

**CHARACTERISATION
OF *LYSINIBACILLUS SPHAERICUS* STRAINS**

by

Hari Purwanto

Being a thesis presented in accordance with the
regulations governing the award of the degree of
Philosophiae Doctor in Cardiff University

Date: April, 2016

**Cardiff School of Biosciences
Cardiff University**

Candidate ID number : 0834759

Candidate's last name : Purwanto

Candidate's first names : Hari

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed Date

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD

SignedDate

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references. The views expressed are my own.

Signed Date

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed Date

STATEMENT 4:

PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed Date

Summary

Some strains of *Lysinibacillus sphaericus* bacteria produce proteins toxic to insects, in particular mosquitoes. The bacteria have been used on a large scale to control these disease vectors. Unfortunately, there have been some reports of insect resistance to *L. sphaericus*. Interestingly, there are also reports of new isolates that may provide an opportunity to exploit these bacteria in other ways. To achieve better control of insect vectors, studies of various aspects of the *L. sphaericus* and its toxins are needed.

This dissertation reports properties of a newly isolated *L. sphaericus* isolate namely, Fang Large. Based on its morphological features and 16S RNA DNA sequence analysis, the Fang Large isolate was confirmed as a *Lysinibacillus sphaericus* strain. This isolate exerts its pathogenicity to Wax moth larvae through septicaemia. Next, regulation of toxin gene expression was performed through analysis of 5' untranslated regions of *L. sphaericus* toxin genes of strain IAB59 using a 5' RLM-RACE method to reveal the promoters of the toxin genes. These investigations did not always produce the expected results but 5'UTR analysis of the *mtx2* sequence confirmed the previously predicted promoter, while for two other genes –*binB* and *cry48*– results indicated that the actual promoters are not those previously predicted and that unreported promoter sequences appear to act in regulating these genes. To study the genomic arrangement of *L. sphaericus* strain NHA15b and its toxin protein genes, next generation sequencing technology was employed. This showed that the chromosomal structure of strain NHA15b is similar to the previously sequenced *L. sphaericus* genomes, but lacked a 35 kb fragment that in strain C3-41 is known to harbour *binA/B* and *mtx4* genes. There are at least five (possibly six) plasmids in *L. sphaericus* strain NHA15b. Two of them, pLsph100 and pLsph200 are similar to other plasmids reported from other *L. sphaericus* strains. Plasmids pLsph300, pLsph400 and pLsph500, in contrast, have never been reported in any *L. sphaericus* strain. The sixth plasmid, pLsph600 was predicted based on the presence of a T4SS system and DNA replication protein genes. The genes may be responsible for the conjugation ability of the particular plasmid. It is predicted that the *cry48/49* genes of the strain NHA15b may be located in this putative plasmid. The prediction is supported by indications from physical separation efforts using Pulsed Field Gel Electrophoresis, that the *cry* genes were located in a plasmid with mobility equivalent to 750 kb linear DNA. This prediction still needs further verification.

Table of Contents

SUMMARY	III
CHAPTER I INTRODUCTION	1
1.1. INSECT BORNE DISEASES	1
1.2. VECTOR CONTROL	2
1.3. <i>LYSINIBACILLUS SPHAERICUS</i> AS A BIOLOGICAL CONTROL AGENT	4
1.4. SYSTEMATICS OF <i>LYSINIBACILLUS SPHAERICUS</i>	4
1.5. <i>LYSINIBACILLUS SPHAERICUS</i> TOXINS	6
1.5.1 Binary toxin	6
1.5.2. Mosquitocidal toxins	7
1.5.3. Cry 48/Cry 49 toxins	8
1.5.4. Sphaericolysin	9
1.5.5. Surface layer protein	10
1.6. <i>L. SPHAERICUS</i> APPLICATION TO CONTROL MOSQUITOES	10
1.7. MOSQUITOES RESISTANT TO <i>L. SPHAERICUS</i>	11
CHAPTER II MATERIALS AND METHODS	13
2.1. MATERIALS	13
2.1. 1. Bacterial strains and plasmids	13
2.1.2. Media	14
2.1.3. Molecular weight marker	14
2.2. METHODS	14
2.2.1. Polymerase chain reaction	14
2.2.2. Agarose gel electrophoresis	15
2.2.3. Gel extraction and purification	15
2.2.4. Preparation of plasmid DNA	15
2.2.5. Restriction digest of plasmid DNA	16
2.2.6. Ligation of DNA into plasmid vectors	16
2.2.7. Preparation of chemically competent <i>E. coli</i> cells	16
2.2.8. Transformation of plasmid DNA into chemically competent <i>E. coli</i>	17
2.2.9. Preparation of electrocompetent <i>L. sphaericus</i> cells	17
2.2.10. Transformation of plasmid DNA into competent <i>L. sphaericus</i> using electroporation	17
2.2.11. DNA Sequencing	18
2.2.12. Induction of recombinant protein expression in <i>E. coli</i>	18
2.2.13. SDS-polyacrylamide gel electrophoresis (SDS-PAGE)	18
2.2.14. Preparation of protein samples for N-terminal sequencing	19
2.2.15. Crystal purification	20
2.2.16. Mega plasmid extraction	20
2.2.17. Pulsed Field Gel Electrophoresis	21
2.2.18. Southern blot	21
2.2.19. Dot blotting of DNA onto membrane	22

2.2.20. PCR generation of DIG labelled probes	22
2.2.21. RNA extraction	23
2.2.21. 5' RNA ligase Mediated Rapid Amplification of cDNA Ends (5'RLM-RACE)	23
2.3.1. Transmission electron microscopy	25
2.3.2. Qualitative bioassay	26
CHAPTER III CHARACTERIZATION OF FANG LARGE ISOLATE	27
3.1. INTRODUCTION	27
3.2. MICROSCOPIC ANALYSIS OF FANG LARGE ISOLATE	27
3.3. 16S rRNA ANALYSIS	30
3.4. PROTEIN IDENTIFICATION	33
3.5. QUALITATIVE BIOASSAY	37
CHAPTER IV DETERMINATION OF LOCATION OF CRY 48/49 GENES	41
IN <i>LYSINIBACILLUS SPHAERICUS</i> STRAINS	41
4.1. INTRODUCTION	41
4.2. PLASMID SEPARATION USING PULSED FIELD GEL ELECTROPHORESIS	42
4.3. DETERMINATION OF LOCATION OF <i>CRY 48/49</i> GENES	44
CHAPTER V 5' UNTRANSLATED REGION ANALYSIS OF <i>LYSINIBACILLUS SPHAERICUS</i> TOXIN GENES	50
5.1. INTRODUCTION	50
5.2. 5'UTR ANALYSIS OF THE <i>CRY48AA1</i> TOXIN GENE	53
5.3. 5'UTR ANALYSIS OF THE <i>CRY49AA1</i> TOXIN GENE	55
5.4. 5'UTR ANALYSIS OF THE <i>MTX1</i> GENE	57
5.5. 5'UTR ANALYSIS OF THE <i>MTX2</i> TOXIN GENE	59
5.6. 5'UTR ANALYSIS OF THE <i>MTX3</i> TOXIN GENE	61
5.7. 5'UTR ANALYSIS OF THE SPHAERICOLYSIN TOXIN GENE	63
5.8. 5'UTR ANALYSIS OF THE <i>BINB</i> TOXIN GENE	63
CHAPTER VI WHOLE GENOME SEQUENCING OF <i>LYSINIBACILLUS SPHAERICUS</i> STRAIN NHA15B	66
6.1. INTRODUCTION	66
6.2. GENERAL NHA15B GENOME COMPARISON	69
6.3. THE PLASMIDS OF STRAIN NHA15B	72
6.3.1. Plasmid <i>pLsph100</i>	73
6.3.2. Plasmid <i>pLsph200</i>	77
6.3.3. Plasmid <i>pLsph300</i>	79
6.3.4. Plasmid <i>pLsph400</i>	82
6.3.5. Plasmid <i>pLsph500</i>	85

6.4. INSECTICIDAL TOXINS OF <i>L. SPHAERICUS</i> STRAIN NHA15B	90
6.4.1. <i>Mtx1</i> toxin	90
6.4.2. <i>Mtx2</i> toxin	95
6.4.4. <i>Mtx4</i> toxin	98
6.4.5. <i>Cry48/49</i> toxins and predicted location of the genes	101
6.4.6. <i>Sphaericolysin</i> toxin	118
CHAPTER VII GENERAL DISCUSSION	119
REFERENCES	127
APPENDIX	135
16S RIBOSOMAL RNA PRIMER:	135

CHAPTER I INTRODUCTION

1.1. Insect borne diseases

Several species of insects including mosquitoes and sand flies are vectors of very important diseases such as malaria, filariasis, dengue/dengue haemorrhagic fever, onchocerciasis (river blindness), lymphatic filariasis, yellow fever, and African trypanosomiasis (sleeping sickness). Every year, these diseases cause a huge amount of morbidity and mortality. Plasmodium parasites are the causal agent for malaria. There are five species that are responsible for human malaria, *Plasmodium falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*. All of them are transmitted through the bites of *Anopheles* mosquitoes. It was estimated that at least 3 billion people in 109 countries/territories are at risk of malaria. There are around 250 million cases per year and over 1 million of them eventually will die of malaria (Cox-Singh et al., 2008; WHO, 2008).

Dengue is a viral disease mainly transmitted by the *Aedes aegypti* mosquito, although in some parts of the world, a second species *A. albopictus* is also becoming an important vector. In the most severe case, the disease will manifest into haemorrhagic fever. The disease has spread to more than 100 countries, placing more than 2.5 billion people at risk. It was estimated that around 500,000 patients with Dengue Haemorrhagic Fever (DHF) must be hospitalized each year. Most of them are children and 2.5% of them will die of DHF (Guzman et al., 2010).

Lymphatic filariasis or elephantiasis is a debilitating disease caused by several species of nematode from the family Filarioidea, particularly *Wuchereria bancrofti*, *Brugia malayi* and *B. timorensis*. Some genera of mosquitoes including *Anopheles*, *Aedes*, *Culex*, *Mansonia* and *Ochlerotatus* are vectors of the disease. The most

prominent sign of the diseases is enlargement of lower limbs or genitals of the infected patient because of fluid retention and tissue swelling due to the blockage of their lymphatic system. There are more than 1 billion people in 80 countries around the world at risk of the disease. It has been estimated that the diseases infected more than 120 million people worldwide, and caused a large disabilities to 40 million of them. It was also estimated that 27 million people, most of them males, have genital problems and another 16 million, mostly are women, have lymphodema or elephantiasis of the leg (WHO, 2010)

1.2. Vector Control

The control of insect borne diseases can be targeted in two broad approaches. Directly, to immunize the population that is potentially exposed to the diseases, and to cure the infected person to eliminate the presence of human reservoir or indirectly, to reduce or stop transmission by insect vectors. In the case of malaria, the problem of drug resistance, lack of an effective vaccine, and the presence of carriers without any sign of infection that can carry and transmit the *Plasmodium*, make vector control a better option (Michalakis and Renaud, 2009). There are several vector control tools for malaria. Takken and Knols (2009), list several tools that are already proven as effective and some prospective measures that still need further testing or improvement. Indoor spraying is effective with those species that rest indoors after biting their prey. Long lasting insecticide-treated nets are useful to kill the mosquito and also protect the target against biting. House improvement that prevents mosquito entry is also useful to lessen malaria transmission. Drainage of breeding sites and modification of paddy culture are environmental sanitation measures that are effective to reduce mosquito

populations. Some larvivorous fish are also useful to check the mosquito population. The potency of repellent, bait, or traps to control mosquitoes is under development.

While genetically engineered mosquitoes are still under development, the sterile insect technique is at the field test stage. One of the sterile insect techniques employed is a transgenic strain of *Aedes aegypti* engineered to carry a dominant, non-sex-specific, late-acting lethal genetic system that is repressed in the presence of tetracycline (Massonnet-Bruneel et al., 2013). The approach was field tested in Malaysia with promising results (Lacroix et al., 2012). In addition, Takken and Knols (2009), mentioned several entomopathogens including viruses, fungi, *B. thuringiensis israelensis* (Bti) and *L. sphaericus* as potential candidates for control measures (Takken and Knols, 2009).

A vaccine for Dengue virus is possible but, since there are four types of Dengue virus, to be effective the vaccination programme must cover all the types. Due to this problem, vector control is more favourable. Several methods are recommended to control *Aedes* mosquitoes. Since *Aedes aegypti* adults use clean, stagnant water, in even small amounts to lay the eggs, the control is focused on elimination of clean water reservoirs in and around the houses. For bigger containers, where cleaning is technically difficult, certain chemical insecticides and *B. thuringiensis israelensis* are recommended. In epidemic situations, or if an epidemic situation is imminent, space spraying using chemical insecticides is also recommended (WHO, 2009).

Mass drug administration to all the population at risk is the current strategy adopted by the WHO to eliminate lymphatic filariasis. Although the method is successful, it was suggested that vector control measures along with mass drug

administration would give a better result to expedite and sustain the interruption of filariasis transmitted by *Anopheline* and *Culex* mosquitoes (Bockarie *et al.*, 2009).

1.3. *Lysinibacillus sphaericus* as a biological control agent

Some bacteria are known to produce toxins that can be exploited to control mosquitoes. A general interest in the use of bacterial pathogens to control mosquitoes was raised after the discovery of a mosquitocidal isolate of *B. thuringiensis* from the Negev dessert in Israel (Goldberg *et al.*, 1977). A strain of *L. sphaericus* pathogenic to mosquito had been isolated several years before but, due to its low toxicity, there was no significant effort to utilise it. It was not until the discovery of several highly toxic *L. sphaericus* strains by a WHO organised programme, that its potential as a biological control agent of insect vectors was developed (Singer, 1988).

In this report, introduction will be focused on *L. sphaericus* and when appropriate, its comparison with *B. thuringiensis*.

1.4. Systematics of *Lysinibacillus sphaericus*

Previously, all mesophilic aerobic rod shaped organisms producing a spherical spore were classified as *Lysinibacillus sphaericus*. This approach was due to a lack of differentiating tests that positively characterised *L. sphaericus*. Using DNA hybridisation techniques, this species can be separated into several groups, with the *L. sphaericus sensu stricto* belonging to group I and the mosquito pathogenic strains belonging to another subgroup, namely subgroup IIA (Krych *et al.*, 1980a). Several techniques were developed to characterise *L. sphaericus* further including, phage typing (Yousten, 1984) and flagellar serotyping with antisera (Debarjac *et al.*, 1985).

Using these systems, strain 1593 M -a pathogenic strain isolated from Indonesia- for example, is grouped into DNA group IIA, phage type 3, flagellar type 5a5b.

Based on 16S rDNA sequence analysis, it was revealed that *L. sphaericus sensu lato* is a diverse group which possibly contains at least 13 *Bacillus/Lysinibacillus* species (Nakamura, 2000). *L. sphaericus sensu stricto* clustered in a separate group, with the pathogenic strains falling into another group closer to *Lysinibacillus fusiformis*. The results were in accordance with the previous numerical taxonomy analysis that suggested the heterogeneity of the group (Alexander and Priest, 1990). This discovery suggests that a different species status is appropriate for pathogenic strains and also gives a practical approach to isolate new pathogenic strains from the field.

The 16S rDNA sequence analysis of *Bacillus spp.* also revealed some unique signatures, sequences of conserved nucleotides that could be utilised to assign a particular strain into its species status. Using MEME (Multiple EM for Motif Elicitation) search there are five signatures that are specific to *L. sphaericus* and differentiate it from other *Bacillus* (Porwal *et al.*, 2009).

Based on distinctive peptidoglycan composition that differs from other *Bacillus* species, *L. sphaericus* along with *L. fusiformis* and a newly discovered species that shares a similar peptidoglycan type A4 α (Lys-asp) cell-wall, was moved into a new Genus *Lysinibacillus* (Ahmed *et al.*, 2007). To date, there are 21 species belonging to the genus *Lysinibacillus* in the LPSN database (<http://www.bacterio.net/lysinibacillus.html>), and one other new species reported as *L. xyleni* appears in publication (Begum *et al.*, 2016). For the purpose of this thesis, the new species name of *Lysinibacillus sphaericus* will be used.

1.5. *Lysinibacillus sphaericus* toxins

Several strains of *L. sphaericus* produce proteins that are toxic to insects particularly mosquitoes. Some strains, only produce one type of toxin, while other strains could produce Binary (Bin), Mtx and Cry toxins altogether.

1.5.1 Binary toxin

Binary toxin is produced by highly pathogenic *L. sphaericus* strains. The toxins consist of two proteins 42 kDa (BinA) and 52 kDa (BinB) in size. It appears as a crystal enclosed in the exosporium formed during the onset of the sporulation stage (Broadwell and Baumann, 1986). After ingestion, the toxin will be solubilised in the midgut and form a heterodimeric protein, the so called binary toxin (Baumann *et al.*, 1991). BinB is the main binding component to brush border membrane that mediates regional binding and internalisation of BinA (Oei *et al.*, 1992). It was shown that BinB binds to brush border membrane, but BinA will bind non-specifically until BinB is present at an equal molar ratio (Charles *et al.*, 1997). Using permeabilisation of model lipid membranes it was suggested that BinA engages in membrane pore formation, while BinB assists and facilitates the formation of the pores (Schwartz *et al.*, 2001). Recently, an alternative mechanism of action involving apoptosis has been proposed to occur following internalisation of both BinA and BinB (Tangsongcharoen *et al.*, 2015). The Bin toxin attaches to the brush border membrane of *Culex pipiens* via an α -glucosidase toxin binding receptor called Cpm1 (*Culex pipiens* maltase 1). This receptor is GPI- anchored to the membrane (Darboux *et al.*, 2001; Silva-Filha *et al.*, 1999). The crystal structure of BinB has recently been published but, at this point, the regions responsible for receptor binding specificity are unknown (Srisucharitpanit *et al.*, 2014)

The binary toxin genes are highly conserved amongst *L. sphaericus* strains. There are six variable amino acids in the known sequences of both BinA and BinB proteins. Based on these differences, the BinA and BinB toxins were designated into four types (Humphreys and Berry, 1998). The genes for binary toxin were thought to be chromosomal (Singer, 1988) but a recent whole genome analysis of strain C3-41, showed that at least in this strain, the genes occur both in the chromosome and plasmid (Hu et al., 2008a). Previous results suggested a similar situation in strain 2297 (Poncet *et al.*, 1997).

1.5.2. Mosquitocidal toxins

During their vegetative growth, some strains of *L. sphaericus* produce other proteins toxic to mosquitoes. A gene encoding a 100 kDa protein from strain SSII-1 was cloned sequenced and expressed (Thanabalu et al., 1991). The product protein called Mtx1 showed toxicity to *C. quinquefasciatus* and *A. aegypti* larvae. Mtx1 showed some degree of homology to ADP-ribosyltransferase toxins. It was shown that the gene also occurred in several other *L. sphaericus* strains. A further analysis of the strain SSII-1 revealed a different kind of mosquitocidal toxin 31.8 kDa in size, designated as Mtx2 (Thanabalu and Porter, 1996). Mtx2 does not show significant homology to Mtx1, instead it shows homology to ϵ -toxin of *Clostridium perfringens* and the cytotoxin of *Pseudomonas aeruginosa*. The strain SSII-1 also revealed another toxin (Mtx3) with a size of 35.8 kDa related to Mtx2 (Liu et al., 1996). The Mtx2 and Mtx3 toxins are also expressed in vegetative stage of several of high and low pathogenic *L. sphaericus* strains. Another Mtx toxin (Mtx4) that is similar to Mtx2/3 is also predicted from the whole genome sequencing of strain C3-41, which also contains another pseudogene gene in this family (Hu et al., 2008a).

The abundance of Mtx1 in sporulated cultures of *L. sphaericus* is very low. It is produced during the vegetative stage and is degraded by proteinases produced by the host strain (Yang et al., 2007). In a proteinase-negative strain (1693), the Mtx1 is better preserved during sporulation, possibly due to differing activity of proteinases (Thanabalu and Porter, 1995). Since the Mtx concentration in *L. sphaericus* spore-based insecticide preparations is undetectable, the contribution of these toxins for pathogenicity of spore preparation used in the field is negligible. The Mtx1 and Mtx2 toxins synergize the activity of *L. sphaericus* Bin toxin and Cry11Aa from Bt on both Bin susceptible and Bin resistant *C. quinquefasciatus* larvae, so the potential of the Mtx toxins to increase toxicity and overcome the problem of *L. sphaericus* resistance should be considered (Wirth et al., 2007).

1.5.3. Cry 48/Cry 49 toxins

The presence of another mosquitocidal toxin apart from Bin and Mtx toxins was predicted in several strains that showed toxicity towards a Bin toxin resistant population of mosquitoes. Strain LP1-G, which is incapable of producing Mtx1 toxin showed a significant toxicity against both Bin-susceptible or resistant mosquito larvae (Yuan et al., 2001). Meanwhile, strain IAB59 also showed a high pathogenicity toward mosquitoes that were resistant to strain 2362 (Nielsen-LeRoux et al., 2001). When a laboratory colony was selected with three strains of *L. sphaericus*, the strain IAB59 also caused a slower development of resistance compared to strains C3-41 and 2362. Therefore, it was predicted that strain IAB59 may produce a different kind of toxin with a different mode of action to Binary toxin (Pei et al., 2002).

Subsequently the new toxin was discovered to be a further two component toxin comprising 49 kDa and 135 kDa proteins, designated formally as Cry49Aa1 and Cry48Aa1 respectively (Jones et al., 2007). The Cry48 forms an amorphous crystal

while, the Cry49 forms a bipyramidal crystal during sporulation. In contrast with Bin toxin, both of the crystals were deposited outside the exosporium. Cry48 shared about 33% sequence similarity with Cry4Aa, a member of three-domain Cry protein family of *B. thuringiensis* that is toxic to mosquitoes. Cry49 is closely related to Cry36, a protein toxic to coleopteran larvae and to both of the Binary toxin proteins BinA and BinB. However, Cry48 and Cry49 are only toxic to *C. quinquefasciatus* larvae when both occur together (Jones et al., 2007). The host range of the Cry48/Cry49 seems very restricted to *C. quinquefasciatus*, and probably to *C. pipiens*, which is unusual compared to a relatively wide host range of Bin and Cry toxins (Jones *et al.*, 2008).

The ultrastructural effects of Cry48/Cry49 treatment on midgut cells of *C. quinquefasciatus* were similar to the effects of Bin/Cry11Aa mixture, suggesting a combination of the effect of Cry-like and Bin-like toxins. The effects included: intense mitochondrial vacuolation, breakdown of endoplasmic reticulum, production of cytoplasmic vacuoles, and microvillus disruption. It was also shown that the Cry48/Cry49 exhibits its effect independent of the presence of Bin receptor, since the toxins cause the same effects both in Bin susceptible and resistant larvae (de Melo *et al.*, 2009).

In addition to strain IAB59 and LP1-G, the Cry48/Cry49 toxins are also known to present in several strains namely, 47-6B, 2173, LB29, LP18, LP148 and NHA15b (Jones and Berry, unpublished data).

1.5.4. Sphaericolysin

Sphaericolysin is a protein first identified from *L. sphaericus* isolated from the crop of ant lion (*Myrmeleon bore*) larvae. The size of the protein is 53 kDa and it has homology to cholesterol binding cytolysin. Thus, the toxicity of this protein is likely

to be exerted by forming pores in the membrane of the target cells. So far, this toxin has been shown to be capable of killing German cockroach and common cutworm *Spodoptera litura* but only when administered by injection (Nishiwaki *et al.*, 2007).

1.5.5. Surface layer protein

The pathogenic capability of the surface layer protein of *L. sphaericus* to kill insects has not yet been explored. There was a report indicating that the surface layer protein of *B. thuringiensis* GP1 is toxic to a coleopteran *Epilachna varivestris* (Pena *et al.*, 2006), but to date, no positive result has been reported for *L. sphaericus*. The surface layer protein C of pathogenic strains of *L. sphaericus*, share a high degree of similarity and the surface layer protein C of strain C3-41, and was shown to have no toxic effect to *C. quinquefasciatus* mosquito larvae (Hu *et al.*, 2008a), but it has not been tested against other insects.

1.6. *L. sphaericus* application to Control Mosquitoes

The progress of *L. sphaericus* utilisation to control mosquitoes has been reviewed recently (Lacey, 2007). There have been several successful trials and mass applications of the pathogen against different target species in different environments and formulations. In general the bacteria had an excellent capability to control *Culex* and *Psorophora* but were less effective against species of *Aedes*, *Ochlerotatus* and *Mansonia* (Lacey, 2007). In a laboratory environment, it was shown that application of a mixture of 10:1 *L. sphaericus* strain 2362 spore/crystal preparation with sporulated *B. thuringiensis* expressing only Cyt1A toxin could enhance the effectiveness of *L. sphaericus* to *A. aegypti* by 3,600 fold (Wirth *et al.*, 2000). In the future, it is possible that the use of such a combination of strains may be able to produce a better control of *A. aegypti*.

1.7. Mosquitoes Resistant to *L. sphaericus*

Several cases of both laboratory and field resistance to *L. sphaericus* have been reported. *L. sphaericus* was used to control populations of *C. pipiens* in the Southern part of France from 1969. After a few generations of strain 2362 application, more than 10,000 fold resistances to this strain was developed (NielsenLeroux *et al.*, 1997). In Southern China, a high level of field resistance to strain C3-41 appeared after 8 years of intensive application of this strain to control *C. quinquefasciatus* (Yuan *et al.*, 2000). After only 5 repeat applications to control *C. quinquefasciatus*, a high level of resistance against strain 2362 was developed in Thailand (Mulla *et al.*, 2001). Some strains e.g. strain C3-41 and 2362 showed a cross-resistance phenomenon, but some strain e.g. LP1-G or IAB59 were able to control the resistant mosquitoes (Yuan *et al.*, 2000) (see section 1.5.3. on Cry48/Cry49 toxins).

There are at least two different mechanisms of resistance to *L. sphaericus*. Working on a laboratory selected resistant colony of *C. pipiens* from California, it was shown that a single point mutation of *cpm1*_{GEO} produced a secreted form of the receptor that has lost its membrane anchor (Darboux *et al.*, 2002). Another mutation occurred in a *C. quinquefasciatus* laboratory selected colony resistant to strain 2362. A 19 nucleotide deletion of the *cqm1* gene, an ortholog of the *cpm1* gene, resulted in a truncated gene that was not expressed (Romao *et al.*, 2006). A non-anchored protein receptor, or even the loss of protein receptor thus leads to resistance.

To overcome the problem of resistance to *L. sphaericus*, several strategies have been explored. There are some efforts to do field exploration and screening in the hope of finding more strains with novel insecticide properties (Dias *et al.*, 1999; Monnerat *et al.*, 2004). The combination of *L. sphaericus* and *B. thuringiensis*

applications could halt the occurrence of resistance (Zahiri and Mulla, 2003). To halt or to delay the occurrence of resistance more systematically, an introduction of *Bt* toxin into *L. sphaericus* was conducted either using recombinant DNA or non-recombinant technology. Introduction of *cryIIA* and *cryIIBa* into the *L. sphaericus* strain 2362 chromosome using *in vivo* recombination produces a recombinant that is toxic to *A. aegypti*, although the product did not show significant toxicity differences towards susceptible *C. quinquefasciatus*, it did perform better towards resistant mosquitoes (Servant *et al.*, 1999). Insertion of the pBtoxis plasmid that carries Cry and Cyt toxins from *Bti* genes into *L. sphaericus* using a conjugal transfer method was successful, although segregational instability caused a substantial loss of plasmid during sporulation (Gammon *et al.*, 2006). However, in other studies, an attempt to produce a recombinant *B. thuringiensis* which combines the presence of the *Bti* Cry toxins, Cyt1A and *L. sphaericus* binary toxin, produced a product with more than 20 fold toxicity compared to its parental bacterium towards *C. quinquefasciatus* and *Culex tarsalis* larvae (Park *et al.*, 2005).

CHAPTER II MATERIALS AND METHODS

2.1. Materials

2.1. 1. Bacterial strains and plasmids

Brevibacillus and *Lysinibacillus spp.* strains used in these experiments are listed in the Table 2.1. below.

Table 2.1. *Brevibacillus* and *Lysinibacillus spp.* strains used in this work

No.	Species	Strain
1.	<i>Lysinibacillus sphaericus</i>	IAB59
2.	<i>Lysinibacillus sphaericus</i>	LP1-G
3.	<i>Lysinibacillus sphaericus</i>	NRS 400
4.	<i>Lysinibacillus sphaericus</i>	47-6B
5.	<i>Lysinibacillus sphaericus</i>	2362
6.	<i>Lysinibacillus sphaericus</i>	1691
7.	<i>Lysinibacillus sphaericus</i>	LP18
8.	<i>Lysinibacillus sphaericus</i>	LP1-48
9.	<i>Lysinibacillus sphaericus</i>	1593
10.	<i>Lysinibacillus sphaericus</i>	31-1
11.	<i>Lysinibacillus sphaericus</i>	2173
12.	<i>Lysinibacillus sphaericus</i>	1691
13.	<i>Lysinibacillus sphaericus</i>	NRS 592
14.	<i>Lysinibacillus sphaericus</i>	KellenQ
15.	<i>Lysinibacillus sphaericus</i>	SSII-1
16.	<i>Lysinibacillus sphaericus</i>	LB29
17.	<i>Lysinibacillus sphaericus</i>	NHA15b
18.	<i>Lysinibacillus sphaericus</i>	PR-1
19.	<i>Lysinibacillus sphaericus</i>	C3-41
20.	<i>Lysinibacillus sphaericus</i>	ATCC 14577
21.	<i>Brevibacillus thuringiensis</i>	HD73
22.	<i>Brevibacillus laterosporus</i>	NRS590
23.	<i>Brevibacillus laterosporus</i>	NRS1642

Lysinibacillus sphaericus isolate Fang Large was obtained from a spider cadaver by Dr. Nick Waterfield, University of Bath. Plasmid pGFP304, which allows production of GFP during vegetative growth was a kind gift from Professor Lin Li (Huazhong Agricultural University, Wuhan, China). *Escherichia coli* strain DH5 α (*fhuA2* Δ (*argF-lacZ*)*UI69 phoA glnV44* Φ 80 Δ (*lacZ*)*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) was used to carry out all genetic manipulation while *E. coli* strain BL21

(DE3)pLysS (*E. coli* BF *dcm ompT hsdS*(r_B⁻ n_B⁻) gal λ (DE3) <pLys Cam^r>) was used to express recombinant protein products.

Plasmid pET 28a (Novagen) was used as the vector to express the gene of interest.

2.1.2. Media

Luria Bertani (LB) medium (Sigma Chemical Ltd.) was used as a general rich medium to culture all bacteria. Embrapa medium was used to culture *Brevibacillus* and *Lysinibacillus spp.* for production of crystals and bioassay preparation. Embrapa medium consists of 8 g/l nutrient broth, 1 g/l of yeast extract, 1 g/l K₂HPO₄ and 10 ml of salt solution (100mM CaCO₃, 40 mM MgSO₄, 3.6 mM FeSO₄, 3.6 mM MnSO₄, 3.5 mM ZnSO₄) was added to 1 litre of medium. The medium was adjusted to pH 7.0 before autoclaving at 121 °C (975 kPa) for 20 minutes. Spizizen broth consist of 0.2% NH₄SO₄, 1.4% K₂HPO₄, 0.6% KH₂PO₄, 0.1% sodium citrate, 0.02% MgSO₄ 7H₂O) supplemented with 0.5% glucose, 0.1% Casamino Acids (Difco), and 0.01% yeast extract (Reyes-Ramirez and Ibarra, 2008).

2.1.3. Molecular weight marker

Molecular weight markers ØX174 DNA/*Hae*III and λDNA/*Hind*III (Promega) were used for measuring DNA fragment sizes on agarose gel electrophoresis, while a pre-stained broad range protein marker (New England Biolabs) was used for sizing protein electrophoresis analysis.

2.2. Methods

2.2.1. Polymerase chain reaction

PCR reactions were carried out in sterile 500 µl PCR tubes. The PCR reaction consisted of 5x GoTaq© buffer, 2 mM MgCl₂, 0.2 mM of dNTPs, 0.5 µM each of appropriate forward and reverse oligonucleotide primers, DNA template and 2.5 U of

GoTaq® flexi DNA polymerase (Promega). The appropriate amount of sterile nuclease free water was then added to a final volume of 50 µl of PCR reaction. The PCR reactions were performed with parameters described in the results. The resulting PCR products were analysed by agarose gel electrophoresis and purified using the appropriate kit if necessary. EasyA® DNA polymerase (Stratagene) was used on occasions, according to the manufacturer's instructions, to reduce mutation of target genes to be cloned.

2.2.2. Agarose gel electrophoresis

For DNA separation, a 1% agarose gel was prepared by dissolving the agarose in TAE buffer (40 mM Tris acetate, 2mM EDTA, pH 8.3) made from a ready-made 50x TAE buffer stock (National Diagnostics). The solution was then allowed to cool (50°C) and ethidium bromide solution was added to a final concentration 0.5 µg/ml. The solution was poured into a gel tray with gel comb in place.

Samples were mixed well with 6x loading buffer (Promega) before loading into the gel. Gels were submerged in 1x TAE buffer and run at a constant voltage of 100 V.

2.2.3. Gel extraction and purification

DNA fragments of interest were cut with a sterile scalpel from agarose gels. DNA extraction was conducted using the Qiaquick® Gel Purification Kit (Qiagen Ltd.) according to the manufacturer's protocol.

2.2.4. Preparation of plasmid DNA

Cultures of *E. coli* strain DH5α containing plasmids of interest were incubated overnight in 5 ml of LB medium at 37°C and 250 rpm with shaking. The plasmids were extracted using the QIAprep®Spin Miniprep Kits (Qiagen Ltd.) according to the manufacturer's protocol.

2.2.5. Restriction digest of plasmid DNA

Several restriction endonucleases (New England Biolabs) were used according to the manufacturer's instructions. The digested products were analysed by agarose gel electrophoresis. The fragment sizes were determined by comparison of their mobility relative to the appropriate DNA markers.

2.2.6. Ligation of DNA into plasmid vectors

DNA inserts and plasmid vectors were digested with the appropriate restriction endonucleases. Both products were then purified by agarose gel electrophoresis and subsequent gel extraction process. A 1:3 molar ratio of vector: insert was added to 5 µl 2x ligase buffer. Three units of T4 DNA ligase (Promega) were added and an amount of sterile water was added to 10 µl of final ligation reaction. The reaction mixture was then incubated at 4°C overnight.

2.2.7. Preparation of chemically competent *E. coli* cells

A 5 ml culture was prepared from a glycerol stock of the *E. coli* strain of interest. This culture was grown in LB medium at 37°C, 250 rpm overnight. and was used to inoculate 1 litre of fresh LB medium. The culture was grown at 37°C, 250 rpm until D_{600} of the culture reached 0.4-0.5. The culture was placed on ice for 10 minutes before harvesting by centrifugation (2,500 x g, 10 minutes, and 4°C). The pellets were re-suspended in half of the original culture volume of ice-cold CaCl₂-glycerol solution (100 mM CaCl₂, 10% (v/v) glycerol) and incubated on ice for 1 hour. The cells were spun again and finally re-suspended in 0.1 original culture volume of ice-cold CaCl₂-glycerol solution. Cells were incubated on ice for 30 minutes before dividing into 50 µl aliquots and flash freezing in liquid nitrogen. The competent cells were stored at -80°C until needed.

2.2.8. Transformation of plasmid DNA into chemically competent E. coli

Stocks of the appropriate competent cells were thawed on ice. The whole ligation product was added to the thawed cells and mixed by gently pipetting. The mixtures were incubated on ice for 30 minutes, and then subjected to 42°C heat shock for 1.5 minutes. After this, the mixtures were returned to the ice for 1 minute before addition of 1 ml of LB medium without any antibiotics and incubated for 1 hour at 37°C. The growing cells were pelleted by centrifugation at 4,000 rpm for 1 minute. The pellets were re-suspended in 100 µl of LB medium and then spread on LB agar plates containing the appropriate antibiotics. Finally, the plates were incubated overnight at 37°C.

The recombinant products were screened by restriction endonuclease digestion of purified plasmids, PCR analysis and sequencing when appropriate.

2.2.9. Preparation of electrocompetent L. sphaericus cells

A single fresh colony was used to inoculate 5 ml of LB medium in a 50 ml falcon tube. The cultures were incubated at 30°C with vigorous shaking until a D_{600} of 0.5 was reached.

The cultures were kept on ice for 10 minutes before spinning down in a bench centrifuge for 5 minutes. The pellets were re-suspended in 1 ml ice-cold 10% sucrose. The spinning and re-suspension process was repeated three times, and finally the pellets were re-suspended in 250 µl 10% sucrose. The suspensions were used for transformation directly.

2.2.10. Transformation of plasmid DNA into competent L. sphaericus using electroporation

A 120 µl sample of fresh electrocompetent cells was transferred to an ice cold electroporation cuvette. Plasmid DNA (10 µl) was added and mixed carefully by

stirring with a pipette tip. The mixture was left on ice for a few minutes before loading with 670 μ l of 10% sucrose. The cuvette was subjected to an electric pulse 1.8 kV, 400 Ω , 25 μ F. The time constant was kept between 7 and 9 ms during the procedure. Pre-warmed LB (1 ml) was added to the cuvette, and immediately all the suspension was removed to a 15 ml falcon tube. The suspension was grown at 30°C for 1 hour, with vigorous shaking. The cells were concentrated by spinning and then plated on LB medium supplemented with appropriate antibiotics.

