of consumption. There is a virtual lack of literature on the direct effects of *Xenorhabdus* spp. on either the maize weevil or the LGB. The most related information is on the effects of *Steinernema* and *Heteroharbditis* entomopathogenic nematodes on coleopteran storage insect pests like; the rice weevil (Curculionidae: *Sitophilus oryzae* (L.)), lesser grain borer (Bostrichidae: *Rhyzopertha dominica* (F.)), saw-toothed grain beetle (Cucujidae: *Oryzaephilus surinamensis* (L.)), the warehouse beetle (Dermestidae: *Trogoderma variabile* (Ballion)), red flour beetle (Tenebrionidae: *Tribolium castaneum* (Herbst)) (Ramos-Rodriguez *et al.,* 2006) and the confused flour beetle (Tenebrionidae: *Tribolium confusum*) (Athanassiou *et al.*, 2008).

These studies have suggested that entomopathogenic nematodes may be useful in controlling stored-product insect pests (Rumbos & Athanassiou, 2012) and all of the nematodes used have been previously reported to host *Xenorhabdus* and *Photorhabdus*

bacterial symbionts in their respective genera. These symbionts play an integral role in the virulence of the nematodes (Boemare & Akhurst, 2006). It therefore follows that, the bacterial symbionts contributed significantly to the infectivity of the nematodes. It is then highly likely that the bacterial symbionts themselves, could be virulent against the target insects as was in this study.

Virulence has been shown in this work, particularly for the maize weevil, demonstrating that, *Xenorhabdus* could serve as a possible biocontrol agent for these pests with an appropriate application method such as the one used in this study. The possible ability of the pellets to support the biological cycle of the insects demonstrates their potential in suppressing progeny production which is as important as parental mortality (Rumbos & Athanassiou, 2012).

As observed in this work, *Xenorhabdus* had changed to the phase II form by the sixth day. This form has been reported to be less potent compared to phase I (Boemare & Akhurst, 2006). However, this was not elucidated by this study. Secondly, bacteria need moist conditions in order to remain viable for long (Rahman & Labuza, 2007; De Goffau *et al*., 2009), these conditions are contrary to what is fit for the insects (Tefera *et al*., 2011). It has been reported that *Xenorhabdus* only occurs in association with its nematode host in nature. It cannot survive in water or soil for long and so it is limited in its use as a biocontrol agent (Morgan *et al.*, 2001).

In this study, PBS was used to make the maize pellets which were then dried at 28°C. The use of PBS and drying at 28°C was aimed at ensuring the viability of the bacteria and viable cells could be re-isolated from the dry pellets up to the sixth day. Phosphate buffered saline (PBS) is an isotonic non-toxic balanced salt solution that ensures optimal pH and

osmolarity required for the viability of most cells. It is isotonic and is generally utilized to maintain cells for the short term in a viable condition while they are manipulated outside of their growth medium (Dulbecco & Vogt, 1954; Medicago AB, 2010). However, it was not established in this study the exact period in which the bacterium can be re-isolated from the pellets (in whichever phase) and thus it is difficult to give the length of time that it would be viable in the pellets. Re-isolation after longer periods of incubation, for example 10, 20, 30 days, would help to answer the question of viability over time. This was not done in this study.

5.4. Bioinformatic analyses of class A genes and proteins

5.4.1. Protein families, domains and motifs of class A proteins

XptA toxin sequences and their homologs obtained in this study showed various conserved domains including VRP1. A BLAST search of all the retrieved toxin sequences yielded this domain with links to protein databases like Interpro and Pfam. This domain was identified at the N-terminal portion of all sequences and was highly associated with the Salmonella virulence plasmid gene A (SpvA) which is reportedly an outer membrane protein (El-Gedaily *et al*., 1997). Salmonella has not been reported to have Tc toxins but it produces toxins similar to the [B] and [C] components of the Tc complex of *Photorhabdus* (Hinchliffe *et al*., 2010).

A specific search for this domain in the aforementioned databases revealed two distinct reference sequences, one in the CDD (NCBI) and the other in the Pfam database; sequence ID: PRK15212 and pfam03538 respectively. Both were consensus sequences derived from multiple sequence alignments of *Salmonella* (in CDD) and Tc sequences (in Pfam). These two consensus sequences shared a 31% amino acid sequence similarity with an E-score of

2e-39 from a 100% query cover and a high bit score (80 – 200). According to Pearson (2013), these results suggested that PRK15212 and pfam03538 are homologous.

All the VRP1 sequences shared homology with these two sequences with a majority showing the greatest homology to the pfam03538 sequence. The TccA and TcaA showed the highest similarity to the PRK15212 sequence, indicating their close relationship to SpvA. The VRP1 domain was identified in all sequences, probably, through comparison to an intermediate sequence (that of pfam03538). Homologous sequences do not always share significant sequence similarity, but are clearly homologous based on statistically significant structural similarity or strong sequence similarity to an intermediate sequence. Therefore, for non-significant alignments, comparisons to an intermediate sequence can be used to demonstrate homology (Pearson, 2013).

TccA in this study showed the highest sequence similarity to the Salmonella sequences thus coinciding with the findings of Hinchliffe *et al*. (2010). Hinchliffe *et al*. (2010) reported that the N-terminus of TccA from *P. luminescens* is similar to the N-terminus of Salmonella plasmid virulence A (SpvA). They also reported that this similarity is not shared by any of the other *Photorhabdus* [A] subunits. However, contrary to their latter observation, some similarity, albeit relatively weak, was found between the TcdA, XptA, other Tca and Salmonella sequences. This could have been because of the role of the intermediate as described by Pearson (2013).

5.4.2. The spread of the VRP1 domain and its seven-motif protein fingerprint

Pfam families showed the VRP1 domain to be widespread among many toxin producing bacteria species. A search for the domain in the NCBI database confirmed it to have homologs within toxin proteins. A search of the FingerPRINTScan database for protein

family fingerprints revealed seven conserved motifs in the SALSPVAPROT fingerprint family. These motifs are typical for Salmonella SpvA sequences and thus explains their presence in the Tc sequences which are homologous to SpvA in the N-terminal region. Their complete conservation across all species could imply a functional importance to these toxins. However further work is required to confirm this.

Many of the VRP1 motifs were conserved despite variations in the sequences of the different species. However, motifs at the terminal ends of the sequences (Motifs 1, 2, 5, 6 and 7) appeared to have been lost in the TcdA and XptA toxins. This was seen in the CDD search and in the multiple sequence alignment (Figure 12B). This indicated that these terminal sequences were not important for the toxins. The TccA toxins on the other hand had many of the motifs from the fingerprint conserved, this explained their close homology to the SpvA protein as had been observed by Hinchliffe *et al*. (2010). Overall, the TcdA and XptA appeared to have the least number of conserved motifs, this was confirmed by their placement on the maximum likelihood tree (Figure 13), they appeared top on the tree, suggesting that they had the most variation. This was in line with the observation by Hinchliffe *et al*. (2010) that their similarities to other known proteins, particularly the SpvA as seen here, are weak.

The Tc genes are longer than the SpvA. Ffrench-Constant & Bowen (2000) reported that in *Salmonella* the Spv genes correspond to small ORFs, whereas in *Photorhabdus* the Spv homologs are fused as part of longer polypeptides. This suggests that in the course of evolution, Tc toxins may have added other parts to their proteins to make them effective, and therefore; may have combined whatever ability they acquired from the SpvA protein, in the form of VRP1, with others, perhaps from elsewhere, to form the class A toxins of

today (Figure 11). The class A proteins also have parts homologous to other less complex pore forming toxins like diphtheria and anthrax (Meusch *et al*., 2014), suggesting that they could have borrowed from these as well.

Comparison of the SpvA protein sequences from *S. enterica* serovar Arizona with previously described subspecies I strains including serovar Dublin revealed that the proteins differed in sequence. The serovar Arizona SpvA protein contained an additional 33-amino acid segment not present in the SpvA proteins of subspecies I strains (Boyd & Hartl, 1998). Additionally, Libby *et al*. (2002) showed that, the complete SpvA protein from serovar Arizona had a frameshift in the C-terminal region resulting in a different protein sequence and slightly longer C terminus than in the subspecies I SpvA proteins.

These two sequence features, the 33-amino acid insert and the C-terminal mutation, demonstrated a common ancestor of TccA and the SpvA proteins which also contained these two features (Libby *et al*., 2002). The findings correlated with the observations made here of the similarity shared between the TccA and SpvA sequences. The conserved amino acids showed by Libby *et al*. (2002) corroborated with those of the SALSPVAPROT fingerprint and others (Figure 12), that were seen to be conserved between the two sequences, as well as the Tca sequences of the other previously described bacteria. The TcdA, XptA and SepA proteins however, appeared to have lost many amino acids present in SpvA and the TccA.

Interestingly, Libby *et al*. (2002) also demonstrated that the Spv locus in serovar Arizona, unlike in the subspecies I strains, is in the chromosome. Further, sequence analysis of the flanking regions led them to propose that the spv locus represents a pathogenicity island in serovar Arizona. This is comparable to the reports that the Tc and Xpt complex proteins

are also contained in pathogenicity islands in their respective parent bacteria (Waterfield *et al*., 2002; Sergeant *et al*., 2006). Additionally, Tc proteins contained in plasmids in some bacteria are highly similar to those in the chromosomes of other bacteria, thus suggesting a possible transfer of these proteins across bacterial species (Hinchliffe *et al*., 2010). Similarities have also been found between the Spv and virulence plasmids of *Yersinia* spp. (Krause *et al*., 1990; 1991).

5.4.3. Structure analyses revealed the most important domains of the [A] protein

The proteins which had the highest sequence homology to the template (1VW1A) also showed the best validation results. This is due to the fact that model building is based on the target-template structural alignment. Where identical residues align, the template's main-chain and side-chain atom coordinates are copied to the model. In regions where aligned residues differ, the main-chain atom coordinates are copied and the side chain is later built. In gap regions of the alignment that have insertions and deletions, loops are modelled (Xiong, 2006).

The models closest to the template in sequence, template-like models, also closely resemble the template in structure. On the other hand, models distantly related to the template sequence-wise, non-template like, have long loop regions which make their structures different from that of the template. This was observed in this study and was further elucidated by the validation results which showed the similarity of the tertiary structures, both to the template and to each other. High RMSD scores, low GDT_TS and few residues in the most favoured region indicated a distant relationship to the template and vice versa.

The structure of the full TcA protomer was initially described to be composed of three large domains: the α-helical, β-sheet and pore-forming domains (Sheets *et al*., 2011; Gatsogiannis *et al*., 2013). These were later further resolved to eight domains comprising of six domains that form the shell and two which form the inner channel. These two domain groups are connected to each other by a 42 amino acid linker (Meusch *et al*., 2014). This description fit to the template-like models generated from TcdA, XptA and SepA sequences. All the other models, TcaA and TccA, deviated from this structure although they had some of the three major domains (mainly α-helical and others the β-sheet as well). They appeared to be incomplete forms of the template, as though they lacked some parts which were present in the template and template-like models

The template-like models generated in this study all closely resembled each other. However, parts of the β-sheet region of TcdA2 and SepA from *Photorhabdus* and *Serratia* spp. respectively, failed to align completely. A similar observation was made by Meusch *et al*. (2014) who identified the missing parts as receptor binding domains A and D in TcdA2 and SepA respectively. The full TcdA1 protomer (similar to XptA1, A2 as well as TcdA1 from *P. syringae*) has, in the α-helical shell domain, three large insertions, forming four receptor-binding domains A–D (A, 298–433; B, 1308–1382, 1491–1580; C, 1383–1490; and D, 1633–1761) which are structurally similar to the diphtheria and anthrax toxin receptor-binding domains, as well as a neuraminidase-like domain (Meusch *et al*., 2014).

Meusch *et al*. (2014) found that, TcA and YenA proteins of *Yersinia* spp., had as many as three domains missing. Others, like TcdA2 from *Photorhabdus* spp., TcdA4 from *P. luminescens,* SppA from *Serratia proteamaculans* and SepA from *Serratia entomophila*, had a single domain missing. Of those missing one domain, the TcdA2 and TcdA4 all

lacked the receptor-binding domain A, while SppA and SepA lacked receptor-binding domain D. YenA and the TcA from *Yersinia* all lacked receptor-binding domains B, C and D (a majority of the β-sheet domain). They therefore concluded that these variations might explain the different host specificities of Tc toxins.

Comparison of the class A proteins generated in this study has therefore revealed the most conserved parts of the A toxins. These include the α-helical shell, the neuraminidase-like domain (electrostatic lock), the pore- forming domain and the TcB-binding domain. Consequently, this implies that the most important abilities for all the toxins, irrespective of their evolutionary position, include, receptor-binding via the neuraminidase-like domain (electrostatic lock), binding with the B components of the complex via the TcB-binding domain and pore formation. The α-helical shell is important in holding the A protein together and, in the A complex, it is also essential for opening the pore during transition from the pre-pore to the pore state (Meusch *et al*., 2014). Hinchcliffe *et al*. (2010) also reported that the N-terminal portion of the class A proteins, which consists of the α-helical shell, is involved in actin condensation. This implies that the class A proteins could have more diverse functions other than pore formation. The other domains appear in different proteins perhaps for host specificity as concluded by Meusch *et al*. (2014).

Experimentally, a crude-lysate native PAGE run consistently showed protein bands ≥ 669 KDa. Similarly, protein samples run after size exclusion chromatography (SEC) also consistently showed bands approximately 669 KDa. These findings were consistent with those of Sheets *et al.* (2011) whose native PAGE analysis of XptA complex revealed that the complex migrated over a significantly less distance in the gel than the highest molecular mass standard of 669 kDa. It is therefore possible that the bands observed in this study

were of XptA proteins, however, more needs to be done in order to identify them conclusively.

5.5. Amplification using VRP1 domain primers yielded possible virulence-related genes

Amplification was done using a two-pronged approach; a set of gene-specific primers designed using the *X. nematophila* ATCC 19061 XptA sequence; and, various sets of degenerate primers designed from multiple sequence alignments of *Xenorhabdus* and *Photorhabdus* spp.. The amplicons obtained using both primer groups were not the expected VRP1 domain. However, the results obtained were of virulence-related DNA. The gene-specific primers yielded a sequence related to an unspecified *Rahnella aquatilis* sequence as determined by the 'somewhat similar' BLAST analysis which uses a lower threshold for sequence similarity. This could be the reason why the 'highly similar' BLAST did not yield results. This is first report to describe this kind of analysis for the VRP1 domain. *R. aquatilis* is a gram-negative bacterium of the family enterobacteriaceae, first isolated from drinking and river water (Izard *et al*., 1979) and subsequently from human clinical specimens (Goubau *et al*., 1988) and rhizospheres of different plants (Berge *et al.*, 1991). It was reported in Pfam families as also having the VRP1 domain ((http://pfam.xfam.org/family/PF03538#tabview=tab7).

Amplification using the degenerate primers yielded genes involved in flagella biosynthesis from different *Xenorhabdus* spp. Among various genes involved in flagella biosynthesis, the most important are the flhDC operon, whose products are required for expression of all other flagella genes (Givaudan *et al*., 2000). This operon is also required for full insect virulence, lipase production, and haemolysis of sheep red blood cells (Givaudan *et al*.,

2000; Kim & Forst, 2005). In *X. nematophila*, there are four clusters of flagella genes located in the same position of the chromosome. In *X. bovienii* cluster three is located in a different region of the chromosome. Non-flagella genes are located between the flagella clusters (Kim & Forst, 2005).

Types of non-flagella genes between the clusters are similar in both *X. nematophila* and *X. bovienii* genomes: region A contains fimbrial genes, region B contains insecticidal genes and invertases, and region C contains genes encoding exceptionally large proteins called polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) that help to produce diverse molecules, including antimicrobial compounds. Having this in mind, it is possible that the degenerate primers could have amplified the flagellar genes that surround the Xpt insecticidal genes. The flagella genes themselves are important due to their importance in virulence against insects. However, more work has to be done in order to isolate the insecticidal genes themselves.

CONCLUSIONS

The *Xenorhabdus* bacterium used in this study was most likely *X. griffinae*. This bacterium caused mortality in the maize weevil and significantly lowered the feeding of the larger grain borer on maize pellets. This implies that the bacterium has a potential for use as a possible biocontrol agent for *P. truncatus* and *S. zeamais*. The homology of the XptA proteins to other established insecticidal class A toxins demonstrated their function as insecticidal protein toxins. Further, this homology indicated activity against a broad-spectrum of insects.

VRP1's widely conserved motifs suggest that it could be important for the functioning of the class A proteins. The VRP1 domain is found in the N-terminal region of class A proteins which is involved in causing actin condensation. The class A proteins could have more diverse functions other than pore formation. Based on structure, the basic required functions of the class A proteins are receptor binding, binding of the B component and pore formation. *Xenorhabdus* and *Photorhabdus* spp. may have added parts onto their genes to make more complex and more efficient proteins.

RECOMMENDATIONS

1. Further characterization of the bacterium be done in order to ascertain its identity at the species level. This includes biochemical tests and/or amplification of other conserved genes like RecA and gyrB. Identification of its nematode host is also recommended.
2. Bioassays to elucidate the lethal dose 50 (LD_{50}) of the bacterium against the two storage pests be carried out.
3. Long term bioassays be done to determine how long the bacterium remains viable in the maize pellets.
4. Bioassays be done in an integrated pest management system where the maize pellets will be combined with other insecticides.
5. Identification of the class A proteins expressed by *Xenorhabdus* sp. either through western blot or mass spectrometry is recommended. This will enable the full analysis of the insecticidal proteins from this species and their comparison to other well-studied insecticidal proteins.
6. Cloning and expression studies of the VRP1 domain, as well as appropriate bioassays, be carried out to elucidate the importance of the domain to the class A proteins.
7. Protein sequencing to identify the proteins in the bands of interest after size exclusion chromatography.

APPENDICES

Appendix 1: Table of top BLAST hits from a similarity search of the NCBI database with the amplified 16S rRNA gene sequence.

Description	Max score	Total score	Query cover	E-value	Identity
Xenorhabdus griffinae strain ID 10 partial sequence	2180	2180	98%	0.0	98%
Xenorhabdus griffinae from Malaysia	2176	2176	98%	0.0	98%
Xenorhabdus sp. MY8	2093	2093	98%	0.0	97%
Xenorhabdus ishibashi	2093	2093	98%	0.0	97%
Xenorhabdus sp GDh7	2087	2087	98%	0.0	97%
Xenorhabdus GDc328	2065	2065	98%	0.0	96%
Xenorhabdus mauleonii strain VC01	2036	2036	98%	0.0	96%
Xenorhabdus poinarii strain NC33	2026	2026	98%	0.0	96%

Appendix 2: Table of A1 toxin genes and proteins.