2.2.11. DNA Sequencing

DNA samples to be sequenced were sent to Lark Technologies or DNA Sequencing Core, Molecular Biology Unit at Cardiff University, either as purified PCR products or plasmid inserts. The sequences obtained were analyzed using the BlastN program or similar programs as described in the results.

2.2.12. Induction of recombinant protein expression in E. coli

A fresh single colony of *E. coli* BL21 (DE3)pLysS containing plasmids with genes of interest was used to make a starter culture. The starters were grown in (LB) medium with 50 μ g/ml of kanamycin and incubated at 37°C, 250 rpm overnight. The starters (500 μ l) were added to 100 ml of fresh LB with antibiotics and incubated at 37°C, 250 rpm until reaching a D_{600} of 0.6 when 1 mM final concentration of IPTG was added to induce the expression. After 5 hours, the cultures were harvested by centrifugation (20,000 g, 15 minutes) and re-suspended in sterile water. Samples were analysed by SDS-PAGE.

2.2.13. SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was carried out using Bio-Rad Mini PROTEAN II systems according to the manufacturer's instructions. Running gels comprised 10% (w/v) acrylamide, 0.27% (w/v) N-N'methylene bisacrylamide, 375 mM Tris-HCl pH 8.8,

0.1% SDS, 0.05% (w/v) APS and 13.2 mM TEMED. Stacking gels comprised 5% (w/v) acrylamide, 0.14% (w/v) N-N'methylene bisacrylamide, 65 mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% (w/v) APS and 13.2 mM TEMED.

Protein samples were mixed with 5x loading buffer (20 mM Tris-HCl pH 6.8, 0.2% (w/v) SDS, 5% (v/v) glycerol, 0.003% Bromophenol Blue, 286 mM β -mercaptoethanol) before heating at 95°C for 5 minutes. The samples were loaded onto the gels along with pre-stained protein markers. The gels were run at 200 V in Tris glycine running buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% (w/v) SDS) until the blue stain line reached the bottom of the gel. Subsequently, the gels were stained with Coomassie Blue for 30 minutes before de-staining in a washing solution (25% (v/v) methanol, 7% (v/v) acetic acid) until the protein bands were clearly visible.

2.2.14. Preparation of protein samples for N-terminal sequencing

Protein samples were first subjected to SDS-PAGE, in conditions similar to those described previously (Section 2.2.13.), except that the running buffer was Tris tricine buffer (25 mM Tris-HCl pH 8.3, 192 mM tricine, 0.1% (w/v) SDS). Proteins were transferred onto PVDF membrane (Bio-Rad) by electroblotting. The membrane was stained with Coomassie Blue (0.1% (w/v) Coomassie Blue R-250, 50% (v/v) methanol) for 5 minutes, de-stained for 2 minutes (50% (v/v) methanol, 10% (v/v), acetic acid) and then washed with de-ionized water three times over a 5 minute period.

The membrane was allowed to dry overnight on the bench. The protein bands of interest were cut with a sterile scalpel and sent to Alta Bioscience for N-terminal sequencing.

2.2.15. Crystal purification

Crystal protein from strains of *Lysinibacillus* were purified using the method described (Silva-Filha *et al.*, 1997). Cultures of bacteria were grown in Embrapa medium with vigorous shaking at 30°C until more than 90% of bacteria reached sporulation as determined by phase contrast microscopy. The sporulated cultures were harvested by centrifugation (20,000x g, 4°C) and washed in ice cold 1 M NaCl containing 10 mM EDTA. The spore suspensions were harvested and washed twice in ice cold 10 mM EDTA, before finally re-suspending in 1/30 of the original culture volume of ice cold sterile water.

The final spore suspensions were sonicated before separation by centrifugation (110,000x g, 15°C, 16 hours) on a discontinuous sucrose gradient (67/72/79/84% (w/v)) using an SW28 rotor in a Beckman LM-8 ultra centrifuge (Beckman Coulter Ltd., Buckinghamshire, UK).

All apparent bands were extracted and thoroughly washed in sterile distilled water before analysis by SDS-PAGE or electron microscopy.

2.2.16. Mega plasmid extraction

L. sphaericus strains were cultured in 50 ml Spizizen broth to a D_{600} of approximately 1.1 at 30°C and 250 rpm shaking. The cells were pelleted by centrifugation. Each pellet was re-suspended in 20 ml of TES buffer (30 mM Tris base, 5 mM EDTA, 50 mM NaCl; pH 8.0 adjusted with 3 N HCl) then spun again under the same conditions. Cells were resuspended in 2 ml lysis buffer (TES buffer containing 20% sucrose, 2 mg/ml lysozyme, and 1 µl of RNase from a 10 mg/ml stock solution) and incubated at 30°C for 3 hours. The cells were supplemented with 3 ml of 8% SDS in TES buffer and then incubated at 68°C for 10 minutes then, 1.5 ml of 3 M sodium acetate (pH 4.8) was added. The suspensions were kept at -20°C for 30 minutes. The

suspensions were centrifuged at 20,000 x g for 20 minutes at 4°C and then, clear supernatants were subjected to filtration through 3M paper. Two volumes of cold absolute ethanol were added to the filtrate and incubated at -20°C overnight. Plasmid DNA was pelleted at 20,000 x g for 20 minutes at 4°C. The pellet was briefly dried on the bench before solubilisation in sterile distilled water. The plasmid DNA was stored at -20°C.

2.2.17. Pulsed Field Gel Electrophoresis

Pulsed Field Gel Electrophoresis (PFGE) was conducted on Bio-Rad CHEF-DR© II systems. A 13 cm x 11 cm gel was made using 1% of PFGE certified agarose (Sigma) by dissolving the gel in 0.5 x TBE. Plasmid DNA (25 µl) mixed with 6 x loading buffer (Promega) was loaded in the agarose gel. A 50 kb lambda ladder (Biorad) was loaded to make comparison of separated plasmid sizes. The electrophoresis was run at 5 Volts/cm, 50 second initial and 90 second final ramped switch time, for 22 hours at 14°C using 0.5 x TBE buffer. The gel was submerged into ethidium bromide solution, before a photograph was taken under UV transillumination.

2.2.18. Southern blot

Resolved DNA fragments were transferred to nylon membrane using a capillary transfer method. Prior to transfer, the DNA in the agarose gels was depurinated by submerging the gels in 250 mM HCl for 10 minutes. Subsequently the DNA in the gels was denatured using denaturation buffer (0.5 M NaOH, 1.5 M NaCl) for 2 x 15 minutes, rinsed in sterile water twice, then subjected to neutralization solution (0.5 M Tris HCl, pH 7.5; 1.5 M NaCl) for 2 x 15 minutes. A positively charged nylon membrane (Roche Applied Sciences) was used to bind the DNA. The transfers were conducted overnight using 20X SSC solution (3 M NaCl, 0.3 M sodium citrate

adjusted to pH 7.0) as transfer buffer. The DNA was fixed to the membrane by baking at 80°C for 2 hours.

Hybridization and detection procedures were performed according to the manufacturer's protocol using DIG Easy Hyb© (Roche Applied Sciences). Pre-hybridization steps were performed at 60°C for 30 minutes on a shaker incubator. Hybridizations were carried out overnight at the same temperature using a digoxigenin-11-dUTP labelled fragment of DNA based on *cry48* or *cry49* sequences. Before detection of the probe, low stringency washes (2x SSC; 0.1% SDS at room temperature) 2x 5 minutes, followed by high stringency washes (0.1 x SSC; 0.1 % SDS at 60°C) 2 x 15 minutes were conducted. Detection of digoxigenin-labeled probe by enzyme immunoassay and enzyme-catalysed colour reaction with NBT/BCIP were carried out using DIG Nucleic Acid Detection Kit (Roche Applied Sciences) according to the manufacturer's protocol.

2.2.19. Dot blotting of DNA onto membrane

DNA solutions were denatured at 95°C for 5 minutes and were transferred quickly to ice. DNA solutions (5 µl) were dotted onto a positively charged membrane (Roche Applied Sciences). DNA was fixed to the membrane by baking at 80°C for 2 hours. Hybridisation of DIG labelled probes was carried out as described in the previous method (method 2.2.18).

2.2.20. PCR generation of DIG labelled probes

DNA probes used in Southern hybridisation were prepared by incorporating digoxigenin-11-dUTP (Roche Applied Sciences) into DNA fragments generated by PCR according to the manufacturer's recommendations. PCR was performed using genomic DNA of *L. sphaericus* strain IAB59 as template and a reaction mixture containing dATP, dCTP, dGTP (200 µM each), dTTP (130 µM) and DIG-dUTP

(70 μ M). Incorporation of DIG was confirmed by analysing the PCR products by agarose gel electrophoresis.

2.2.21. RNA extraction

RNA extraction was performed using a hot phenol extraction method. One day or 2 day old *L. sphaericus* cultures grown in 20 ml of LB medium were harvested by centrifugation at 20,000 x g, 15 minutes 4°C. The pellets were re-suspended in 750 μ l lysis buffer (2% sodium dodecyl sulphate, 16 mM EDTA, pH 8.0; 20 mM NaCl) and subsequently put in boiling water for 3 min. Following lysis, the RNA was extracted in 750 μ l of hot (65°C) acid buffered phenol twice, then this was repeated with the same amount of acid buffered phenol (22°C) twice, again extracted with a 25:24:1 mix of phenol:CHCl₃: isoamyl alcohol, and finally extracted with 24:1 CHCl₃: isoamylalcohol. RNA was then precipitated by addition of 2.5 volumes of cold absolute ethanol and incubated at -20°C overnight. Pellets were washed with 70% ethanol and re-suspended in 100 μ l sterile nuclease free water. The resulting RNA was further purified using RNeasy© kit (QIAGEN). The concentration and quality of RNA samples were measured using a Nanodrop© machine. The RNA samples were kept at -20°C until used.

The possibility of the presence of contaminating DNA in the RNA preparations was eliminated by DNase I treatment according to the manufacturer's instruction. One μ l 10x reaction buffer and 1 μ l of DNase I (Sigma) were added to 10 μ l of RNA and then incubated at 37°C for 15 minutes. To stop the reaction, 1 μ l of stopping solution was added and the mixture was heated at 70°C for 10 minutes.

2.2.21. 5' RNA ligase Mediated Rapid Amplification of cDNA Ends (5'RLM-RACE)

The 5 prime RNA ligase Mediated Rapid Amplification of cDNA ends was performed using FirstChoice©RLM-RACE kit (Ambion) according to the

manufacturer's protocol. At first, a 5'RACE adapter was ligated to the RNA as follows. In a PCR tube, 1 μ l 10x RNA Ligase buffer was added to 6 μ l of RNA, 1 μ l 5'RACE adapter and 1 μ l of T4 RNA Ligase. All the components were mixed gently, spun briefly and then incubated at 37°C for 1 hour.

RNA ligated with 5'RACE adapter was reverse transcribed into single strand DNA using M-MLV Reverse Transcriptase. An assemblage of 2 μ l ligated RNA, 4 μ l dNTP mix, 2 μ l random decamers, 2 μ l 10x RT buffer, 1 μ l RNase inhibitor, 1 μ l M-MLV Reverse Transcriptase and 8 μ l of nuclease free water in an RNA-free microfuge tube was mixed and then spun briefly. The reaction mixture was incubated at 42°C for 1 hour.

A nested PCR was conducted to amplify 5' un-translated parts of genes of interest (Figure 2.1). A set of primers, with the forward primer complementary to the 5'RACE adapter and a reverse primer complementary to the known coding sequence of the gene of interest, was used for the first round of amplification. Second round PCR used the primer complimentary to the 5'RACE adaptor and another reverse primer that bound within the first amplicon.

The PCR products were analyzed by agarose gel electrophoresis, purified and cloned. The fragments of interest were ligated into pGEM-T vector (Promega) and then transformed to competent *E. coli* cells (NEB Turbo, New England Biolabs) using standard cloning procedures. The plasmids containing the inserted fragments of interest were sent for sequencing. Finally, the sequence data were analyzed. At first, the sequence identity was searched using the BLAST program and the NCBI database. After the identity was confirmed, the sequence was aligned with the corresponding gene

of strain C3-41 in an attempt to determine the transcription start of the gene., Promoter prediction was then undertaken on the upstream fragment of C3-41 toxin genes.

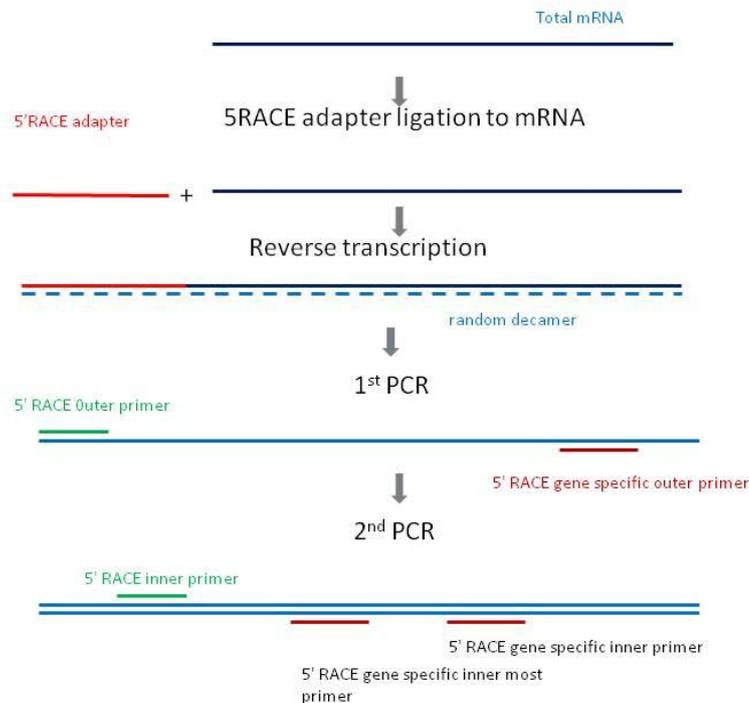


Figure 2.1. Overview of the RLM-RACE procedure

2.3.1. Transmission electron microscopy

A small drop of bacterial suspension (10 μ l) was applied to the surface of a pioloform coated copper grid. After 5 minutes, the suspension was absorbed into a piece of filter paper then, substituted immediately with a drop of 2% methylamine tungstate. The negative stain was then partially removed with another piece of filter paper and the grids were allowed to dry on the bench. The grids were observed using a Philips TEM 208 transmission electron microscope.

2.3.2. Qualitative bioassay

Waxmoth (*Galleria mellonella*) larvae (Livefood UK) were used as primary insects for bioassay. Late instar larvae were injected with 25 µl bacterial suspension, containing approximately 1×10^6 spores using an insulin syringe with small needle (BD). Ten larvae were used for each strain tested. Larval mortality was observed at 24 and 28 hours after injection.

In addition to waxmoth larvae, adult cockroach (*Blaberus discoidalis*), mealworm larvae (*Tenebrio molitor*) and butterworm larvae (*Chilecomadia moorei*) were also tested for their susceptibility to isolate Fang Large. All of the insects were obtained from Livefood UK. The insects were used as soon as they arrived and injected as described above. After the treatment, the insects were kept in falcon tubes (50 ml) and maintained at room temperature on the laboratory bench.

CHAPTER III CHARACTERIZATION OF FANG LARGE ISOLATE

3.1. Introduction

A strain with morphological features of *Lysinibacillus sphaericus* was isolated from a spider (Dr. Nick Waterfield, University of Bath, personal communication). Hence, it is called Fang Large isolate. This isolate was shown to produce mortality when administered to wax worm larvae by injection. To investigate the nature of its pathogenicity, a further characterisation and attempts to understand factors responsible for its pathogenicity were conducted.

3.2. Microscopic Analysis of Fang Large Isolate

A two day old culture of Fang Large isolate grown on LB medium was prepared for transmission electron microscopy. Grids from *L. sphaericus* strain NHA15b were also prepared for a comparison. *L. sphaericus* is a bacilliform bacterium that will produce a terminal or subterminal round shaped endospore. On sporulation, the bacteria will assume the shape of a club. On the completion of sporulation, the cell walls will rupture and the round spore will be easily distinguished from vegetative cells by its round shape. Some isolates will also produce a noticeable crystal protein, namely BinA/BinB, and/or Cry48/49 toxin.

Figure 3.1A shows a Fang Large isolate preparation at different life stages of bacilli. The **a** arrow shows a vegetative stage of bacilliform bacterium. Meanwhile, **b** arrow points to a sporulating cell with a terminally located round shape spore. After the sporulation process is completed, the cell walls will disintegrate and the spores will be freed in the medium and can be seen as round shaped objects (c arrow). An attempt to locate the presence of any crystal as shown in the figure 3.1B (**a** arrow) for

L. sphaericus strain NHA15b in the preparation of Fang Large isolate, failed. The lack of crystal protein in the Fang Large isolate confirmed the result of sucrose gradient ultra centrifugation which also failed to separate any crystal from the spore preparation of Fang Large isolate (data not shown).

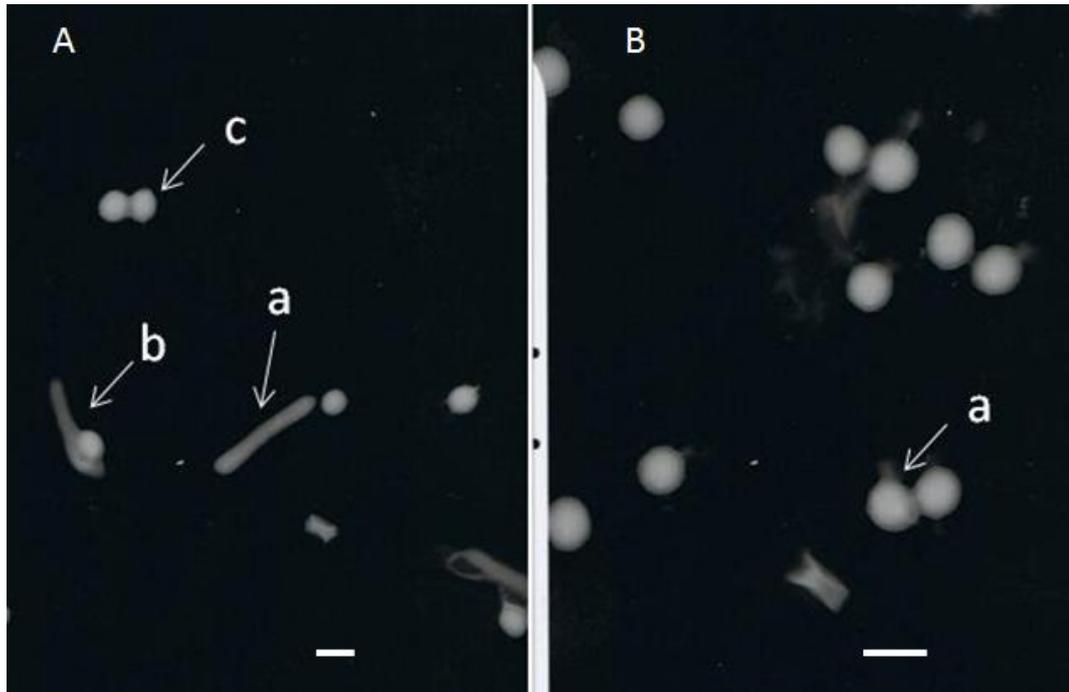


Figure 3.1. Electron micrograph of Fang Large (A) and NHA15b strain (B). Negative staining, bar size is 1 μm . (A) Fang Large isolate showing rod shaped vegetative cell (a), sporulated cell with round terminal spore (b), free spores (c). (B) NHA15b strain showing crystal toxin close to the spore (a).

3.3. 16S rRNA Analysis

A further confirmation of the identity of the Fang Large strain was conducted by analyzing its 16S rRNA DNA sequence. A pair of specific primers to amplify segments of bacterial 16S rRNA DNA which consist of forward primer 63f (5'-CAG GCC TAA CAC ATG CAA GTC-3') and reverse primer 1387r (5'-GGG CGG WGT GTA CAA GGC-3') was used in this study (Marchesi *et al.*, 1998). PCR was run for 30 cycles of 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1.5 minutes, followed by a final extension step of 72°C for 5 minutes. A 1,323 base pair fragment was successfully produced and directly sequenced (Figure 3.2.). The sequence was submitted for BLAST analysis (Blastn; www.ncbi.nlm.nih.gov). The highest five hits were of *Lysinibacillus fusiformis* strain G1-1 16S ribosomal RNA gene, partial sequence (GQ927149.1); *Lysinibacillus sp.* PCSAS2-35 16S ribosomal RNA gene, partial sequence (GQ284494.1); *Lysinibacillus sp.* FBW-3 16S ribosomal RNA gene, partial sequence (GQ287636.1); *Lysinibacillus sp.* R-31191 partial 16S rRNA gene, strain R-31191 (AM910304.1); *Bacillus sp.* CO64 16S ribosomal RNA gene, partial sequence (DQ643066.1) all with the same values for Query Coverage (95%), E Value (0.0) and Maximum Identity (98%). *Lysinibacillus* is a new proposed genus, which contains *L. sphaericus* and a few of other species including *L. fusiformis*, which were characterized by the presence of a Lys-Asp type of peptidoglycan in their cell walls (Ahmed *et al.*, 2007). The results imply that the Fang Large isolate either belongs to *Lysinibacillus (Bacillus) sphaericus* or *Lysinibacillus (Bacillus) fusiformis*.

To assign the Fang Large isolate to species level, the presence of *Bacillus* species-specific signatures within the 16S rRNA fragment obtained from Fang Large isolate was examined. Using MEME analysis, Porwal *et al.* (2009), recognised five

L. sphaericus -specific signatures. Four of these signatures were present in the Fang Large isolate sequence (Figure 3.1.). The fifth signature was also recognised in the sequence (nucleotide 48 to 78) with differences at bases 59 (A instead of G) and 62 (C instead of A). So, instead of the recognised species-specific signature reading AGTAACACGTGGGCAACCTACCTTATAGTT, it read AGTAACACGT-GAGCCACCTACCTTATAGTT. The chromatogram reading showed that for both of the bases, there are secondary peaks matched with the altered bases. If the two bases were considered an artefact, then there is a fifth species-specific signature in the 16S rRNA DNA of the Fang Large isolate. The results of 16S rRNA DNA analysis showed that Fang Large isolate is likely to belong to the species of *L. sphaericus*.

```

>species specific signature 5?
  >base difference
  >base difference
GGACGCAGTAAAGCGCTAAAGCCCAAACAGGTAAGTGCAGCGGGTAGTAACACGTGAGCCACCTACCT < 70
TATAGTTTGGGATAACTCCGGGAAACCGGGGCTAATACCGAATAATCTGTTTCACCTCATGGTGAACAC < 140
TGAAAGACGGTTTCGGCTGTCGCTATAGGATGGGCCCGGGCGCATTAGCTAGTTGGTGAGGTAACGGCT < 210
CACCAAGGCAGCATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCAGA < 280
CTCCTACGGGAGGCAGCAGTAGGAATCTTCCACAATGGGCGAAAGCCTGATGGAGCAACGCCGCGTGAG < 350
  >Species specific signature 1
TGAAAGAAGGATTTTCGGTTCGTAAAACTCTGTTGTAAGGGAAGAACAAGTACAGTAGTAAGTGGCTGTACC < 420
>Species specific signature 2
TTGACGGTACCTTATTAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG < 490
CGTTGTCCGGAATTATTTGGGCGTAAAGCGCGCAGGTGGTTTCTTAAGTCTGATGTGAAAGCCACGGC < 560
  >Species specific signature 3
TCAACCGTGGAGGGTCATTGGAACTGGGAGACTTGAGTGCAGAAGAGGATAGTGGAAATCCAAGTGTAG < 630
CGGTGAAATGCGTAGAGATTTGGAGGAACACCAGTGGCGAAGGCGACTATCTGGTCTGTAAGTACTGACTG < 700
AGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACGATGAGTGCTA < 770
AGTGTTAGGGGTTTCGCCCCCTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCCTGGGGAGTACGGTC < 840
GCAAGACTGAAACTCAAGGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTAATTGCAAGCA < 910
ACGCGAAGAACCTTACCAGGTCTTGACATCCCCTGACCCTGTAGAGATATGGTTTCCCCTTCCGGGGC < 980
AACGGTGACAGGTGGTGCATGGTTGTCGTACCTCGTGTCTGAGATGTTGGGTTAAGTCCCCAACGAG < 1050
CGCAACCCCTGAACCTAATTGCCATCATTTAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAG < 1120
GAAGGTGGGGATGACGTCAAATCATCATGCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGAT < 1190
  >Species specific signature 4
ACAAACGGTTGCCAACTCGCGAGAGGGAGCTAATCCGATAAAGTCGTTCTCAGTTCGGATTGTAGGCTGC < 1260
AACTCGCCTACAGAAGCCGGAATCGGTAGTTTCGCGGCCCTTCTGCCCTCTGCCCCGCTGT < 1323

```

Figure 3.1. 16S rRNA DNA of *L. sphaericus* Fang Large isolates showing the Species Specific Signatures of *L. sphaericus*. Species-specific signature number 5 is shown with 2 base differences highlighted in brown colour.

3.4. Protein Identification

Spore associated mortality caused by *Bacillus/Lysinibacillus* is often associated with major spore proteins. As a result, we examined the profile of spore proteins from the Fang Large isolate.

Fang Large isolate was grown in Embrapa medium until sporulation was complete. The spores were harvested by centrifugation and the pellet was washed. The pelleted spore preparation was then subjected to SDS-PAGE using a tricine running buffer. The proteins separated in the gel were transferred to nitrocellulose membrane. The major protein band observed (approximately 130 kDa) was cut out and sent to Alta Biosciences for N-terminal sequencing. The sequencing result showed that the first 15 N-terminal amino acids were: AQLNDFNIKSGYAKE. The protein sequence was submitted for BLAST analysis (BLASTp, <http://blast.ncbi.nlm.nih.gov>). The ten highest hits were surface layer protein SlpC [*Lysinibacillus sphaericus*] (accession no. ABQ00416.1; ABQ00417.1; ABQ00418.1; ABQ00414.1), putative S-layer protein/N-acetylmuramoyl-L-alanine amidase [*Bacillus sp.* B14905] >gb|EAZ83781.1| putative S-layer protein/N-acetylmuramoyl-L-alanine amidase [*Bacillus sp.* B14905] (accession no. ZP01725715.1), surface layer protein [*Lysinibacillus fusiformis*] (accession no. CAH0324.1; CAH0323.1), putative S-layer protein/N-acetylmuramoyl-L-alanine amidase [*Lysinibacillus sphaericus* C3-41] >sp|P38537.1|SLAP_BACSH RecName: Full=Surface-layer 125 kDa protein; Flags: Precursor >gb|AAA50256.1| surface layer protein [*Lysinibacillus sphaericus*] >gb|ABQ00414.1| surface layer protein SlpC [*Lysinibacillus sphaericus* C3-41] >gb|ACA38653.1| putative S-layer protein/N-acetylmuramoyl-L-alanine amidase [*Lysinibacillus sphaericus* C3-41] (accession no. YP_001696783.1), S-layer protein

[*Lysinibacillus fusiformis* ZC1] >gb|EFI67190.1| S-layer protein [*Lysinibacillus fusiformis* ZC1] (accession no. ZP 07051305.1) and surface layer protein precursor [*Lysinibacillus sphaericus*] (accession no. AAF22978.1). All the above hits indicate that the sequenced protein was a surface layer protein. The surface layer protein of *L. sphaericus* had never been reported to be toxic to any insect, but cell wall-associated toxicity has been reported (Klein *et al.*, 2002; Sgarrella and Szulmajster, 1987) but the protein in their report was a 38 kDa protein. In addition, at least one *B. thuringiensis* isolate produces surface layer protein that has been reported to be toxic for a coleopteran species *Epilachna varivestris* (Pena *et al.*, 2006).

To study the role of surface layer protein of the Fang Large isolate in its pathogenicity against wax worm, an attempt to clone this protein into *E. coli* was conducted. A pair of primers was used based on the surface layer protein C sequence from *L. sphaericus* strain 2362 as designed by Hu *et al.* (2008b). The forward primer is 4 *Bam*HI F (GCAGGATCCGCAAAGCAAACAAAGGCCGTA), and the reverse primer is 3594 *Sac*I R (GTAGAGCTCGGTTTTTCATTTCCACTCCA) (Hu *et al.*, 2008b). Along with the Fang Large isolate, the C3-41 strain was also used as template for surface layer protein gene amplification.

The attempt to amplify surface layer protein gene from Fang Large failed although several modifications to the PCR parameters were made. Since the isolate C3-41 had been amplified easily and isolate C3-41 was also shown to cause mortality when injected into wax worm larvae (result shown below), only the fragment from this isolate was cloned further. The full DNA fragment encoding surface layer protein was double digested with *Bam*HI and *Sac*I and then ligated into the pET28a vector. The

vector was transformed into *E. coli* strain DH5 α and finally into strain BL21(DE3)pLysS to express the protein.

Cells were grown in LB medium containing kanamycin to a D₆₀₀ of 0.6 before induction of protein production with addition of 1 mM IPTG (final concentration). After 4 hours, the cultures were harvested by centrifugation. Analysis of protein expression showed two bands of 130 and 114 kDa which are similar to the 130 kDa expected size for the surface layer protein SlpC and the size of a known product of its partial digestion (Hu *et al.*, 2008b) (Figure 3.3.).

Wax worm larvae injected with a suspension of strain BL21(DE3)pLysS containing the cloned gene induced to express the surface layer protein did not show any mortality when administered at a dosage of 25 μ l of bacterial suspension at D₆₀₀ = 1.0. This indicates that the surface layer protein of isolate C3-41 was not the protein responsible for the pathogenicity of *L. sphaericus* to wax worm larvae.

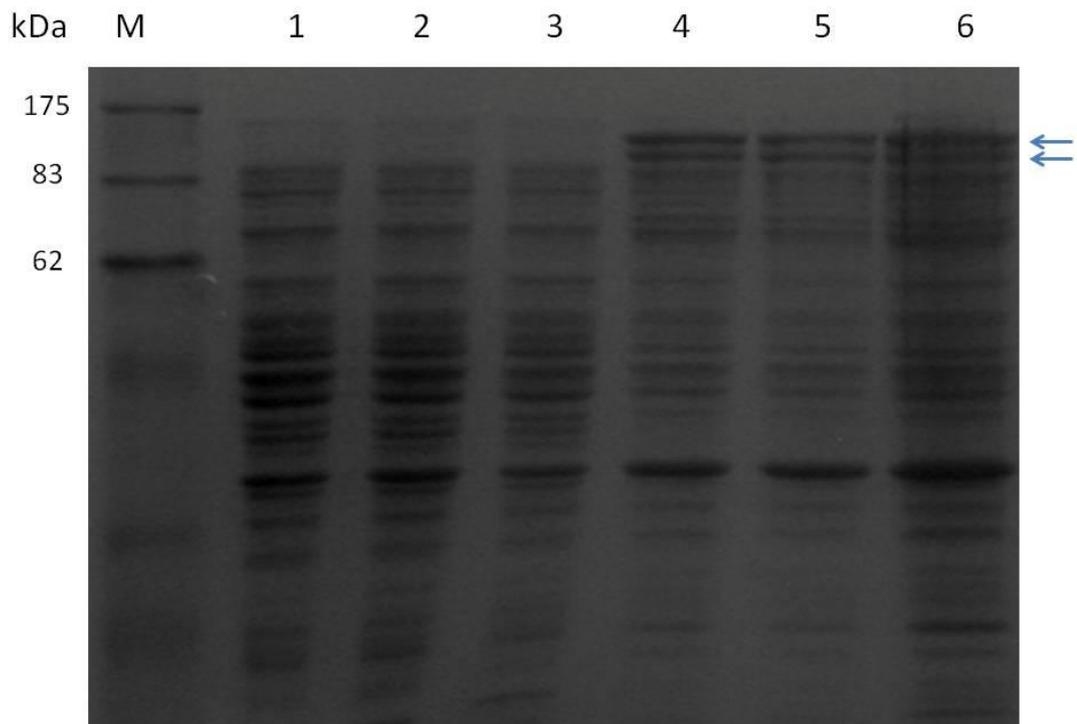


Figure 3.3. Induction of recombinant surface layer protein (SLP).

Lane M; broad range protein marker, lane 1,2,3 non-induced *E. coli* BL21(DE3)pLysS from clone a, b, c respectively, and lane 4,5,6 the same *E. coli* BL21(DE3)pLysS clones induced with IPTG and harvested after 5 hours incubation. Arrows shows the expressed SLP protein.

3.5. Qualitative Bioassay

To obtain information on the pathogenic characteristics of the Fang Large isolate, several treatments were conducted using bacterial suspensions injected into wax worm larvae. At first, the bacterial suspensions were prepared from a one day-old vegetative culture and also from a three day-old sporulated culture. Second, the spore suspensions were heated for 5, 10, and 15 minutes at 95°C. Third, the spore suspensions were subjected into UV light for 10, 20, 30 minutes. Fourth, the spore suspensions were treated with Proteinase K (0.05 mg/ml) with incubation at 37°C for 30 minutes.

Suspensions of Fang Large isolate were able to kill wax worm larvae either in vegetative or sporulated form. This killing capacity was not affected by enzyme treatment. The ability was degraded when the suspension was subjected to heating. The pathogenicity was reduced to 40% when the suspension was heated to 95°C for 5 minutes, and totally lost when heated for 10 minutes. The pathogenicity was also lost when the suspension was exposed to UV light for at least 10 minutes. Plate cultures of the same suspension showed that after the suspension was heated to 95°C for 5 minutes, the colony number observed was low, and there were no colonies growing on the plates when the suspensions were subjected to heat or UV light for 10 minutes. All of the above observations indicate that the Fang Large pathogenicity is correlated with its viability.

To assess whether the injection-pathogenicity was restricted to Fang Large, a survey, limited by the availability of the larvae, was conducted on several *Bacillus*, *Brevibacillus* and *Lysinibacillus sphaericus* strains. The strains selected included representatives of Group IIA *L. sphaericus* strains from a range of serotypes (H3: LP1-

G, LP148, 47-6B; H5a,5b: 1593,2362,C3-41; H6: IAB59; H25: 2297; H48:PR1); type strains of other *L. sphaericus* groups (Groups I: ATCC14577; Group IIB: 7055; Group III: NRS592; Group IV: NRS400; Group V: NRS1198); two *Brevibacillus laterosporus* strains (NRS590 and NRS1642); and a *B. thuringiensis kurstaki* strain HD73 that encodes a single insecticidal protein Cry1Ac (Hofte and Whiteley, 1989). The bacteria were grown in Embrapa medium for 3 days (30°C, 250 rpm). The cultures were pelleted, washed, suspended in ¼ strength Ringer's solution and counted using plate counts. Each wax worm larva was injected with 25 µl of bacterial suspension. Control larvae were injected with ¼ strength Ringer's solution alone. Larval mortality was counted on the first and second day after injection. Since the Fang Large isolate produces 100% mortality at 10x spore suspension dilution and shows no observed mortality at 1/10x spore suspension dilution, these two dilutions were used to test bacterial pathogenicity. *Brevibacillus laterosporus* NRS590 and *B. thuringiensis* HD 73 produced mortality at 1/10x concentration, therefore further serial dilutions (1/1,000x, and 1/100,000x) were also used for both strains. Spore concentrations of the various dilutions are shown in Table 3.1.

All 14 bacterial strains tested could produce 100% mortality at the highest concentration, but there was no mortality observed when the concentration was lowered 1/100 times, except for *B. laterosporus* NRS590 and *B. thuringiensis* HD 73. Meanwhile, there was no mortality observed on control treatment (Table 3.1.).

B. laterosporus strain NRS590 is the most pathogenic strain tested, which was able to produce 40% mortality at 1/1,000 dilution, followed by *B. thuringiensis* HD 73 which was able to kill 20% of the wax worm larvae at 1/10x spore suspension dilution. No mortality was observed on a lower dilution. Considering the high spore

concentration of the Fang Large isolate needed to produce larval mortality it is more likely that the Fang Large is causing septicaemia rather than toxaemia. The *B. laterosporus* NRS590 strain is more than 1,000 times more pathogenic than Fang Large and was more pathogenic than HD73 even though it is known that HD73 produces Cry1Ac a toxin active against Lepidopteron insects. It is possible that solubilisation or activation of toxin strain HD73 in this experiment was not optimal and this likely as Bt Cry toxins normally require high pH for solubilisation and gut proteinases for their activation (Palma et al., 2014).

While it seems that the Fang Large strain exerts its pathogenicity through septicaemia, it is still not known whether *B. laterosporus* NRS590 killed the wax moth larvae by septicaemia or a toxic mechanism. This subject warrants further study that has not been covered in this thesis.

Table 3.1. Mortality of waxmoth *Galleria mellonella* larvae injected with 25 µl of spore suspension in different suspension dilution (%).