No.	Gene/ Protein name	Organism	Accession number	UniProtKB/ TrEMBL	Length of the gene	Number of amino acid residues
1.	TcaA1	*Pseudomonas fluorescens* A506	YP_006324534	I2BTH6	2737 nt	878
2.	TccA1	*Photorhabdus luminescens*	NP_931352	Q7MZV5	2601 nt	966
3.	TcaA1	*Yersinia pestis* 91001 (biovar Microtus)	WP_011171976.1	Q74PN4	2331 nt	776
4.	TcdA1	*Photorhabdus luminescens*	NP_928296	Q7N7Y9	7578 nt	2525
5.	TcdA1	*Pseudomonas syringae pv. syringae* B728a	YP_237273	Q4ZNN7	7509 nt	2502
6.	XptA1	*X. nematophila* ATCC 19061		D3VHH9	7,572 nt	2,523
7.	SepA	*Serratia entomophila*	WP_010895734	Q9F9Z3	7128 nt	2376

Appendix 3: Table of A2 toxin genes and proteins.

No.	Gene/Protein name	Organism	Accession number	UniProtKB /TrEMBL	Length of the gene	Number of amino acid residues
1.	TccA2	*Photorhabdus luminescens*	NP_929697	Q7N492	3522 nt	1173
2.	TccA2	*Photorhabdus asymbiotica*	YP_003040893	C7BJ81	3552 nt	1183
3.	TccA2	*Pseudomonas fluorescens* F113	YP_005207871	G8QCB7	3591 nt	1196
4.	TccA2	*Xenorhabdus bovienii* SS-2004	YP_003466509	D3UWD9	3555 nt	1184
5.	TcdA2	*Photorhabdus luminescens*	NP_928304	Q7N7Y1	7335 nt	2444
6.	TcdA2	*Photorhabdus asymbiotica*	YP_003039736	C7BNH9	7194 nt	2397
7.	XptA2	*X. nematophila* ATCC 19061	YP_003712781	D3VHI2	7,575 nt	2,524
8.	XptA2	*Xenorhabdus bovienii* SS-2004	YP_003467478	D3V0P9	7,554 nt	2,517

Appendix 4: Table of VRP1 domains retrieved from A1 sequences found in the database and their lengths.

No.	Gene/Protein name	Organism	Length of VRP1 domain gene	Number of amino acid residues
1.	TcaA1	*Pseudomonas fluorescens* A506	675 nt	236 aa
2.	TccA1	*Photorhabdus luminescens*	939 nt	313 aa
3.	TcdA1	*Photorhabdus luminescens*	654 nt	211 aa
4.	TcdA1	*Pseudomonas syringae pv. syringae* B728a	420 nt	140 aa
5.	XptA1	*Xenorhabdus nematophila* ATCC 19061	387 nt	129 aa
6.	SepA	*Serratia entomophila*	522 nt	174 aa
7.	TcaA1	*Yersinia pestis* 91001 (biovar Microtus)	156 nt	138 aa

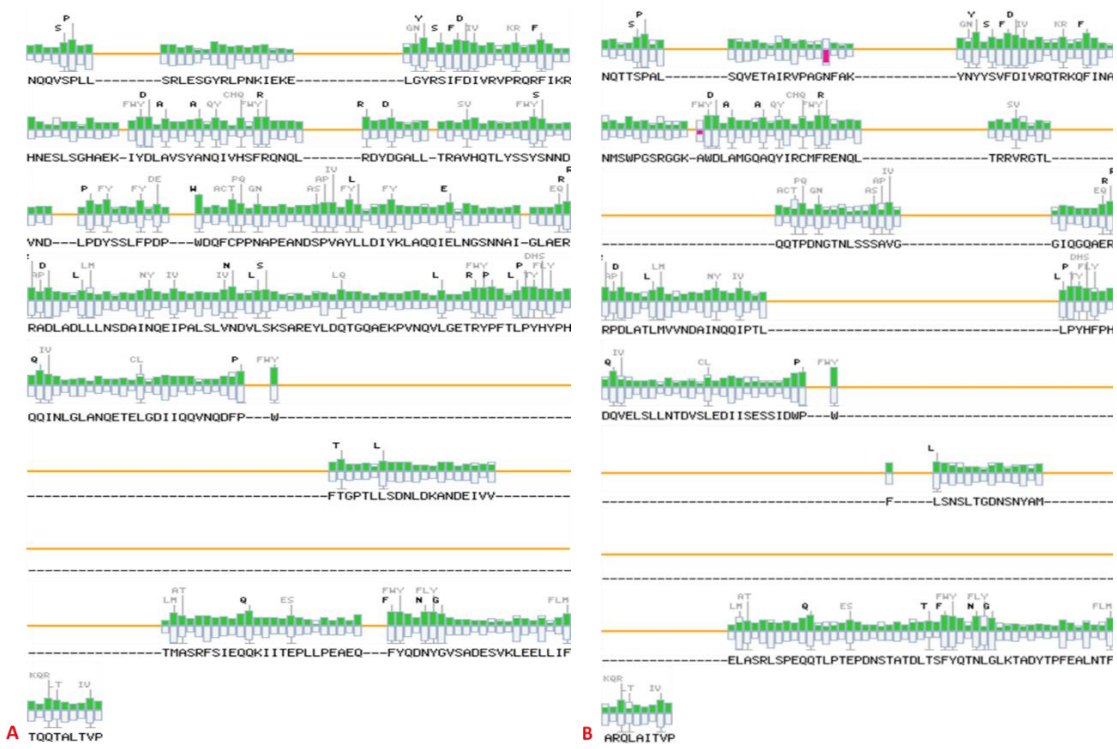
Appendix 5: Table of VRP1 domains retrieved from A2 sequences found in the database and their lengths.

No.	Gene/Protein name	Organism	Length of VRP1 domain	Number of amino acid residues
1.	TccA2	*Photorhabdus luminescens*	1530 nt	510 aa
2.	TccA2	*Xenorhabdus bovienii*	1521 nt	507 aa
3.	TccA2	*Photorhabdus asymbiotica*	1520 nt	507 aa
4.	TccA2	*Pseudomonas fluorescens* F113	1365 nt	455 aa
5.	TcdA2	*Photorhabdus luminescens*	429 nt	143 aa
6.	TcdA2	*Photorhabdus asymbiotica*	426 nt	142 aa
7.	XptA2	*Xenorhabdus bovienii*	399 nt	132 aa
8.	XptA2	*Xenorhabdus nematophila* ATCC 19061	386 nt	129 aa

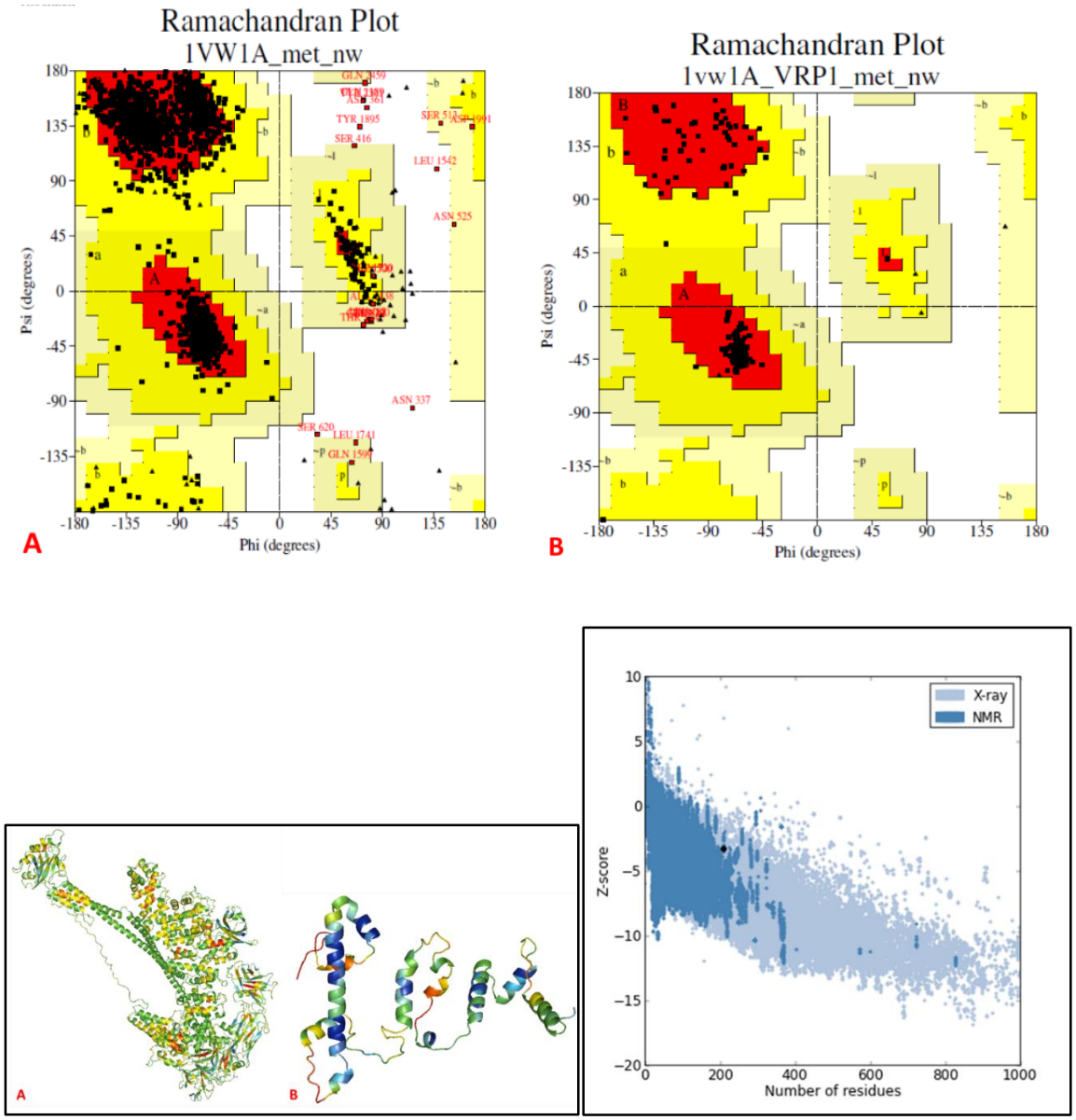
Appendix 6: Table of VRP1 domain motifs of the SALSPVAPROT fingerprint obtained from the PRINTS database. The highlighted motif sequences are those also observed in the motif prediction by Motifscan.

Motif	Sequence	Motif Length	Motif Position
Motif 1 of 7	ETAIRVPAGNFAKYNYYSVFD	21	15
Motif 2 of 7	WPGSRGGKTWDLAMGQAQY	19	51
Motif 3 of 7	TPDNGTNLSSSAVGGIQGQAE	21	90
Motif 4 of 7	NQQIPTLLPYHFPHDQVELS	20	126
Motif 5 of 7	DWPWFLSNSLTGDNSNYAMELASR	24	163
Motif 6 of 7	PTEPDNSTATDLTSFYQTNLGLK	23	195
Motif 7 of 7	QLAITVP	7	233

Appendix 7: Figure of Predicted motifs in the two reference VRP1 domain sequences. The bold letters above the green bars show the strongest motif matches while the grey ones show the weaker matches.

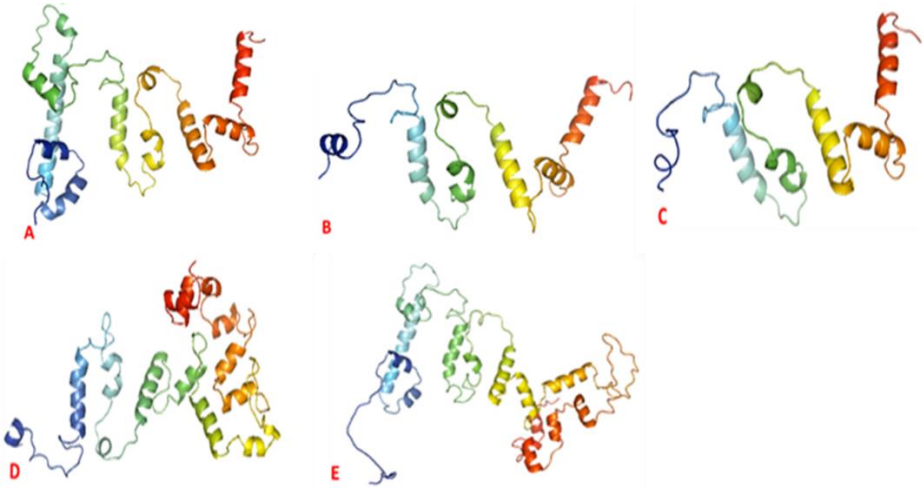


Appendix 8: Figure of Model quality validation of the 1VW1A full protein and VRP1 domain templates.



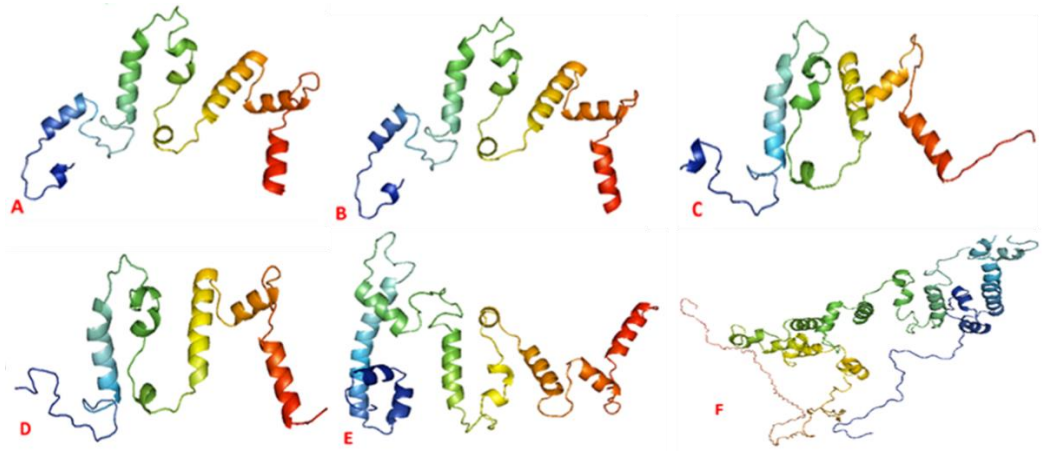
A. PROCHECK. **B.** MetaMQAPII. The validation is for the full protein (A) and VRP1 domain (B). **C.** PROSA validation for the VRP1 domain.

Appendix 9: Figure of A1 VRP1 models from different proteins.



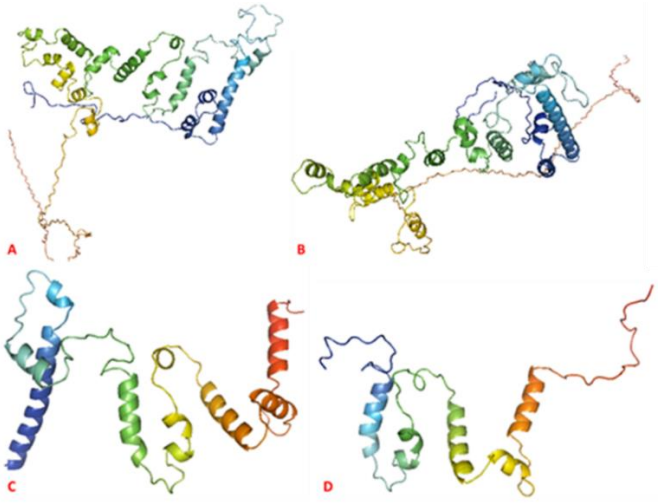
A. TcdA1 from *P. luminescens*. **B.** TcdA1 from *P. syringae*. **C.** XptA1 from *X. nematophila*. **D.** TcaA1 from *P. fluorescens*. **E.** TccA1 from *P. luminescens*. The models are coloured in a spectrum of colours ranging from blue in the N-Terminal region to red in the C-terminal.

Appendix 10: Figure of A2 VRP1 models from different proteins.



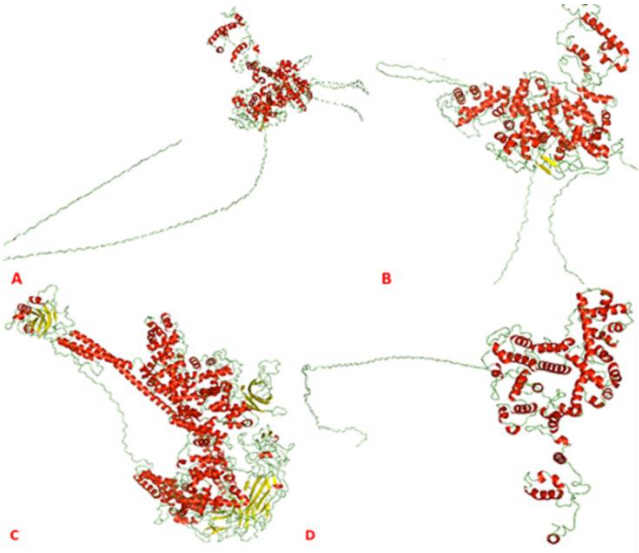
A. TcdA2 from *P. asymbiotica*. B. TcdA2 from *P. luminescens*. C. XptA2 from *X. bovienii*. D. XptA2 from *X. nematophila*. E. TccA2 from *P. fluorescens*. F. TccA2 from *X. bovienii*. The models are coloured in a spectrum of colours ranging from blue in the N-Terminal region to red in the C-terminal.

Appendix 11: Figure of 3D structures of VRP1 models from different class A proteins.



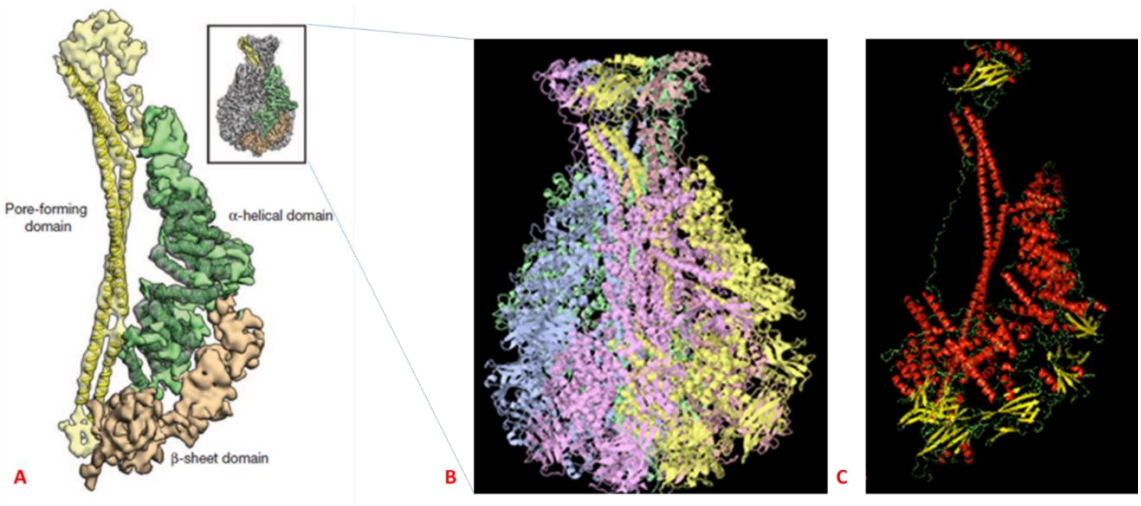
A. TccA2 VRP1 domain from *P. asymbiotica*. **B.** TccA2 VRP1 domain from *P. luminescens*. **C.** SepA VRP1 domain from *S. entomophila*. **D.** TcaA1 VRP1 domain from *Y. pestis* 91001. The models are coloured in a spectrum of colours ranging from blue in the N-Terminal region to red in the C-terminal.