STRAIN	Spore counts (spore/ml)	% Mortality				
		Spore Suspension Dilution				
		10 ¹	10 ⁰	10 ⁻¹	10 ⁻³	10 ⁻⁵
Ls 1593	9.67 x10 ⁷	100	Nt	0	Nt	Nt
Ls 2297	9.67 x10 ⁷	100	Nt	0	Nt	Nt
Ls 2362	6.33 x10 ⁷	95	Nt	0	Nt	Nt
Ls 47-6B	7.33 x10 ⁷	100	Nt	0	Nt	Nt
Ls ATTC 14577	1.60 x10 ⁸	100	Nt	0	Nt	Nt
Bl NRS1642	6.33 x10 ⁷	95	Nt	0	Nt	Nt
Bl NRS590	8.00 x10 ⁷	100	Nt	100	40	0
Ls C3-41	2.67 x10 ⁷	100	Nt	0	Nt	Nt
Ls Fang Large	4.33 x10 ⁷	100	25	0	Nt	Nt
Bt HD73	2.33 x10 ⁷	100	Nt	20	Nt	Nt
Ls LP148	1.97 x10 ⁸	95	Nt	0	Nt	Nt
Ls LP1G	2.00 x10 ⁷	95	Nt	0	Nt	Nt
Ls NRS400	5.00 x10 ⁷	100	Nt	0	Nt	Nt
Ls NRS592	1.67 x10 ⁷	100	Nt	0	Nt	Nt
Ls PR1	1.33 x10 ⁸	95	Nt	0	Nt	Nt
Induced <i>E. coli</i> BL21(DE3)pLysS containing SLP insert	2.99 x10 ⁸	Nt	0	Nt	Nt	Nt
Control	0		0			Nt

Note: Nt = not tested ; Ls= *L. sphaericus*; Bl= *B. laterosporus*; Bt= *B. thuringiensis*

CHAPTER IV
DETERMINATION OF LOCATION OF *CRY 48/49* GENES
IN *LYSINIBACILLUS SPHAERICUS* STRAINS

4.1. Introduction

Genes for *cry 48/49* are found in a variety of *L. sphaericus* strains of different serotypes and with different profiles of other known toxins as shown in Table 4.1 (Berry and Jones, unpublished results). Since their presence seems not to be correlated with serotype or the presence of other toxin proteins, it is possible that the *cry 48/49* genes are located on a mobile element. To investigate this possibility, the plasmid profiles of strains were analysed and probed for the presence of the genes.

Table 4.1. *Lysinibacillus sphaericus* strains used in the experiment to determine the location of *cry 48/49* genes and their toxin profiles.

Where profiles are unknown, table entries are left blank.

No.	Strain	DNA Group	Serotype	Phage group	Cry 48/49	Mtx1	Mtx2	Mtx3	Bin
1.	IAB 59	IIA	H5a,5b	3	+	+	+	+	+
2.	NHA15b				+				-
3.	LP1-G		H3		+	-			+
4.	LP18				+				+
5.	LP148		H3		+	-			+
6.	LB29		H26a,26b		+				
7.	47-6B		H3		+				+
8.	2173	IIA	H26a,26b		+	-	-	-	-
9.	C3-41	IIA	H5a,5b		-	+ pseudo	+	+	+

4.2. Plasmid separation Using Pulsed Field Gel Electrophoresis

Nine isolates of *L. sphaericus* namely C3-41, IAB 59, NHA15b, LP1-G, LP18, LP148 LB29, 47-6B and 2173 were used to determine the location of *cry48/cry 49* genes. Based on previous work (Berry and Jones, unpublished), all these strains except C3-41, were known to carry *cry48/cry 49* genes. Bacteria were grown on Spizizen medium (section 2.2.16.) but since *L. sphaericus* cannot utilize sugars, the glucose component was omitted from the composition.

The plasmid extraction procedure gave reproducible results, showing a consistent set of plasmid bands for each strain in repeated preparations (section 2.2.16). Of the nine isolates of *L. sphaericus* checked for the location of the *cry48/cry 49* genes, 6 of them consistently showed 1 to 3 bands above the chromosomal band, meanwhile strains *Cry48/49* producing strains 2173 and LB29 and the non-producing control C3-41, did not show any of these bands (Fig. 4.1).

It is unclear why the preparation procedure could not produce any plasmid band from strain C3-41, since it is known that strain C3-41 has at least one megaplasmid with a size of 177,642 bp (Hu *et al.*, 2008a). It is possible that the plasmid is co-migrating with the chromosomal band. Isolate IAB59, LP18, LP148 and LP1-G only show one plasmid band that co-migrated with a 750 kb fragment of the 50 kb lambda DNA ladder used as a marker, with a slightly bigger size for the IAB59 band. There were two plasmid bands seen in isolate 47-6B, meanwhile in isolate NHA15b, there were three discernible plasmid bands. In addition, strain 47-6B was also shown to have a noticeable plasmid below the chromosomal band.

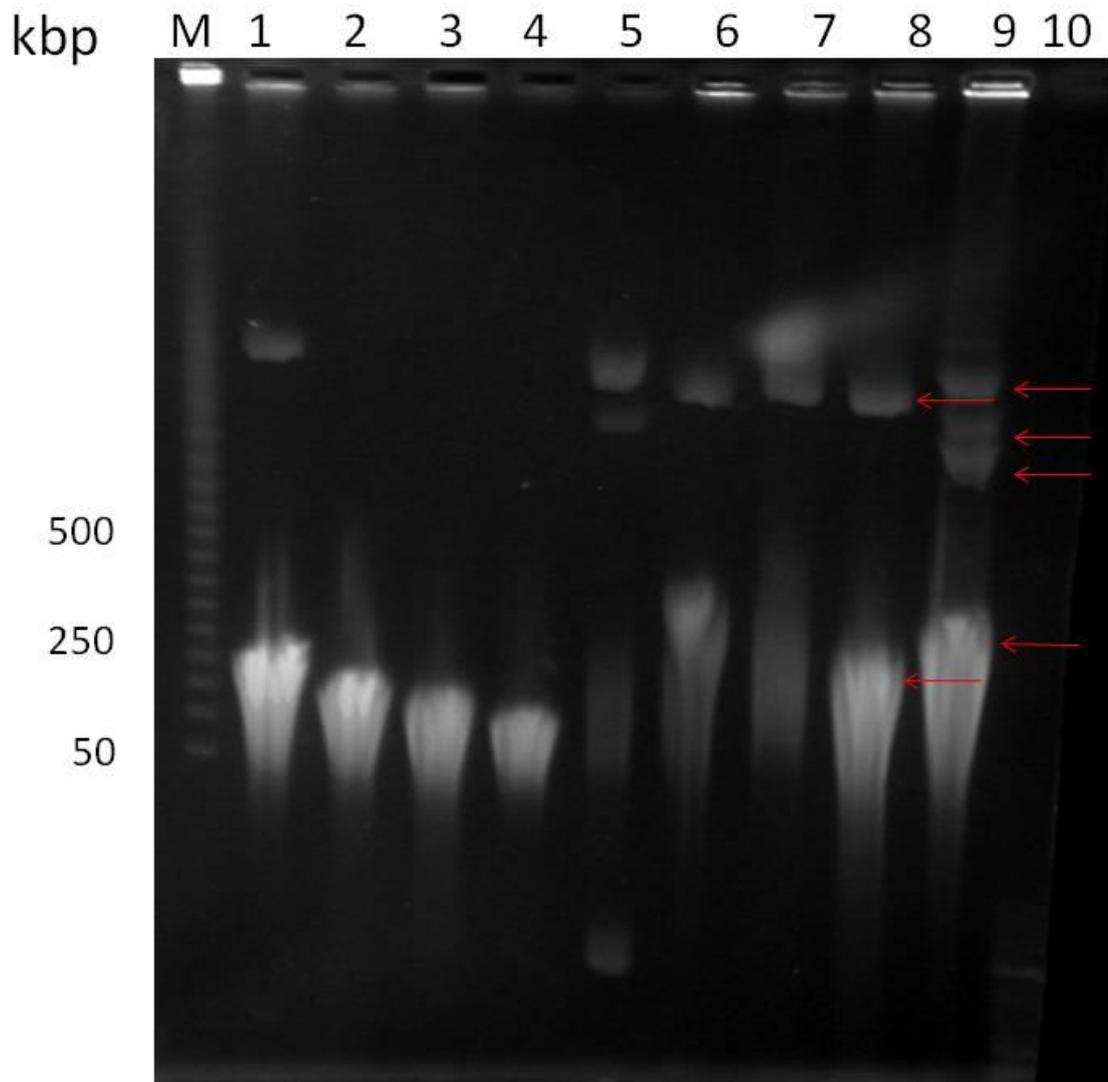


Figure 4.1. Plasmid patterns from *L. sphaericus* strains. M, 50 kb lambda DNA ladder; 1, IAB59; 2, C3-41; 3, 2173; 4, LB29; 5, 47-6B; 6, LP18; 7, LP148; 8, LP1-G; 9, NHA15B; 10, C3-41 total DNA digested with *Hind*III restriction enzyme. Red arrows point to bands extracted for DNA gel purification and PCR testing for *cry*48/49 genes.

4.3. Determination of Location of *Cry 48/49* Genes

Southern blotting and in gel hybridization techniques were employed to determine the location of *cry48/cry49* genes in the plasmid preparation that had been separated using PFGE techniques. Probes for *cry48/49* genes were PCR products, which incorporated DIG-11-dUTP instead of dTTP as substrate for Taq DNA polymerase. The primers specific for each probe are listed in the appendix. The hybridisation probes were prepared according to the manufacturer's protocol (Roche Diagnostics) (method 2.2.19). Unfortunately, despite repeated attempts, no hybridising bands were obtained from these procedures even in the genomic DNA. This appears to indicate that the probing process did not work properly.

DNA was gel purified from NHA15b and IAB59 plasmid and chromosomal bands using Wizard® Plus SV Minipreps kit (Promega), and PCR was used to amplify *cry48*, *cry49* and *bin* genes using these templates. The result showed that none of the DNA bands used as PCR templates produced positive results (data not shown).

To verify the presence of the genes in the original plasmid preparations, which consist of plasmid and chromosomal DNA, samples of the whole preparations were used in dot-blot hybridisation (method 2.2.20). Figure 4.2 shows that the *bin* probe could detect the presence of *bin* genes in the total DNA preparation of the strains C3-41, and LP1-G. The lack of hybridisation to NHA15b is expected as this strain is known not to have a *bin* gene.

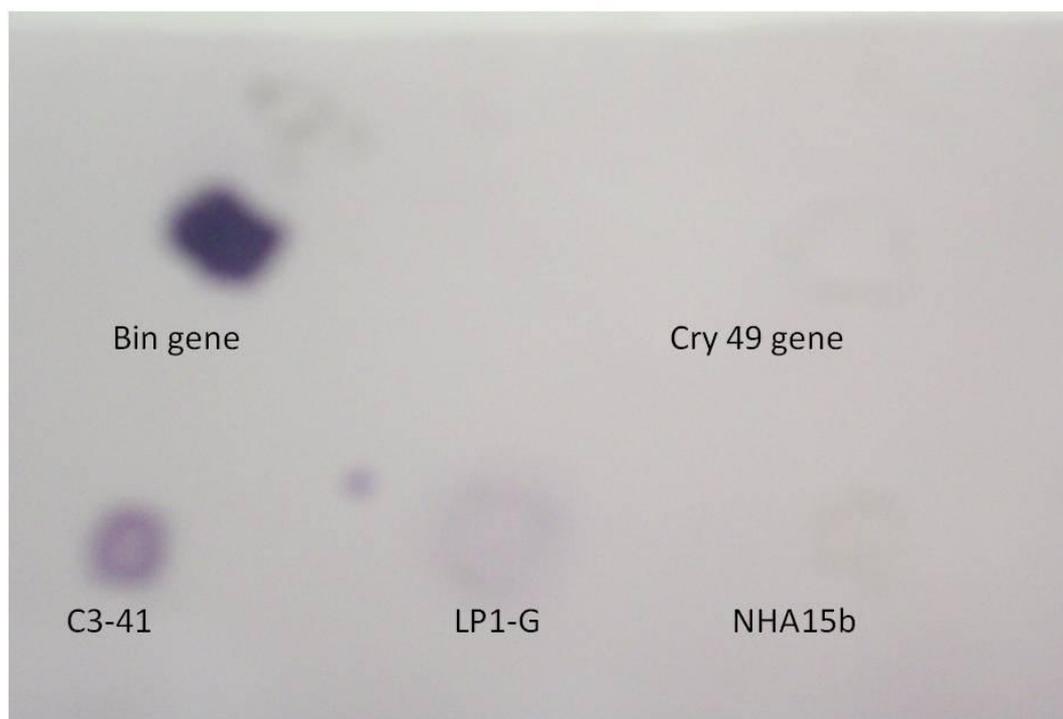


Figure 4.2. Dot blot using *bin* probe.

The *bin* and *cry49* genes on the blot were PCR products of genes of strain IAB59 amplified with the same primers as the primers for probe preparation. C3-41, LP1-G and NHA15b were total DNA obtained using plasmid preparation procedures.

In an attempt to prepare DNA samples that would reflect the occurrence of the known plasmid in strain C3-41 and *cry* 48/49 genes, a new modified, plasmid preparation was conducted with 90 minutes of incubation with lysozyme instead of 3 hours. The result is shown in Figure 4.3. There were no plasmid bands visible in the strain C3-41, but there were two bands instead of one in some strains, including strain IAB59. Both bands of strain IAB59 were cut from the agarose gel and extracted using a gel-electroelution method. The gel fragments were cut out, and electrophoresed for 30 minutes at 100 V. The eluent was cleaned with phenol:chloroform:IAA, and chloroform:IAA extraction steps and then ethanol precipitated. The DNA obtained was used as template for PCR reactions using *cry49* and *bin* toxin gene primers. Pipette tip-stabbed gel plugs were also used as templates. The plasmid preparation of strain IAB59 was used as a positive control. The result is shown on Figure 4.4. Both of the extracted bands produced positive results for the presence of the *cry49* gene but were negative for the presence of *bin* toxin genes, although the first band produced a much fainter amplicon, but pipette tip-stabbed gel plugs did not produce products. These results show that it is most likely that the *cry48/49* genes of strain IAB59 are on a plasmid with a size around 750 kb based on comparison to 50 kb DNA ladder and distinct from the location of the *bin* gene. It should be noted that the DNA ladder is in linear form, while the samples were circular plasmid DNA preparations. The mobility of linear DNA is faster that of circular DNA.

It is possible that the two bands in the last preparation of strain IAB59 represent one plasmid with different mobility due to its coiling state, and identical with the single plasmid observed in previous plasmid preparation. Alternatively, the second band may represent the plasmid containing the *cry49* gene while contamination of the

first band may explain the fainter band observed for band 1. It is clear from the above results that even minor variations in plasmid preparation methods can produce significant differences in plasmid patterns observed. This may explain the unexpected results with respect to strain C3-41 and also requires some caution in the interpretation of the results in this section. While no single, conserved plasmid band could be identified in all Cry48/49 expressing strains as a common factor that might carry the genes, however, some indications of localisation were obtained. This issue will be further explored with the examination of the genome sequence of strain NHA15B in a subsequent chapter.

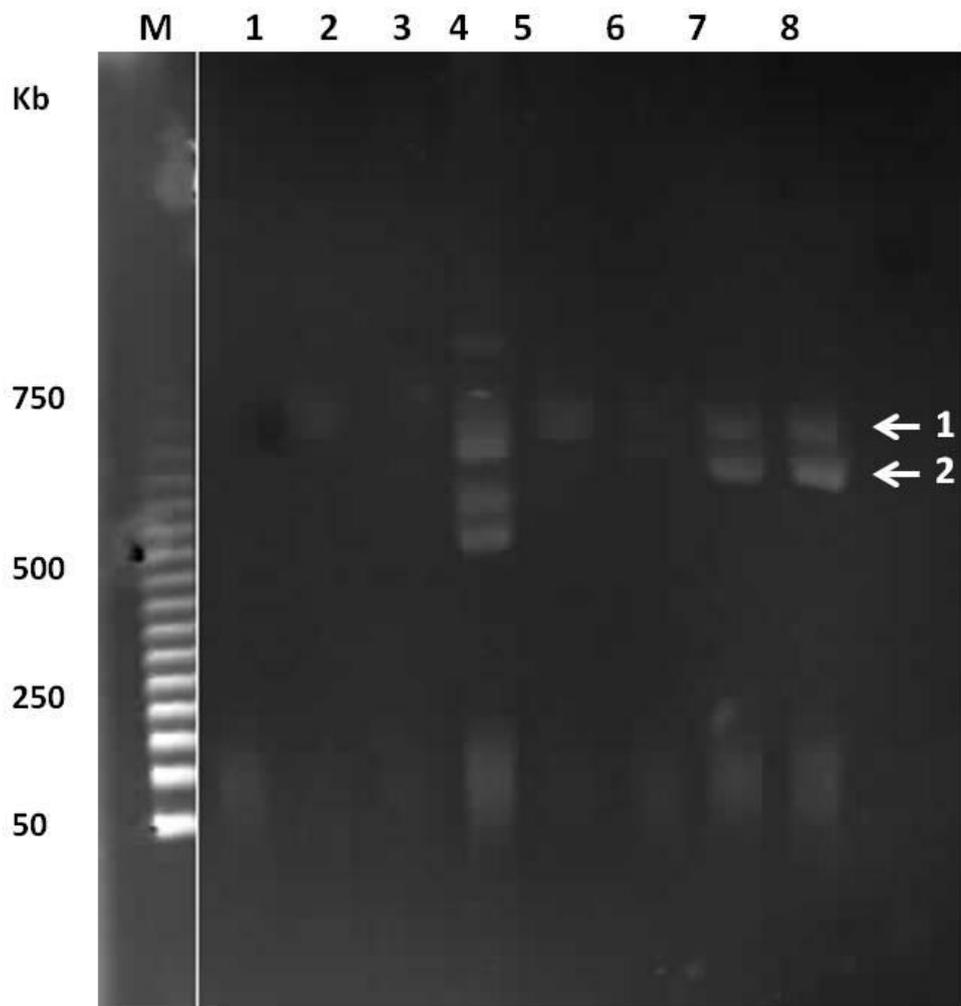


Figure 4.3. Plasmid patterns from *L. sphaericus* strains. M, 50 kb lambda DNA ladder; 1, NHA15B; 2, 2173; 3, C3-41; 4, LP1-G; 5, LP148; 6, LP18; 7, IAB59; 8 IAB59. The arrows point to bands cut for DNA gel purification using gel electroelution method.



Figure 4.4. PCR product of *cry49* and *bin* genes from plasmid bands of strain IAB59. M, 100bp DNA ladder marker; 1, 1st band gel electroelution; 2, 2nd band gel electroelution; 3, pipette tip stabbed from 1st band; 4, pipette tip stabbed from 2nd band; 5, total plasmid preparation DNA, , 6, 1st band gel electroelution; 7, 2nd band gel electroelution; 8, pipette tip stabbed from 1st band; 9, pipette tip stabbed from 2nd band; 10, total plasmid preparation DNA. 1-5 amplified with *cry49* primers and 6-10 amplified with *bin* primers.

CHAPTER V
5' UNTRANSLATED REGION ANALYSIS
OF *LYSINIBACILLUS SPHAERICUS* TOXIN GENES

5.1. Introduction

The expression level of a particular gene depends on its regulatory factors. The 5' untranslated region of (5'UTR) upstream of the coding region contains information important to regulate the gene such as the transcription start and ribosome binding site (RBS). When the transcription start of a particular gene can be determined, then we may be able to predict the promoter sequence more accurately as this should be located upstream from the transcription start. Some prediction of such regulatory factors for the toxin genes of *L. sphaericus* had been done previously. For example, it is predicted that in the IAB59 strain, there are sigma-E like sequences as the putative promoter upstream of *cry48Aa1* and *cry49Aa1* genes (Jones and Berry, unpublished result). The sequences TTTAAT and TTGATT found upstream of the *mtx2* gene are almost identical to the vegetative promoter of RNA polymerase of *B. subtilis*, and were proposed as the putative promoter (Thanabalu and Porter, 1996). However, the exact nature of these regulatory sequences has not yet been determined. Since there are whole genome sequences of *L. sphaericus* available in the database, e.g. *L. sphaericus* isolate C3-41, in this experiment the C3-41 sequence was used to predict the toxin gene promoters once the transcription start was known.

In these experiments an RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) was employed to obtain sequences of the 5' untranslated region of *binB*, *cry48Aa1/49Aa1*, *mtx1*, *mtx2*, *mtx3*, and the *sphaericolysin* gene of *L. sphaericus* IAB59. Total RNA was obtained initially from 2 day old cultures using the hot phenol method. The RNA Easy® kit (Qiagen) was employed to clean the RNA before it was

subjected to DNA digestion using DNase. As Total RNA extraction using the hot phenol method was not successful in obtaining a good template for the 5' RLM-RACE experiment, a second Total RNA extraction using the RNeasy Protect Bacteria Mini Kit® (Qiagen) was employed.

The FirstChoice RLM-RACE kit (Ambion) was employed to obtain the 5'UTR of genes of interest according to the manufacturer's protocol with the omission of Calf Intestine Alkaline Protease and Tobacco Acid Pyrophosphatase treatments. To the RNA sample that had been treated with DNase, a 45 base RNA adapter oligonucleotide was ligated over night at 4°C using T4 RNA ligase. The ligated RNA was then used as the template for producing cDNA using a random-primed reverse transcription reaction. To amplify the targeted genes, a nested PCR reaction was conducted using the cDNA as shown schematically in figure 5.1. The first round of PCR reaction was conducted using a pair of primers corresponding to RNA adapter (forward), designated as outer RACE primer and specific toxin gene sequences designed based on published sequences (reverse) designated as outer gene specific primer/TS4. A pair of primers corresponding to the RNA adapter downstream of the outer RACE primer (designated as inner RACE primer) and specific toxin gene sequences upstream of the outer gene specific primer (designated as inner gene specific primer/TS5) was employed for the second round of the nested PCR reaction. With some genes, a third reverse primer located upstream of the inner gene specific primer, designated as the innermost gene specific primer/TS6, produced a sharper band than the PCR product using the inner gene specific primer. Negative controls were set up without reverse transcription to ensure that there was no contamination with bacterial DNA. The PCR products were cut from the agarose gels, purified, cloned

into pGEM-T Vector System and sequenced. The sequences were submitted to BLAST analysis (<http://blast.ncbi.nlm.nih.gov>). If the BLAST analysis showed that the sequence obtained was related to the targeted toxin genes, the sequence was aligned to the toxin sequences of strain C3-41 along with a couple hundred nucleotides upstream of the gene. The detail of the experiment is outlined in section 2.2.20. The primers used in 5'RLM-RACE experiments are listed in the appendix.

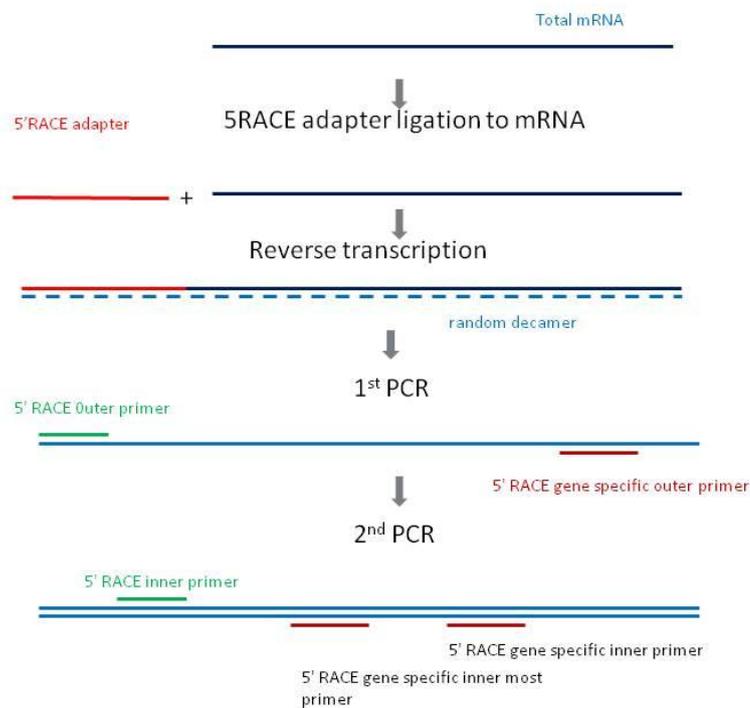


Figure 5.1. 5' RLM-RACE procedure

5.2. 5'UTR Analysis of the *cry48Aa1* Toxin Gene

RLM-RACE of strain IAB59 produced a 466 bp fragment. The first five hits of BLAST analysis results showed that this fragment was identical to part of the *cry48Aa1* toxin gene *L. sphaericus* strain IAB59, NHA15b, 47-6B, 2173, LP1G (accession number AJ841948.3, AM237206.1, AM237205.1, AM237208.1, AM237207.1 respectively). When the *cry48* RLM-RACE product was aligned to the *cry48* sequence of reference strain IAB59, it was apparent that the product is truncated. The sequence begins 7 nucleotides downstream of the initiation codon. It is likely that the poor result was produced due to degradation of the RNA sample.

A second attempt using total RNA extracted using RNeasy Protect Bacteria Mini Kit® (Qiagen), produced a longer product. To make sure that the amplified product of 5'RLM-RACE is the targeted gene, a third round of PCR was conducted using an innermost primer. The 267 bp PCR product was cloned into pGEM-T vector and sequenced. The results showed that the transcript extended 68 nucleotides upstream of the initiation codon -further upstream than expected (Figure 5.2). As it had been predicted that gene might be controlled by a putative sigma-E promoter beginning at 33 nucleotides upstream the initiation codon (-10 sequence) and 54 nucleotides upstream the initiation codon (-35 sequence), but the result showed that this cannot be the promoter, since this putative promoter is located downstream from the transcript start. A further analysis of the region upstream of the transcript was carried out by searching the promoter database at DBTBS (<http://dbtbs.hgc.jp>) but although putative promoters were found, none were close to the transcript start. It is possible that the actual promoter for the gene is different from any recognised promoter in the database.

```

                                >SigF
tcaaatcgcgcatctaccagctgtaaaggaaaaattgaacgtgttaaagaagctgtagaagatacgcaa < 70
gaacatctaaattattcaacagacaccgatgcacgagtaggacagaaaacagcagacagttctttctttg < 140
gttataaaacacatgtatcaattagtgatgaacgtataattacagcagcagtcggttacaacaggtgaaaa < 210
atcagatggacattatctacaagatctgattaaaaaaagtaaagtaacaggaatagaagttgaaacagtc < 280
atcgctgacacggcttattcagggaaaaacaatttactttatacaaaaagaacaaagcattcaactgattt < 350
ctaagttaaatccaatgattactcaaaatgggtgaggaacggaaaaggatttgactttaataaagggtgc < 420
aggtttatttgtatgtccagcaggacatttagcaaatagaaaaagtaaaaaaccagattcacaccagga < 490
gcaagtgcagaatggtttatTTTTTgacattgaaaaatgtaaggctgtccactccgagaaggacaag < 560
aaaccggttcatttctgccaagataattgctctgaagaagataacgaactagctggattacacaaaaga < 630
aaaccaacgtttactgatggagaatgatatttttaagcaagcagcgcgtgatatgggacgaaaaatagatgt < 700
gattcgaaataacgctcacaaatattcagtatcagaaatgtgtgacgtccttgatattcgaaaagtacgt < 770
                                >SigA
attatattcatgccgatgtttggcgaacgtgcgctaaaaactgaagataaaagagatttccaagagat < 840
tgctcatattttcaagaaagccgtaacaattatggtagcgcgtaaaatcaagaagaactaagtaaatta < 910
                                >SigX
ccagagccaaaaatgtatcacgctcgtcgaattggctgtaatagaaaggctctggactagtatcaaaact < 980
                                >SigD                                >SigD
atacagtagcacagtataagctgcataactcaacatgtaatagaagcgcgataaaaaatgaattacaggg < 1050
                                >SigL
ccaatttgagcaagtagagccgttggtgtcatcgtgattaacatatgtccgtgtaaattggcattatgta < 1120
tgcattttgtcgacctttcaatcgtgaaatcattggccgtagtgctggtgcaaataagacagcagagc < 1190
tggtttatcaagcactagtaagtagaggaacttgaaacgacatacaaatgtaacatacagaccgagg < 1260
aaaagggtttgataatacagatcatttcaaaagccctagaaacatttggaatccagcaatccctgagcatg < 1330
aagggatgcccttatgataacgcccgtggctgaggcgtgtttaaggattttccaagcaacagaccttata < 1400
atTTTTgtccagtttagtggtgacatatcagttcttcttccgaacaattgaaggatatgtaattttaacc < 1470
gattggatgtttacttcgccccaaaaaatgattcccttgataaacagattatTTTTTTaaagctgtt < 1540
                                >Transcript product 2nd
                                >SigH                                >-35 sigE
aggagcatalcactctcatgaatatttagctTTTTTTattaatgtaggaattaagaaaaaacctagta < 1610
                                >SigL                                >-10 SigE                                >Transcript product 1st
tattgttatttcacctgcataatattatatttaggaaaaaggaggactgtgtatggatcaacaataataa < 1680

                                >Cry48Aa1
tgaaaaggaaattattaattctcatttactacat < 1715

```

Figure 5.2. Part of *cry48* gene and its upstream sequence.

Transcript start from the second experiment overlaps with the predicted sigmaH promoter sequence. The transcript start is located upstream to the putative sigmaE promoter. Red = putative promoter sequence; blue = transcript start from 5'RLM-RACE product; green = part of gene coding sequence; underlined= initiation codon.

5.3. 5'UTR Analysis of the *cry49Aa1* Toxin Gene

A first attempt using total RNA from the hot phenol extraction method and *cry49* inner primer produced a 962 bp fragment but sequencing showed that the product contained the RACE primer at both ends. The remaining sequence was submitted for BLAST analysis, the result showed that this fragment was not a *cry49* gene sequence and instead, it was a 16S rRNA DNA. The first five hits referring to the 16S RNA DNA of *Lysinibacillus sphaericus* C3-41, complete genome (Accession CP000817.1), *L. sphaericus* strain PRE16, strain Lma3, strain S8, strain S2 16S ribosomal RNA gene, partial sequence respectively (accession EU880531.1, GU204967.1, GQ889259.1, GQ889257.1). Further analysis did not show any sequence similarities between the specific primers employed in this experiment and the 16S rDNA of *L. sphaericus*. This is discussed further in section 5.4 below.

An RNA preparation using the RNeasy Protect Bacteria kit was used in the second attempt to produce 5'RLM-RACE of the *cry49* gene. The second PCR round was conducted using an innermost primer. The 149 bp product was cloned using pGEM-T vector and sequenced. The BLAST analysis showed that the 5'RLM-RACE product is a truncated *cry49* sequence identical to the *cry49* sequence of IAB59 strain in the database (accession number AJ841948.3) (Figure 5.3). The transcript obtained started 29 nucleotides downstream of the initiation codon. Further attempts failed to identify larger transcripts from this gene.

5.4. 5'UTR Analysis of the *mtx1* Gene

A very similar result to the attempt to obtain 5' UTR product of the *cry49* gene was also found in the experiment to obtain the 5'UTR of the *mtx1* gene. Total RNA obtained using RNeasy Protect Bacteria mini kit was used and produced a 531 nucleotide fragment using the inner primer. When submitted for BLAST analysis, the sequence was recognised as a fragment of 16S rDNA. There were no similarities observed between the primer of *mtx1* and 16S rDNA of *Lysinibacillus sphaericus*.

The RNeasy kit is supposed to clean the whole total RNA free from ribosomal RNA but, as two results had showed, there was enough ribosomal RNA to interfere with RLM-RACE procedure. Both of the RNA extraction methods face the same problem.

Another attempt to produce 5'RLM-RACE of *mtx1* gene product was done with the second round of PCR reaction using an innermost primer. The PCR product was subsequently cloned and sequenced. BLAST analysis of the sequence showed that it was a truncated version of the *mtx1* gene (Figure 5.4).

The transcript obtained was 6 nucleotides downstream of the initiation codon. The result suggests that the innermost primer is more useful to ensure production of an amplicon for the correct gene than the inner primer but to date the transcript obtained is still in the truncated form.

```

                                >SigB
atagaagccactactggaaagatccgagggtttgaggcgttagtgcgctggaaaagccctgagcttggac < 70
                                >SigH
ttgtttctccagaagatTTTTATTCCGATAGCAGAAAAAACAGGETTGATTACACAAATTGACGAATGGGT < 140
catgtatcaagcatgcttgaaaaatgtcgaattacaacatcaattcggttatcccttctctgatgtcggtt < 210
aatatttcggctttacaattaggaagagcagatTTTGTGGATAAAGTCCAGCTTATTTGAATGAAACTA < 280
agatgaagccggaacatttagaaaattgagattactgagagtatcttagttgagtcatttgaaagctctat < 350
atgcatactgcgaaaattaagaacctaggggtgaaaattgccaggatgattttggtaccggttattct < 420
tcacttaactacttaacattattgccgattcatacccttaaaattgatagatcccttattcaaaatga < 490
catcggcaacagcagaaaaaacgataatcgaatcgattatccaccttgccataaattaggtcatgatgt < 560
agttgcagaaggagtggagacaaaagaacagtatatcttgctcaaagaatggaactgcgattttgtccag < 630
ggatactatttcagtaggcctgtttcgtcagacatacttgcgaatgtctgcaggggaaaagaactcca < 700
caaatcagtaaaaaataaagctctgtaattacaaaaacaacaacctgccaatttagattataaattat < 770
>SigG >Thanabalu et al.-35 >Thanabalu et al.-10
tgtagaaattcactatcatcccttcaaaataaataataaacattatttttatttaattatttaa < 840
                                >Mtx1>Transcript product
atacataatttaataacattaaataaataaataacaaaaagaggtgcaattgatatggctataaaaaaag < 910

```

Figure 5.4. Part of *mtx1* gene and its upstream sequence. The truncated transcript start is 6 nucleotides downstream the initiation codon. Red = putative promoter sequences; red and underlined = promoter predicted by Thanabalu *et al.*(1991); blue = transcript start from 5'RLM-RACE product; green = part of gene coding sequence; green and underlined= initiation codon.

5.5. 5'UTR Analysis of the *mtx2* Toxin Gene

A 235 nucleotide sequence was obtained when a 5'RLM-RACE using an innermost *mtx2* primer was conducted. BLAST analysis showed that the sequence is identical to a length of the *mtx2* gene of *L. sphaericus* strain SSII-1, including some nucleotides upstream of the initiation codon. The transcript started 49 nucleotides upstream of the initiation codon (Figure 5.5).

The sequences 94 nucleotides upstream of the initiation codon (TTTAAT) and 67 nucleotides upstream of the initiation codon (TTGAAT) were proposed as a possible promoter (Thanabalu and Porter, 1996). The result of the experiment is consistent with the prediction, since it is the closest promoter predicted upstream of the transcript and the distance between the promoter sequence and the beginning of the transcript is 18 nucleotides.

```

                                >SigB
aaaattcgtgtggttagcagactcatctgggggaattaagcaaagtatttatagtaaagaatactgttaa < 70
aacaattctgtcacggagaattgttttcaatatgaaatcagattttttataaaagagtttgaaagtgttga < 140
caaaattaaaacggaattaggaataatttactttattacaatacaaaacaaaattaaggcaaaataaaag < 210
gcatgagcctgatacttaccatatcgaactccttgcccaaacactgactaatgaaataaccgtgtctaac < 280
                                >SigA
ttttaggggtcatttcttttatttaattctataagcttagaaaatttttgaagtgattttaacactgtata < 350
                                >SigB+>SigA (overlapped)                                >Thanabalu and Porter -35
attcttgatactagtatatttaataagatttggaattcagtttaattattctaaaggagactaaattg < 420
                                >Thanabalu and Porter-10>transcript product
aattttcggtttccagttatcgtttaattagcgcgccatttatcattatcaaagcatttaaaaaataa < 490
                                >Mtx2
gatattggagagtgaaaaataatgaaaagga < 521

```

Figure 5.5. Part of *mtx2* gene and its upstream sequence.

The transcript product started 49 nucleotides upstream of the initiation codon. The promoter predicted by Thanabalu and Porter (1996) is located just upstream of the transcript. Red = putative promoter sequences; blue = transcript start from 5'RLM-RACE product; green = part of gene coding sequence; underlined= initiation codon.

5.6. 5'UTR Analysis of the *mtx3* Toxin Gene

A 402 nucleotide sequence was obtained when a 5'RLM-RACE using an inner *mtx3* primer was conducted. BLAST analysis showed that the sequence is identical to a length of the *mtx3* gene of *L. sphaericus* strain SSII (accession number U42328.1). The transcript product started 80 nucleotides downstream of the initiation codon (Figure 5.6.). No other PCR products were obtained with this gene to allow identification of the promoter region.

```

                >Liu et al. -35                >Liu et al. -10
attaaaaattatattattctaccaattatttattttatggattatgtattataaaattaatatttttg < 70
tttattaatccaaaaataataacaactaaacttagtatgtagccatagcgtaaattatacactactaaaat < 140
                >Mtx3
tggaggagaaattatgaaaaacaagcaaaagtaattattaatgggagctactattggtctaagttactt < 210
                >Transcript product
tcatcaccaatagcaatggcgcaaatggagatagtaatgttaagagaatcaaagtatagcaaatttta < 280

```

Figure 5.6. Part of *mtx3* gene and its upstream sequence.