Appendix 12: Figure of 3D structures of full protein models from different A proteins.



A. TccA2 full protein from *P. asymbiotica*. **B.** TccA2 full protein from *P. luminescens*. **C.** SepA full protein from *Serratia entomophila*. **D.** TcaA1 full protein from *Yersinia pestis* 91001. The models are coloured to show the different secondary structures; red for the α-helix, yellow for the β-sheet and green for the loops.

Appendix 13: Figure of Crystal structure of TcdA1 toxin from *P. luminescens* (PDB ID: 1VW1).



A. Extracted electron density of a protomer with identified α-helices. The outlined box shows the localization of the subunit in the complete structure. Densities corresponding to the N-terminal α-helical domain, the central β-sheet domain and the C-terminal pore-forming domain are depicted in green, brown and yellow, respectively (adapted from Gatsogiannis *et al*., 2013). **B**. TcdA1 complex of five identical protein subunits (protomers). The different colours represent the five identical TcdA1 subunits forming the complex. **C.** One of the five TcdA1 protomers chain A in PDB. The red structures are α-helices, the yellow are β-sheets and the green lines are loops.

Appendix 14: Figure of Multiple sequence alignment of class A VRP1 models.

A. Template-like VRP1 domains of XptA and TcdA proteins; TcdA1 from *P. syringae* and SepA from *S. entomophila*. The sequences are numbered 1 – 8 respectively. The sequences are represented by numbers 1-10 respectively. The row [AVE_SESTR] is the average secondary structure of each site and the completely conserved sites are shown in upper letters. The letters H, S, T and C in this row represent alpha helix, bend, hydrogen-bonded turn and coil respectively. Completely and incompletely conserved sites are shown in upper and lower case letters, respectively. The bottom line shows the sequence identity (*) and the conserved residues (**:**) and semi-conserved residues (**.**).

Appendix 15: Table of Sequences of degenerate primers used in the PCR amplification of VRP1 domain.

No.	Primer name	Primer sequence (5' to 3')
1.	A1VRP1Set1XenorF	TTTGTMAAWCCGGGMTCMRTKGCYTCCATYTT
2.	A1VRP1Set1XenorR	TTACGGWSMAATGCKGSMAAGGKTTTTATCT
3.	A1VRPSet2PhotoF	CCSYYYAAYSYYRACCYRATAGGYAACAG
4.	A1VRPSet2PhotoR	AATAWCCARTAGWGAKGYTKGRTWCATTA
5.	A2VRP1Set1XenorF	ATGTAYAGCACGGCTGTRTAACTMAAYAAAAT CAGT
6.	A2VRP1Set1XenorR	GCCAGAATCGGTAATGAAGCWCCTTCYGCCT
7.	A2VRPSet2PhotoF	ATGAAYACACTCAARCCYGARTAYCAACAAGC RTTAGGA
8.	A2VRPSet2PhotoR	TAATMTCCAACAGRGAARYTTGATAC

Appendix 16: Edited PCR sequence results of VRP1 amplification obtained using gene-specific and degenerate primers.

a) VRP1 gene-specific primers, 680bp

5’AAAACCGGTTTTATATCGAATAAATGCCCTTGATAGCCATCATACCCATCAGAATATGAAGGG
GTTTTATCAATCGGGGATCACGGGTAAATAACGGTTCAATGAAGAAAGGCTTATCTGCGACCAC
AATAAAATCAATCCATGATCCTGGAATATCAACACGTGGCAGGTCAGATTCATCGTCGACCAAT
TCGTTAATTTGAGCAATAACGATACCGTTTTTGAATGCAGTGGCTTCTACCAGAGGTGGGGTAT
CTTCTGTACTGGGGCCAGTATAGAGATTTCCATGTCGGTCAGCTTTATAGGCAGCGACTAATGC
GACATTGGGCACTAAATCCACAAATAATCGGGAATAAAGCTCCAGCCTGTATGAATAGCGCCAA
TTTCTTCCGCCCGCCTTCCAGCCCCTGTGATAATGGCCCGGTTTGCTCTCAGCATATGACTGCTCC
TACTGAGAGCAAATGCGGAGTTCGATCAATTCTACATCCTTTCGACCGGTGCTCCTTTTCCCGCC
ATACGAAGATCGCCGACCCTGTTCCCATCACCCTCCGAGGATCACTTCGTTAAGCTCTTCGCGAC
TGCCTGTTACCTCTCCCAACTACTGCTCTCAAAACGCTCACGTCCAGCACCTCTCCAACATTCTGC
TTAAGATCTCCCTTCTTCTCACGCGCTCGTGCTCT3’

b) A1VRPSet2, 525bp

5’ATCGCTAACGCCAAGCCCATTTTAACTTTTATCGGTATCTGTTTTTCACTGAAAATAGGGGCTG
TGCTGAACAAAGCCAGCAGGCGCACCAATGGCCAGAAAAAATCACTGACGAGGCGGGAAAGT
GTTTCGCTATCAAATGGGATCATAACTAACCGATAATAGCCGGAATACTGCTAAACAGGGTACG
CATGTAATCCAGCAACAAATTTAACATCCATGGTCCGGCAGTGACGATAGTCACGAAAACGGCC
AGGATTTTTGGAATAAACGATAAGGTCATCTCGTTTACCTGGGTTGCCGCCTGTAATAAGCTGA
TGATCAAGCCACTAATTAAGGCCGATAAGAGAGGAGGGGCGGCCAGTGCTAACGCCACTTTCA
TTGCTTCAACACCGAGGGCCATTACCGATTCTGGAGTCATAATTTCCTCCAACCGCTTTTTGATCT
TTATTCCCATAAGATTTAGGTTTCACGCTAAACATAAAAACTTTGTGCTTATGAACCCAACCTCAC
TACTGGATATTA 3’

c) A2VRPSet1, 799bp

5’ATGAACTAAGCCAGCCCACACCTCCCACGTGGCGGCATCAAATGAGACATTGGCGCAATGGG
CAATACAGTCGTCAGGTCCTATATCGGCAAAGCCATTATTGATGATAAGGCTAACGACGTTGCG
ATGTTCGACCATCACCCCTTTGGGCTGGCCGGTCGAACCCGATGTGTAAATGACATAGGCTAAA
TGACGTGATGTCAGTCCCAGTGCTTGGGCGTTTGGGGTATCCGTCGGTTGTGTCGTCAGCAGG
GTTTCATTGCGGTCACATAAAGAGTCAAGCAAGATCGTCGGCACGGTGTCAGGCAGCTTGTTGA
CTTGTGCTGTCTGGGTCAATAAAACAACGGGAGCTGAATCCCTGAGCATATAGGCTAACCGTTC
AATCGGATAGGTTGGATCGAGCGGGACATAGGCACCGCCTGCCTTGAGGATCGCCAGTAAGCC
GACCACCATCTCCAGACTGCGTTCGATACAGATAGCGACGCGATCATCTGGACGCACCCCTAGC
GTAATCAAATGATGAGCCAGACGATTGGCATGTTGGTTCAATTCTACGTAGCTCAATGTCTGGC
CTTCGCAAATCACGGCGGTGGCATGGGGGCGTTGTAATGCCTGAGCTTCAATGAGTTGATGGA
TCAGGGCATTTTGCGGAAAATCCACTTGGGTTGGGTTGAAGCCTACCAGCAATTGTTGACGCTC
TGGTTCCGGCAAGATTGGAACATTTAGGATCAATTGTTGTGGGTTGTGGATCAGGGCATCGATC
AGGCCGTTAATGGCCGTTTCCAGATAGGCAGAAGGA3’

REFERENCES

Abebe, F., Tefera, T., Mugo, S., Beyene, Y. & Vidal, S., 2009. Resistance of maize varieties to the maize weevil Sitophilus zeamais (Motsch.)(Coleoptera: Curculionidae). *African Journal of Biotechnology*, **8**: 5937–5943.

Akhurst, R., 1980. Morphological and functional dimorphism in *Xenorhabdus* spp., bacteria symbiotically associated with the insect pathogenic nematodes Neoaplectana and Heterorhabditis. *Journal of General Microbiology*, **121**: 303–309.

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

Akhurst, R. J., 1983. Taxonomic Study of Xenorhabdus, a Genus of Bacteria Symbiotically Associated with Insect Pathogenic Nematodes. *International Journal of Systematic Bacteriology*, **33**: 38–45.

Akhurst, R. J. & Boemare, N.E., 1988. A numerical taxonomic study of the genus Xenorhabdus (Enterobacteriaceae) and proposed elevation of the subspecies of X. nematophilus to species. *Journal of General Microbiology*, **134**: 1835–1845.