The promoter was predicted by Liu *et al.* (1996). Transcript product from RLM-RACE started 80 nucleotides downstream of the initiation codon. Red = putative promoter sequence; blue = transcript start from 5'RLM-RACE product; green = part of gene coding sequence; underlined= initiation codon.

5.7. 5'UTR Analysis of the Sphaericolysin Toxin Gene

A 146 bp amplicon was obtained from 5'RLM-RACE of sphaericolysin gene using the inner primer that was designed to produce around 500 bp PCR product plus some length of 5'untranslated region. When the sequence was sent for BLAST analysis, the result showed that the sequence was a truncated version of the sphaericolysin gene (accession number AB273179.1) (data not shown). Although the result did not reveal any part of 5' untranslated region of the sphaericolysin gene, it showed that the gene is present in *L. sphaericus* strain IAB59 and transcribed into mRNA. No further analysis of this transcript was undertaken.

5.8. 5'UTR Analysis of the *binB* Toxin Gene

The Bin proteins are produced from the *bin* operon (Berry, 2012) and, therefore, the genes must be transcribed from a promoter upstream of the *binB* gene. Total RNA obtained using RNeasy Protect Bacteria mini kit was used to analyse the 5' UTR sequence of the *binB* gene using RLM-RACE. The result produced a 484 nucleotide fragment using the *bin* inner primer. When submitted for BLAST analysis, the sequence was recognised as *binB* sequence (51.4 kDa protein of accession number X14964.1). It seems that this 5' UTR product of the *binB* gene contains 5'UTR sequence and may represent full length transcript, since three separate 5'RLM-RACE of *binB* gene produce almost identical 5'UTR transcripts starting at 67, 66 or 65 nucleotides upstream of the initiation codon (Figure 5.7.).

The transcript starts overlap with a putative promoter, which begins at 70 nucleotides upstream of the initiation codon. This promoter was proposed as similar to the consensus for a vegetative *B. subtilis/E. coli* promoter (Ahmed et al., 1995) . The predicted promoter is not consistent with our results for the *binB* gene, since the

5'untranscribed region obtained in these experiments extends into the predicted promoter. Baumann and Baumann (1989), suggested that the *binB* promoter is located between the *KpnI* site and the initiation codon, a region 96 nucleotides long (Baumann and Baumann, 1989) because the fragment gave Bin production when it was cloned. Our apparent initiation site is consistent with this proposal since it is downstream of the *KpnI* site although the 32 nucleotides (including the site) up to the start of transcription does not appear sufficient to contain a complete promoter. It is possible that transcription in the work of Baumann and Baumann was driven by plasmid-based factors or a combination of sequences derived from the plasmid and the 32 nucleotides upstream of the initiation point. Analysis of the region upstream of the gene using DBTBS database (<http://dbtbs.hgc.jp>) cannot find any putative promoter beyond that proposed by Ahmed *et al.*. As a result, we can speculate that a novel promoter, unrecognised in the databases, may exist upstream of the *binB* gene. Further investigation of the upstream region of the *bin* genes will be necessary to characterise the promoter responsible for transcription of the operon.

```

cttgccaacatgtgaagattaaggttaactttcagtttctttctgtgtaacaatacacgaagtaatat < 70
gtatctctatagaaattaatcaaaaaagacctagtgaactgcaccctgtcaagtagatagtggaataa < 140
taaaaatgtattaagcggcttgtgctcttaattctaagaaggtgatgaaaaacttgactctagggcgta < 210
ccaagcaatcaaaggagaaaacgttgtatctccgattaattcgaactacaacgtctattgtagattaac < 280
cgtaagagattatataattattataacaatactttaattaattcaattgaaattaacaacaacttttt < 350
cggtatatactatctaacaatcgtttcagttccatcgttctaaaggtagccttcttttgggtaacggt < 420
    >KpnI recognition seq.
    >Transcript start
    >-35 promoter      >-10 promoter
atthaatgaactttttaggttttaataataataatgagaagtatttttatcaatgataaggagatgaag < 490
    >BinB
aaagcatgtgcgattcaaaagacaattctggcgtttcagaaaaatgcg < 539

```

Figure 5.7. Part of *binB* gene and its upstream sequence.

Transcript starts from three separate experiments at 67, 66 or 65 nucleotides upstream of the initiation codon, overlapped with the -35 promoter cited by Ahmed *et al.* (1995). Red = putative promoter sequence; blue = transcript start from 5'RLM-RACE product; brown = *KpnI* recognition sequence; green = part of gene coding sequence; underlined= initiation codon.

CHAPTER VI WHOLE GENOME SEQUENCING OF *LYSINIBACILLUS SPHAERICUS* STRAIN NHA15b

6.1. Introduction

The potency of *Lysinibacillus sphaericus* to control mosquitoes has been known for several decades and some strains have already been applied in the field to control these insects. However, a thorough understanding of the diversity of *L. sphaericus* toxins was still needed especially in the wake of resistance reported from several field application and laboratory settings (e.g. (Nielsen-Leroux et al., 1995; Yuan et al., 2000) The Next Generation Sequencing techniques provide a new approach to the study of toxin diversity of this bacterium. In the beginning of this study, there was only one completed genome sequencing project on a pathogenic strain of *L. sphaericus* C3-41 (Hu et al., 2008a), although, during the period of the project, there have been several studies to sequence (not in completed form) other strains of *L. sphaericus* (Xu et al., 2015)(Pena-Montenegro et al., 2015)(Jeong et al., 2013). In this study, we focus on the strain NHA15b that differs from the C3-41 strain; particularly, as it does not have the *binA/B* genes and instead encodes the Cry48/49 binary toxins.

This study was started attempting to use the HiSeq 2000 Sequencing System (Illumina Sequencing) in a single read mode with a read length of 50 bases (GATC Biotech, Konstanz, Germany), but then there was an opportunity to exploit another strategy that would give a better result. The whole study in this chapter is mostly derived from paired-end sequencing data using The Illumina System. The contigs that came from *de novo* assembly of the single reads data were still utilized when there was a need, e.g. to validate the circularity of the plasmids identified.

NHA15b total DNA was prepared using an Imp DNA mini kit (Qiagen) and was sequenced at the Biomedical Laboratories, Institute of Life Science, Medical School, Swansea University, Swansea. The Illumina sequencing of strain NHA15b produced paired-end reads and contigs with statistics as shown in Table 6.1. below:

Table 6.1. Sequence and assembly data for the genome of *L. sphaericus* NHA15b

Number of sequences	1,027,654
Best kmer	95
Number of contigs for best assembly	140
N50	87,923
Estimated coverage	19.1
Length of longest contig	457,255
Total bases in contigs	4,975,675
Number of contigs > 1k	101
Total bases in contigs > 1k:	4,963,146

The sequence data were received from the sequencing facility at Swansea as contigs fasta format and paired end fastq format data. Contig sequences were assembled using the Velvet program (Zerbino, 2010) by Dr Ben Pascoe (Swansea University) with kmer 95 as the best kmer chosen based on the Velvet Optimizer result.

To improve the quality of the assembly and to annotate the genome, a software package called PAGIT (post-assembly genome-improvement kit) (Swain et al., 2012) was used. PAGIT consist of 4 tools, namely ABACAS (algorithm-based automatic contiguation of assembled sequences), IMAGE, ICORN, RATT. To assist the

assembly improvement process, three available *L. sphaericus* sequences were used as references, namely the complete chromosome sequence *L. sphaericus* strain C3-41 (reference: NC_010382.1), plasmid pBsph sequence from *L. sphaericus* strain C3-41 (reference: NC_010381.1)(Hu et al., 2008a) and plasmid pLG sequence from *L. sphaericus* strain LP1-G (reference: AY325804.1) (Wu et al., 2007) . Using ABACAS, the contigs were ordered and oriented to *L. sphaericus* C3-41 and plasmid pLG sequences. The results showed that most of the NHA15b contigs could be aligned to the C3-41 reference genome, hence showing that there is a high genome similarity between C3-41 and NHA15b genomes. This is consistent with the fact that all mosquito pathogenic strains are known to belong to the same DNA homology subgroup of *L. sphaericus*, GroupIIA (Krych et al., 1980b) . The analysis also showed the presence of pBsph-like and pLG-like plasmids in the NHA15b strain. The ordered contigs were then passed into the IMAGE program to extend contig ends into gaps. From 158 gaps produced from the ABACAS processes, the IMAGE tools were able to fill some of the gaps, so at the end of iteration, there were 114 gaps left unconnected. The IMAGE products were directly fed to the RATT program to transfer the annotation of reference sequences to the NHA15b sequences. In the three reference sequences, 4285 putative genes were previously identified. By comparison of the NHA15b data with these CDSs, 3756 were successfully transferred to NHA15b sequences. Most of these CDSs corresponded to genomic sequences. Of 179 annotated CDSs from plasmid Bsph of C3-41, 155 were successfully transferred and from 23 annotated CDSs of plasmid pLG, 15 were transferred (see later sections on plasmid composition).

In parallel, the ABACAS products were submitted to the Basys (Bacterial annotation system) website (<https://www.basys.ca>) for an automatic annotation (Van Domselaar et al., 2005). The annotation results were used to complement the results from the annotation transfer procedure using the RATT tool, especially for the regions that did not share any similarity to the reference genomes.

The BLAST Ring Image Generator (BRIG) was used for genome comparison and visualization. The tool was employed to generate circular images of either chromosome or plasmids of *L. sphaericus* strain NHA15b in comparison to other bacterial genomes that share similarity to the genome of NHA15b or its plasmids. The genomes for comparison were downloaded from the NCBI website. These genomes were selected on the basis of Blast search showing a high degree of sequence identity.

6.2. General NHA15b Genome Comparison

IMAGE tools produced a 3,992,820 bp chromosome scaffold. The size did not reflect the real size of the chromosome due to a lack of synteny between NHA15b and C3-41 for some regions, which meant that some NHA15b sequences were not incorporated into this chromosome scaffold. However, the RATT tool was later able to annotate further sequences in comparison to the reference genome meaning that the non incorporated contigs indeed are part of chromosome fragments. To provide general comparison, the NHA15b genome was compared to C3-41 and several other genomes of *Lysinibacillus sp.* available to date ((Jeong et al., 2013); (Pena-Montenegro et al., 2015); (Xu et al., 2015)). Figure 6.1. shows that the strain NHA15b shared great similarity to the C3-41 genome and the other genomes being compared. There was a prominent gap in the NHA15b chromosome compared with

the C3-41 chromosome, covering a region that was known as the location of the *bin* toxin genes (BSPH_RS14880 and BSPH_RS14885 in the C3-41 annotation) and the *mtx4* gene (BSPH_RS14895). This region represents a 35 kb duplicated region in the chromosome and plasmid pBsph of strain C3-41 (Hu et al., 2008a). On the other hand, NHA15b showed the presence of genes equivalent to the C3-41 *mtx1* (BSPH_RS05330), *mtx2* (BSPH_RS05295), and *mtx3* genes (BSPH_RS13250). Also detected was the sphaericolysin gene (BSPH_RS19015). The occurrence of these toxin genes was also confirmed from the annotation transfer results using the RATT tool. The *mtx4* gene was found later, when contigs corresponding to a 15 kb region known to harbour *cry48/49* genes was examined. The toxin genes found in the NHA15b genome will be presented in a separate subsection.

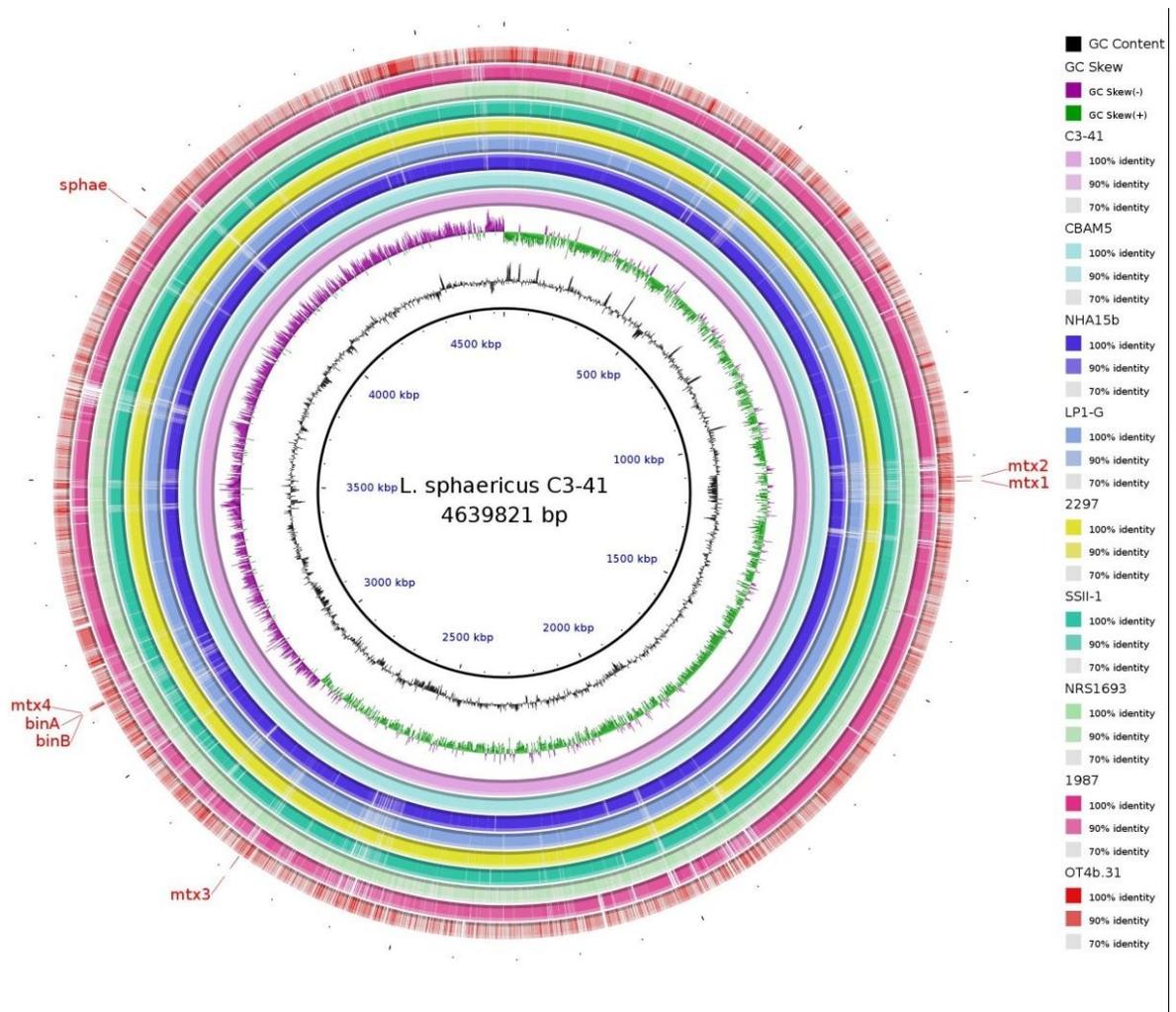


Figure 6.1. Comparison of chromosome of *L. sphaericus* strain C3-41 to several *Lysinibacillus* genomes

Red blocks showed location of toxin genes on strain C3-41. **C3-41**: chromosome of *L. sphaericus* strain C3-41 (NC_010382.1); **CBAM5**: whole genome sequence of strain CBAM5 (WGS: AYKQ01); **NHA15b**: whole genome sequence of strain NHA15b; **LP1-G**: whole genome sequence of strain LP1-G (WGS: JPDL01); **2297**: whole genome sequence of strain 2297 (WGS: JPDJ01); **SSII-1**: whole genome sequence of strain SSII-1 (WGS: JPDK01); **NRS1693**: whole genome sequence of strain NRS1693 (WGS: JPDM01); **1987**: whole genome sequence of strain 1987 (WGS: JMMU01); **OT4b.31**: whole genome sequence of strain OT4b.31 (AQPX01)

6.3. The plasmids of strain NHA15b

In addition to pBsph-like and pLG-like plasmids that were discovered using the alignment process of ABACAS (above), further plasmid sequences were also identified in strain NHA15b. Several criteria were used to identify plasmids. These included the identification of NHA15b contigs showing an above average level of sequence coverage (general level of coverage = 19.1x) indicating that these contigs may represent repetitive sequences or plasmids with a copy number greater than one. The prediction that some contigs may represent plasmids was supported by the fact that they could be circularized, and contained a gene or genes predicting the presence of a plasmid. The genes used are the genes that were used to classify and identify microbial plasmids, namely relaxase, type IV coupling protein (T4CP), type IV secretion system (T4SS), and FtsK-homologous protein genes (Shintani et al., 2015). Some discretion was applied in this process so that Contig 26, despite having an average coverage close to the general level (20.28x compared to 19.1x), is proposed as a plasmid since it also contains genes frequently associated with the presence of a plasmid and could be circularized. In all, five plasmids were identified in NHA15b (see below).

There are different suggestions on how properly to name a new plasmid (for example instructions to authors for numerous journals) but no officially recognized system. The plasmids previously discovered in *L. sphaericus* strains were named as: pBsph - that was first discovered in strain C3-41, pBsph-2 -that was discovered in strain SSII-1 and is similar to pBsph, and pLG –because it was discovered in the strain LP1-G. Based on the suggestion in the Journal of Bacteriology (Instructions to

Authors), that plasmid naming should consist of the name of the species in which the plasmid was found, plus a reference to the laboratory that worked on it plus a series of digits, it was decided to name the five plasmids as pLsph100, pLsph200, pLsph300, pLsph400, pLsph500, using a format of p = plasmid; Lsph = specific name; xxx = number. Reference to the originating laboratory was omitted.

6.3.1. Plasmid *pLsph100*

During the sequencing of the C3-41 genome, a megaplasmid, designated pBsph, was identified (Hu et al., 2008a) and this was used as one of the reference sequences for the assembly of the NHA15b data. In NHA15b we have identified a plasmid, here designated pLsph100 with clear homology with pBsph. The size of the pLsph100 plasmid is 151,059 bp, somewhat smaller than pBsph (177642 base pair). As mentioned before, RATT annotation was able to transfer 155 out of 179 of the pBsph genes to pLsph100. The success of the annotation transfer reflects the degree of similarity between the two plasmids. The BLAST analysis using the BRIG tool showed a high similarity to the pBsph of C3-41 but the 35 kb region that carries the *bin* genes (BSPH_RS23180; BSPH_RS23185) and *mtx4* gene (BSPH_RS23195) in pBsph was absent from pLsph100 (Figure 6.2.). Conversely, small regions are present in pLsph100 that contain CDSs not present in pBsph of C3-41 but these CDSs were annotated as encoding ical proteins so at this point it is not possible to assign functional significance to the presence of these genes. Recently, a variant of the pBsph plasmid was also shown to be present in the SSII-1 genome and this was named pBsph-2 (Xu et al., 2015). This plasmid, like pLsph100, was also reported to lack the so-called 35 kb region that carries the *bin* genes and *mtx4* gene, however, the SSII-1 sequence was not assembled to complete the pBsph-2 sequence so further comparisons with pLsph100 are not possible.

It was shown that pBsph-1 and pBsph2 harbour genes that encode proteins related to a type IV secretion system (T4SS), (e.g. VirD4, VirB4 and VirB6) even though the degree of similarity of these genes was low compared to the type IV secretion system genes from other bacteria. The T4SS protein are involved in mating pair formation during plasmid conjugation but based on the distant relationship of these pBsph genes to known T4SS, it was speculated that the pBsph plasmids may have lost their ability to act as conjugative plasmids (Xu et al., 2015) . There are no reports in the literature of these plasmids acting as conjugative plasmids. Blastx search for the three T4SS genes from the pBsph plasmids on pLsph100 found that the *virD4* (BSPH_RS22450) *virB4* (BSPH_RS22695) and *virB6* (BSPH_RS22700) genes were also present in the plasmid pLsph100. Similar to pBsph-1 and pBsph-2, the gene encoding an FtsZ/tubulin family protein known to be involved in translocation of double-stranded DNA to the recipient cell in gram positive bacteria (Goessweiner-Mohr et al., 2013) was also detected in the CDSs of plasmid pLsph100.

A Blast search was conducted against the genomes deposited in the NCBI database and this revealed that *Lysinibacillus fusiformis* strain SW-B9 genome (reference: JRBA01000000) might also contain a similar plasmid since there is much sequence similarity, however, as for strain SSII-1 above, a plasmid similar to pBsph was not assembled or reported for this strain. With the exception of pBsph, all of the other plasmids lacked the 35 kb region that harbours the *bin* and *mtx4* genes. The *rep*, *virD4* and *virB4* and also the *virB6* genes were shown to be in the conserved region of the plasmids.

Overall, our work and previous studies appear to indicate that pBsph-like sequences are widespread in the *Lysinibacillus* genus but that in many cases this is not associated with the presence of insecticidal toxin genes.

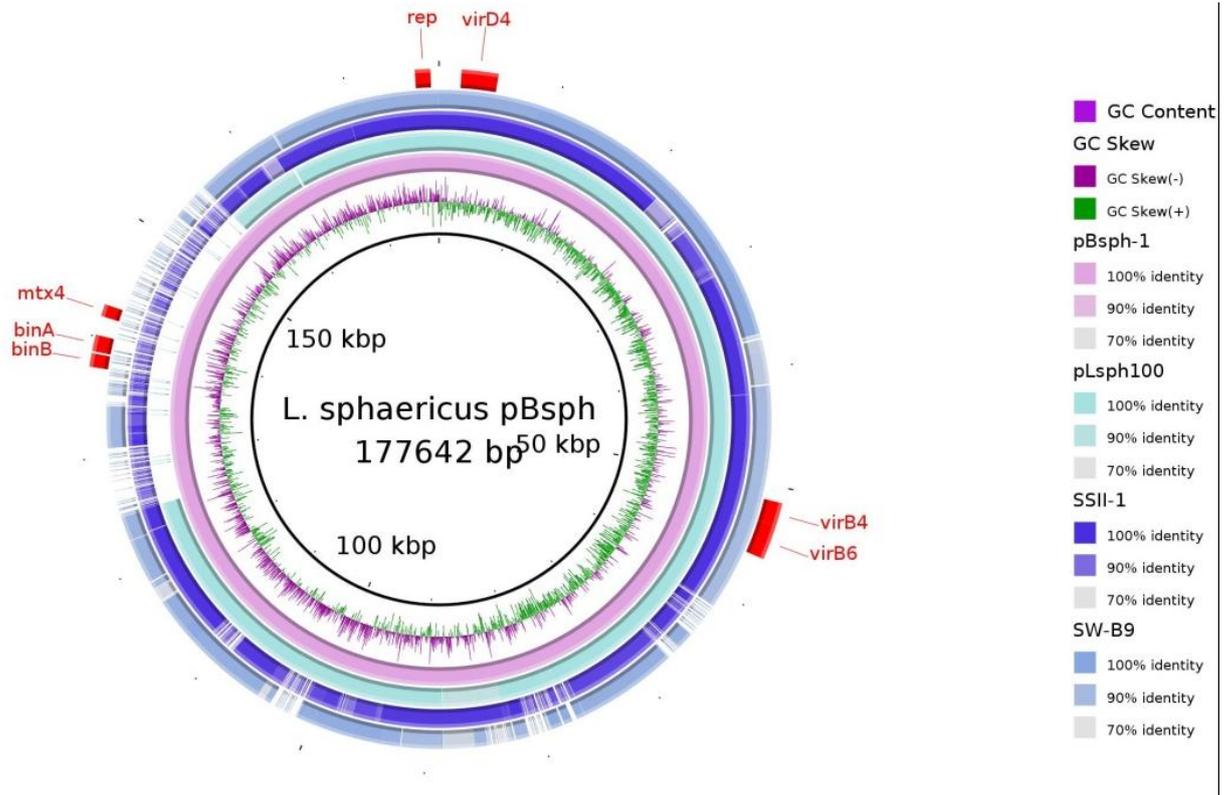


Figure 6.2 . Comparison of pBsph-1 of *L. sphaericus* strain C3-41 to pLsph100 and *Lysinibacillus* genomes containing similar plasmids. Red blocks show location of toxin, replication initiation protein (rep) and T4SS genes on pBsph-1. **pBsph-1:** plasmid pBsph of *L. sphaericus* strain C3-41 (NC_010381.1); **pLsph100:** plasmid pLsph100 from *L. sphaericus* strain NHA15b; **SSII-1:** plasmid predicted from the whole genome of *L. sphaericus* strain SSII-1 (WGS: JPDK01); **SW-B9:** plasmid predicted from the whole genome of *L. fusiformis* SW-B9 (WGS: JRBA01)

6.3.2. Plasmid pLsph200

A cryptic 11,066 bp plasmid, pLG was previously described from *L. sphaericus* strain LP1-G (Wu et al., 2007) and was used as a template sequence for assembly in this work. A homologous sequence in NHA15b, here named as pLsph200, was detected from a contig with 87.32x coverage and contained total of 12,466 base pairs, making it larger than the pLG plasmid (11,066 bp). The pLsph200 plasmid was shown to harbour at least 15 CDSs identical to the 23 CDSs of the pLG plasmid, leaving only 8 CDSs that could not be transferred to the pLsph200 plasmid using the RATT tool. The un-transferred CDSs included the *rep* CDS that encodes a replication initiation protein for the pLG plasmid. This could not be transferred because the degree of similarity of that particular region with pLsph200 was very low. The Basys automatic annotation tool produced 17 CDS, most of them are encoding hypothetical proteins. Nevertheless, using blastx of the NCBI websites, a CDS (BASYS00017) in the equivalent region of pLsph200 was found to have some degree of similarity to the replication initiation protein (partial) of *B. amyloliquefaciens* (WP_032859203.1 E = 7e-130 Ident = 50%). Thus, it appears that while there is considerable conservation between many parts of the plasmids, the regions putatively encoding the replication initiation protein appear to be quite different. Since all the rest of the CDS of the two plasmids are predicted to encode hypothetical proteins, it is not known whether there is any difference in function of the plasmids. Blastn comparison between the two plasmids using the Easyfig tool showed that, despite clear homology, the plasmids were significantly different (Figure 6.3.).

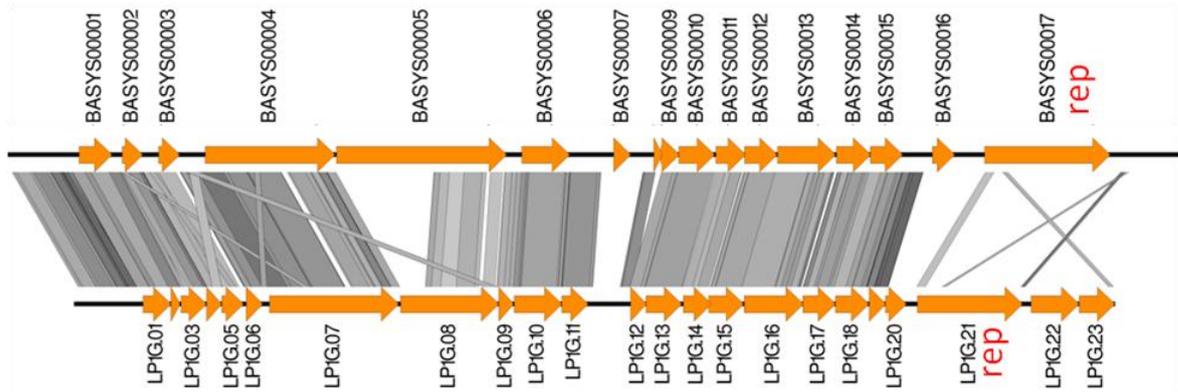


Figure 6.3. Comparison of plasmid pLG of *L. sphaericus* strain LP1-G and plasmid pLsph200 of *L. sphaericus* strain NHA15b
 Blastn plasmid comparison between plasmid pLG (AY325804.1) (line below) and pLsph200 using Easyfig tool. Darker shade indicates a higher degree of similarity. Most of the CDSs show synteny between the two sequences. The LP1G.21 CDS and BASYS017 CDS are both predicted to encode replication initiation proteins (rep).

6.3.3. Plasmid pLsph300

Plasmid pLsph300 was detected from a contig with a size of 10,341 base pairs and 307.24x coverage. Using the data from single read sequencing, the plasmid was circularized and trimmed to 10,336 base pairs. The Basys automatic annotation predicted that the plasmid contained 20 CDSs. Blastx search on NCBI databases showed that most of the CDSs are predicted to produce hypothetical proteins, so the plasmid is a cryptic plasmid. However, one of the CDSs showed similarity to the ribbon-helix-helix, CopG family protein of *B. coagulans* DSM 1 = ATCC 7050 (reference: AJH79919.1 E = 1e-14. Ident = 73%). CopG/Ribbon-Helix-Helix proteins function to control the copy number of a plasmid and was previously called repA protein (del Solar et al., 1998), hence it indicated further that the contig was indeed a plasmid, even though, at present, no CDSs can be predicted as replication initiation proteins or any other genes necessary for the replication of the plasmid. The list of the blast results for the CDSs of pLsph300 is presented in Table 6.2. Blast search to discover whether homolog of this plasmid occur in other bacterial genomes did not encounter any similar sequences.

Table 6.2. CDS of plasmid pLsph300 of *L. sphaericus* strain NHA15b

Name	Position (base)	Size (aa)	Product (Basys Annotation)	Best BLASTx match
BASYS00932	29..403	124	Hypothetical Protein	hypothetical protein [<i>Sporolactobacillus terrae</i>] ref WP_037580981.1 E= 5e-07 Ident = 30%
BASYS00933	407..751	114	Hypothetical Protein	hypothetical protein [<i>Amphibacillus jilinensis</i>] WP_017473875.1 E = 6e-15 Ident = 36%
BASYS00934	783..917	44	Hypothetical Protein	no hit
BASYS00935	c(1014..3914)	966	Helicase Domain-Containing Protein	hypothetical protein [<i>L. varians</i>] WP_025219684.1 E = 0.0 Ident = 60%
BASYS00936	3971..4072	33	Hypothetical Protein	no hit
BASYS00937	c(4023..4187)	54	Hypothetical Protein	no hit
BASYS00938	c(4537..5259)	240	Transposon Tn552 resolvase	resolvase [<i>L. xylanilyticus</i>] WP_049669020.1 E = 6e-97 Ident = 74%
BASYS00939	c(5291..5575)	94	Hypothetical Protein	MULTISPECIES: hypothetical protein [<i>Paenibacillus</i>] WP_024633656.1 E = 1e-11 Ident = 40%
BASYS00940	c(5635..6699)	354	Recombinase	integrase [<i>Solibacillus silvestris</i>] WP_014829545.1 E = 1e-136 Ident = 56%
BASYS00941	6889..7029	46	Hypothetical Protein	no hit
BASYS00942	c(7026..7343)	105	Hypothetical Protein	MULTISPECIES: hypothetical protein [<i>Geobacillus</i>] WP_033016628.1 E = 0.004 Ident = 31%

BASYS00943	c(7333..7455)	40	Hypothetical Protein	no hit
BASYS00944	7398..7532	44	Hypothetical Protein	no hit
BASYS00945	7622..7783	53	Hypothetical Protein	no hit
BASYS00946	7774..8220	148	Hypothetical Protein	hypothetical protein BB14905_22673 EAZ83476.1 E = 70% Ident = 70%
BASYS00947	8302..8457	51	Hypothetical Protein	no hit
BASYS00948	c(8387..8641)	84	Hypothetical Protein	hypothetical protein [<i>B. cereus</i>] WP_046941590.1 E = 2e-15 Ident = 59%
BASYS00949	c(8753..9346)	197	Hypothetical Protein	feS assembly ATPase SufC [<i>Coprobacillus</i> sp. CAG:235] CCZ23995.1 E = 2.6 Ident = 23%
BASYS00950	9448..9603	51	Hypothetical Protein	hypothetical protein BN1050_02324 CEA05130.1 E = 6e-18 Ident = 84%; ribbon-helix-helix, copG family protein [<i>Bacillus coagulans</i> DSM 1 = ATCC 7050] AJH79919.1 E = 1e-14 Ident = 73%
BASYS00951	9678..10133	151	Hypothetical Protein	hypothetical protein [<i>B. megaterium</i>] WP_047934255.1 E = 0.003 Ident = 28%
BASYS00952	10130..10249	39	Hypothetical Protein	no hit

6.3.4. Plasmid pLsph400

A contig with 10,687 base pairs and 43.33x coverage was designated as plasmid pLsph400. It was circularized with the help of the single read sequencing data and trimmed to 10,593 bp. The Basys annotation produced 14 CDSs for plasmid pLsph400. Using Blast search on the NCBI websites, a 975 base pair CDS found downstream of CDS BASYS00014 was predicted to encode a protein similar to a DNA replication protein of the bacterium *Vibrio anguillarum* (reference : WP_015061687.1 E = 0.0 Ident = 85%) and a hypothetical protein from *Lysinibacillus sp.* strain LK3 (reference WP_048395416.1 E = 0.0 Ident = 89%). This supports the suggestion that the NHA15b contig is a plasmid because it has a *rep* gene (although the *Vibrio* sequence exists in isolation in the database and its association with a plasmid has not been described). The rest of the pLsph400 CDSs appear to encode hypothetical proteins, suggesting that the plasmid is a cryptic plasmid. The list of its CDSs is presented in Table 6.3. Blast search of plasmid pLsph400 showed similarity to sequences from the total genomic sequences of *L. fusiformis* strain ZC1 (ADJR01) and *Lysinibacillus sp.* strain LK3 (LDUJ01) with a lower degree of similarity to the *Bacillus sp.* strain Ged10 genome (CAVI01). Alignment using BRIG shows that the strains share some degree of similarity throughout the plasmid region especially in the *rep* gene location. The plasmid alignment is presented in Figure 6.4

Table 6.3. CDS of plasmid pLsph400 of *L. sphaericus* strain NHA15b

Name	Position (base)	Size (aa)	Product (Basys Annotation)	Best BLASTx match
BASYS000001	c(362..1027)	221	Hypothetical Protein	hypothetical protein [<i>Lysinibacillus</i> sp. LK3] WP_048395414.1 E = 4e-148 Ident = 96%
BASYS000002	c(1352..1960)	202	Hypothetical Protein ABC	hypothetical protein [<i>L. sphaericus</i>] WP_036128137.1 E = 3e-122 Ident = 97%
BASYS000003	c(2037..2273)	78	Hypothetical Protein	hypothetical protein [<i>Lysinibacillus</i> sp. LK3] WP_048395410.1 E = 8e-32 Ident = 98%
BASYS000004	2381..2881	166	Hypothetical Protein	hypothetical protein [<i>L. fusiformis</i>] WP_004234047.1 E = 3e-94 Ident = 83%
BASYS000007	c(4367..4723)	118	Hypothetical Protein	hypothetical protein [<i>L. sphaericus</i>] WP_024361730.1 E = 7e-64 Ident = 81%
BASYS000009	5160..5315	51	Hypothetical Protein	acetyl-CoA carboxylase [<i>Rhodococcus fascians</i>] WP_048317114.1 E = 2.1 Ident = 37%
BASYS000011	6672..7130	152	Hypothetical Protein	hypothetical protein [<i>B. marisflavi</i>] WP_048005416.1 E = 1e-15 Ident = 37%
BASYS000012	c(6617..7123)	168	Hypothetical Protein	hypothetical protein [<i>B. marisflavi</i>] WP_048005416.1 E = 3e-15 Ident = 37%
BASYS000013	8441..8662	73	Hypothetical Protein	hypothetical protein [<i>L. fusiformis</i>] WP_004233073.1 WP_004233073.1 E = 3e-34 Ident = 90%
BASYS000014	c(9212..9472)	86	Hypothetical Protein	hypothetical protein [<i>Kurthia huakuii</i>] WP_029501244.1 E = 5e-35 Ident = 69%

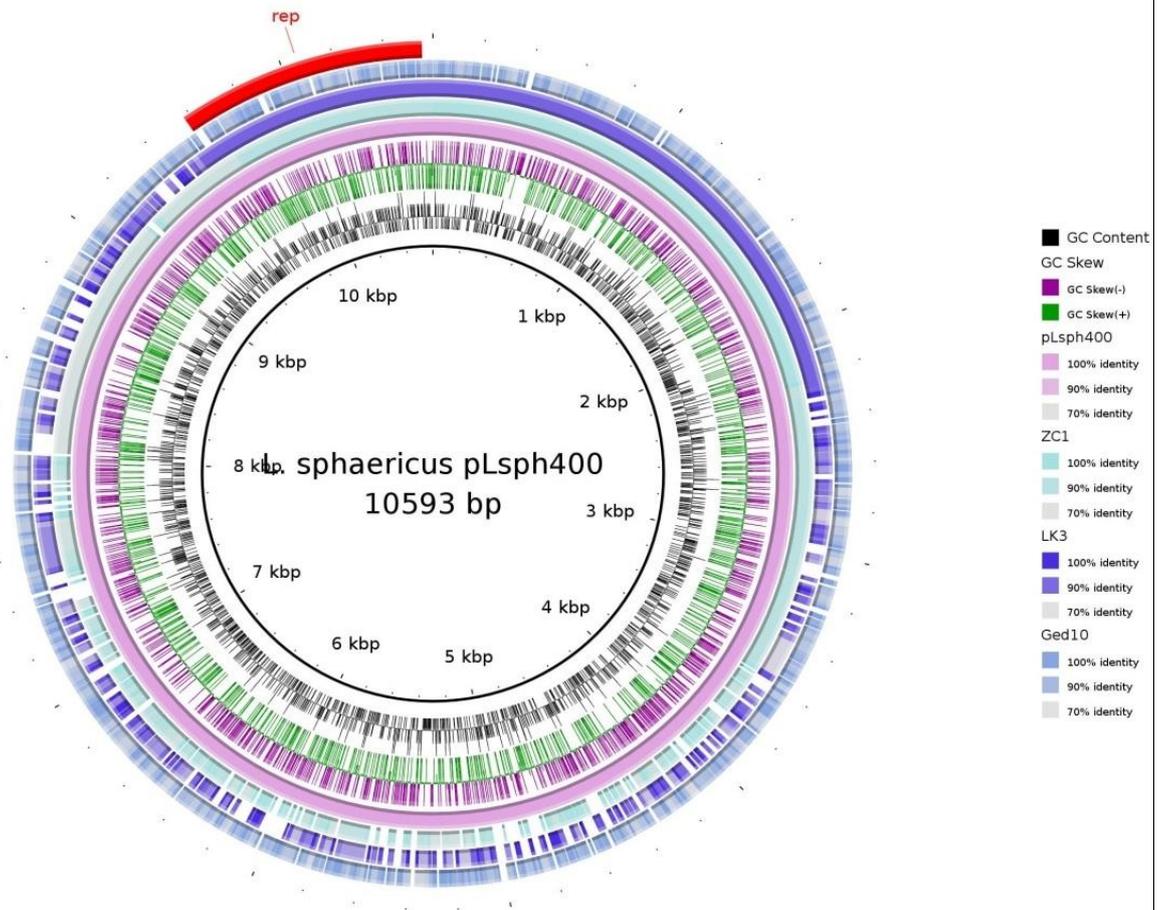


Figure 6.4. Comparison of pLsph400 of *L. sphaericus* strain NHA15b and plasmids predicted from the whole genome sequences
 Red blocks show location of replication initiation protein (rep) pLsph400. **pLsph400**: plasmid pLsph400 from *L. sphaericus* strain NHA15b; **ZC1**: plasmid predicted from the whole genome of *L. fusiformis* strain ZC1 (WGS: ADJR01); **LK3**: plasmid predicted from the whole genome of *Lysinibacillus sp.* LK3 (WGS: LDUJ01) **Ged10**: plasmid predicted from the whole genome of *Bacillus sp.* GeD10 (WGS: CAVI01)

6.3.5. Plasmid pLsph500

Plasmid pLsph500 was designated from a contig with a size of 50,525 base pairs with 20.28x coverage. Again, using the single read sequencing data, the contig was circularized and trimmed into a 50,459 base pair plasmid. Basys automatic annotation produced 182 CDSs. Some of these appear to represent genes that function in plasmid replication and conjugation, hence supporting the designation of this element as a plasmid. The BASYS000052 CDS, which was predicted by Basys to produce a hypothetical protein, was found to be related to a DNA replication protein in *Vibrio anguillarum* when compared using tblastn to NCBI databases. The plasmid also appears to encode a putative ATPase that may contribute to plasmid mobility as shown from the tblastn result (BASYS000051 CDS). It also harbours some genes related to Type IV Secretion Systems (T4SS). Basys annotation predicts that CDS BASYS000017 produces a product related to an uncharacterized protein of pXO2 plasmid, a plasmid that was first found in *B. anthracis* and which carries capsule genes as one of its virulence factors (Read et al., 2003). However, tblastn search on NCBI databases suggested that the CDS encoded a type IV secretory pathway protein, VirD4 of *Lysinibacillus*. In BASYS 000024, the automatic annotation tool found a match to another uncharacterized protein of pXO2, that on the tblastn search was revealed as a type IV secretion system VirB4 protein.

The *VirB4* and *VirD4* genes of pLsph500 shared some degree of similarity to the homologs of these genes on pLsph100, but are more closely related to the homologous genes of the predicted plasmid containing *cry* genes. Further analysis of the T4SS genes will be discussed in section 6.4.5.

Furthermore, an FtsZ/Tubulin-Related Protein (BASYS000045) was also predicted on the plasmid. Based on the presence of the IV secretion system (T4SS)

genes, both pLsph100 and pLsph500 are predicted as conjugative plasmids. Similar to the pBsph situation, since there are significant differences to the reference genes, it is predicted that the plasmids are no longer capable of self conjugation (Xu et al., 2015) . Most of the rest of the CDSs of pLsph500 were predicted to encode hypothetical proteins with unknown functions. A representative list of CDSs related to genes with a known function is presented in Table 6.4.

Blast search on the NCBI website suggested that this plasmid is also present in the *L. sphaericus* strain LP1-G, *L. fusiformis* strain RB-21, and *L. fusiformis* ZC1 genomes. The comparison of plasmid pLsph500 and its predicted similar plasmids from the three genomes is presented in Figure 6.5.

Table 6.4. CDS of plasmid pLsph500 of *L. sphaericus* strain NHA15b

Start	End	Strand	Accession	Protein Function (BASYS)	Blastx Result
642	917	+	BASYS00002	Transition state regulatory protein AbrB [H]	
2004	2372	+	BASYS00005	Transcriptional Regulator	
2820	3932	+	BASYS00006	Surface-layer 125 kDa protein [H]	
12926	15214	+	BASYS00017	Uncharacterized protein pXO2-15/BXB0014/GBAA_pXO2_0014 [H]	type IV secretory pathway, VirD4 protein <i>Lysinibacillus sphaericus</i> E:0.0; Ident: 75%
16225	17334	+	BASYS00020	Uncharacterized protein pXO2-25/BXB0023/GBAA_pXO2_0023 [H]	type II/IV secretion system protein [<i>Lysinibacillus fusiformis</i>] WP_004230935 E;0.0; Ident: 79%
18179	20506	+	BASYS00022	Uncharacterized protein pXO2-14/BXB0013/GBAA_pXO2_0013 [H]	
21496	23451	+	BASYS00024	Uncharacterized protein pXO2-09/BXB0008/GBAA_pXO2_0008 [H]	MULTISPECIES: type IV secretion system protein VirB4 [<i>Lysinibacillus</i>] WP_052323568.1 E+ 0.0; Ident: 77%
24028	25098	+	BASYS00026	Uncharacterized protein yddH [H]	hypothetical conjugation protein [<i>Lysinibacillus fusiformis</i> ZC1] EFI66364 E:0.0; Ident: 86%
31482	33668	+	BASYS00038	DNA topoisomerase 3 [H]	DNA topoisomerase III [<i>Lysinibacillus sphaericus</i>] WP_036165815 E:0.0; Ident: 60%
35992	36345	+	BASYS00041	Hypothetical Protein Bsph_p	
36329	36691	+	BASYS00042	Hypothetical Protein Bsph_p	
39011	37455	-	BASYS00044	Hypothetical Protein Bpr_I	

40944	39523	-	BASYS00045	FtsZ/Tubulin-Related Protein	cell division protein FtsZ <i>Bacillus</i> sp. FJAT-14515 WP_028393938.1 E: 1e-142; Ident: 48%
42477	41569	-	BASYS00047	Hypothetical Protein Bsph	
42764	43330	+	BASYS00048	Transposon Tn917 resolvase [H]	
46038	45193	-	BASYS00051	Sporulation initiation inhibitor protein soj [H]	ATPase [<i>Lysinibacillus odyseyi</i>] WP_036150309 E: 2e-172; Ident: 87%
46702	47787	+	BASYS00052	Hypothetical Protein BASYS00052	DNA replication protein <i>Vibrio anguillarum</i> WP_015061687.1 E: 0.0 ; Ident 86%
48453	47956	-	BASYS00053	Phage-like element PBSX protein xkdA [H]	
48909	48556	-	BASYS00054	HTH-type transcriptional regulator xre [H]	
49064	49333	+	BASYS00055	Uncharacterized HTH-type transcriptional regulator yqaF [H]	

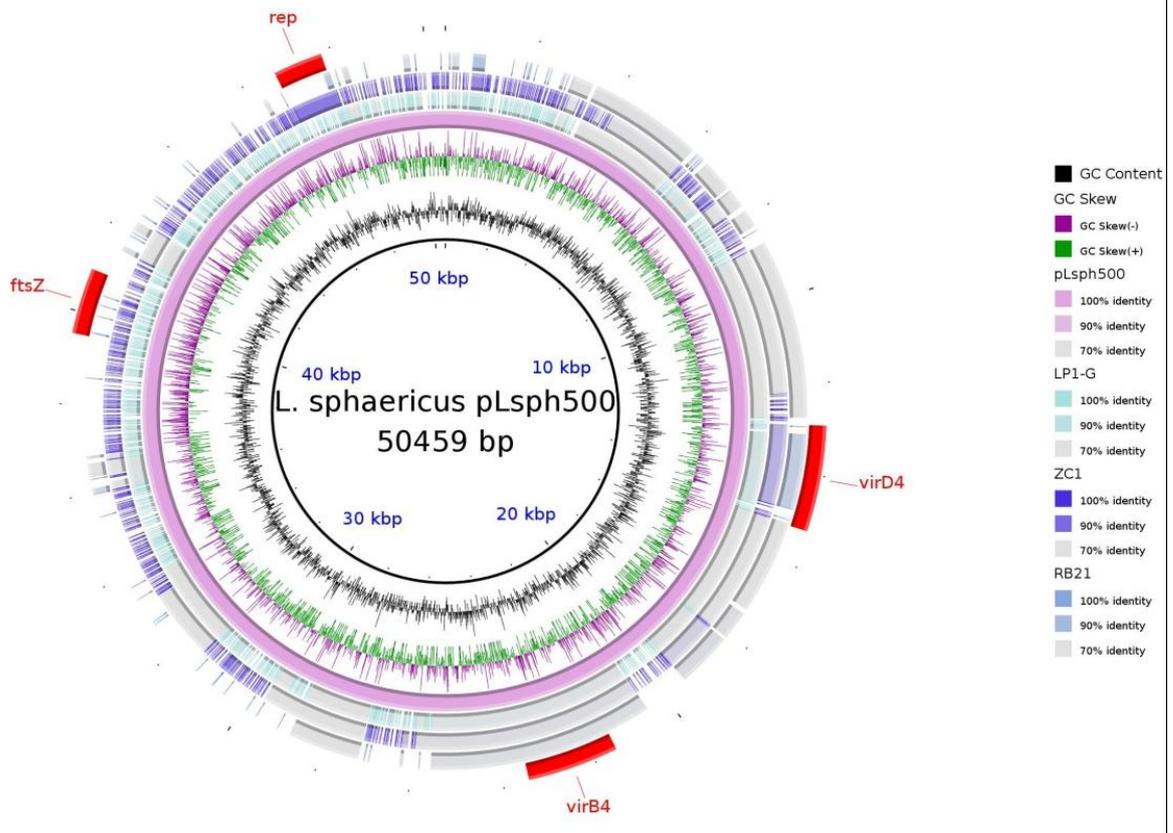


Figure 6.5. Comparison of pLsph500 of *L. sphaericus* strain NHA15b and plasmids predicted from the whole genome sequences of *Lysinibacillus*. Red blocks showed location of replication initiation protein (*rep*) and T4SS genes on pLsph500. **pLsph500**: plasmid pLsph500 from *L. sphaericus* strain NHA15b; **LP1-G**: plasmid predicted from the whole genome of *L. sphaericus* strain LP1-G (WGS: JPDL01); **ZC1**: plasmid predicted from the whole genome of *L. fusiformis* strain ZC1 (WGS: ADJR01); **RB21**: plasmid predicted from the whole genome of *L. fusiformis* strain RB-21 (CP010820.1)

6.4. Insecticidal toxins of *L. sphaericus* strain NHA15b

A number of mosquitocidal toxins have been discovered in *L. sphaericus* strains and in this section the presence of these genes in NHA15b will be reviewed.

6.4.1. *Mtx1* toxin

The 100 kDa Mtx1 protein is produced during the vegetative stages of bacterial growth and is widely distributed among highly-toxic and low toxicity strains of *L. sphaericus*.

The *mtx1* gene of strain NHA15b was detected on the genome using RATT annotation as BSPH_RS05330 based on its identity and synteny to the *mtx1* gene of the strain C3-41. The *mtx1* gene of C3-41 was predicted to be a pseudogene because it has frame shifts within the gene (Yuan et al., 2008). The *mtx1* gene of the NHA15b strain seems to be functional since it does not have similar frame shifts and shared a high similarity to the other *mtx1* genes from other strains including the first-characterised gene from strain SSII-1, which is known to produce an active toxin (Thanabalu et al., 1992). There are several amino acid differences between the NHA15b and SSII-1 sequences, in particular in the region from amino acid 93 to 231 that is considered as the catalytic core (Reinert et al., 2006), but since the amino acid changes are all conservative and do not affect actual catalytic residues, it is predicted that these changes would not affect the toxicity of the NHA15b protein. The Mtx1 sequence alignment is presented in Figure 6.6. The amino acid differences are shown in yellow highlight.

In strain SSII-1, there is an inverted repeat sequence upstream of the *mtx1* gene that is predicted to be involved in regulating its expression. The repeat is located at

nucleotide 1135 to 1179 between the proposed -10 region and the ribosome binding site (Thanabalu et al., 1991). However, in the NHA15b strain, such a repeat was not detected when a 2000 base pair sequence upstream of the *mtxI* gene was checked using Tandem Repeat Finder Program (Benson, 1999). An alignment of the regions is shown in Figure 6.7. Based on this situation, it is predicted that the *mtxI* gene in this strain could be expressed constitutively without the regulation proposed in SSII-1.

AAN04906.1	1	MAIKKVLKIIILAI III I I SCQLPLNQKTVYASPNSPKDN TWIQAASLTWLM DMSSLLYQLI STRIP SFASPNGLHMREQTI	80
BAD01571.1	1	MAIKKVLKIIILAI III I I SCQLPLNQKTVYASPNSPKDN TWIQAASLTWLM DMSSLLYQLI STRIP SFASPNGLYMREQTI	80
ADC91856.1		-----	
Mtx1 NHA15b	1	MAIKKVIKIIISLII III I I SCQLPLNQKTVYASPNPPKDN TWIQAASLTWLM DMSSLLYQLI STRIP SFASPNGLHMREQTI	80
AAN04906.1	81	DSNTGQIQIDNEHRLLRWDRRPPNDIFLNGFI PRVTNQNLSPVEDTHLLNYLRTNSPSIFVSTTRARYNNLGLG EITPWP T	160
BAD01571.1	81	DSNTGQIQIDNEHRLLRWDRRPPNDIFLNGFI PRVTNQNLSPVEDTHLLNYLRTNSPSIFVSTTRARYNNLGLG EITPWP T	160
ADC91856.1		-----	
Mtx1 NHA15b	81	DSNTGQIQIDNEHSANNNI IYRYEIFAPGGIDINASESRNHNPFPNEDEITFPGGIRPEFIRSTY EYHNGEIVRIWINPNFINPSTLNDVSG	160
AAN04906.1	161	HSANNNI IYRYEIFAPGGIDINASESRNHNPFPNEDEITFPGGIRPEFIRSTY EYHNGEIVRIWINPNFINPSTLNDVSG	240
BAD01571.1	161	HSANNNI IYRYEIFAPGGIDINASESRNHNPFPNEDEITFPGGIRPEFIRSTY EYHNGEIVRIWINPNFINPSTLNDVSG	240
ADC91856.1	1	-----MNFFGRFSKKYVQEVNWIENNSYR--FPEPSACNHTDFEPIRP	41
Mtx1 NHA15b	161	YSANNNI IYRYEIFAPGGIDINASESRNHNPFPNEDEITFPGGIRPEFIRSTY EYHNGEIVRIWINPNFINPSTLNDVSG	240
AAN04906.1	241	PSNISKVFWHENHSEG---NNMDSKGFILDLDYNQDFDMFAPNGEIPNNLLNNSLNVIQNSEYQIKNKKDRNIVVTLD	317
BAD01571.1	241	PSNISKVFWHENHSEG---NNMDSKGFILDLDYNQDFDMFAPNGEIPNNLLNNSLNVIQNSEYQIKNKKDRNIVVTLD	317
ADC91856.1	42	YIADKRIVWIGENSHGvaqNNMLKAKLIAFLHQELDFKV VAFESGLSECYSVNG-----LKGRLDAEEMM---	106
tx1 NHA15b	241	SSNISKVFWHENHSEG---NNMDSKGFILDLDYNQDFDMFAPNGEIPNNLLNNSLNVIQNSEYQIKNKKDRNIVVTLD	317
AAN04906.1	318	SDYGGSPVESYKNFGFENQKWN IKYDSKKNAYKIYNRETPTLLL SWNSNSSNGEQVIRGYTESGSNNQYWTIEKNVNGFY	397
BAD01571.1	318	SDYGGSPVESYKNFGFENQKWN IKYDSKKNAYKIYNRETPTLLL SWNSNSSNGEQVIRGYTESGSNNQYWTIEKNVNGFY	397
ADC91856.1	107	-----KQSI FSLWRTEETLPLFQLLKSTDLTLAGFDFQPSATVHPLREMLQRQDGLGIDTIEE----LH	166
Mtx1 NHA15b	318	SDYGGSPVESYKNFGFENQKWN IKYDSKKNAYKIYNRETPTLLL SWNSNSSNGEQVIRGYTESGSNNQYWKIEKNLNGSY	397
AAN04906.1	398	KFRNLSDP SKILDLKDGNTLNKTPLVVSSSENSSSSQEWLIEKTNYQTVKDGTYQVSSKLNENKVIEQIST-NKVHIFSN S	476
BAD01571.1	398	KFRNLSDP SKILDLKDGNTLNKTPLVVSSSENSSSSQEWLIEKTNYQTVKDGTYQVSSKLNENKVIEQIST-NKIHIFSN S	476
ADC91856.1	167	QLAEYSNQWYYRIGKFRANRKRI PKELLMEFEDSKAEKLR TIVQLRSALESYPKNQVLLMLGRFLDNTAI--FLHCLACS	244
Mtx1 NHA15b	398	KFRNLSDP SKVLDLKDGNTLNKTPLVVSSEKSSSQEWIIEKTNYQTVKDGTYQLASKLNENKVIEHFPTnNKIHIFSN S	477
AAN04906.1	477	DKENQVWNLIYNPILKAYKIKSLKYPNYSLAWDSNNTRTIVAATGDYNDQYWLIERNEDNTYIIRNYENRKIVLDLSNGS	556
BAD01571.1	477	DKENQVWNLIYNPILKAYKIKSLKYPNYSLAWDSNNTRTIVAATGDYNDQYWLIERNEDNTYIIRNYENRKIVLDLSNGS	556
ADC91856.1	245	DR-----KYGKYRDQVMADNLEWLLT-----	265
Mtx1 NHA15b	478	DKENQVWNLIYNPVLKAYKIKSLKYPNYSLAWDSNNTRTIVAATGDYNDQYWLIERNEDNTYIIRNYENRKIVLDLSNGS	557

AAN04906.1	557	TTDGNGLLGFEFHGGINQRWIIKPFSSFNISIQDGIYQFMTVINQDLIADLTTNNYTIATKTNNYSSNQKWTVTYNDKKRAY	636
BAD01571.1	557	TTDGNGLLGFEFHGGINQRWIIKPFSSFNISIQDGIYQFMTVINQDLIADLTTNNYTIATKTNNYSSNQKWTVTYNDKKRAY	636
ADC91856.1		-----	
Mtx1 NHA15b	558	TTDGNGLLGFEFHGGINQRWIIKPFSSFNISIPDGIYQFMTVINQDLIADLTTNNYTIATKINSYSSNQKWTVTYNDKKRAY	637
AAN04906.1	637	KIRNLQHAHLSLAWDSNHSKIFGATGDYDDQYWIPILQTDGSFIFRNYKNPNKIFGTNGQPINDIPLKAQDVTGQNNQK	716
BAD01571.1	637	KIRNLQHAHLSLAWDSNHSKIFGATGDYDDQYWIPILQTDGSFIFRNYKNPNKIFGTNGQPINDIPLKAQDVTGQNNQK	716
ADC91856.1		-----	
Mtx1 NHA15b	638	KIRNLQHAHLSLAWDSNHSKIFGATGDYDDQYWIPILQTDGSFIFRNYKNPNKIFGTNSQPINDIPLKTQDLTGQSNQK	717
AAN04906.1	717	WYLRHLNSSNNFTGYFNISSKKNFNKIITMNSNKTQAVIFDNIGINNQSWKLYNDNKNAYQIHIHLDNFLYFQGGHNIVA	796
BAD01571.1	717	WYLRHLNSSNNFTGYFNISSKKNFNKIITMNSNKTQAVIFDNIGINNQSWKLYNDNKNAYQIHIHLDNFLYFQGGHNIVA	796
ADC91856.1		-----	
Mtx1 NHA15b	718	WYLRHLNSSNNLKGYFTISSKKNFNKTIITMNSNKTQAVIFDNIGIINQSWKLYNDNKNAYQIHIHSDNFLYFQGGHNIVA	797
AAN04906.1	797	TMRNVTNDDLRSYWYVEYNFNKDGFIIRNAFDTSYVLDVDFQGNFANNTPIITYQNYLNDNQLWNFIPSLGVEPR	870
BAD01571.1	797	TMQNVTNDDLRSYWYVEYNFNKDGFIIRNAFDTSYVLDVDFQGNFANNTPIITYQNYLNDNQLWNFIPSLGVEPR	870
ADC91856.1		-----	
Mtx1 NHA15b	798	TMQNVTNDDLRSYWYVEYNFNKGGFIIRSAFDTSYVLDVDFQGNFANNTPIITYQNYLNDNQLWYFIPSSGVEPR	871

Figure 6.6. Alignment of Mtx1 of *L. sphaericus* strain NHA15b protein with related proteins.

The protein is shown with the equivalent toxins from *L. sphaericus* LP1-G (accession number: AAN04906.1); *L. sphaericus* 2297 (BAD01571.1); erythromycin esterase-like protein, partial, from *Paenibacillus popilliae* (ADC91856); and Mtx1 NHA15b. Alignment was produced using Cobalt (<http://www.st-va.ncbi.nlm.nih.gov>). Colouring scheme taken from the program as follows: red for conserved residues, blue for columns with no gaps. Grey is for columns containing gaps. Where less than 50% of the sequences contain gaps, they are shown in grey uppercase, greater than 50% will be grey lowercase. The green highlight denotes the active core of Mtx1 protein. The yellow highlight marks the amino acid differences in the active core among the proteins.

6.4.2. Mtx2 toxin

Mtx2 and Mtx3 toxins are related to each other and to other beta-pore forming toxins such as epsilon toxin from *Clostridium perfringens*. Mtx2 is a 32 kDa protein and shows some amino acid sequence variation between strains (Chan et al., 1996) , meanwhile the Mtx3 is a 36 kDa protein that is highly conserved across several strains (Liu et al., 1996).

The *mtx2* gene of NHA15b is located in the genome at BSPH_RS05295 CDS. It is known that the Mtx2 toxin is the most diverse Mtx in *L. sphaericus* and that variations in the amino acid sequence determine its host range (Chan et al., 1996) . Amino acid position 224 is critical in this: if it is Lysine, as in the Mtx2 toxin from strain SSII-1, the toxin is more potent to *C. quinquefasciatus*, whereas if the amino acid in this position is Threonine, then the toxin is more potent to *A. aegypti* (Chan et al., 1996). Amino acid sequence analysis of the protein encoded in NHA15b showed that the entire Mtx2 sequence in this strain is the same as the Mtx2 of strain IAB59. Furthermore, since the amino acid at position 224 is Threonine, it is predicted that the Mtx2 of NHA15b would be more toxic to *A. aegypti* mosquitoes than to *C. quinquefasciatus*. The Mtx2 protein alignment is presented in Figure 6.8.

AAC44127.1	1	MKRTKLLFYIMIASFLFVNGSIYTAKATTIDENNHDIIKQGVSIEDIDKIDNMIASIPPLFGFLYSRFPYIFGESVD	80
AAC44126.1	1	MKRTKLLFYIMIASFLFVNGSIYTAKATTIDENNHDIIKQGVSIEDIDKIDNMIASIPPLFGFLYSRFPYIFGESVD	80
AAC44125.1	1	MKRTKLLFYIMIASFLFVNGSIYTAKATTIDENNHDIIKQGVSIEDIDKIDNMIASIPPLFGFLYSRFPYIFGESVD	80
AAC44124.1	1	MKRTKLLFYIMIASFLFVNGSIYTAKATTIDENNHDIIKQGVSIEDIDKIDNMIASIPPLFGFLYSRFPYIFGESVD	80
AAC44120.1	1	MKRTKLLFYIMIASFLFVNGSIYTAKATTIDENNHDIIKQGVSIEDIDKIDNMIASIPPLFGFLYSRFPYIFGESVD	80
AAC44105.1	1	MKRTKLLFYIMIASFLFVNGSIYTAKATTIDENNHDIIKQGVSIEDIDKIDNMIASIPPLFGFLYSRFPYIFGESVD	80
MTX2 NHA15b	1	-----MIASFLFVNGSIYTAKATTIDENNHDIIKQGVSIEDIDKIDNMIASIPPLFGFLYSRFPYIFGESVD	70
AAC44127.1	81	VSGINIENNTVTSVVPLFIGSNTFENTTDRMTFNTVVSFSKSIDSTTTQTLNGFKTAFEASGKVGIPLVAEGQIKTTLE	160
AAC44126.1	81	VSGINIENNTVTSVVPLFIGSNTFENTTDRMTFNTVVSFSKSIDSTTTQTLNGFKTAFEASGKVGIPLVAEGQIKTTLE	160
AAC44125.1	81	VSGINIENNTVTSVVPLFIGSNTFENTTDRMTFNTVVSFSKSIDSTTTQTLNGFKTAFEASGKVGIPLVAEGQIKTTLE	160
AAC44124.1	81	VSGINIENNTVTSVVPLFIGSNTFENTTDRMTFNTVVSFSKSIDSTTTQTLNGFKTAFEASGKVGIPLVAEGQIKTTLE	160
AAC44120.1	81	VSGINIENNTVTSVVPLFIGSNTFENTTDRMTFNTVVSFSKSIDSTTTQTLNGFKTAFEASGKVGIPLVAEGQIKTTLE	160
AAC44105.1	81	VSGINIENNTVTSVVPLFIGSNTFENTTDRMTFNTVVSFSKSIDSTTTQTLNGFKTAFEASGKVGIPLVAEGQIKTTLE	160
MTX2 NHA15b	71	VSGINIENNTVTSVVPLFIGSNTFENTTDRMTFNTVVSFSKSIDSTTTQTLNGFKTAFEASGKVGIPLVAEGQIKTTLE	150
AAC44127.1	161	YNFSHTNSNTKSVTTTTYTVPPQPIPVPPHTKTRTDVYLNQVSIISGNVEIYADAITGIKAESSGTVISIGDGLNLSNTYG	240
AAC44126.1	161	YNFSHTNSNTKSVTTTTYTVPPQPIPVPPHTKTRTDVYLNQVSIISGNVEIYADAITGIKAESSGTVISIGDGLNLSNTYG	240
AAC44125.1	161	YNFSHTNSNTKSVTTTTYTVPPQPIPVPPHTKTRTDVYLNQVSIISGNVEIYADAITGIKAESSGTVISIGDGLNLSNTYG	240
AAC44124.1	161	YNFSHTNSNTKSVTTTTYTVPPQPIPVPPHTKTRTDVYLNQVSIISGNVEIYADAITGIKAESSGTVISIGDGLNLSNTYG	240
AAC44120.1	161	YNFSHTNSNTKSVTTTTYTVPPQPIPVPPHTKTRTDVYLNQVSIISGNVEIYADAITGIKAESSGTVISIGDGLNLSNTYG	240
AAC44105.1	161	YNFSHTNSNTKSVTTTTYTVPPQPIPVPPHTKTRTDVYLNQVSIISGNVEIYADAITGIKAESSGTVISIGDGLNLSNTYG	240
MTX2 NHA15b	151	YNFSHTNSNTKSVTTTTYTVPPQPIPVPPHTKTRTDVYLNQVSIISGNVEIYADAITGIKAESSGTVISIGDGLNLSNTYG	230

AAC44127.1	241	LIRSPQDPDRVRAIGSGKFNLHGADFTAITYDITSGEASARIIDVKEISFK	292
AAC44126.1	241	LIRSPQDPDRVRAIGSGKFNLHGADFTAITYDITSGEASARIIDVKEISFK	292
AAC44125.1	241	LIRSPQDPDRVRAIGSGKFNLHGADFTAITYDITSGEASARIIDVKEISFK	292
AAC44124.1	241	LIRSPQDPDRVRAIGSGKFNLHGADFTAITYDITSGEASARIIDVKEISFK	292
AAC44120.1	241	LIRSPQDPDRVRAIGSGKFNLHGADFTAITYDITSGEASARIIDVKEISFK	292
AAC44105.1	241	LIRSPQDPDRVRAIGSGKFNLHGADFTAITYDITSGEASARIIDVKEISFK	292
MTX2 NHA15b	231	LIRSPQDPDRVRAIGSGKFNLHGADFTAITYDITSGEASARIIDVKEISFK	282

Figure 6.8. Alignment of Mtx2 of *L. sphaericus* strain NHA15b protein sequence with proteins sequences from *L. sphaericus* strain IAB59 (accession number: AAC44127.1); 2362 (AAC44126.1); 31-2 (AAC44125.1); 2297(AAC44124.1); Kellen Q (AAC44120.1); SSII-1(AAC44105.1). Alignment was produced using Cobalt (<http://www.st-va.ncbi.nlm.nih.gov>). Colouring scheme taken from the program as follows: red for conserved residues, blue for columns with no gaps. Grey is for columns containing gaps. Where less 50% of the sequence contains gaps, they are shown in grey uppercase, greater than 50% will be grey lowercase. The highlighted amino acids showed differences among the Mtx2 proteins.

6.4.3. Mtx3

The *mtx3* gene of strain NHA15b is located in the genome at BSPH_RS13250 CDS and encodes a protein of 326 amino acids. It is known that the Mtx3 toxin is much conserved. Among nine *mtx3* genes sequenced, it was found that most of the encoded protein sequences were the same except amino acid number 45, which in the strains 2297, Kellen Q and IAB59 is Isoleucine but in strains SSII-1, 1593, 2362, 1691, 2317.3, 31-2 is Valine (Liu et al., 1996). In the NHA15b strain, the amino acid in this position is Isoleucine, and in C3-41 strain, the amino acid is Valine. No studies investigating the effect of this conservative amino acid replacement have been published and, as a result, it is unknown whether this difference affects the expression or the activity of the Mtx3 toxin.

6.4.4. Mtx4 toxin

The *mtx4* gene is the most recently predicted toxin gene that belongs to the *mtx2* family. The gene was previously reported in the genome of strain C3-41 and is located downstream of the binary toxins in the 35 kb duplicated region in both the chromosome and the plasmid pBsph (Hu et al., 2008a). In contrast, in the strain NHA15b, the gene (BASYS00030), was found downstream of *cry48/cry49* genes (BSYS00014; BASYS00009), and directly upstream of the gene encoding tyrocidine synthase 3 (BASYS00031 –an enzyme that may be involved in the synthesis of a cyclic decapeptide antibiotic). The *mtx4* of strain NHA15b, like the *mtx4* gene from C3-41, encodes a protein of 337 amino acids. The amino acid sequences differ by only a few residues, most of them conservative changes. It seems that the Mtx4 toxin occurs in a wide range of bacteria. Blast search on the NCBI database showed that it is present in bacteria such as *Listeria grandensis* FSL F6-971 (EUJ24953.1) and *Paenibacillus dendritiformis* (WP_006675070.1). A gene alignment of the *mtx4* of strain NHA15b and the other *mtx4* genes from the NCBI database is presented in Figure 6.9. The amino acid differences among the protein sequences are highlighted in green.

ACA40704.1	1	MEGERDMRSKTLRNALIIGMVFITLGGTTTIP	PSQTYAESNSMQEKNIGITNVNQVLEKIGS	YYYQRNLSL	TWYEAPNSI	80		
ACA42387.1	1	MEGERDMRSKTLRNALIIGMVFITLGGTTTIP	PSQTYAESNSMQEKNIGITNVNQVLEKIGS	YYYQRNLSL	TWYEAPNSI	80		
EWH32665.1	1	MEGERDMRSKTLRNALIIGMVFITLGGTTTIP	PSQTYAESNSMQEKNIGITNVNQVLEKIGS	YYYQRNLSL	TWYEAPNSI	80		
EUJ24953.1	1	-----MKKLIILLFVFCG---YTIGITSFSN	TASASSELT	PA	TK-----IDWIGNTYYKNVL-----EHGNIK	56		
WP_006675070.1	1	-----MRKHVLRNGLVIGLSALMLGSASV--	PPQAYAEASIAQEK	SIGIVDVNEVLK	KGTYYYKNHLANT---VANPE	69		
Mtx4 NHA15b	1	LEGESDMRSKTLRNALIIGMVFITLGGTTTIP	PSQTYAESNSMQEKNIGITNVNQVLEKIGS	YYYQRNLSL	TWYEAPNSI	80		
ACA40704.1	81	GVNHDFQLRKNPDISSEVDLSISGEEIESL	YDSTIPKFLGENFENNTNQEQT	YNTSKFSE	TYTESTSTSVSKGFKIN	160		
ACA42387.1	81	GVNHDFQLRKNPDISSEVDLSISGEEIESL	YDSTIPKFLGENFENNTNQEQT	YNTSKFSE	TYTESTSTSVSKGFKIN	160		
EWH32665.1	81	GVNHDFQLRKNPDISSEVDLSISGEEIESL	YDSTIPKFLGENFENNTNQEQT	YNTSKFSE	TYTESTSTSVSKGFKIN	160		
EUJ24953.1	57	RLKYN-TFNTDLFNVSGTKNGEPIRTDIEPI	YFGETSKGLV-----NNTDEEQTF	SVPSFSK	SVAYSKSTSISKNASFG	129		
WP_006675070.1	70	SNGFRYTLRANPDSVTPVIDLSIEART-ENI	HYDSTMPVFIGENTFPNTLDEEQTF	NLKYSK	TYTESNSTSMSSGFKVG	148		
Mtx4 NHA15b	81	GVNHDFQLRKNPDISSEVDLSISGEEIESL	YDSTIPKFLGENFENNTDQEQT	YNTSKFSE	TYTESTSTSVSKGFKIN	160		
ACA40704.1	161	VGRD----FIIPLILNEGKINLEYNSGSTNT	NLTKTYTLEAPSQPVKVP	PNKIYKAV	VEYSQRTYKGT	VKFYGRNTHN	236	
ACA42387.1	161	VGRD----FIIPLILNEGKINLEYNSGSTNT	NLTKTYTLEAPSQPVKVP	PNKIYKAV	VEYSQRTYKGT	VKFYGRNTHN	236	
EWH32665.1	161	VGRD----FIIPLILNEGKINLEYNSGSTNT	NLTKTYTLEAPSQPVKVP	PNKIYKAV	VEYSQRTYKGT	VKFYGRNTHN	236	
EUJ24953.1	130	LGKVe-----IPLL----NELSATVNTS	STDNTETETVTVTSPSQSVKVP	PAHKT	YGVKYYLNRIKYTGQ	FKLKASTWRN	200	
WP_006675070.1	149	GAGDlgnkFILPLILEGAGKLNLEYNTG	STETHTTSESYTLEVPSQPVKVP	PHKTYKT	IVEFSQRTYRGDVVFT	GRGTGST	228	
Mtx4 NHA15b	161	VGRD----FIIPLILNEGKINLEYNSGSTNT	NLTKTYTLEAPSQPVKVP	PNKIYKAV	VEYSQRTYKGT	VKFYGRNTHN	236	
ACA40704.1	237	PPINTIKT----TGSYTGWMGMQEIKQFT	FNDPLYKHYDGLSDSQKKEVEN-----	DGVVIEPFLGATE	VVVEGKGSF	306		
ACA42387.1	237	PPINTIKT----TGSYTGWMGMQEIKQFT	FNDPLYKHYDGLSDSQKKEVEN-----	DGVVIEPFLGATE	VVVEGKGSF	306		
EWH32665.1	237	PPINTIKT----TGSYTGWMGMQEIKQFT	FNDPLYKHYDGLSDSQKKEVEN-----	DGVVIEPFLGATE	VVVEGKGSF	306		
EUJ24953.1	201	YEEITATVVrrvpTGHHTGY--LDEKITY	PIDYAYKNTANFKENG	VNYSSNstsled	AQEKYKNGTLR	TTLTSL	LEGDGEY	278
WP_006675070.1	229	MANRSIINA----TGIYTGWMGMQDSKK	FTYDNFLGRHYAGLTASQQQEIK	agvqit	TSGYGAAANAY	SSVEVTGKGSF	304	
Mtx4 NHA15b	237	PPINTIKT----TGSYTGWMGMQEIKQFT	FNDPLYKHYDGLSDSQKKEVEN-----	DGVVIEPFLGATE	VVVEGKGSF	306		

ACA40704.1	307	EGVYGAKLNVKTYD	TD	QNKV	KLVD	SR	SIDL	---	337	
ACA42387.1	307	EGVYGAKLNVKTYD	TD	QNKV	KLVD	SR	SIDL	---	337	
EWH32665.1	307	EGVYGAKLNVKTYD	TD	QNKV	KLVD	SR	SIDL	---	337	
EUJ24953.1	279	TIENGTDFHVVIYDMSN	-	NRTIKQYDKS	IIE	----			308	
WP_006675070.1	305	SGVIGAE	LIVTTTYD	VTD	aKN	-	ARLVNQR	TVP	Idld	338
Mtx4 NHA15b	307	EGVYGAKLNVKTYD	TD	QNKV	KLVD	SR	SIDL	---	337	

Figure 6.9. Alignment of Mtx4 protein of *L. sphaericus* strain NHA15b with related proteins.