Aktar, M. W., Sengupta, D. & Chowdhury, A., 2009. Impact of pesticides use in agriculture: their benefits and hazards. *Interdisciplinary Toxicology*, **2**: 1–12.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, **25**: 3389–3402.

Athanassiou, C.G., Palyvos, N.E. & Kakouli-duarte, T., 2008. Insecticidal effect of Steinernema feltiae (Filipjev) (Nematoda : Steinernematidae) against Tribolium confusum du Val (Coleoptera : Tenebrionidae) and Ephestia kuehniella (Zeller) (Lepidoptera : Pyralidae) in stored wheat. *Journal of Stored Products Research*, **44**: 52–57.

Ausubel, F.M., Brent, R., Kingston, R. E., Moore, D. E., Seidman, J. G., Smith, J. A. & Struhl, K., 1987. *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Interscience. pp. 1.2.2.

Bajorath, J., Stenkamp, R. & Aruffo, A., 1993. Knowledge-based model building of proteins: concepts and examples. *Protein science : a publication of the Protein Society*, **2**: 1798–810.

Berge, O., Heulin, Thierry, Achouak, Wafa, Richard, Claude, Bally, Rene & Balandreau, Jacques, 1991. Rahnella aquatilis , a nitrogen-fixing enteric bacterium associated with the rhizosphere of wheat and maize. *Canadian Journal of Microbiology*, **37**: 195–203.

Boemare, N. & Akhurst, R., 2006. The Genera Photorhabdus and Xenorhabdus. *Prokaryotes*, **6**: 451–494.

Bosque-Perez, N.A., 1995. *Major insect pests of maize in Africa: biology and control IITA Research guide 30*, Ibadan, Nigeria.

Bowen, D., Rocheleau, T., Blackburn, M., Andreev, O., Golubeva, E., Bhartia, R. & Ffrench-Constant, R. H., 1998. Insecticidal toxins from the bacterium Photorhabdus luminescens. *Science*, **280**: 2129–2132.

Bowen, D. & Ensign, J., 1998. Purification and characterization of a high-molecular-weight insecticidal protein complex produced by the entomopathogenic bacterium photorhabdus luminescens. *Applied and Environmental Microbiology*, **64**: 3029–3035.

Boxall, R. A., 2002. Damage and Loss Caused by the Larger Grain Borer Prostephanus truncatus. *Integrated Pest Management Reviews*, **7**: 105–121.

Boyd, E. F. & Hartl, D. L., 1998. Salmonella virulence plasmid: Modular acquisition of the spv virulence region by an F-plasmid in Salmonella enterica subspecies I and insertion into the chromosome of subspecies II, IIIa, IV and VII isolates. *Genetics*, **149**: 1183–1190.

Brunel, B., Givaudan, A., Lanois, A., Arkhust, R. J. & 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**: 574–580.

Chaisaeng, P., Chongrattanameteekul, W., Visarathanonth, P. & Vajarasathiara, B., 2010. Laboratory studies on control of the maize weevil Sitophilus zeamais by the parasitoid Anisopteromalus calandrae. *Science Asia*, **36**: 6–11.

Chaston, J.M. Suen, Garret Tucker, Sarah, L, Andersen, Aaron, W., Bhasin, Archna, Bode, Edna, Bode, Helge, B., Brachmann, Alexander, O., Cowles, Charles, E., Cowles, Kimberly, N., Darby, Creg, de Léon, Limaris, Drace, Kevin, Du, Zijin, Givaudan, Alain, Herbert Tran, Erin E., Jewell, Kelsea, Knack, Jennifer, J., Krasomil-Osterfeld, Karina, C., Kukor, Ryan, Lanois, Anne, Latreille, Phil, Leimgruber, Nancy, K., Lipke, Carolyn, M., Liu, Renyi, Lu, Xiaojun, Martens, Eric, C., Marri, Pradeep, R., Médigue, Claudine, Menard, Megan, L., Miller, Nancy, M., Morales-Soto, Nydia, Norton, Stacie, Ogier, Jean-Claude, Orchard, Samantha, S., Park, Dongjin, Park, Youngjin, Qurollo,

Barbara, Sugar, Darby Renneckar, Richards, Gregory, R., Rouy, Zoé, Slominski, Brad, Slominski, Kathryn, Snyder, Holly, Tjaden, Brian, C., van der Hoeven, ansome, Welch, Roy, D., Wheeler, Cathy, Xiang, Bosong, Barbazuk, Brad, Gaudriault, Sophie, Goodner, Brad, Slater, Steven, C., Forst, Steven, Goldman, Barry, S. & Goodrich-Blair, Heidi, 2011. The entomopathogenic bacterial endosymbionts Xenorhabdus and Photorhabdus: convergent lifestyles from divergent genomes. *PloS one*, **6**: 1-13.

Chattopadhyay, A., Bhatnagar, N.B. & Bhatnagar, R., 2004. Bacterial insecticidal toxins. *Critical reviews in microbiology*, **30**: 33–54.

Chen, W. & Kuo, T., 1993. A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. *Nucleic acids research*, **21**: 2260.

Clair, Emilie, Mesnage, Robin, Travert, Carine, Séralini & Gilles-Éric, 2012. A glyphosate-based herbicide induces necrosis and apoptosis in mature rat testicular cells in vitro, and testosterone decrease at lower levels. *Toxicology in vitro*, **26**: 269–279.

Cory, J. & Franklin, M., 2012. Evolution and the microbial control of insects. *Evolutionary Applications*, **5**: 455–469.

Danho, M., Gaspar, C. & Haubruge, E., 2002. The impact of grain quantity on the biology of Sitophilus zeamais Motschulsky (Coleoptera: Curculionidae): oviposition, distribution of eggs, adult emergence, body weight and sex ratio. *Journal of Stored Products Research*, **38**: 259–266.

De Goffau, Marcus, C., Yang Xiaomei, Van Dijl, Jan Maarten & Harmsen, Hermie, J. M., 2009. Bacterial pleomorphism and competition in a relative humidity gradient. *Environmental Microbiology*, **11**: 809–822.

Demissie, G., Tefera, T. & Tadesse, A., 2008. Importance of husk covering on field infestation of maize by Sitophilus zeamais Motsch (Coleoptera: Curculionidea) at Bako, Western Ethiopia. *African Journal of Biotechnology*, **7**: 3777–3782.

Di Luccio, E. & Koehl, P., 2011. A quality metric for homology modeling: the H-factor. *BMC Bioinformatics*, **12**: 48-67.

Donnelly, S. F. H., Pocklington, M. J., Pallotta, D. & Orr, E., 1993. A proline-rich protein, verprolin, involved in cytoskeletal organization and cellular growth in the yeast Saccharomyces cerevisiae. *Molecular Microbiology*, **10**: 585–596.

El-Gedaily, A., Paesold, G. & Krause, M., 1997. Expression profile and subcellular location of the plasmid-encoded virulence (Spv) proteins in wild-type Salmonella dublin. *Infection and Immunity*, **65**: 3406–3411.

Estruch, J., Carozzi, N., Desai, N., Duck, N, B., Warren, G. W. & Koziel, M. G., 1997. Transgenic plants: an emerging approach to pest control. *Nature Biotechnology*, **15**: 137–141.

Estruch, J., Warren, G., Mullins, M., Nye, G. J., Craig, J. A. & Koziel, G., 1996. Vip3A, a novel Bacillus thuringiensis vegetative insecticidal protein with a wide spectrum of activities against lepidopteran insects. *Proceedings of the National Academy of Sciences, USA*, **93**: 5389–5394.

Ffrench-Constant, R. & Waterfield, N., 2005. An ABC guide to the bacterial toxin complexes. *Advances in Applied Microbiology*, **58**: 169–183.

Ffrench-Constant, R. H. & Bowen, D. J., 2000. Novel insecticidal toxins from nematode-symbiotic bacteria. *Cellular and molecular life sciences : CMLS*, **57**: 828–833.

Forst, S. & Nealson, K., 1996. Molecular biology of the symbiotic-pathogenic bacteria Xenorhabdus spp. and Photorhabdus spp. *Microbiological reviews*, **60**: 21–43.

Gatsogiannis, C., Lang, A. E., Meusch, D., Pfaumann, V., Hofnagel, O., Benz, R., Aktories, K. & Raunser, S., 2013. A syringe-like injection mechanism in Photorhabdus luminescens toxins. *Nature*, **495**: 520–523.

Gauler, R. & Boemare, N., 2002. *Entomopathogenic Nematology*, CABI, pp. 1-388.

Ghazal, S. Hurst, Sheldon, G., Morris, Krystalynne, Abebe-Akele, Feseha, Thomas, W., Kelley, Badr, Usama, M., Hussein, Mona, Abouzaied, Mohamed, Khalil, Kamal, M. & Tisa, Louis, S., 2014. Draft Genome Sequence of Photorhabdus luminescens Strain BA1, an Entomopathogenic Bacterium Isolated from Nematodes Found in Egypt. *Genome Announcements*, **2**: 1-2.

Gherardini, P. F. & Helmer-Citterich, M., 2008. Structure-based function prediction: approaches and applications. *Briefings in Functional Genomics & Proteomics*, **7**: 291–302.

Givaudan, A., Baghdiguian, S., Lanois, A. & Boemare, N., 1995. Swarming and Swimming Changes Concomitant with Phase Variation in Xenorhabdus nematophilus. *Applied and environmental microbiology*, **61**: 1408–1413.

Givaudan, A., Lanois, A. & Compare, L. D. P., 2000. flhDC , the Flagellar Master Operon ofXenorhabdus nematophilus : Requirement for Motility , Lipolysis , Extracellular Hemolysis , and Full Virulence in Insects flhDC , the Flagellar Master Operon of Xenorhabdus nematophilus : Requirement for Motility , Li. *Journal of Bacteriology*, **182**: 107–115.

Goubau, P., Van Aelst, F., Verhaegen, J. & Boogaerts, M., 1988. Septicaemia caused by Rahnella aquatilis in an immunocompromised patient. *European Journal of Clinical Microbiology & Infectious Diseases*, **7**: 697–699.

Gouy, M., Guindon, S. & Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, **27**: 221–224.

Guindon, S. & Gascuel, O., 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, **52**: 696–704.

Guo, L., Fatig, R. O., Orr, G. L., Schafer, B. W., Strickland, J. A., Sukhapinda, K., Woodsworth, A. T. & Petell, J. K., 1999. Photorhabdus luminescens W-14 Insecticidal Activity Consists of at Least Two Similar but Distinct Proteins: Purification and characterization of toxin a and toxin b. *Journal of Biological Chemistry*, **274**: 9836–9842.

Haines, C. P., 1991. *Insects and arachnids of tropical stored products: their biology and identification : a training manual* 2, illustr. C. P. Haines, ed., Natural Resources Institute. pp. 1-246.

Hanahan, D., 1983. Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology*, **166**: 557–580.

Herbert, E. E. & Goodrich-Blair, H., 2007. Friend and foe: the two faces of Xenorhabdus nematophila. *Nature reviews. Microbiology*, **5**: 634–646.

Hill, D. S., 1983. *Agricultural Insect Pests of the Tropics and Their Control* illustrate., CUP. pp. 1-746.

Hillisch, A., Pineda, L. F. & Hilgenfeld, R., 2004. Utility of homology models in the drug discovery process. *Drug Discovery Today*, **9**: 659–669.

Hinchliffe, S. J., Hares, M., Dowling, A. & Ffrench-Constant, R. H., 2010. Insecticidal toxins from the Photorhabdus and Xenorhabdus bacteria. *The Open Toxicology Journal*, **3**: 101–118.

Izard, D., Gavini, F., Trinel, P. A. & Leclere, H., 1979. Rahnella aquatilis, a new member of the Enterobacteriaceae (author's transl). *Annales de Microbiologie*, **130**: 163–177.

Joo Lee. P., Ahn, J., Kim, Y., Wook Kim, S., Kim, J., Park, J. & Lee, J., 2004. Cloning and heterologous expression of a novel insecticidal gene (tccC1) from Xenorhabdus nematophilus strain. *Biochemical and Biophysical Research Communications*, **319**: 1110–1116.

Kawabata, T., 2003. MATRAS: a program for protein 3D structure comparison. *Nucleic Acids Research*, **31**: 3367–3369.

Kawaka, F., Kimenju, J., Okoth, S. A., Ayodo, G., Mwaniki, S., Muoma, J. & Orinda, G., 2014. Effects of Soil Chemical Characteristics on the Occurrence of Entomopathogenic Nematodes. *British Journal of Applied Science & Technology*, **4**: 2333–2343.

Kawaka, F. J., Kimenju, J. W., Ayodo, G., Mwaniki, S. W., Muoma, J. O., Okoth, S. A. & Orinda, G. O., 2011. Impact of land use on the distribution and diversity of entomopathogenic nematodes in Embu and Taita Districts, Kenya. *Tropical and Subtropical Agroecosystems*, **13**: 59–63.

Kim, D. & Forst, S., 2005. Xenorhabdus nematophila: Mutualist and Pathogen. *ASM News*, **71**: 174–178.

Kimatu, J. N., Mcconchie, R., Xie, X. & Nguluu, S. N., 2012. The Significant Role of Post-Harvest Management in Farm Management , Aflatoxin Mitigation and Food Security in Sub-Saharan Africa. *Greener Journal of Agricultural Sciences*, **2**: 279–288.

Kosmoliaptsis, V., Dafforn, T. R., Chaudhry, A. N., Halsall, D. J., Bradley, J. A. & Taylor, C. J., 2011. High-resolution, three-dimensional modeling of human leukocyte antigen class I structure and surface electrostatic potential reveals the molecular basis for alloantibody binding epitopes. *Human Immunology*, **72**: 1049–1059.

Kranz, J. K., Schmutterer, H. & Koch, W., 1977. *Diseases, Pests and Weeds in Tropical Crops*, Parey Publishers, pp. 400-430.

Krause, M., Fierer, J. & Guiney, D., 1990. Homologous DNA sequences on the virulence plasmids of pathogenic Yersinia and Salmonella dublin lane. *Molecular Microbiology*, **4**: 905–911.

Lane, D. J., 1991. 16S/23S rRNA sequencing. In E. Stackebrandt & M. Goodfellow, eds. *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons, pp. 115 – 175.

Larrain, P. I., Araya, J. E. & Paschke, J. D., 1995. Methods of infestation of sorghum lines for the evaluation of resistance to the maize weevil, Sitophilus zeamais Motschulsky (Coleoptera: Curculionidae). *Crop Protection*, **14**: 561–564.

Lee, M. R., Tsai, J., Baker, D. & Kollman, P. A., 2001. Molecular dynamics in the endgame of protein structure prediction. *Journal of Molecular Biology*, **313**: 417–430.

Lee, S. C., Stoilova-McPhie, S., Baxter, L., Fülöp, V., Henderson, J., Rodger, A., Roper, D. I., Scott, D. J., Smith, C.J. & Morgan, J. A. W., 2007. Structural characterisation of the insecticidal toxin XptA1, reveals a 1.15 MDa tetramer with a cage-like structure. *Journal of Molecular Biology*, **366**: 1558–1568.

Lesieur, C., Vécsey-Semjén, B., Abrami, L., Fivaz, M. & Gisou van der Goot, F., 1997. Membrane insertion: The strategies of toxins (Review). *Molecular Membrane Biology*, **14**: 45–64.

Levitt, M. & Lifson, S., 1969. Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of molecular biology*, **46**: 269–279.

Li, L., 1988. *Behavioural ecology and life history evolution in the larger grain borer, Prostephanus truncatus (Horn)*. Thesis. Department of Pure and Applied Zoology, University of Reading. British Library Document Supply Centre.

Libby, S. J., Lesnick, M., Hasegawa, P., Kurth, M., Belcher, C., Fierer, J. & Guiney, D. G., 2002. Characterization of the spv locus in Salmonella enterica serovar Arizona. *Infection and Immunity*, **70**: 3290–3294.

Liu, D., Burton, S., Glancy, T., Li, Z., Hampton, R., Meade, T. & Merlo, D.J., 2003. Insect resistance conferred by 283-kDa Photorhabdus luminescens protein TcdA in Arabidopsis thaliana. *Nature Biotechnology*, **21**: 1222–1228.

Mahar, A., Munir, M. & Mahar, A., 2004. Studies of different application methods of Xenorhabdus and Photorhabdus cells and their toxin in broth solution to control locust (Schistocerca gregaria). *Asian Journal of Plant Sciences*, **3**: 690–695.

Meusch, D., Gatsogiannis, C., Efremov, R. G., Lang, A. E., Hofnagel, O., Vetter, I. R., Aktories, K. & Raunser, S., 2014. Mechanism of Tc toxin action revealed in molecular detail. *Nature*, **508**: 61–65.

Morgan, J., Sergent, M., Ellis, D., Ousley, M. & Jarrett, P., 2001. Sequence Analysis of Insecticidal Genes fromXenorhabdus nematophilus PMFI296. *Applied and environmental microbiology*, **67**: 2062–2069.

Mortz, E., Krogh, T. N., Vorum, H. & Görg, A., 2001. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics*, **1**: 1359–1363.

Moult, J., 2005. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Current opinion in structural biology*, **15**: 285–289.

Munn, A. L. & Thanabalu, T., 2009. Verprolin: A cool set of actin-binding sites and some very HOT prolines. *IUBMB Life*, **61**: 707–712.

Murphy, J. R., 2011. Mechanism of diphtheria toxin catalytic domain delivery to the eukaryotic cell cytosol and the cellular factors that directly participate in the process. *Toxins*, **3**: 294–308.

Mutambuki, K., Ngatia, C. M. & Mbugua, J. N., 2010. Post-harvest technology transfer to reduce on farm grain losses in Kitui district , Kenya. *10th international working conference on stored product protection*. Estoril, Portugal: Julius-Kühn-Archiv, pp. 984–990.

Mwaniki, S. W., Nderitu, J. H., Olubayo, F., Kimenju, J. W. & Nguyen, Khuong, 2008. Factors influencing the occurrence of entomopathogenic nematodes in the Central Rift Valley Region of Kenya. *African Journal of Ecology*, **46**: 79–84.

Mwololo, J. K., Mugo, S. N., Tefera, T., Okori, P., Munyiri, S. W., Semagn, K., Otim, M., Beyene, Y., 2012. Resistance of tropical maize genotypes to the larger grain borer. *Journal of Pest Science*, **85**: 267–275.

Nayeem, A., Sitkoff, D. & Jr, S. K., 2006. A comparative study of available software for high-accuracy homology modeling : From sequence alignments to structural models. *Protein Science*, **15**: 808–824.

Njoroge, A. W., Affognon, H. D., Mutungi, C. M., Manono, J., Lamuka, P. O. & Murdock, L. L., 2014. Triple bag hermetic storage delivers a lethal punch to Prostephanus truncatus (Horn) (Coleoptera: Bostrichidae) in stored maize. *Journal of Stored Products Research*, **58**: 12–19.