The protein is shown with the equivalent Mtx4 toxin from the genome of *L. sphaericus C3-41* (accession number: ACA40704.1); from *L. sphaericus C3-41* plasmid pBsph (ACA42387.1); *L. sphaericus* strain CBAM5 (EWH32665.1); Mtx2/3 toxin-like protein from *Listeria grandensis* FSL F6-971 (EUJ24953.1); Mtx2/3 toxin-like protein *Paenibacillus dendritiformis* (WP_006675070.1); and Mtx4 of *L. sphaericus* NHA15b. Alignment was produced using Cobalt (<http://www.st-va.ncbi.nlm.nih.gov>). Colouring scheme taken from the program as follows: red for conserved residues, blue for columns with no gaps. Grey is for columns containing gaps. Where less 50% of the sequences contain gaps, they are shown in grey uppercase, greater than 50% will be grey lowercase. 6.3.5.

6.4.5. *Cry48/49* toxins and predicted location of the genes

The annotation transfer (RATT) and Basys annotation both showed that the *cry48/49* toxin genes were located on the contigs no. 81 (22.480 bp) of the NHA15b genome. Since the genes are not present in the genomes of other currently sequenced strains, the actual location of the genes is unknown (including whether the genes are located in the chromosome or in the plasmid). To date, the only report on the sequencing of *cry48/49* genes is from Jones *et al.* (2008) who sequenced a long fragment containing the *cry* genes and several other adjacent genes. To predict contigs adjoining the contig 81, a seeding process using IMAGE software was conducted using contig 81 as the seed. In the second round, then the contig 41 and 85 were also used as a seeding sequences, since the two were shown as the adjacent contigs of contig 81. It was shown that several contigs, i.e. contigs 23 (36.401 bp), 41 (9.694 bp), 60 (178.328 bp), 85 (770 bp), 103 (104.770 bp), and 113 (78.959 bp) were possible candidate contigs to flank contig 81, either on its 5' end or 3' end. To verify the contig arrangement in this fragment, a series of PCRs was conducted using several primers designed to produce amplicons if the contigs are linked. The results showed that the contig arrangement is 5'-contig 85 (reverse complemented)--contig 81--contig 41 (reverse complemented) --contig 60 (reverse complemented) -3'. The final arrangement of fragments containing the *cry* genes of strain NHA15b is 211,486 base pairs long (the sum of the length of contigs 41, 60, 81 and 85 plus gap filling sequences minus some overlapping sequence that occurs between contig 81 and 85). Despite the success in extending the region on which the *cry* genes are located, their localisation on the chromosome or a plasmid is still not proven. The problem is increased by the fact that the extended fragment that contains the *cry* genes and *mtx4* is flanked by transposase genes that exist in multiple locations in the genome. This situation makes it difficult to ascertain the precise location of the *cry* genes.

Annotation transfer of chromosome of strain C3-41 into Contig 60 using RATT software revealed that approximately 125,000 bp of its 178,234 bp long Contig 60 sequence is identical to the chromosomal region of C3-41 that harbour CDS BSPH_RS11385 to BSPH_RS12080, and

the rest of Contig 60 (approximately 53,000 bp) does not have any similarity to the C3-41 genome. The 53,000 bp side of the Contig 60 is the side that connected to the contigs forming the extended *cry* fragment. When the CDSs of the 53,000 bp region was analysed by blast (blastx) against the NCBI database it revealed that some the CDSs shared some degree of similarities with genes characteristic of plasmids, among others are: DNA primase [*Virgibacillus halodenitrificans*] Sequence ID: emb|CDQ37269.1; type IV secretion protein Rhs, partial [*Paenibacillus sanguinis*]; type IV secretory pathway, VirD4 protein [*Lysinibacillus sphaericus*] Sequence ID: ref|WP_031417586.1; type II/IV secretion system protein [*Lysinibacillus fusiformis* ZC1] Sequence ID: gb|EFI67494.1; type IV secretion system VirB4 protein [*Lysinibacillus sphaericus*] Sequence ID: ref|WP_031417600.1; and putative conjugation protein [*Bacillus subtilis*] Sequence ID: gb|KIU04550.1|. The presence of the above genes means that Contig 60 is likely to be a fragment of a plasmid, or, at least some part of it is a fragment of a plasmid that by some computational error during contig assembly joined into chromosomal sequences. Based on the indications presented, it is predicted that this represents a sixth plasmid in the NHA15b strain, hence named pLsph600.

Analysis of T4SS genes of the plasmids of NHA15b strain revealed that the *VirD4* gene of the plasmid *pLsph100* is similar to either the gene of pBsph-1 and pBsph-2, but differs considerably from the *VirD4* genes of pLsph500 and pLsph600. Meanwhile, the latter two plasmids have a considerable similarity in their *VirD4* genes. All of the *VirD4* genes of *L. sphaericus* differ from the reference gene. Comparison of the *VirD6* gene also produced a similar picture. The *VirB6* gene of pBsph-1 and pBsph-2 are similar to each other, but the gene was not found on the plasmid pLsph500 or on pLsph600. Protein alignments of the predicted products of the above genes (*VirD4*, *VirD6*, and *VirB6*) are presented in the Figures 6.10 ,6.11, and 6.12 respectively.

VirD4_ref		-----	
VirD4_pBsph_1	1	-LKQLKKKKEGKIENRFARLNRIAKIEEFIRIKFHLGAVTSLSFINVFILVPIALCMWLGAAILGNPSMMLMQDWSWGITIF	79
VirD4_pBsph_2	1	rLKQLKKKKEGKIENRFARLNRIAKIEEFIRIKFHLGAVTSLSFINVFILVPIALCMWLGAAILGNRSMMLMQDWSWGITIF	80
VirD4_pLsph100	1	-LKQLKKKKEGKIENRFARLNRIAKIEEFIRIKFHLGAVTSLSFINVFILVPIALCMWLGAAILGNRSMMLMQDWSWGITIF	79
VirD4_pLsph500		-----	
VirD4_pLsph600		-----	
VirD4_ref		-----M	
VirD4_pBsph_1	80	FWILPLLTWYYSTITIPYQKDHKYILDFNDTVRGRWISAIANVGALHVMTWAVFAKYVVEYYLGYLASIR-LDAAYKPFL	158
VirD4_pBsph_2	81	FWILPLLTWYYSTITIPYQKDHKYILDFNDTVRGRWISAIANVGALHVMTWAVFAKYVVEYYLGYLASIR-LDAAYKPFL	159
VirD4_pLsph100	80	FWILPLLTWYYSTITIPYQKDHKYILDFNDTVRGRWISAIANVGALHVMTWAVFAKYVVEYYLGYLASIR-LDAAYKPFL	158
VirD4_pLsph500	1	-----LLLKRRAYLLNIYGKEKEVYMYENKyLQRLSTPKM	35
VirD4_pLsph600	1	-----MNETKyLQRLAQPRG	15
VirD4_ref	2	NSSKTPQRLAVSIVCSLAAGFCaasLYVTFRHGFNGEAMMFSVFAFWYETPLYMGH----ATPVFYCGLAIVVSTSIV	77
VirD4_pBsph_1	159	LTDDLDSFILILFVLPALISGLM---IFIQVRDFMVHKDMLSESFMTWKAPIIRRFTHKKLLGTADIIIGYQFKTMPIIV	235
VirD4_pBsph_2	160	LTDDLDSFILILFVLPALISGLM---IFIQVRDFMVHKDMLSESFMTWKAPIIRRFTHKKLLGTADIIIGYQFKTMPIIV	236
VirD4_pLsph100	159	LTDDLDSFILILFVLPALISGLM---IFIQVRDFMVHKDMLSESFMTWKAPIIRRFTHKKLLGTADIIIGYQFKTMPIIV	235
VirD4_pLsph500	36	LIIFSLAYTVVMTLLANLTLAFIqa-VMITIEQFRIDPKSTSFSLD-WRSIFDFQES-----WGFFYLGFALLSVIGLF	107
VirD4_pLsph600	16	LLVFSIVFWSLMFFVSNFVLAFLqs-LLASFQQFQADPKSVSFSID-WRSFLEFQSN-----WIPFYIGFAILSTIGFV	87
VirD4_ref	78	V-----LLSQLIISFR--NHEHGTARWAGFGE-----MRHAGYLQRYNRIKG-PIFGKTCGPrwfgsYLTNGEQPH	141
VirD4_pBsph_1	236	LKEAQRFLHEAVIGAT--GSGKTSTALLLRQAQDLINIATGRRKMGLVFLEPKGDGVDDVLMCKK-----LKIPDEKIK	308
VirD4_pBsph_2	237	LKEAQRFLHEAVIGAT--GSGKTSTALLLRQAQDLINIATGRRKMGLVFLEPKGDGVDDVLMCKK-----LKIPDEKIK	309
VirD4_pLsph100	236	LKEAQRFLHEAVIGAT--GSGKTSTALLLRQAQDLINIATGRRKMGLVFLEPKGDGVDDVLMCKK-----LKIPDEKIK	308
VirD4_pLsph500	108	-----KFIYNIRMSYSdlNKGQHGTSEFEQVRE----LKKQYRIIPAVEKEYDGDGGVIIISALQEKgrpyrLLIDDGPIH	178
VirD4_pLsph600	88	-----KFIYIRLNFVdlNKGQHGTSEFEQPRE----LKKQYKIIPAAEKEYDGDGGVIIISALQEKgrpyrLLIDEGPIH	158
VirD4_ref	142	SLVVAPTRAGKG---VGVVIPTLLTFKG--SVIALDV----KGELFELTSRARK----AGGDAVFKFSP----LDPERRT	204
VirD4_pBsph_1	309	VIDATKAFTIKFNPFIGPSAPAAATFEGTINALSGDQDDFYKGQQNEAASTLTKLAKIAFGEKTNIFHIQRMFSDPRYLA	388
VirD4_pBsph_2	310	VIDATKAFTIKFNPFIGPSAPAAATFEGTVNALAGDQDEFFKGQQNEAASTYTKLAKIAFGEKTNIFHIQRMFTDPRYLA	389
VirD4_pLsph100	309	VIDATKAFTIKFNPFIGPSAPAAATFEGTINALSGDQDDFYKGQQNEAASTLTKLAKIAFGEKTNIFHIQRMFSDPRYLA	388
VirD4_pLsph500	179	TMAIGITRSKGGETFVFPMDVQSRANKPISLVINDP----KGELAAASYD----TLVARGYDVYVFNL----IQ-QYMG	245
VirD4_pLsph600	159	TMVIGITRSKGGETFVFPMDVLSRANKPISLVVNDP----KGELAAASYD----TLIARGYEVHVFNL----IQ-QYMG	225

VirD4_ref	205	HCYNPVLDAIALPPERQFTETRRLaan----LITAKGKGAEGFIDGARDL FVAGILTCIERGTP TIG-----AVYDLF	273
VirD4_pBsph_1	389	NIVESIREQITTNREKQFREQKDL-----IMLLDSLGT SVEHIEQTEHELMFKLKNLSNS SAINENQ-----KKLNEI	456
VirD4_pBsph_2	390	NIVESIREQITTNREKQFREQKEL-----NMILDSFGTNVEHIEQTEHELMFKLKNLSNS SAINENH-----SKLNEI	457
VirD4_pLsph100	389	NIVESIREQITTNREKQFREQKDL-----NMLLDSLGT SVEHIEQTEHELMFKLKNLSNS SAINENQ-----KKLNEI	456
VirD4_pLsph500	246	MGFNPLQLVVD AWKSGNTSLAQQYansiaF MLYDDPSAKDKFWSNSAKSLVTAI I LALTEDALAI GKddtinlySVANFL	325
VirD4_pLsph600	226	MGFNPLQVV IDAWKSGNTSLAQQYansvgF MLYNDPSAKDKFWSNSAQSLVTAI I LALTEDALLTGKedtinlySVANFL	305
VirD4_ref	274	AQPGEKYKLFAHLAEE SRNKEA-----QRIFDNMAGNDTKILTSYTSVLGDG-----GLNLWADP	328
VirD4_pBsph_1	457	QELKKRNNLLRAQQIEIDNTEQIIIFYFENEVLLYKVDQRTQQPI ILYPKNHIYANQQMVESK KDKFVTGAKKY LNDIALND	536
VirD4_pBsph_2	458	QELKKRNNLLRAQQIEVDNTEQIIIFYFENEVLLYKVDQRTQQPI ILYPKNHIYANQQMVESK KDKFVTGAKKY LNDIALND	537
VirD4_pLsph100	457	QELKKRNNLLRAQQIEIDNTEQIIIFYFENEVLLYKVDQRTQQPI ILYPKNHIYANQQMVESK KDKFVTGAKKY LNDIALND	536
VirD4_pLsph500	326	ATLGSND-----EETEENALDQFFKARDDI-----NPARMMYATSNFAAGNTRAGIFSTAMDKLQIF TLE	386
VirD4_pLsph600	306	STLGSDNS-----EEGD-NALDQFFQARDEI-----NPARMMYATSNFAAGNTRAGIFSSAMDKLQIF TLE	365
VirD4_ref	329	LVKAA-TSRSDFSVYDLRR--KRTC VYLCVSP-NDLEVva-plM RLLFQQVVSILQRS L---PGKDERHEVLFLLDEFK h	400
VirD4_pBsph_1	537	LLKNLFIGAEGEEVFDAD EFLREGGVLLVNTS LAELDE-----LSLMFGQFFIRQFQSAIFRRPQEGRIPIFFYID EFP-	610
VirD4_pBsph_2	538	LLKNLFIGADGEEVFDAD EFLREGGVLLVNTS LAELDE-----LSLMFGQFFIRQFQSAIFRRPQEGRIPIFFYID EFP-	611
VirD4_pLsph100	537	LLKNLFIGAEGEEVFDAD EFLREGGVLLVNTS LAELDE-----LSLMFGQFFIRQFQSAIFRRPQEGRIPIFFYID EFP-	610
VirD4_pLsph500	387	PNAKL-TSYNSLDLTDVGFGEKPIAVFM-VTP--DFDEsnhvlASMFVSQLYRVNAEKATMSPSGKMKRQVHMLLDEF G n	462
VirD4_pLsph600	366	PNAKL-TSYNSLDLTEIGFGDKPIAVFM-VTP--DYDKsnhvlASIFVSQLYRVNAEKATMSPSGKMKRHVHALLDEF G n	441
VirD4_ref	401	LGKLEAIETAITTIAGYKGRFMFIIQSLSALTGIYddAGKQN-FLSNTGVQVFMATADDETP TYISKAIGDYTFKARSTS	479
VirD4_pBsph_1	611	LYVNEAFERILTLGRSYNVGAVIAMQSIGQLEGVK--AGYQDIILGNASSKTVFGRGPNKDNEYFSLEFGEKEINEESLN	688
VirD4_pBsph_2	612	LYVNEAFERILTLGRSYNVGAVIAMQSIGQLEGVK--AGYQDIILGNASSKTVFGRGPNKDNEYFSLEFGEKEINEESLN	689
VirD4_pLsph100	611	LYVNEAFERILTLGRSYNVGAVIAMQSIGQLEGVK--AGYQDIILGNASSKTVFGRGPNKDNEYFSLEFGEKEINEESLN	688
VirD4_pLsph500	463	MPVIEGMAGMVTVGAGRGFRHLIIQAYSQ LSKY--GDDADTIIGNCSNQIYILTKDKATAEQFSALIGTKTITDISRS	540
VirD4_pLsph600	442	MPVVDGMDGMVTVGAGRGFRHLIIQAYSQMKSLY--GDDADTIIGNCSNQIYILTKDKATAEQFSALIGTKTITDVRS	519
VirD4_ref	480	-----YSQARMFHDHNIQISDQGAPLLRPEQVRLLEDNNEIVL I Kgh-----PPLK-----LRKVRYYS DRMLRRL	539
VirD4_pBsph_1	689	ESASPMTSEDQKWGYRLNTAKKKVARFSTTAIRELPFKHMIVQIV----DETNSIAPPLKAVGRFVN EARFIKPYLNLKK	764
VirD4_pBsph_2	690	ESASPMTSEDQKWGYRLNTAKKKVARFSTTAIRELPFKHMIVQIV----DETNSIAPPLKAVGRFVN EARFIKPYLKLKK	765
VirD4_pLsph100	689	ESASPMTSEDQKWGYRLNTAKKKVARFSTTAIRELPFKHMIVQIV----DETNSIAPPLKAVGRFVN EARFIKPYLKLKK	764
VirD4_pLsph500	541	-GKLLSTD KSH-----SESVKERALLMPDELMQLKEGESVVVRAnkrqDNKRKKI VPKPIFNKGVTAAKFRYEYLSQYF	613
VirD4_pLsph600	520	-GKLLSTD KSH-----SESVKERALLMPDELMQLKEGESVVVRAnkrqDNERKKI VPKPIFNTGVTAAKFRWEYLAEDF	592

VirD4_ref	540	FECQIGALPEP-----ASLMLSEGVHRDGDLSQqaavteaagglGDIDSIPNNMEAATPQNSEMDDEQDS-----L	605
VirD4_pBsph_1	765	SDIKSNQEKEV-----GGLVENAIKGFASKPEIKE-----GDIVHIPIESMPKLEDEVKPFDKSED----ELLEIL	825
VirD4_pBsph_2	766	SDIKSNQEKEV-----GGLVENAIKGFASKPEIKE-----GDIVHIPIESMPKLEDEVKPFDKSED----ELLEIL	826
VirD4_pLsph100	765	SDIKSNQEKEV-----GGLVENAIKGFASKPEIKE-----GDIVHIPIESMPKLEDEVKPFDKSED----ELLEIL	825
VirD4_pLsph500	614	DNSKSILSLPImsttyHNLDMNKMVFTSKNDNNQylkmlvvmgaEAFSKLLAMLRKEIQVD----NEIEDwayiHLLSFL	689
VirD4_pLsph600	593	DNSKSILSLPImnsqyHNLKMSDMVFTSKSERDLyvcmslamsaAEFSKLTRTLYQVLRTNDDYVDEIPTwsflQFFSYL	672
VirD4_ref	606	PTGIDVPQGLIESDEVKEDAGGvvPDFGVSAEMAPAMIAQQQLLEQIIALQQRygPASSHSVK-----	668
VirD4_pBsph_1	826	KYVEEVVEEFIEQKTIAQQENSW--EQILYSSDNKEDVSSVATNENKVELEPIH-IESTDEIKVSKITKEGEQNIENLMK	902
VirD4_pBsph_2	827	KYVEEVVEEFIEQKTIAQQENSW--EQILYSSDNKEDVSSVATTENKVELEPIH-IESTDEIKVSKITKEGEQNIENLMK	903
VirD4_pLsph100	826	KYVEEVVEEFIEQKTIAQQENSW--EQILYSSDNKEDVSSVATTENKVELEPIH-IESTDEIKVSKITKEGEQNIENLMK	902
VirD4_pLsph500	690	AYESDISRSLFEKSILP LLEKhlpKSVIEEWKEEIKHQRDERKPGVFANI ENTN-TEEDIQEKMKRLITEG---EDN---	762
VirD4_pLsph600	673	TYGSSIDSVVVERRIFPDLKKYlpNKVLRWTDILREQRAKTMLD-----ENVA-TQDEKITTMQKLIMEGSLKEDV---	743
VirD4_ref		-----	
VirD4_pBsph_1	903	ELHEAVEGSGTGNKQVYNPLDSEELVVDYLPPEENEELPPLEHAYLAALPTEVNHINKQINEPQTKKTVADSNSIEKPA	982
VirD4_pBsph_2	904	ELHEAVEGSGTGNKKVYNPLDSEELVINDYLPPEENEELPPLEHAYLAALPTEVHHINKQINEPQTKKTVADSNSIEKPA	983
VirD4_pLsph100	903	ELHEAVEGSGTGNKKVYNPLDSEELVINDYLPPEENEELPPLEHAYLAALPTEVHHINKQINEPQTKKTVADSNSIEKPA	982
VirD4_pLsph500		-----	
VirD4_pLsph600		-----	
VirD4_ref		-----	
VirD4_pBsph_1	983	VQPYVTMGTMIEDDF	997
VirD4_pBsph_2	984	VQPYVTMGTMIEDDF	998
VirD4_pLsph100	983	VQPYVTMGTMIEDDF	997
VirD4_pLsph500		-----	
VirD4_pLsph600		-----	

Figure 6.10. Alignment of VirD4 proteins encoded by plasmids **pLsph_100**, **pLsph_500** and **pLsph_600** *L. sphaericus* strain NHA15b protein sequences with related sequences .

The sequences are shown with the genes VirD4_ref: reference protein type IV secretion system protein VirD4 (plasmid) *Agrobacterium fabrum* str. C58 (NP_396505.2); VirD4_pBsph_1: VirD4 from *L. sphaericus* strain C3-41 pBsph plasmid (NC_010381.1); VirD4_pBsph_2: VirD4 from strain SSII-1 pBsph2 plasmid (WGS: JPDK01). Alignment was done using Cobalt (<http://www.st-va.ncbi.nlm.nih.gov>). Colouring scheme taken from the program as follows: red for conserved residues, blue for columns with no gaps. Grey is for columns containing gaps. Where less 50% of the sequences contain gaps, they are shown in grey uppercase, greater than 50% will be grey lowercase.

VirB4_ref	1	mstlfkgltrpalirglgvplypflgmciicvllgwiheamytlilpgwyairrvtqfderffdllylrtlvkgyplsn	80
VirB4_pBsph_1		-----	
VirB4_pBsph_2		-----	
VirB4_pLsph_100		-----	
VirB4_pLsph_500		-----	
VirB4_pLsph_600		-----	
VirB4_ref	81	krfsavhyagsqynevdiskvndfmklkdqssveelipysshitdniivtknrllatwqidgayfecvdsedlsilt dq	160
VirB4_pBsph_1		-----	
VirB4_pBsph_2		-----	
VirB4_pLsph_100		-----	
VirB4_pLsph_500		-----	
VirB4_pLsph_600		-----	
VirB4_ref	161	lntlirsfegksvtlyphrirckkgvrpvfnskipfvnrvmndyyesfpqseffenkflfticfkpfttedkvthffsrs	240
VirB4_pBsph_1		-----	
VirB4_pBsph_2		-----	
VirB4_pLsph_100		-----	
VirB4_pLsph_500		-----	
VirB4_pLsph_600		-----	
VirB4_ref	241	kkqkdifkepvnemneicdrln tylsr fhsrrlglyedhGVVYSDQLSLFQYLLSGRWQKVRVSSSPFYTYLGGK-DLFF	319
VirB4_pBsph_1	1	-----VINE--GIEF	8
VirB4_pBsph_2	1	-----illgklpsFPILRSPFKRKKVIVPNRHESEFSFPTLIDNVINE--GIEF	48
VirB4_pLsph_100	1	-----VINE--GIEF	8
VirB4_pLsph_500	1	-----MKLLEKIKGNNKQESSTNKESKKKYNAAF LAIQPQgGVIF	41
VirB4_pLsph_600	1	-----tMKLLDSVMKFKQPDKSKREK---EHNPTFLAIQPPGILF	39
VirB4_ref	320	GNDAGQITASDHARYFR--CIEIKDYFQETDAGILDALMYLPvEYVVTSSFTAMDKQSAIKALDDQIDKLEMTDDAAKS	396
VirB4_pBsph_1	9	HSDYFIIQIGIGTKKYGrsfYIKPSGYPRKVRIGWLVLGFKAE-DINVSVHSKPFDRTDAVRQLKNKLDDYETVVISAQK	87
VirB4_pBsph_2	49	HSDYFIIQIGIGTKKYGrsfYIKPSGYPRKVRIGWLVLGFKAE-DINVSVHSKPFDRTDAVRQLKNKLDDYETVVISAQK	127
VirB4_pLsph_100	9	HSDYFIIQIGIGTKKYGrsfYIKPSGYPRKVRIGWLVLGFKAE-DINVSVHSKPFDRTDAVRQLKNKLDDYETVVISAQK	87
VirB4_pLsph_500	42	KDKY--IQKGDGYEAC-----IQIWDYPSSGSELWLEKIYSMR-DVIVVEDNATMPKSEAVSTINKSMLEQDVRYRSAKN	113
VirB4_pLsph_600	40	KDKY--IQKGDGYEAC-----IHIWDYPTAVSKLWLEKIFSMR-DVISTMDIATMPKDETVSAINKSMVEQTVRYSSGKQ	111

VirB4_ref	397	LLA-----DLKVGDMVSSGYISFGKSHQTLVVFADSPERLVKDTNIVTSTLEDLGLIVTYSTLSLG----AA	460
VirB4_pBsph_1	88	QDNagKIEAiGQQVRDAKTLREIQNNMNSLHFTSISATIYADTLEELNEKSVAVERALAAESIEVVNAYDRQK----EG	163
VirB4_pBsph_2	128	QDNagKIDAiGQQVRDAKTLREIQNNMNSLHFTSISATIYADTLEELNEKSVAVERALAAESIEVVNAYDRQK----EG	203
VirB4_pLsph_100	88	QDNagKIEAiGQQVRDAKTLREIQNNMNSLHFTSISATIYADTLEELNEKSVAVERALAAESIEVVNAYDRQK----EG	163
VirB4_pLsph_500	114	KSE--EIDA-ENNYQELERLYKVISQMGEVRLIDLRLRFVHAKTIHELEQKIKTVESDLISYGYKGTVFLNEAKwewqSL	190
VirB4_pLsph_600	112	QSE--KMDA-QNSYQELLESYQGISQMGEVVKLVDIRLRFVHAKTMHELEQKVQTVETDLISYGFKGTVFLNESSwewqSL	188
VirB4_ref	461	YFAQLP-GNYTLRPRSLTSSLNFAEMESFHNFFSGKEKGNWGEKLitlrgSGNDIYHLNYHMTTEHQNFfgknptlgH	539
VirB4_pBsph_1	164	FLSTQPLGANLYLKNSERNLDENALAAMFPHNSSTLNHTGGMPIGV-Y-----NNEYIYFNNF--DSKLESF-----S	227
VirB4_pBsph_2	204	FLSTQPLGANLYLKNSERNLDENALAAMFPHNSSTLNHTGGMPIGV-Y-----NNEYIYFNNF--DSKLESF-----S	267
VirB4_pLsph_100	164	FLSTQPLGANLYLKNSERNLDENALAAMFPHNSSTLNHTGGMPIGV-Y-----NNEYIYFNNF--DSKLESF-----S	227
VirB4_pLsph_500	191	YLPYDEQQEFPNKREKGLHGLTLAAGLPYHFSELNDPTGSYLGTTF-----SGGNVVFDFLHRDKKRRYY-----N	257
VirB4_pLsph_600	189	FVPYEEQQTFPNKRESKGI PALSLAAGLPYHFSELNDQTSYLGTTF-----SGGNVIYDLFHRDKKRRYY-----N	255
VirB4_ref	540	TEILGTSNVGKTVLLMTKAFAAQQFGTPESFP----ANRKLKLLTT-----VFFDKDRAGEVGIRAMGGSYYRVKE--	606
VirB4_pBsph_1	228	MGIFGISGAGKGVFVKTVIGRGYLDGITKNII--LDVEPEYTELTY-ALGGIVIELRSDDVVSNF SRINPLDVYVEKEvk	304
VirB4_pBsph_2	268	MGIFGISGAGKGVFVKTVIGRGYLDGITKNII--LDVEPEYTELTY-ALGGIVIELRSDDVVSNF SRINPLDVYVEKEvk	344
VirB4_pLsph_100	228	MGIFGISGAGKGVFVKTVIGRGYLDGITKNII--LDVEPEYTELTY-ALGGIVIELRSDDVVSNF SRINPLDVYVEKEvk	304
VirB4_pLsph_500	258	AVAVGTMGAGKSTLLKKL---GLDNGCRGNFIrgFDVTGEFATLVMdSLNGQLISLDGTQGI-----INPLQIYRVQD--	327
VirB4_pLsph_600	256	AVAVGTMGAGKSTLLKKL---FTDNACRGNFIrgFDVTGEFETTVK-TLNGQLISLDGSQGI-----INPLHIYVRD--	324
VirB4_ref	607	-----GEPTGWNPAALPPTKRNI AFMKDMVRLLC----TLNSEPLDD-----YQNSLISDAVERL---MQRSDRSYP	666
VirB4_pBsph_1	305	rrytPFEYIVEKVNINEKVKEAIEFFKVMKESTYpkdpNLDPHELDALDKILKKLYAKCGITEDPESLYYHEEGMDEQGN	384
VirB4_pBsph_2	345	rrytPFEYIVEKVNINEKVKEAIEFFKVMKESTYpkdpNLDPHELDALDKILKKLYAKCGITEDPESLYYHEEGMDEQGN	424
VirB4_pLsph_100	305	rrytPFEYIVEKVNINEKVKEAIEFFKVMKESTYpkdpNLDPHELDALDKILKKLYAKCGITEDPESLYYHEEGMDEQGN	384
VirB4_pLsph_500	328	----GTDQKSEELSFMQHISKLATFYQFLAH-----DPATEEMEEFKRISRSFYESCGFIERIKTTGVTQLKNEEYPI	396
VirB4_pLsph_600	325	----GSDQKSEKLSFMQHLSKLATFYQFLTV-----NPSTEEMEEFKKITREFYHSLGFEDLIETTGVTQLKNDEYPI	393
VirB4_ref	667	ISKLRPLIQEPDDTETKRHGLKARL-----KPWTQGEFFGWVFDNREDT-----FDVDNLDV	718
VirB4_pBsph_1	385	IIFNRVYKKMPTISDVHRDI---QALVENGNEMLKNLLSVVSLFLAGRSFG-MFDCQTDIrdkdgnlpPTALDESPIVT	460
VirB4_pBsph_2	425	IIFNRVYKKMPTISDVHRDI---QALVENGNEMLKNLLSVVSLFLAGRSFG-MFDCQTDIrdkdgnlpPTALDESPIVT	500
VirB4_pLsph_100	385	IIFNRVYKKMPTISDVHRDI---QALVENGNEMLKNLLSVVSLFLAGRSFG-MFDCQTDIrdkdgnlpPTALDESPIVT	460
VirB4_pLsph_500	397	FTDFLSYIKGELYENPHKRIIRPELSIARSERLEKIELVIDNLV---NTYAHLFNGHTSI-----PMFTDEE-VIL	463
VirB4_pLsph_600	394	FSDFLVHVRDELYENAERRIIRPELSIARANRLEKIELVIDNLV---NTYGHLFNGHTTL-----PAFTDEQ-IIM	460

```

VirB4_ref      719  FGIDG-----TEFLDNKVLASAASFYLIYRVTML--ADGRRLLIYMDEFWQWINNEAFR--DFVYN  775
VirB4_pBsph_1 461  FDI---SKLSKNGIERPLaQHVLMTWVWNRLIVNPNPKEKKRVIQD-----EAWMMLKFPSM--IEFFKT  519
VirB4_pBsph_2 501  FDI---SKLSKNGIERPLaQHVLMTWVWNRLIVNPNPKEKKRVIQD-----EAWMMLKFPSM--IEFFKT  559
VirB4_pLsph_100 461  FDI---SKLSKNGIERPLaQHVLMTWVWNRLIVNPNPKEKKRVIQD-----EAWMMLKFPSM--IEFFKT  519
VirB4_pLsph_500 464  FSIRNLTQLSKNIFNAQM--HNALSLIWDNLIQVGAPQLKAIYEDPNFneEDAIRTLILIDEAHRLINADNMLaVSFLTD  541
VirB4_pLsph_600 461  FSIRNLTRLSKNIFNAQM--FSALSLIWDNLIQIGAPQLKAIYEDPNFneDDALRFVIFIDESHRIVNAENVLaVDYLIE  538

VirB4_ref      776  KLKTGRKLDMVLVVATQSPDELIKSPIAAAVREQCATHIYLANPKAKCSEYVDGLQVRELYFDkikaiDPLSRQFLVVK  855
VirB4_pBsph_1  520  LSARGRKWNVSLTLVSQRYEMFHRNEEARDIIAQLGTVA-----FMKQSSSDVEKILEAFH-----FSDDVGQVIR  585
VirB4_pBsph_2  560  LSARGRKWNVSLTLVSQRYEMFHRNEEARDIIAQLGTVA-----FMKQSSSDVEKILEAFH-----FSDDVGQVIR  625
VirB4_pLsph_100 520  LSARGRKWNVSLTLVSQRYEMFHRNEEARDIIAQLGTVA-----FMKQSSSDVEKILEAFH-----FSDDVGQVIR  585
VirB4_pLsph_500 542  FAREARKYLGLLVFATQSIIRDVVPEHADSEVVEKIRTLFELTQYKFIGQDANTLDVMRNIFE-----GQLSESELAQIP  616
VirB4_pLsph_600 539  FERARKYLGLLVFATQSIIRDVVPEHADSEVVKIRTLFELTQYKFIGQDSNTLDAMRTIFE-----GQISESELAQIP  613

VirB4_ref      856  PQRKGEsddfaafARLELGKAAAYLPVLSASKPQLELFdeiwKEGMKP--EEWLDTYLeqanli  917
VirB4_pBsph_1  586  TAEQGD-----VVMK-AGKEIVYFHVVPTESEWQYI----NTNMNVtvDELIAKGG-----  631
VirB4_pBsph_2  626  TAEQGD-----VVMK-AGKEIVYFHVVPTESEWQYI----NTNMNVtvDELIAKGG-----  671
VirB4_pLsph_100 586  TAEQGD-----VVMK-AGKEIVYFHVVPTESEWQYI----NTNMNVtvDELIAKGG-----  631
VirB4_pLsph_500 617  QFEQGE-----CLLSINGVGNLMFTVEASDEELAMF----RGGM-----  651
VirB4_pLsph_600 614  QLQQGD-----CLLSISGVGNLMFSVEASEEELAI F----RGGM-----  648

```

Figure 6.11. Alignment of VirB4 of plasmid **pLsph_100**, **pLsph_500** and **pLsph_600** *L. sphaericus* strain NHA15b protein sequences with related sequences .

The sequences are shown with **VirB4_ref**: reference protein type IncX type IV secretion system protein VirB4 (plasmid) *Escherichia coli* (YP_006953788.1); **VirB4_pBsph_1**: VirB4 protein sequence from *L. sphaericus* strain C3-41 pBsph plasmid (NC_010381.1); **VirB4_pBsph_2**: VirB4 protein sequence from strain SSII-1 pBsph2 plasmid (WGS: JPDK01). Alignment was done using Cobalt (<http://www.st-va.ncbi.nlm.nih.gov>). Colouring scheme taken from the program as follows: red for conserved residues, blue for columns with no gaps. Grey is for columns containing gaps. Where less 50% of the sequences contain gaps, they are shown in grey uppercase, greater than 50% will be grey lowercase.

VirB6_ref		-----	
VirB6_Lf	1	MRILKIVMLLFFCFFISPFSANARVDEGAGGGVGNPNTSQADKTTTFNCPGGALISTDVNNNAVDFKFKDEKANLSEKFIAS	80
VirB6_pBsph_1	1	MRILKIVMLLFFCFFISPFSANARVDEGAGGGVGNQNTSEADKTTTFNCPGGALISKDVNNNAVDFKFKDEKANLPEKFIAS	80
VirB6_pBsph_2	1	MRILKIVMLLFFCFFISPFSANARVDEGAGGGVGNQNTSEADKATTFNCPGGALISKDVNNNAVDFKFKDEKANLPEKFIAS	80
VirB6_pLsph_100	1	MRILKIVMLLFFCFFISPFSANARVDEGAGGGVGNQNTSEADETTTFNCPGGALISKDVNNNAVDFKFKDEKANLPEKFIAS	80
VirB6_ref	1	-----MGIVTGIpdGTDLSGNVTSpANAGGVPGFDSDFQTTHEVIFNILNKSISGKLSEYSDVAYTlgkYGVSLYVL	73
VirB6_Lf	81	QLQNLWNIGDINGL--STLIYGNPYC-VWADSILRKESIVMSSNGIFTQAEKEHVIDPILKMFQASYI---YILVLAIM	154
VirB6_pBsph_1	81	QLQNFWNIGDINSL--STLIYGNPYC-VWADSILRKESIVMSSNGIFTQAEKEHVIDPILKMFQASYI---YILVLAIM	154
VirB6_pBsph_2	81	QLQNFWNIGDINSL--STLIYGNPYC-VWADSILRKESIVMSSNGIFTQAEKEHVIDPILKMFQASYI---YILVLAIM	154
VirB6_pLsph_100	81	QLQNFWNIGDINSL--STLIYGNPYC-VWADSILRKESIVMSSNGIFTQAEKEHVIDPILKMFQASYI---YILVLAIM	154
VirB6_ref	74	WYAFTVLARKQQTVPDFIWNICRFYI-----ILLFVKNTGGILTSATDAIDGLKNTLAGGDPwVWMDQLwvKVIQVATL	148
VirB6_Lf	155	FASVKTMISSTKGRAMSDLGDDARMWIFSLVFILFYPQITNIFELNAAVVISLKSLESKG--VKVDSF--SLIASWKD	230
VirB6_pBsph_1	155	VSSVKTMLNSTRGRAMSDLGDDAKMWVASAGFILFYPKITGILFELNAAVLVLSLKSLESKG--VNLDSF--SLIASWKD	230
VirB6_pBsph_2	155	VSSVKTMLNSTRGRAMSDLGDDAKMWVASAGFILFYPKITGILFELNAAVLVLSLKSLESKG--VNLDSF--SLIASWKD	230
VirB6_pLsph_100	155	VSSVKTMLNSTRGRAMSDLGDDAKMWVASAGFILFYPKITGILFELNAAVLVLSLKSLESKG--VNLDSF--SLIASWKD	230
VirB6_ref	149	IFDKDTSTVPvaggIGALLTYVGGVLALLLCSIVFAAAELTLLLLSVTAPIFIMCLMFGFLRQMFNSWLQLNFSLLLVL	228
VirB6_Lf	231	LFVGDYPIPIP----LSYIIVIFAEWVLA AVLNGVYILRKITIGLLLVLGFVAAFSLLFARTRPFFGAWLKELTSSVFLQS	306
VirB6_pBsph_1	231	LFIGDVSPLP----LSYIMVIFAEWVLA AVLNGVYILRKILIGLLLVLGFVAAFSLLFARTRPFFGAWLKELCSNVFLQS	306
VirB6_pBsph_2	231	LFIGDVSPLP----LSYIMVIFAEWVLA AVLNGVYILRKILIGLLLVLGFVAAFSLLFARTRPFFGAWLKELCSNVFLQS	306
VirB6_pLsph_100	231	LFIGDVSPLP----LSYIMVIFAEWVLA AVLNGVYILRKILIGLLLVLGFVAAFSLLFARTRPFFGAWLKELCSNVFLQS	306
VirB6_ref	229	FAALALRAGTWQLNMALSTSI-----ATASENNLLQTGVTSLAAGIFMAWIIwQAQSYASQIAGVGVEGA	293
VirB6_Lf	307	VHAIVLYTMVLMSSVGVSVSANFISPMFDSGPVLLTAAATDEYDMTTTGAGTFFKLIMMAAFI-PVTGLITRMLKMGDSSS	385
VirB6_pBsph_1	307	VHAIVLYTMILMSSVGVSVSTNFISPLFDSGPVLLTATTDEYDMTKTGAGTFFKLIMMAAFI-PVTGLITRMLKMGDSTS	385
VirB6_pBsph_2	307	VHAIVLYTMILMSSVGVSVSTNFISPLFDSGPVLLTATTDEYDMTKTGAGTFFKLIMMAAFI-PVTGLITRMLKMGDSTS	385
VirB6_pLsph_100	307	VHAIVLYTMILMSSVGVSVSTNFISPLFDSGPVLLTATTDEYDMTKTGAGTFFKLIMMAAFI-PVTGLITRMLKMGDSTS	385
VirB6_ref	294	MQGAAAMGIGAGVF-----GASRMARGALGMGRNAGIGAWKGLRRQEG---GFGQSPGITGKTANLAGQGVNIGAK	361
VirB6_Lf	386	SLGTAMTMLGVGGFMSTFNLAQAGSVITGGSVFSMSTSSLSGGDSTNQSNNTTVGMNSASDSNATSI TMNALGNNSTPW	465
VirB6_pBsph_1	386	KLGTAMTMLGVGGFMSTYNLTQAGSVITGGSVFSMSNSLSGGDSTNQSFNNTVGTNSPDSNTTSTI TMNALGNNSTPW	465
VirB6_pBsph_2	386	KLGTAMTMLGVGGFMSTYNLTQAGSVITGGSVFSMSNSLSGGDSTNQSFNNTVGTNSPDSNTTSTI TMNALGNNSTPW	465
VirB6_pLsph_100	386	KLGTAMTMLGVGGFMSTYNLTQAGSVITGGSVFSMSNSLSGGDSTNQSFNNTVGTNSPDSNTTSTI TMNALGNNSTPW	465

VirB6_ref	362	KLRQAAIERA---KKMYGG-----	377
VirB6_Lf	466	KSAKNAVSKTAGI STAIGGGMILGPMGAQAGYKFGSKSAELMMQTTPRNVGFGIKGIADTLKNAKQYSGPNGVGMKGVMSN	545
VirB6_pBsph_1	466	KATKDAISKGVGI STAIGGGMILGPPGAQAGFKLGSKSAELLMQTTPRSIGFGVKGISDTLKNAKQYSGPNGVGMKGVMGN	545
VirB6_pBsph_2	466	KATKDAISKGVGI STAIGGGMILGPPGAQAGFKLGSKSAELLMQTTPRSIGFGVKGISDTLKNAKQYSGPNGVGMKGVMGN	545
VirB6_pLsph_100	466	KATKDAISKGVGI STAIGGGMILGPPGAQAGFKLGSKSAELLMQTTPRSIGFGVKGISDTLKNAKQYSGPNGVGMKGVMGN	545
VirB6_ref		-----	
VirB6_Lf	546	LASRRQFFGQMGESVGMVVGQGVFGKQLGQIASGVSRQAIASLPMEQGGMMFRNSNGLPALGTFQNLASHYPNEKLQFIQ	625
VirB6_pBsph_1	546	LAARRQFFGNMGESVGMVVGQGAFAFRQLGQIASGVSRQAIASLPMEQGGMMFKNSNGLPALGTFQNLAHHPNEKIQFIQ	625
VirB6_pBsph_2	546	LAARRQFFGNMGESVGMVVGQGAFAFRQLGQIASGVSRQAIASLPMEQGGMMFKNSNGLPALGTFQNLAHHPNEKIQFIQ	625
VirB6_pLsph_100	546	LAARRQFFGNMGESVGMVVGQGAFAFRQLGQIASGVSRQAIASLPMEQGGMMFKNSNGLPALGTFQNLAHHPNEKIQFIQ	625
VirB6_ref		-----	
VirB6_Lf	626	NNQSGGFYLNDRNGSLRQIGLTGAADPTLDNGVSRVIDYQFKAPNSDIQLNNGGSYSPGSQIALNNSALPSYNEALNSTSN	705
VirB6_pBsph_1	626	NNQSGGFYLNDRNGTLRQIGLTGAADPSLENGVSRVIDYKFRAPNSDIQLDNKGSYSPGSQIALNNSALPSYNDALSRMSS	705
VirB6_pBsph_2	626	NNQSGGFYLNDRNGTLRQIGLTGAADPSLENGVSRVIDYKFRAPNSDIQLDNKGSYSPGSQIALNNSALPSYNDALSRMSS	705
VirB6_pLsph_100	626	NNQSGGFYLNDRNGTLRQIGLTGAADPSLENGVSRVIDYKFRAPNSDIQLDNKGSYSPGSQIALNNSALPSYNDALSRMSS	705
VirB6_ref		-----	
VirB6_Lf	706	TTTASN---GTL SINPIGQSMKGINGYPIPLRAKNRWEHYVE---PSSNLGNSVAPGALSTGSTAPFAAMS NLSVSR	777
VirB6_pBsph_1	706	TTTASN VGGLPSSMINPIGQPMKAVNGYPLPLRAQNRWEGHVEHIAS PATNIGNPMARSSLSTGTTT PFAAMSNVTAP	785
VirB6_pBsph_2	706	TTTASN VGGLPSSMINPIGQPMKAVNGYPLPLRAQNRWEGHVEHIAS PATNIGNPMARSSLSTGTTT PFAAMSNVTAP	785
VirB6_pLsph_100	706	TTTASN VGGLPSSMINPIGQPMKAVNGYPLPLRAQNRWEGHVEHIAS PATNIGNPMARSSLSTGTTT PFAAMSNVTAP	785
VirB6_ref		-----	
VirB6_Lf	778	A-TQPSYGHGASIALD TYKSNLAQFSSNGSSIPSLNSSTTSISGGGNRAVGLPGSTDYVSRTSGAYFVGG LSSSGQITMD	856
VirB6_pBsph_1	786	ASTQSSYQVASTALNSYQS NLAQYSSHSSSI PSLDLSTSSISG VGNRVVGLPGSTDVSR TSGAYFVGG LSPGTGQITTD	865
VirB6_pBsph_2	786	ASTQSSYQVASTALNSYQS NLAQYSSHSSSI PSLDLSTSSISG VGNRVVGLPGSTDVSR TSGAYFVGG LSPGTGQITTD	865
VirB6_pLsph_100	786	ASTQSSYQVASTALNSYQS NLAQYSSHSSSI PSLDLT TSSISG VGNRVVGLPGSTDVSR TSGAYFVGG LSPGTGQITTD	865
VirB6_ref		-----	
VirB6_Lf	857	SMNMINNPLTPKYDNPSYRADHINPDSFVYHLS PNADLRNNVDIGADAVNKL TQ--KRETS DVQTGWAQLLKKEFSEERQ	934
VirB6_pBsph_1	866	SVNMINNPLTPKFDNPSFRADHVNPD SFVYHISP NADLRNNVDIGADKVNQITQRVKKAPSDVQSGWAQLLKNDYSEERK	945
VirB6_pBsph_2	866	SVNMINNPLTPKFDNPSFRADHVNPD SFVYHISP NADLRNNVDIGADKVNQITQRVKKAPSDVQSGWAQLLKNDYSEERK	945
VirB6_pLsph_100	866	SVNMINNPLTPKFDNPSFRADHVNPD SFVYHISP NADLRNNVDIGADKVNQITQRVKKAPSDVQSGWAQLLKNDYSEERK	945

```

VirB6_ref          -----
VirB6_Lf           935 SNIS  938
VirB6_pBsph_1     946 SNIS  949
VirB6_pBsph_2     946 SNIS  949
VirB6_pLsph_100  946 SNIS  949

```

Figure 6.12 Alignment of VirB6 of plasmid pLsph_100 of *L. sphaericus* strain NHA15b protein sequences with related sequences . The sequences are shown with the **VirB6_ref**: VirB6 (plasmid) *Escherichia coli* (YP_006953062.1); **VirB6_Lf**: hypothetical protein KQ41_06695 *Lysinibacillus fusiformis* (KGA83722.1); **VirB6_pBsph_1**: VirB6 from *L. sphaericus* strain C3-41 pBsph plasmid (NC_010381.1), **VirB6_pBsph_2**: VirB6 from strain SSII-1 pBsph2 plasmid (WGS: JPDK01). Alignment was done using Cobalt (<http://www.st-va.ncbi.nlm.nih.gov>). Colouring scheme taken from the program as follows: red for conserved residues, blue for columns with no gaps. Grey is for columns containing gaps. Where less 50% of the sequences contain gaps, they are shown in grey uppercase, greater than 50% will be grey lowercase.

During the search on the whole genome sequences on the database, it was found that in the LP1-G genome, there is a CDS that encodes similar *cry48* and *cry49* genes. This is an indication of *cry* gene duplication that warrants a further analysis that could give a clue to the *cry* genes' location. The nucleotide sequences of *cry* genes of the NHA15b strain from this study are identical to the previously reported study (Jones et al., 2007). Alignment of Cry48 and Cry49 protein are presented on Figures 6.13. and 6.14.

CAJ86549.1	1	MDINNNNEKEI INSHLLPASLLKK--HPIKS-----LQSTNYKDWLNLQCDFNKDIESYDLVTAVSSGTIVVGTMLSAIY	73
CAJ86548.1	1	MDINNNNEKEI INSHLLPASLLKK--HPIKS-----LQSTNYKDWLNLQCDFNKDIESYDLVTAVSSGTIVVGTMLSAIY	73
CAJ86545.1	1	MDINNNNEKEI INSHLLPYSLKK--YPIKS-----LQSTNYKDWLNLQCDFNKDIESYDLVTAVSSGTIVVGTMLSAIY	73
Cry48 LP1-G_2	1	MEKNKESENDVTNSYLSPOSYLKKpiHPILAnpnfpLQSNKYKNWINMCNE-SSVLESFDSITAI STGSIVFGTMLAGVF	79
Cry48 NHA15b	1	MDINNNNEKEI INSHLLP[REDACTED]SLKK--[REDACTED]PIKS-----LQSTNYKDWLNLQCDFNKDIESYDLVTAVSSGTIVVGTMLSAIY	73
CAJ86549.1	74	APAI IAGPIGVIGAI IISFGTLLPLLWSEDENNPKTVWIEFIRMGERLVDKTI SQTVLNILESYLKDLKVNLI DYEKAKQ	153
CAJ86548.1	74	APAI IAGPIGVIGAI IISFGTLLPLLWSEDENNPKTVWIEFIRMGERLVDKTI SQTVLNILESYLKDLKVNLI DYEKAKQ	153
CAJ86545.1	74	APALIAGPIGII GAI IISFGTLLPLLWNESENNPKTTWIEFIRMGEQLVDKTI SQTVFNILESYLKDLKVNLDVYEKAKQ	153
Cry48 LP1-G_2	80	ATASFAGPIGII GAI IISFGTILPLLFTEDQNQ-KNSWTEFIRMGERYVDIT ISETVKAQVLGKLN DIKENLMNYEEAFC	158
Cry48 NHA15b	74	APA[REDACTED]IAGPIGII GAI IISFGTLLPLLW[REDACTED]SENNPKT[REDACTED]WIEFIRMGE[REDACTED]LVDKTI SQTV[REDACTED]NILESYLKDLKVNLI[REDACTED]DYEKAKQ	153
CAJ86549.1	154	DWIELKKQQLPGSPPSINLRNAADIAHQRLDSLHNKFAELNVFKVASYETILLPVYAQAANLHLNLLQQGAMFADQWIED	233
CAJ86548.1	154	DWIELKKQQLPGSPPSINLRNAADIAHQRLDSLHNKFAELNVFKVASYETILLPVYAQAANLHLNLLQQGAMFADQWIED	233
CAJ86545.1	154	DWIELKKQQLPGSPPSTKLRNAADIAHQRLDSLHNKFAELNKFVPEPYETILLPVYAQAANLHLNLLQQGAMFADQWIED	233
Cry48 LP1-G_2	159	DWIELKKQQLPGSPPSAALRNAADIVHDRLEGINRDFGALGMFSISPYQTILLPVYAQAANLHLNLLQQGSMFADKWIQD	238
Cry48 NHA15b	154	DWIELKKQQLPGSPPSTKLRNAADIAHQRLDSLHNKFAELNKFVPEPYETILLPVYAQAANLHLNLLQQGAMFADQWIED	233
CAJ86549.1	234	KYSPRNDTFAGNSNDYQDLLKSRTITYINHIENTYKDGLNYLWNQPEMTWDIYNEYRTNMTLTALDLLPLFPFYNKELYD	313
CAJ86548.1	234	KYSPRNDTFAGNSNDYQDLLKSRTITYINHIENTYKDGLNYLWNQPEMTWDIYNEYRTNMTLTALDLLPLFPFYNKELYD	313
CAJ86545.1	234	KYSSRNDTFAGNSNDYQDLLKSRTITYINHIENTYQGLNYLWNQPEMTWDIYNEYRTKMTITALDLMALFPFYNKELYD	313
Cry48 LP1-G_2	239	KHPTKNEFAGTSDNYQKKLESKIKEYITYVEKTYKDGLFSLSNPNMTWSIYNDYRREMTLTALDIITIFPFYNRKEYD	318
Cry48 NHA15b	234	KYSSRNDTFAGNSNDYQDLLKSRTITYINHIENTYQGLNYLWNQPEMTWDIYNEYRTKMTITALDLMALFPFYNKELYD	313
CAJ86549.1	314	PRVGIKSELTREVIINTPV-DPHLHRYFKLGETEDKLTNNSSELFKWLTS LKFRTFNQPGFPFLIGNMNYFKKTNG-TQLI	391
CAJ86548.1	314	PRVGIKSELTREVIINTPV-DPHLHRYFKLGETEDKLTNNSSELFKWLTS LKFRTFNQPGFPFPFLIGNMNYFKKTNG-TQLI	391
CAJ86545.1	314	PTVGIKSELTREIFINTPV-EPHLHRYFKLSETEEKL TNNSDLFKWLTS LKFRTLYQPGFPFLIGNMNSFTNTNG-TQLI	391
Cry48 LP1-G_2	319	TTIGVKTEL TRELYLSTSVrDSAINEFFAVDNTEAILKNKPSLFYWLKQLDIKTLLTPKGNILIGNRVFQPTNNLSREL	398
Cry48 NHA15b	314	PTVGIKSELTREIFINTPV-EPHLHRYFKLSETEEKL TNNSDLFKWLTS LKFRTLYQPGFPFPFLIGNMNSFTNTNG-TQLI	391
CAJ86549.1	392	NNQQQLWSF-PGTTE-IEKLFPS PANIDKVTMYIYYGSGWEVPEPISITINKLI FNHHKHGLITEYDAGNTNAPT MGIYV	469
CAJ86548.1	392	NNQQQLWSF-PGTTE-IEKLFPS PANIDKVTMYIYYGSGWEVPEPISITINKLI FNHHKHGLITEYDAGNTNAPT MGIYV	469
CAJ86545.1	392	NNQQQLWSF-PGTTENEEKLFPS PANIDQVTMYIYYGSGWGIPEPISTTINKLI FNHDKHELISEYDAGNTNAPT RSLSL	470
Cry48 LP1-G_2	399	HNYQQLWNYqPPPYELMLQEFRFN SHINKIVMS-----EISDAVTTNITLLRFYLD-DELI ALYNSGDSN-PHTSTTL	469
Cry48 NHA15b	392	NNQQQLWSF-PGTTENEEKLFPS PANIDQVTMYIYYGSGWGIPEPISTTINKLI FNHDKHELISEYDAGNTNAPT RSLSL	470

CAJ86549.1	470	NLPKHYLSCLNSYYPLTATTNGM-GKEELKMYSFGWTHESVDFLNEISNDKITQIPAVKAYNLNSNSRVIKGPGHIGGNL	548
CAJ86548.1	470	NLPKHYLSCLNSYYPLTATTNGM-GKEELKMYSFGWTHESVDFLNEISNDKITQIPAVKAYNLNSNSRVIKGPGHIGGNL	548
CAJ86545.1	471	GLPNHYLSCLNSYYPLTATTDGM-NKEELKMYSFGWTHNSVDFLNEISKDKITQIPAVKAYRLTSNSRVIKGPGSHIGGNL	549
Cry48 LP1-G_2	470	EIPNHLYSCLNSYYPIRIKKNKGI _d GPASVKLQAFGWTHNCVDYDNEISKDRITQIPAVKGNFIGNRSRVIKGSGHHGGDL	549
Cry48 NHA15b	471	GLPNHYLSCLNSYYPLTATTDGM-NKEELKMYSFGWTHNSVDFLNEISKDKITQIPAVKAYRLTSNSRVIKGPGSHIGGNL	549
CAJ86549.1	549	VYLSDKSQLSLACRYTNSSPQDFLIRIRYASNKRNMVQLFTPFSTH-QFVLPQTFNHLNIEQTKYEDYEYAQLPGSLTIN	627
CAJ86548.1	549	VYLSDKSQLSLACRYTNSSPQDFLIRIRYASNKRNMVQLFTPFSTH-QFVLPQTFNHLNIEQTKYEDYEYAQLPGSLTIN	627
CAJ86545.1	550	VYLSSENSQMALTCRYTNSSPQEQYKIRIRYASNRLNMGQLFTTFFSSH-QFVLPPTFNHFNIEQAKYEDYAYAEFPESMSIR	628
Cry48 LP1-G_2	550	VLLNDR--LDISCLYRNPSPQEFRI RVRYASNHFQAQVSILTNNNNRvNGILRPT--NFDLAEELNYDDFLTLELNGTFTLN	625
Cry48 NHA15b	550	VYLSSENSQMALTCRYTNSSPQEQYKIRIRYASNRLNMGQLFTTFFSSH-QFVLPPTFNHFNIEQAKYEDYAYAEFPESMSIR	628
CAJ86549.1	628	GNVNIDLLFLLNVLDDGG-----ELLLDKIEFIPLTQKVVDNLEKEKIDMLKKNLTDSLFNPAKDTLTKIDSTDYQIDQI	700
CAJ86548.1	628	GNVNIDLLFLLNVLDDGG-----ELLLDKIEFIPLTQKVVDNLEKEKIDMLKKNLTDSLFNPAKDTLTKINSTDYQIDQI	700
CAJ86545.1	629	GNLNSDILLILNILAGG-----ELLLDKIEFIPLTQKVVDNLEKEKIDMLKKNLTDSLFNPSKDTLTKIDSTDYQIDQI	701
Cry48 LP1-G_2	626	-SVPVGLPYIMSISKIN _{istdeke} NLIIDKIEFVPIT-----QKEKLEQIQIATNSLFTSPTKNTLKIESTDYQINQI	697
Cry48 NHA15b	629	GNLNSDILLILNILAGG-----ELLLDKIEFIPLTQKVVDNLEKEKIDMLKKNLTDSLFNPSKDTLTKIDSTDYQIDQI	701
CAJ86549.1	701	AFQIESINEEINTQEKMKLLDNIKYAKKLNQLRNLLYSRESQAQIDWVTSNDVSIYHGKPPFNEYTLVMS--GSSLKIT	778
CAJ86548.1	701	AFQIESINEEINTQEKMKLLDNIKYAKKLNQLRNLLYSRESQAQIDWVTSNDVSIYHGKPPFNEYTLVMS--GSSLKIT	778
CAJ86545.1	702	AFQIESINEEINPQEKMELLDNIKYAKKLNQLRNLLYSRESQAQIDWVTSNDVSIYHGKPPFNDYTLVMS _{rt} SSSLSEIT	781
Cry48 LP1-G_2	698	AMQIESLPEEFYSQEKMLSDDIKHAKHLNQRSNLLQIEDFQSLSGWFISNSVSVVERGNQNFKDYILLMP--GASTSEIN	775
Cry48 NHA15b	702	AFQIESINEEINPQEKMELLDNIKYAKKLNQLRNLLYSRESQAQIDWVTSNDVSIYHGKPPFNDYTLVMS _{rt} SSSLSEIT	781
CAJ86549.1	779	SSNYPTYIYKKIEESKLKPYTRYLVGRFISNSDNLEIFISRYENEIHTNMNVHVD _{DD} TLNLSYKRQNECESKLP _{IV} FDET	858
CAJ86548.1	779	SSNYPTYIYKKIEESKLKPYTRYLVGRFISNSDNLEIFISRYENEIHTNMNVHVD _{DD} TLNLSYKRQNECESKLP _{IV} FDET	858
CAJ86545.1	782	ATNYQTYIYKKIEESKLKPYTRYLVGRFISNSEDLEIFISRYENEIHTNMNVHG _{DD} TLNLSYKRQNECESKLP _{II} FDAT	861
Cry48 LP1-G_2	776	NYNFPTYIYQQIDESKLKPYTRYLVGRFVKDCKELEVFVSRYNKEIQTIMNVRG _{DD} INLLNTNEQP _{ND} C-----	844
Cry48 NHA15b	782	ATNYQTYIYKKIEESKLKPYTRYLVGRFISNSEDLEIFISRYENEIHTNMNVHG _{DD} TLNLSYKRQNECESKLP _{II} FDAT	861
CAJ86549.1	859	SQFPLSPSRTSGISNHSYNGAQSSCHDTQIFSFSIDTGDVDFNEYPGIEILFKLSNSNGYASISNLEVIEERLLTEEE	938
CAJ86548.1	859	SQYPLSPSRTSGISNHSYNGAQSSCHDTQIFSFSIDTGDVDFNEYPGIEILFKLSNSNGYASISNLEVIEERLLTEEE	938
CAJ86545.1	862	SQYSLSPSRTSGISNHSYNNGHQSSCNDTHIFSFSIDTGEVDFNNYPGIEILFKLSNTNGYASISNLEVIEERLLTEEE	941
Cry48 LP1-G_2	845	-----HNTQTF _S FSIDTGNLDFNECPGFEILFQISNSEGYASLSNLEVLEERLLTEKE	897
Cry48 NHA15b	862	SQYSLSPSRTSGISNHSYNNGHQSSCNDTHIFSFSIDTGEVDFNNYPGIEILFKLSNTNGYASISNLEVIEERLLTEEE	941

CAJ86549.1	939	KRHIIEIENRWKAKKEIQRNETEKETTQAQQAINNLFTDTQYSKLKFETTKQSI SKANAILENI PYVYNSLLPTEPGMNF	1018
CAJ86548.1	939	KRHIIEIENRWKAKKEIQRNETEKETTQAQQAINNLFTDTQYSKLKFETTKQSI SKANAILENI PYVYNSLLPTEPGMNF	1018
CAJ86545.1	942	KRQIIQIENRWKAKKESQRNETEKITTTQAQQAINSLFTDTQYSNLKFETTKQNITEANTILENI PYVYNALLPTEPGMNF	1021
Cry48 LP1-G_2	2898	KQHIIEIEERWRRAKKVNQHIETEKVYNQAQQAIDNLFIDTQYSKLKFETTKHNINTADAILDITIPYVYNSLLPNEPGMNY	977
Cry48 NHA15b	942	KRQIIQIENRWKAKKESQRNETEKITTTQAQQAINSLFTDTQYSNLKFETTKQNITEANTILENI PYVYNALLPTEPGMNF	1021
CAJ86549.1	1019	ELFNSEFKDQINKAHTLYKMRNSIKNGDFINGTKYWSISTDVKLEKTNIETILVMSSWSAQSSQQILVQKQNRILLRVIK	1098
CAJ86548.1	1019	ELFNSEFKDQINKAHTLYKMRNSIKNGDFINGTEYWSISTDVKLEKTNIETILVMSSWSAQSSQQILVQKQNRILLRVIK	1098
CAJ86545.1	1022	VLFNSEFKDQINKAHALYKMRNLIKNGDFINDTKYWSISTDVKLEKVNKETILVLSSWEAQASQQILVQKQKRYLLRVIK	1101
Cry48 LP1-G_2	978	DLFKNLKLIYKAYTSFKLRNTIKNGDFITNGTENWSISAGIKIEKTDNEHILVIPNWSAQVSQQIQLEKQNRILLRVIK	1057
Cry48 NHA15b	1022	VLFNSEFKDQINKAHALYKMRNLIKNGDFINDTKYWSISTDVKLEKVNKETILVLSSWEAQASQQILVQKQKRYLLRVIK	1101
CAJ86549.1	1099	KEDMGSGNVTISDCLNNAKIEFIPHDCNMN-----	1129
CAJ86548.1	1099	KEDMGSGNVTISDCLNNAKIEFIPHDCNMN-----	1129
CAJ86545.1	1102	KEDMGRGNVVIISDCLNNAKIDFTPHDCNMNHIQNSSEFIIKTIHFSPNTEQVRIDIGQSDGVFKVESIELICVNY	1177
Cry48 LP1-G_2	21058	KEGMGNGYVKISDCLYNIKIDYISSEYYTINKQNSLEYITKTIQFIPNTDKVHIELGETEGVFKIKSIELICINN	1133
Cry48 NHA15b	1102	KEDMGRGNVVIISDCLNNAKIDFTPHDCNMNHIQNSSEFIIKTIHFSPNTEQVRIDIGQSDGVFKVESIELICVNY	1177

Figure 6.13. Alignment of Cry48Aa of *L. sphaericus* strain NHA15b protein with related proteins.

The protein is shown with the same toxin from strain 2173 (accession number: CAJ86549.1), LP1-G (CAJ86548.1), 47-6B (CAJ86545.1). The Cry48 LP1-G_2 protein is predicted from the whole genome sequence of LP1-G (JPDL01000001.1). Alignment was done using Cobalt (<http://www.st-va.ncbi.nlm.nih.gov>). Colouring scheme taken from the program as follows: red for conserved residues, blue for columns with no gaps. Grey is for columns containing gaps. Where less 50% of the sequences contain gaps, they are shown in grey uppercase, greater than 50% will be grey lowercase.

CAJ86544.1	1	MENQIKEEFNKNNHGIPSDCSCIKEGDDYNSLT--E----INAKEFSYCSPNMFNLNLPEQSTRFQTIASIHNNCSFEI	74
CAJ86542.1	1	MENQIKEEFNKNNHGIPSDCSCIKEGDDYNSLT--EvpseINAKEFSYCSPNMFNLNLPEQSTRFQTIASIHNNCSFEI	78
CAJ86541.1	1	MENQIKEEFNKNNHGIPSDCSCIKEGDDYNSLT--E----INAKEFSYCSPNMFNLNLPEQSTRFQTIASIHNNCSFEI	74
LP1-G_2	1	-----VFKSFTkkE-pheINNGVCCYGNNTMFDNLPPLSTRFQTTAATLSDNRSFEI	52
Cry49 NHA15b	1	LENQIKEEFNKNNHGIPSDCSCIKEGDDYNSLT--E----INAKEFSYCSPNMFNLNLPEQSTRFQTIASIHNNCSFEI	74
CAJ86544.1	75	LNNDPGYIYGDSVDGECRIAVAHRELGNGLERTGDDRFLFI FYALDNNNFI IANRHDGFVLQFLIANGQGVIVSREYQPN	154
CAJ86542.1	79	LNNDPGYIYGDSVAGECRIAVAHRELGNGLERTGDDRFLFI FYALDNNNFI IANRHDGFVLQFLIANGQGVIVSREYQPN	158
CAJ86541.1	75	LNNDPGYIYGDSVDGECRIAVAHRELGNGLERTGDDRFLFI FYALDNNNFI IANRHDGFVLQFLIANGQGVIVSREYQPN	154
LP1-G_2	53	LFNTPGYKYEGAQRGECSTAVAHREIGNGVERS GDDRFLFI FYLLDNNNFI VANRYDGFVLQYWIQEGQVIVSREYEPN	132
Cry49 NHA15b	75	LNNDPGYIYGDSVDGECRIAVAHRELGNGLERTGDDRFLFI FYALDNNNFI IANRHDGFVLQFLIANGQGVIVSREYQPN	154
CAJ86544.1	155	IHQEFTIQSINSDTFRLHSRDTNTFATVCWAQFNSWTKIVSRVDNPGAPNANLKHRSLLT DINMPQLPSLTPLOPLPRLT	234
CAJ86542.1	159	IRQEFTIQSINSDTFRLHSRDTNTFATVCWAQFNSWTKIVSRVDNPGAPNADLRHRSFL-DINMPQLPSLTPLOPLPRLT	237
CAJ86541.1	155	IHQEFTIQSINSDTFRLHSRDTNTFATVCWAQFNSWTKIVSRVDNPGAPNANLKHRSLLT DINMPQLPSLTPLOPLPRLT	234
LP1-G_2	133	IYQEFsirLLAQNNFRL--RTGNTFATVCWGFNTWTKIVSRVDQPTAPNASLRHRSFL-NIDMPQLPSITPLQPLPQLT	209
Cry49 NHA15b	155	IHQEFTIQSINSDTFRLHSRDTNTFATVCWAQFNSWTKIVSRVDNPGAPNANLKHRSLLT DINMPQLPSLTPLOPLPRLT	234
CAJ86544.1	235	ELEDGGLSPAQAPRAIIGRTLIPCLFVNDPVLRLLENRIKQSPYYVLEHRQYWHRIWTDIF TAGERREYREVTGINNNAQN	314
CAJ86542.1	238	GLEDGGLSPAQAPRAIIGRTLIPCLFVNDPVLRLLESRIKQSPYYVLEHRQYWHRLWTDIFNAGERREYREVTGINHNAQN	317
CAJ86541.1	235	ELEDGGLSPAQAPRAIIGRTLIPCLFVNDPVLRLLENRIKQSPYYVLEHRQYWHRIWTDIF TAGERREYREVTGINNNAQN	314
LP1-G_2	210	GLEDGGLSSDQAPRAIIGRTLIPCLFVNDPVLRLLESRIKQSPYYVLEHRQYWHRLWTDIFSPGERREYHEVTGINHNAQN	289
Cry49 NHA15b	235	ELEDGGLSPAQAPRAIIGRTLIPCLFVNDPVLRLLENRIKQSPYYVLEHRQYWHRIWTDIF TAGERREYREVTGINNNAQN	314
CAJ86544.1	315	DMNKMINITIGADGNRLRFGNLSTPFRQQIIDSNSNTLGSFANTNYGTRTDIVNVFNSEFHQVRYARFVKAYEYRLTRAD	394
CAJ86542.1	318	DMNNMINITIGSDGNRLRLFHLSTPFRQQIIDSNSNTLGSFANSNYSSRTESITYLNTEFHQVRFARFVKAYEYRLTRAD	397
CAJ86541.1	315	DMNKMINITIGADGNRLRFGNLSTPFRQQIIDSNSNTLGSFANTNYGTRTDIVNVFNSEFHQVRYARFVKAYEYRLTRAD	394
LP1-G_2	290	DMNNMINITIGSDGNRLRLF-----PWR--VIGNWET-----	319
Cry49 NHA15b	315	DMNKMINITIGADGNRLRFGNLSTPFRQQIIDSNSNTLGSFANTNYGTRTDIVNVFNSEFHQVRYARFVKAYEYRLTRAD	394

CAJ86544.1	395	GSQVGT	PWVVLDRKEMDLR	TYPHNMAITL	ENVKIDNADNSYDLS	IWKTP	PLKLDGKII	IENHENS	KPYYN	464
CAJ86542.1	398	GTLVGT	PWVVLDRKEMDLR	TTFPHNMTLN	ENVKIVNADNSYDLSV	WKTP	PLKLDGKII	IENHENS	KPYYN	467
CAJ86541.1	395	GSQVGT	PWVVLDRKEMDLR	TYPHNMAITL	ENVKIDNADNSYDLS	IWKTP	PLKLDGKII	IENHENS	KPYYN	464
LP1-G_2		-----	-----	-----	-----	-----	-----	-----	-----	
Cry49 NHA15b	395	GSQVGT	PWVVLDRKEMDLR	TYPHNMAITL	ENVKIDNADNSYDLS	IWKTP	PLKLDGKII	IENHENS	KPYYN	464

Figure 6.14. Alignment of Cry49Aa of *L. sphaericus* strain NHA15b protein with related proteins.

The protein is shown with the same toxin from strain 2173 (accession number: CAJ86544), LP1-G (CAJ86542.1), 47-6B (CAJ86541.1). The Cry49 LP1-G_2 protein is predicted from the whole genome sequence of LP1-G (JPDL01000001.1). Alignment was done using Cobalt (<http://www.st-va.ncbi.nlm.nih.gov>). Colouring scheme taken from the program as follows: red for conserved residues, blue for columns with no gaps. Grey is for columns containing gaps. Where less 50% of the sequences contain gaps, they are shown in grey uppercase, greater than 50% will be grey lowercase.

6.4.6. *Sphaericolysin toxin*

The sphaericolysin was first identified in the *L. sphaericus* strain A3-2 from Japan. It does not have any toxicity toward mosquitoes but is toxic to the German cockroach (*Blattella germanica*) and common cutworm (*Spodoptera litura*) (Nishiwaki et al., 2007). The sphaericolysin toxin of NHA15b was encoded in the genome by BSPH_RS19015 CDS. The gene was predicted to produce a 506 amino acid long protein. The protein appears to be highly widespread in insect-associated gram positive bacteria and is highly conserved across this range of bacteria (Palma et al., 2014). The NHA15b sphaericolysin protein sequence is identical to the sphaericolysin protein from A3-2 and from C3-41.

CHAPTER VII GENERAL DISCUSSION

New isolate activity

Lysinibacillus sphaericus bacteria are already known to harbour a range of toxin proteins pathogenic to insects. When we received a newly isolated bacteria from a dead spider- hence named Fang Large isolate- known to kill the wax moth larva, it was interesting to investigate the toxin proteins carried by this strain to ascertain whether this strain represented a new class of lepidopteran active *L. sphaericus*. Using transmission electron microscopy it was shown that the isolate was a bacilliform bacterium. It changed to a club shaped rod, with a round terminal spore when sporulating, characteristic of the genus *Lysinibacillus*. The micrograph did not show the presence of any crystal either inside the bacteria or released after sporulation. The 16S RNA DNA sequence suggested that the isolate is either a *L. sphaericus* or *Lysinibacillus fusiformis* species. Based on the presence of a *L. sphaericus* (*Bacillus sphaericus*)-specific signature, it was confirmed that the bacterium is from the *L. sphaericus* species. A qualitative bioassay of the Fang Large isolate on the wax moth larvae *Galleria melonella* showed that the isolate killed its target by causing septicaemia when injected in relatively large quantity. Based on its mode of action and a low pathogenicity, no further work was attempted on this isolate.