Omondi, B. A., Van den Berg, J., Masiga, D. & Schulthess, F., 2011. Phylogeographic structure of Teretrius nigrescens (Coleoptera: Histeridae) predator of the invasive post harvest pest Prostephanus truncatus (Coleoptera: Bostrichidae). *Bulletin of Entomological Research*, **101**: 521–532.

Omondi, B. A., Van den Berg, J., Masiga, D. & Schulthess, F., 2014. Molecular markers reveal narrow genetic base and culturing-associated genetic drift in Teretrius nigrescens Lewis populations released for the biological control of the larger grain borer in Africa. *Bulletin of Entomological Research*, **104**: 143–154.

Park, D., Ciezki, K., van der Hoeven, R., Singh, S., Reimer, D., Bode, H. B. & Forst, S., 2009. Genetic analysis of xenocoumacin antibiotic production in the mutualistic bacterium Xenorhabdus nematophila. *Molecular Microbiology*, **73**: 938–949.

Parugrug, M. & Roxas, A., 2008. Insecticidal action of five plants against maize weevil, Sitophilus zeamais Motsch.(Coleoptera: Curculionidae). *KMITL Science and Technology Journal*, **8**: 24–38.

Pawlowski, M., Gajda, M. J., Matlak, R. & Bujnicki, J. M., 2008. MetaMQAP: a meta-server for the quality assessment of protein models. *BMC bioinformatics*, **9**: 403.

Pearson, W. R., 2013. An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*: 1–9.

Pei, J., Kim, B. H. & Grishin, N. V., 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research*, **36**: 2295–2300.

Pierri, C. L., Parisi, G. & Porcelli, V., 2010. Computational approaches for protein function prediction: a combined strategy from multiple sequence alignment to molecular docking-based virtual screening. *Biochimica et Biophysica Acta*, **1804**: 1695–1712.

Rahman, M. S. & Labuza, T. P., 2007. *Handbook of Food Preservation, Second Edition* S. M. Rahman, ed., CRC Press. pp. 456-461.

Ramos-Rodriguez, O., Campbell, J. F. & Ramaswamy, S. B., 2006. Pathogenicity of three species of entomopathogenic nematodes to some major stored-product insect pests. *Journal of Stored Products Research*, **42**: 241–252.

Rumbos, C. I. & Athanassiou, C. G., 2012. Insecticidal effect of six entomopathogenic nematode strains against Lasioderma serricorne (F .) (Coleoptera : Anobiidae) and Tribolium confusum Jacquelin du Val (Coleoptera : Tenebrionidae). *Journal of Stored Products Research*, **50**: 21–26.

Ryu, K. Bae, J. S., Yu, Y. S., Park, S. H., 2000. Insecticidal toxin fromXenorhabdus nematophilus, symbiotic bacterium associated with entomopathogenic nematode Steinernema glaseri. *Biotechnology and Bioprocess Engineering*, **5**: 141–145.

Sali, A. & Blundell, T. L., 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, **234**: 779–815.

Saux, M. F. Viallardt, V., Brunelt, B., Normand, P. & Boemarel, N. E., 1999. Polyphasic classification of the genus Photorhabdus and proposal of new taxa : P. Iuminescens subsp. luminescens subsp. nov., P. luminescens subsp. akhurstii subsp. nov., P. luminescens subsp. laumondii subsp. nov., P. temperata sp. nov., P. temperata sub. *International Journal of Systematic Bacteriology*, **49**: 1645–1656.

Scholz, D., Tchabi, A., Markham, R. H., Poehling, H. M. & Borgemeister, C., 1998. Factors affecting pheromone production and behavioral responses by Prostephanus truncatus (Coleoptera: Brostrichidae). *Annals of the Entomological Society of America*, **91**: 872–878.

Sergeant, M., Jarrett, P., Ousley, M. & Morgan, J. A. W., 2003. Interactions of insecticidal toxin gene products from Xenorhabdus nematophilus PMFI296. *Applied and Environmental Microbiology*, **69**: 3344–3349.

Sergeant, M., Baxter, L., Jarrett, P., Shaw, E., Ousley, M., Winstanley, C. & Morgan, J. A. W., 2006. Identification, typing, and insecticidal activity of Xenorhabdus isolates from entomopathogenic nematodes in United Kingdom soil and characterization of the xpt toxin loci. *Applied and Environmental Microbiology*, **72**: 5895–5907.

Sheets, J. J., Hey, T. D., Fencil, K. J., Burton, S.L., Ni, W., Lang, A.E., Benz, R. & Aktories, K., 2011. Insecticidal toxin complex proteins from Xenorhabdus nematophilus: structure and pore formation. *The Journal of Biological Chemistry*, **286**: 22742–9.

Shendure, J. & Ji, H., 2008. Next-generation DNA sequencing. *Nature Biotechnology*, **26**: 1135–1145.

Shikano, I. & Cory, J. S., 2014. Genetic resistance to Bacillus thuringiensis alters feeding behaviour in the cabbage looper, Trichoplusia ni. *PloS one*, **9**: 1-6.

Short, C., Mulinge, W. & Witwer, M., 2012. Analysis of Incentives and Disincentives for Cattle in Kenya. Monitoring African Food and Agricultural Policies project (MAFAP), Food and Agriculture Organization (FAO) Technical Note: pp. 1–49.

Sippl, M. J., 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins*, **17**: 355–362.

Sippl, M. J., 1995. Knowledge-based potentials for proteins. *Current Opinion in Structural Biology*, **5**: 229–235.

Siwale, J., Mbata, K., Microbert, J., Microbert, J. & Lungu, D., 2009. Comparative resistance of improved maize genotypes and Landraces to maize weevil. *African Crop Science Journal*, **17**: 1–16.

Söding, J., Biegert, A. & Lupas, A. N., 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, **33**: 244–248.

Stackebrandt, E. & Goebel, B. M., 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology*, **44**: 846–849.

Tabashnik, B. E., Van Rensburg, J. B. J. & Carrière, Y., 2009. Field-evolved insect resistance to Bt crops: definition, theory, and data. *Journal of Economic Entomology*, **102**: 2011–2025.

Tailliez, P., Pagès, S., Ginibre, N. & Boemare, N., 2006. New insight into diversity in the genus Xenorhabdus, including the description of ten novel species. *International Journal of Systematic and Evolutionary Microbiology*, **56**: 2805–2818.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**: 2731–2739.

Tefera, T., Mugo, S., Tende, R. & Likhayo, P., 2010. *Mass rearing of stem borers, maize weevil, and larger grain borer insect pests of maize*, Nairobi: CIMMYT. pp. 19-32.

Tefera, T., Mugo, S. & Likhayo, P., 2011. Effects of insect population density and storage time on grain damage and weight loss in maize due to the maize weevil Sitophilus zeamais and the larger grain. *African Journal of Agricultural research*, **6**: 2249–2254.

Tooke, F. G.. & Scott, M., 1944. *Wood-boring beetles in South Africa, preventive and remedial measures*, Pretoria. Bulletin No. 247. pp. 1-37.

Venclovas, C., 2003. Comparative modeling in CASP5: progress is evident, but alignment errors remain a significant hindrance. *Proteins*, **53**: 380–388.

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

Waterfield, N. Hares, M., Yang, G., Dowling, A. & Ffrench-Constant, R., 2005. Potentiation and cellular phenotypes of the insecticidal toxin complexes of Photorhabdus bacteria. *Cellular Microbiology*, **7**: 373–382.

Waterfield, N. R., Daborn, P. J. & ffrench-Constant, R. H., 2002. Genomic islands in Photorhabdus. *Trends in microbiology*, **10**: 541–545.

Waturu, C., Hunt, D. & Reid, A., 1997. Steinernema karii sp. n.(Nematoda: Steinernematidae), a new entomopathogenic nematode from Kenya. *International Journal of Nematology*, **7**: 68–75.

Wiederstein, M. & Sippl, M. J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, **35**: 407–410.

Wilkinson, P., Waterfield, N. R., Crossman, L., Corton, C., Sanchez-Contreras, M., Vlisidou, I., Barron, A., Bignell, A., Clark, L., Ormond, D., Mayho, M., Bason, N., Smith, F., Simmonds, M., Churcher, C., Harris, D., Thompson, N. R., Quail, M., Parkhill, J. & Ffrench-Constant, R. H., 2009. Comparative genomics of the emerging human pathogen Photorhabdus asymbiotica with the insect pathogen Photorhabdus luminescens. *BMC Genomics*, **10**: 302-324.

Xiong, J., 2006. *Essential Bioinformatics.* Genomics, bioinformatics and systems biology. Cambridge University Press, New York, USA.

Yadav, M., Singh, A., Rathaur, S. & Liebau, E., 2010. Structural modeling and simulation studies of Brugia malayi glutathione-S-transferase with compounds exhibiting antifilarial activity: implications in drug targeting and designing. *Journal of Molecular Graphics & Modelling*, **28**: 435–445.

Yang, J., Zeng, H., Lin, H., Yang, X., Liu, Z., Guo, L., Yuan, J., Qiu, D., 2012. An insecticidal protein from Xenorhabdus budapestensis that results in prophenoloxidase activation in the wax moth, Galleria mellonella. *Journal of Invertebrate Pathology*, **110**: 60–67.

Young, J. A. T. & Collier, R. J., 2007. Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annual Review of Biochemistry*, **76**: 243–265.

Zemla, A., 2003. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Research*, **31**: 3370–3374.