Plasmid profiles and location of *cry48/cry49*

Apart from Bin toxin proteins, some isolates of highly pathogenic *Lysinibacillus sphaericus* bacteria were known produce Cry48 and Cry49 proteins (Jones et al., 2007). It was known from this previous work that the genes for the *cry48*

and *cry49* toxin genes were located side by side on an approximately 15 kb DNA fragment but it was not known if this was of genomic or chromosomal origin. The exact location of the genes is still undetermined but it is predicted that these genes are located on a plasmids similar to some of the *cry* genes of *Bacillus thuringiensis*. It is also possible that the *cry 48/49* genes are located in both chromosome and plasmid similar to the *bin* toxin genes of *L. sphaericus* isolate C3-41 (Hu et al., 2008a). To determine to location of the *cry 48/49* genes in the *L. sphaericus* isolates known to harbour the *cry* genes, a plasmid separation procedure was conducted using Pulsed Field Gel Electrophoresis (PFGE) techniques. The procedure indicated that the bacterial isolates possessed several plasmids with different sizes. Several attempts to directly probe the occurrence of the *cry* genes on these plasmids failed. However, when a PCR DNA from electroeluted bands of the plasmids from isolate IAB59 was conducted, the result showed that one of these plasmid bands of IAB59 with a size approximately of 750 kb (based on comparison to 50 kb lambda DNA ladder) produced a positive PCR product with *cry49* primers. Although the results are still not conclusive as contaminating DNA cannot be ruled out, this result indicates that the *cry48/49* genes, particularly in the *L. sphaericus* IAB59 isolate, are most likely to be located on a large plasmid.

Transcript mapping

To attempt to identify promoters driving toxin gene expression in *L. sphaericus*, mRNA was isolated from *L. sphaericus* isolates; turned into cDNA using an RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE). The cDNA products were then cloned and sequenced. These experiments were able to

obtain sequences of the toxin genes including part of its 5' untranslated region for *cry48*, *mtx2* and *binB* toxin genes. While for other genes, such as *cry49*, *mtx1*, *mtx3* and *sphaericolysin*, the procedure did not produce the expected results because the sequenced cDNA obtained had not produced long enough fragments to reveal the 5'UTR of the genes. The results showed that the predicted promoter for *cry48*, a putative sigma-E promoter beginning at 33 nucleotides upstream the initiation codon (-10 sequence) and 54 nucleotides upstream the initiation codon (-35 sequence) (Jones and Berry, unpublished data), is not likely to be the promoter of the *cry48* toxin gene, since this putative promoter is located downstream from the transcript start. Database searches for a likely promoter candidate did not produce any results. It is possible that the promoter of the *cry48* gene is of a type not previously recognised. Further transcript mapping and mutagenesis of the region surrounding the transcript start identified may be necessary to resolve this issue.

The sequence 94 nucleotides upstream of the initiation codon (TTTAAT) and 67 nucleotides upstream of the initiation codon (TTGAAT) were proposed as a possible promoters of the *mtx2* toxin gene (Thanabalu and Porter, 1996). This work is consistent with the prediction that had been made in showing a likely transcription start site 49 nucleotides upstream of the start codon. These sequences appear to be similar to vegetative promoters, consistent with vegetative production of the Mtx2 toxin.

It had been proposed that the *binB* gene promoter exists in the region between the initiation codon and the *KpnI* site 99 nucleotides upstream (Baumann and Baumann, 1989). Another proposal, predicted that the promoter started 70 nucleotides upstream of the initiation codon (Ahmed et al., 1995). The present work succeeded in obtaining a 5'UTR as long as 67 nucleotides upstream the initiation codon, between

the putative -35 and -10 regions of the predicted promoter of Ahmed et al., indicating that this was not the correct promoter and suggesting that the promoter should be in the region between 96 nucleotides and 67 nucleotide upstream of the initiation codon. Unfortunately, database searches for a likely promoter candidate did not produce any results. It is possible that, similar to the promoter of the *cry48* gene, the *binB* promoter is a novel one and has not been registered yet.

Our work showed that the 5'RLM-RACE method combined with previous sequence data upstream of the gene of interest, could be employed to reveal the nature of some of the toxin gene promoters by identification of the start point of transcription. Out of eight targeted genes, 4 produced sequences without 5'UTR sequences. The work on *mtx4* gene did not produce any cDNA product. It was assumed that strain IAB59 would contain this gene but this has not been confirmed so this may explain this result. Three targeted genes produced 5'UTR sequences. In these cases, we were able to confirm the previously predicted promoter for the *mtx2* gene, but were still unable to resolve the exact promoters for the *cry48* and *binB* genes due to the lack of recognisable promoter sequences upstream of the start points discovered. To reveal the nature of the toxin gene promoters of *L. sphaericus*, a more direct method such cloning and expression of the toxin genes or reporter proteins linked to their promoter regions followed by mutagenesis may be required to define the promoters more accurately.

Genome sequencing

A draft chromosome sequence of *L. sphaericus* strain NHA15b was successfully constructed using paired-end sequencing data from the Illumina System.

The chromosome size of the strain was estimated to be at least 3,992,820 bp and shares a high similarity with the previously sequenced *L. sphaericus* genomes, in particular with the strain C3-41 (Hu et al., 2008a). However, in comparing the sequences, there is a prominent gap in the chromosome of strain NHA15b of approximately 35 kb in the region where C3-41 carries the *bin* toxins genes, which are known to be absent from NHA15b.

The whole genome sequencing of the NHA15b strain also detects the occurrence of 5 plasmids, hence named pLsph100 to pLsph500. The size of pLsph100 is 151,059 bp, somewhat smaller than plasmid pBsph of C3-41 (177,642 bp), since it lacks an approximately 35 kb region that harbour the *bin* toxin and *mtx4* genes in the C3-4 strain. The plasmid pLsph100 also looks similar to the recently reported plasmid pBsph-2 of *L. sphaericus* strain SSII-1 (Xu et al., 2015). Our analysis also predicted the occurrence of the plasmid in the closely related species of *L. fusiformis* strain SW-B9. This was the first time that the occurrence of pLsph or a similar plasmid was predicted to exist in a bacterium other than *L. sphaericus*. Based on the occurrence of T4SS genes, pLsph100, like pBsph and pBsph2, may be predicted to be a conjugative plasmid. According to the previous analysis, the T4SS genes on these plasmids are very different to the reference genes, so it is possible that they are no longer functional so that transfer of the plasmid from one bacterium to another would need some conjugation factors from other plasmids.

In the genome of NHA15b there is also a plasmid similar to the plasmid pLG of *L. sphaericus* of strain LP1-G. The plasmid was named as pLsph200. The plasmid size is slightly larger than pLG (12,466 bp compared to 11,066 bp of the pLG). Out of 23 CDSs of pLG, 15 CDSs were successfully transferred using RATT tool, meaning

that the two plasmids are similar. Interestingly enough, the predicted *rep* genes between the two plasmids are quite dissimilar. It would take a further analysis to confirm that the predicted *rep* genes in both of the plasmids are the true *rep* genes for the particular plasmid. Since the CDSs of the pLsph200 do not indicate any functional genes, it is predicted that the small plasmid pLsph200 is a cryptic plasmid.

There are two other cryptic plasmids predicted in the genome of *L. sphaericus* strain NHA15b, namely pLsph300 and pLsph400. The plasmid pLsph300 is 10,336 bp in size and it is predicted to contain 20 CDSs. Blast search showed that most CDS predicted a hypothetical proteins. One CDS showed similarity to a CopG family protein. The plasmid pLsph400 has a similar size (10,593 bp), and is predicted by BASYS annotation to contain 14 CDSs. Additional CDS next to BASYS00014 was similar to DNA replication protein *Vibrio anguillarum* (gi 504874585; WP_015061687.1) (E value = 0.0; Ident. = 85%). The presence of the DNA replication protein indicates that this fragment is indeed a plasmid.

The fifth plasmid found in the genome of NHA15b is a mega plasmid with a size of 50,525 bp. It is predicted to contain 182 CDSs. Some of the CDSs appear to indicate the occurrence of Type IV Secretion System (T4SS) characteristic of a conjugation plasmid. Similar to pLsph100, the T4SS genes on this plasmid are quite dissimilar to the reference gene; hence the ability of this plasmid to conjugate by itself is questionable. Based on the Basys annotation result and blastx analysis, the function of this plasmid cannot determined. Further experiments would be needed to understand its role on the bacteria that harbour this plasmid. Based on the sequencing coverage (20.28x) the plasmid is predicted to occur singly in the genome, a character common in the mega plasmids.

There are a number of proteins isolated from *L. sphaericus* that are toxic to insects. Some of the toxins are already known to occur in the bacteria *L. sphaericus* strain NHA15b. The genome sequencing showed the occurrence the *mtx* genes, particularly *mtx1*, *mtx2*, *mtx3*, and *mtx4* genes in strain NHA15b. It is also showed the occurrence of *sphaericolysin* gene with sequence similar to the previously reported data from *L. sphaericus* strains A3-2 and C3-41.

The whole genome sequencing of the NHA15b isolate confirmed that this isolate does not harbour any *bin* genes, in fact, the genome does not have a 35 kb duplicated DNA fragments containing the *bin* genes that in strain C3-41 exist in both of its chromosome and plasmid.

One of the main objectives of the genome sequencing of strain NHA15b was to attempt to resolve the location of the *cry* genes, and confirm whether the genes are in the chromosome or on a plasmid. The previous work of this thesis, using PFGE techniques to separate plasmids of various *L. sphaericus* strains and then probe the occurrence of the genes did not provide a firm conclusion. Previous work (Jones and Berry, unpublished data), succeeded in sequencing a 15 kb region of DNA containing the *cry* genes and a number of surrounding genes. In this work, the 15 kb fragment was able to be extended to 211,486 bp by PCR, linking several contigs that were predicted to be adjacent to each other. Annotation of contig 60 showed that part of the contig is identical to chromosomal sequence of strain C3-41, and another part of the contig harboured several genes indicating the existence of a conjugation plasmid. The two parts of the contig were connected by the assembly software. I speculate that the *cry* genes and their neighbour, the *mtx4* gene are located on a conjugation plasmid with a size bigger than 100 kb but less than 750 kb. The predicted plasmid would be likely

to occur singly in the genome of *L. sphaericus* strain NHA15b. To prove the prediction, we can try to physically isolate the plasmid using improved methods to detect the presence of the *cry* genes in the plasmid. Alternatively, we can PCR link each of the contigs available in the whole genome sequencing that are still not integrated into either chromosome or plasmid sequences and by trial and error, hopefully we will either assemble the elusive plasmid or confirm that the genes are located chromosomally. As new whole genome sequencing data from *L. sphaericus* and related bacteria appear in the databases we can continue to mine the *L. sphaericus* the genomes to discover the location of the *cry* genes.

Overall

The work that has been done in this study provides several new insights into the characteristics of *L. sphaericus*, a group of bacteria useful in the control of mosquitoes. In this work, a variety of methods were employed, ranging from electron microscopy, insect bioassay, biochemical and molecular analyses through to whole genome sequencing. In particular, the whole genome sequencing using the latest technology provided us with an opportunity to explore deeper into the biology of the *L. sphaericus* bacteria.

REFERENCES

- Ahmed, H.K., Mitchell, W.J., Priest, F.G., 1995, Regulation of mosquitocidal toxin synthesis in *Bacillus sphaericus*. *Appl Microbiol Biotechnol* 43, 310-314.
- Ahmed, I., Yokota, A., Yamazoe, A., Fujiwara, T., 2007, Proposal of *Lysinibacillus boronitolerans* gen. nov sp nov., and transfer of *Bacillus fusiformis* to *Lysinibacillus fusiformis* comb. nov and *Bacillus sphaericus* to *Lysinibacillus sphaericus* comb. nov. *Int J Syst Evol Micr* 57, 1117-1125.
- Alexander, B., Priest, F.G., 1990, Numerical Classification and Identification of *Bacillus-Sphaericus* Including Some Strains Pathogenic for Mosquito Larvae. *J Gen Microbiol* 136, 367-376.
- Baumann, L., Baumann, P., 1989, Expression in *Bacillus subtilis* of the 51- and 42-kilodalton mosquitocidal toxin genes of *Bacillus sphaericus*. *Appl Environ Microbiol* 55, 252-253.
- Baumann, P., Clark, M.A., Baumann, L., Broadwell, A.H., 1991, *Bacillus-Sphaericus* as a Mosquito Pathogen - Properties of the Organism and Its Toxins. *Microbiol Rev* 55, 425-436.
- Begum, M.A., Rahul, K., Sasikala, C., Ramana, C.V., 2016, *Lysinibacillus xyleni* sp. nov., isolated from a bottle of xylene. *Arch Microbiol*.
- Benson, G., 1999, Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27, 573-580.
- Berry, C., 2012, The bacterium, *Lysinibacillus sphaericus*, as an insect pathogen. *J Invertebr Pathol* 109, 1-10.
- Bockarie, M.J., Pedersen, E.M., White, G.B., Michael, E., 2009, Role of vector control in the global program to eliminate lymphatic filariasis. *Annu Rev Entomol* 54, 469-487.
- Broadwell, A.H., Baumann, P., 1986, Sporulation-Associated Activation of *Bacillus-Sphaericus* Larvicide. *Appl Environ Microb* 52, 758-764.
- Chan, S.W., Thanabalu, T., Wee, B.Y., Porter, A.G., 1996, Unusual amino acid determinants of host range in the Mtx2 family of mosquitocidal toxins. *J Biol Chem* 271, 14183-14187.
- Charles, J.F., Silva-Filha, M.H., Nielsen-LeRoux, C., Humphreys, M.J., Berry, C., 1997, Binding of the 51- and 42-kDa individual components from the *Bacillus sphaericus* crystal toxin to mosquito larval midgut membranes from *Culex* and *Anopheles* sp. (Diptera: Culicidae). *Fems Microbiol Lett* 156, 153-159.
- Cox-Singh, J., Davis, T.M., Lee, K.S., Shamsul, S.S., Matusop, A., Ratnam, S., Rahman, H.A., Conway, D.J., Singh, B., 2008, *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clin Infect Dis* 46, 165-171.
- Darboux, I., Nielsen-LeRoux, C., Charles, J.F., Pauron, D., 2001, The receptor of *Bacillus sphaericus* binary toxin in *Culex pipiens* (Diptera : Culicidae) midgut: molecular cloning and expression. *Insect Biochem Molec* 31, 981-990.

- Darboux, I., Pauchet, Y., Castella, C., Silva-Filha, M.H., Nielsen-LeRoux, C., Charles, J.F., Pauron, D., 2002, Loss of the membrane anchor of the target receptor is a mechanism of bioinsecticide resistance. *P Natl Acad Sci USA* 99, 5830-5835.
- de Melo, J.V., Jones, G.W., Berry, C., Vasconcelos, R.H.T., de Oliveira, C.M.F., Furtado, A.F., Peixoto, C.A., Silva-Filha, M.H.N.L., 2009, Cytopathological effects of *Bacillus sphaericus* Cry48Aa/Cry49Aa toxin on binary toxin-susceptible and -resistant *Culex quinquefasciatus* larvae. *Appl Environ Microb* 75, 4782-4789.
- Debarjac, H., Largetthiery, I., Dumanoir, V.C., Ripouteau, H., 1985, Serological Classification of *Bacillus-Sphaericus* Strains on the Basis of Toxicity to Mosquito Larvae. *Appl Microbiol Biot* 21, 85-90.
- del Solar, G., Giraldo, R., Ruiz-Echevarría, M.J., Espinosa, M., Díaz-Orejas, R., 1998, Replication and Control of Circular Bacterial Plasmids. *Microbiol Mol Biol R* 62, 434-464.
- Dias, S.C., Sagardoy, M.A., Silva, S.F., Dias, J.M.C.S., 1999, Characterization and pathogenic evaluation of *Bacillus thuringiensis* and *Bacillus sphaericus* isolates from Argentinean soils. *Biocontrol* 44, 59-71.
- Gammon, K., Jones, G.W., Hope, S.J., de Oliveira, C.M.F., Regis, L., Silva, M.H.N.L., Dancer, B.N., Berry, C., 2006, Conjugal transfer of a toxin-coding megaplasmid from *Bacillus thuringiensis* subsp *israelensis* to mosquitocidal strains of *Bacillus sphaericus*. *Appl Environ Microb* 72, 1766-1770.
- Goessweiner-Mohr, N., Arends, K., Keller, W., Grohmann, E., 2013, Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid* 70, 289-302.
- Goldberg, L.J., Ford, I., Tanabe, A.M., Watkins, H.M.S., 1977, Effectiveness of *Bacillus-Sphaericus* Var *Fusiformis* (Ssii-1) as a Potential Mosquito Larval Control Agent - Role of Variations in Natural Microbial-Flora in Larval Environment. *Mosq News* 37, 465-470.
- Guzman, M.G., Halstead, S.B., Artsob, H., Buchy, P., Farrar, J., Gubler, D.J., Hunsperger, E., Kroeger, A., Margolis, H.S., Martinez, E., Nathan, M.B., Pelegriño, J.L., Simmons, C., Yoksan, S., Peeling, R.W., 2010, Dengue: a continuing global threat. *Nat Rev Microbiol* 8, S7-16.
- Hofte, H., Whiteley, H.R., 1989, Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol Rev* 53, 242-255.
- Hu, X., Fan, W., Han, B., Liu, H., Zheng, D., Li, Q., Dong, W., Yan, J., Gao, M., Berry, C., Yuan, Z., 2008a, Complete genome sequence of the mosquitocidal bacterium *Bacillus sphaericus* C3-41 and comparison with those of closely related *Bacillus* species. *J Bacteriol* 190, 2892-2902.

- Hu, X., Li, J., Hansen, B.M., Yuan, Z., 2008b, Phylogenetic analysis and heterologous expression of surface layer protein SlpC of *Bacillus sphaericus* C3-41. *Biosci Biotechnol Biochem* 72, 1257-1263.
- Humphreys, M.J., Berry, C., 1998, Variants of the *Bacillus sphaericus* binary toxins: Implications for differential toxicity of strains. *J Invertebr Pathol* 71, 184-185.
- Jeong, H., Jeong, D.E., Sim, Y.M., Park, S.H., Choi, S.K., 2013, Genome Sequence of *Lysinibacillus sphaericus* Strain KCTC 3346T. *Genome announcements* 1.
- Jones, G.W., Nielsen-Leroux, C., Yang, Y., Yuan, Z., Dumas, V.F., Monnerat, R.G., Berry, C., 2007, A new Cry toxin with a unique two-component dependency from *Bacillus sphaericus*. *FASEB J* 21, 4112-4120.
- Jones, G.W., Wirth, M.C., Monnerat, R.G., Berry, C., 2008, The Cry48Aa-Cry49Aa binary toxin from *Bacillus sphaericus* exhibits highly restricted target specificity. *Environ Microbiol* 10, 2418-2424.
- Klein, D., Uspensky, I., Braun, S., 2002, Tightly bound binary toxin in the cell wall of *Bacillus sphaericus*. *Appl Environ Microbiol* 68, 3300-3307.
- Krych, V.K., Johnson, J.L., Yousten, A.A., 1980a, Deoxyribonucleic-Acid Homologies among Strains of *Bacillus-Sphaericus*. *Int J Syst Bacteriol* 30, 476-484.
- Krych, V.K., Johnson, J.L., Yousten, A.A., 1980b, Deoxyribonucleic acid homologies among strains of *Bacillus sphaericus*. *Int. J. Sys. Bacteriol.* 30, 476-484.
- Lacey, L.A., 2007, *Bacillus thuringiensis* serovariety *israelensis* and *Bacillus sphaericus* for mosquito control. *J Am Mosquito Contr* 23, 133-163.
- Lacroix, R., McKemey, A.R., Raduan, N., Kwee Wee, L., Hong Ming, W., Guat Ney, T., Rahidah, A.A.S., Salman, S., Subramaniam, S., Nordin, O., Hanum, A.T.N., Angamuthu, C., Marlina Mansor, S., Lees, R.S., Naish, N., Scaife, S., Gray, P., Labbe, G., Beech, C., Nimmo, D., Alphey, L., Vasan, S.S., Han Lim, L., Wasi, A.N., Murad, S., 2012, Open field release of genetically engineered sterile male *Aedes aegypti* in Malaysia. *PLoS One* 7, e42771.
- Liu, J.-W., Porter, A.G., Wee, B.Y., Thanabalu, T., 1996, New gene from nine *Bacillus sphaericus* strains encoding highly conserved 35.8-kilodalton mosquitocidal toxins. *Appl. Environ. Microbiol.* 62, 2174-2176.
- Marchesi, J.R., Sato, T., Weightman, A.J., Martin, T.A., Fry, J.C., Hiom, S.J., Wade, W.G., 1998, Design and Evaluation of Useful Bacterium-Specific PCR Primers That Amplify Genes Coding for Bacterial 16S rRNA. *Appl. Environ. Microbiol.* 64, 795-799.
- Massonnet-Bruneel, B., Corre-Catelin, N., Lacroix, R., Lees, R.S., Hoang, K.P., Nimmo, D., Alphey, L., Reiter, P., 2013, Fitness of transgenic

- mosquito *Aedes aegypti* males carrying a dominant lethal genetic system. PLoS One 8, e62711.
- Michalakakis, Y., Renaud, F., 2009, Malaria: Evolution in vector control. Nature 462, 298-300.
- Monnerat, R., da Silva, S.F., Dias, D.S., Martins, E.S., Praca, L.B., Jones, G.W., Soares, C.M., Dias, J.M.C.S., Berry, C., 2004, Screening of Brazilian *Bacillus sphaericus* strains for high toxicity against *Culex quinquefasciatus* and *Aedes aegypti*. J Appl Entomol 128, 469-473.
- Mulla, M.S., Thavara, U., Tawatsin, A., Kong-ngamsuk, W., Chompoonsri, J., Su, T., 2001, Mosquito larval control with *Bacillus sphaericus*: reduction in adult populations in low-income communities in Nonthaburi Province, Thailand. J Vector Ecol 26, 221-231.
- Nakamura, L.K., 2000, Phylogeny of *Bacillus sphaericus*-like organisms. Int J Syst Evol Micr 50, 1715-1722.
- Nielsen-Leroux, C., Charles, J.-F., Thiéry, I., Georghiou, G.P., 1995, Resistance in a laboratory population of *Culex quinquefasciatus* (Diptera: Culicidae) to *Bacillus sphaericus* binary toxin is due to a change in the receptor on midgut brush-border membranes. Eur. J. Biochem. 228, 206-210.
- Nielsen-LeRoux, C., Rao, D.R., Murphy, J.R., Carron, A., Mani, T.R., Hamon, S., Mulla, M.S., 2001, Various levels of cross-resistance to *Bacillus sphaericus* strains in *Culex pipiens* (Diptera : Culicidae) colonies resistant to B-sphaericus strain 2362. Appl Environ Microb 67, 5049-5054.
- NielsenLeroux, C., Pasquier, F., Charles, J.F., Sinigre, G., Gaven, B., Pasteur, N., 1997, Resistance to *Bacillus sphaericus* involves different mechanisms in *Culex pipiens* (Diptera: Culicidae) larvae. J Med Entomol 34, 321-327.
- Nishiwaki, H., Nakashima, K., Ishida, C., Kawamura, T., Matsuda, K., 2007, Cloning, functional characterization, and mode of action of a novel insecticidal pore-forming toxin, sphaericolysin, produced by *Bacillus sphaericus*. Appl Environ Microbiol 73, 3404-3411.
- Oei, C., Hindley, J., Berry, C., 1992, Binding of purified *Bacillus sphaericus* binary toxin and its deletion derivatives to *Culex quinquefasciatus* gut: elucidation of functional binding domains. J Gen Microbiol 138, 1515-1526.
- Palma, L., Munoz, D., Berry, C., Murillo, J., Caballero, P., 2014, *Bacillus thuringiensis* toxins: an overview of their biocidal activity. Toxins (Basel) 6, 3296-3325.
- Park, H.W., Bideshi, D.K., Wirth, M.C., Johnson, J.J., Walton, W.E., Federici, B.A., 2005, Recombinant larvicidal bacteria with markedly improved efficacy against culex vectors of west Nile virus. Am J Trop Med Hyg 72, 732-738.
- Pei, G.F., Oliveira, C.M.F., Yuan, Z.M., Nielsen-LeRoux, C., Silva-Filha, M.H., Yan, J.P., Regis, L., 2002, A strain of *Bacillus sphaericus*

- causes slower development of resistance in *Culex quinquefasciatus*. *Appl Environ Microb* 68, 3003-3009.
- Pena-Montenegro, T.D., Lozano, L., Dussan, J., 2015, Genome sequence and description of the mosquitocidal and heavy metal tolerant strain *Lysinibacillus sphaericus* CBAM5. *Stand Genomic Sci* 10, 2.
- Pena, G., Miranda-Rios, J., de la Riva, G., Pardo-Lopez, L., Soberon, M., Bravo, A., 2006, A *Bacillus thuringiensis* S-layer protein involved in toxicity against *Epilachna varivestis* (Coleoptera: Coccinellidae). *Appl Environ Microbiol* 72, 353-360.
- Poncet, S., Bernard, C., Dervyn, E., Cayley, J., Klier, A., Rapoport, G., 1997, Improvement of *Bacillus sphaericus* toxicity against dipteran larvae by integration, via homologous recombination, of the Cry11A toxin gene from *Bacillus thuringiensis* subsp. israelensis. *Appl Environ Microbiol* 63, 4413-4420.
- Porwal, S., Lal, S., Cheema, S., Kalia, V.C., 2009, Phylogeny in Aid of the Present and Novel Microbial Lineages: Diversity in *Bacillus*. *Plos One* 4, -.
- Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R., Holtzapple, E.K., Okstad, O.A., Helgason, E., Rilstone, J., Wu, M., Kolonay, J.F., Beanan, M.J., Dodson, R.J., Brinkac, L.M., Gwinn, M., DeBoy, R.T., Madpu, R., Daugherty, S.C., Durkin, A.S., Haft, D.H., Nelson, W.C., Peterson, J.D., Pop, M., Khouri, H.M., Radune, D., Benton, J.L., Mahamoud, Y., Jiang, L.X., Hance, I.R., Weidman, J.F., Berry, K.J., Plaut, R.D., Wolf, A.M., Watkins, K.L., Nierman, W.C., Hazen, A., Cline, R., Redmond, C., Thwaite, J.E., White, O., Salzberg, S.L., Thomason, B., Friedlander, A.M., Koehler, T.M., Hanna, P.C., Kolsto, A.B., Fraser, C.M., 2003, The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 423, 81-86.
- Reinert, D.J., Carpusca, I., Aktories, K., Schulz, G.E., 2006, Structure of the mosquitocidal toxin from *Bacillus sphaericus*. *J Mol Biol* 357, 1226-1236.
- Reyes-Ramirez, A., Ibarra, J.E., 2008, Plasmid patterns of *Bacillus thuringiensis* type strains. *Appl Environ Microbiol* 74, 125-129.
- Romao, T.P., de Melo Chalegre, K.D., Key, S., Ayres, C.F., Fontes de Oliveira, C.M., de-Melo-Neto, O.P., Silva-Filha, M.H., 2006, A second independent resistance mechanism to *Bacillus sphaericus* binary toxin targets its alpha-glucosidase receptor in *Culex quinquefasciatus*. *Febs J* 273, 1556-1568.
- Schwartz, J.L., Potvin, L., Coux, F., Charles, J.F., Berry, C., Humphreys, M.J., Jones, A.F., Bernhart, I., Dalla Serra, M., Menestrina, G., 2001, Permeabilization of model lipid membranes by *Bacillus sphaericus* mosquitocidal binary toxin and its individual components. *J Membrane Biol* 184, 171-183.

- Servant, P., Rosso, M.L., Hamon, S., Poncet, S., Delecluse, A., Rapoport, G., 1999, Production of Cry11A and Cry11Ba toxins in *Bacillus sphaericus* confers toxicity towards *Aedes aegypti* and resistant *Culex* populations. *Appl Environ Microb* 65, 3021-3026.
- Sgarrella, F., Szulmajster, J., 1987, Purification and characterization of the larvicidal toxin of *Bacillus sphaericus* 1593M. *Biochem Biophys Res Commun* 143, 901-907.
- Shintani, M., Sanchez, Z.K., Kimbara, K., 2015, Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front Microbiol* 6, 242.
- Silva-Filha, M.H., Nielsen-Leroux, C., Charles, J.F., 1997, Binding kinetics of *Bacillus sphaericus* binary toxin to midgut brush-border membranes of *Anopheles* and *Culex* sp. mosquito larvae. *Eur J Biochem* 247, 754-761.
- Silva-Filha, M.H., Nielsen-LeRoux, C., Charles, J.F., 1999, Identification of the receptor for *Bacillus sphaericus* crystal toxin in the brush border membrane of the mosquito *Culex pipiens* (Diptera : Culicidae). *Insect Biochem Molec* 29, 711-721.
- Singer, S., 1988, Clonal Populations with Special Reference to *Bacillus-Sphaericus*. *Adv Appl Microbiol* 33, 47-74.
- Srisucharitpanit, K., Yao, M., Promdonkoy, B., Chimnaronk, S., Tanaka, I., Boonserm, P., 2014, Crystal structure of BinB: a receptor binding component of the binary toxin from *Lysinibacillus sphaericus*. *Proteins* 82, 2703-2712.
- Swain, M.T., Tsai, I.J., Assefa, S.A., Newbold, C., Berriman, M., Otto, T.D., 2012, A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc* 7, 1260-1284.
- Takken, W., Knols, B.G., 2009, Malaria vector control: current and future strategies. *Trends Parasitol* 25, 101-104.
- Tangsongcharoen, C., Chomanee, N., Promdonkoy, B., Boonserm, P., 2015, *Lysinibacillus sphaericus* binary toxin induces apoptosis in susceptible *Culex quinquefasciatus* larvae. *J Invertebr Pathol* 128, 57-63.
- Thanabalu, T., Hindley, J., Brenner, S., Oei, C., Berry, C., 1992, Expression of the mosquitocidal toxins of *Bacillus sphaericus* and *Bacillus thuringiensis* subsp. *israelensis* by recombinant *Caulobacter crescentus*, a vehicle for biological control of aquatic insect larvae. *Appl Environ Microbiol* 58, 905-910.
- Thanabalu, T., Hindley, J., Jacksonyap, J., Berry, C., 1991, Cloning, Sequencing, and Expression of a Gene Encoding a 100-Kilodalton Mosquitocidal Toxin from *Bacillus-Sphaericus* Ssii-1. *Journal of Bacteriology* 173, 2776-2785.
- Thanabalu, T., Porter, A.G., 1995, Efficient Expression of a 100-Kilodalton Mosquitocidal Toxin in Protease-Deficient

- Recombinant *Bacillus-Sphaericus*. *Appl Environ Microb* 61, 4031-4036.
- Thanabalu, T., Porter, A.G., 1996, A *Bacillus sphaericus* gene encoding a novel type of mosquitocidal toxin of 31.8 kDa. *Gene* 170, 85-89.
- Van Domselaar, G.H., Stothard, P., Shrivastava, S., Cruz, J.A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R., Wishart, D.S., 2005, BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res* 33, W455-459.
- WHO 2008. World Malaria Report 2008 (WHO).
- WHO 2009. Dengue: guidelines for diagnosis, treatment, prevention and control -- New edition (Geneva, World Health Organization).
- WHO 2010. Progress report 2000–2009 and strategic plan 2010–2020 of the global programme to eliminate lymphatic filariasis: halfway towards eliminating lymphatic filariasis (Geneva, World Health Organization).
- Wirth, M.C., Federici, B.A., Walton, W.E., 2000, Cyt1A from *Bacillus thuringiensis* synergizes activity of *Bacillus sphaericus* against *Aedes aegypti* (Diptera : Culicidae). *Appl Environ Microb* 66, 1093-1097.
- Wirth, M.C., Yang, Y., Walton, W.E., Federici, B.A., Berry, C., 2007, Mtx toxins synergize *Bacillus sphaericus* and Cry11Aa against susceptible and insecticide-resistant *Culex quinquefasciatus* larvae. *Appl Environ Microb* 73, 6066-6071.
- Wu, E., Jun, L., Yuan, Y., Yan, J., Berry, C., Yuan, Z., 2007, Characterization of a cryptic plasmid from *Bacillus sphaericus* strain LP1-G. *Plasmid* 57, 296-305.
- Xu, K., Yuan, Z., Rayner, S., Hu, X., 2015, Genome comparison provides molecular insights into the phylogeny of the reassigned new genus *Lysinibacillus*. *BMC Genomics* 16, 140.
- Yang, Y., Wang, L., Gaviria, A., Yuan, Z., Berry, C., 2007, Proteolytic stability of insecticidal toxins expressed in recombinant bacilli. *Appl Environ Microbiol* 73, 218-225.
- Yousten, A.A., 1984, Bacteriophage-Typing of Mosquito Pathogenic Strains of *Bacillus-Sphaericus*. *J Invertebr Pathol* 43, 124-125.
- Yuan, Z., Hu, X.M., Fan, W., Han, B., Liu, H.Z., Zheng, D., Li, Q., Dong, W., Yan, J., Gao, M., Berry, C., 2008, Complete genome sequence of the mosquitocidal bacterium *Bacillus sphaericus* C3-41 and comparison with those of closely related bacillus species. *Journal of Bacteriology* 190, 2892-2902.
- Yuan, Z., Rang, C., Maroun, R.C., Juarez-Perez, V., Frutos, R., Pasteur, N., Vendrely, C., Charles, J.F., Nielsen-Leroux, C., 2001, Identification and molecular structural prediction analysis of a toxicity determinant in the *Bacillus sphaericus* crystal larvicidal toxin. *Eur J Biochem* 268, 2751-2760.

- Yuan, Z., Zhang, Y.M., Cai, Q.X., Liu, E.Y., 2000, High-level field resistance to *Bacillus sphaericus* C3-41 in *Culex quinquefasciatus* from Southern China. *Biocontrol Sci. Techn.* 10, 41-49.
- Zahiri, N.S., Mulla, M.S., 2003, Susceptibility profile of *Culex quinquefasciatus* (Diptera: Culicidae) to *Bacillus sphaericus* on selection with rotation and mixture of *B. sphaericus* and *B. thuringiensis israelensis*. *J Med Entomol* 40, 672-677.
- Zerbino, D.R., 2010, Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics* Chapter 11, Unit 11 15.

APPENDIX

16S ribosomal RNA

Forward primer 63f 5'-CAG GCC TAA CAC ATG CAA GTC-3'

Reverse primer 1387r 5'-GGG CGG WGT GTA CAA GGC-3'

SSurface layer protein C primer:

Forward primer 4 *Bam*HI F

5'-GCAGGATCCGCAAAGCAAACAAAGGCCGTA-3'

Reverse primer 3594 *Sac*I R

5'-GTAGAGCTCGGTTTTTCATTTCCACTCCA-3'

55'RLM-RACE primer:

Outer primer

5'-GCTGATGGCGATGAATGAACACTG-3'

Inner primer

5'-CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG-3'

Sphaericolysin

Outer primer 5'-GTTCTACCATAAGCTACATTCG-3'

Inner primer 5'-CATTCAGTCTAGTACTTCCC-3'

Mtx1

Outer primer 5'-TTAGGTATTTCTCCATTGGGG-3'

Inner primer 5'-TGTATCTAGCTCTAGTTGTGG-3'

Inner most primer 5'-GATACAATAAACTCGACATATCC-3'

Mtx2

Outer primer 5'-TCTTTAACAATCTATTATTCTGGC-3'

Inner primer 5'-AATTCCCACCTTTTCCTGATGC-3'

Inner most primer 5'-AAAAAGTGGTGGGATACTAGC-3'

Mtx3

Outer primer 5'-CCGTATTCTGATTCGAACCG-3'

Inner primer 5'-GAAAAGGTATCCCTGTTGCC-3'

Inner most primer 5'-AATTAAGTGGTGGGATACTAGC-3'

Mtx4

Outer primer 5'-CTACAACAACAAAAGTTGCACC-3'

Inner primer 5'-ACACTAGTTGATGTGGATTTCG-3'

Inner most primer 5'-TAATAGCTCCCTATCTTCTCG-3'

BinB

Outer primer 5'-TCAGATATTCCTGTTTCGTTCC-3'

Inner primer 5'-AATCTCCACTACCTACCTG-3'

Inner most primer 5'-TTGATAAACCATTAACCATTCGG-3'

Cry48Aa1

Outer primer 5'-ATATCCCACGTCCATTTCCGG-3'

Inner primer 5'-ATTCAATCCAATCTTGTTTCGC-3'

Inner most primer 5'-TACGACAATAGTTCCGCTCG-3'

Cry49Aa1

Outer primer 5'-TATCTGTCCATATTCTGTGCC-3'

Inner primer 5'-TATTCTCTAGATACTATTACACC-3'

Inner most primer 5'-AAATCTAGTACTTTGTTCTGGC-3'