THE MOLECULAR BASIS OF *PASTEURIA*-NEMATODE INTERACTIONS USING CLOSELY RELATED *BACILLUS* SPP.

Arohi Srivastava, B.Sc., M.Sc.

Submitted to the University of Hertfordshire in partial fulfilment of the requirements of the degree of Doctor of Philosophy

School of Life and Medical Sciences, University of Hertfordshire, Hatfield, Herts, AL10 9AB

April 2017

To Mumma

for being my first teacher

and Papa

for his strong believe in me.

Acknowledgments

> "You can teach a student a lesson for a day; but if you can teach him to learn
> by creating curiosity, he will continue the learning process as long as he lives."
>
> -- Clay P. Bedford

Contents

Abstract

Phytonematodes are known to cause substantial losses in crop yields across the world. Since the middle of the last century, these pests have been adequately controlled by chemical nematicides. However, due to increasing public health concern, strict regulations in the EU and elsewhere have significantly reduced the usage of these environmentally not-so-safe chemicals. This has led us to look for reliable biological alternatives. The *Pasteuria* group of Gram-positive endospore-forming bacteria (phylum: Firmicutes) often associated with nematode-suppressive soils are potentially reliable nematode biocontrol agents. However, the highly specific interaction of *Pasteuria* to their nematode hosts poses a challenge to the management of heterogeneous populations of nematodes in the field; the mechanism behind this specificity remains unclear. One of the fundamental basis of host specificity is the attachment of *Pasteuria* endospores to the cuticle of their host nematodes which is the first and essential step in the infection process. Thus, understanding the molecular mechanisms that govern the attachment process is important in identifying suitable populations of *Pasteuria* for effective broad-range management of plant parasitic nematodes in soil. Previous studies suggest the presence of immunogenic collagen-like fibres and carbohydrates on the endospore coat of *Pasteuria* that may have a role in the initial interaction of the endospores with their nematode hosts. Published work on phylogeny relates *Pasteuria* to *Bacillus* spp. most of which have well annotated and characterized genomes while the genome of *Pasteuria* remains to be sequenced completely.

In this thesis, I attempt to explore the endospore biology of obligate and fastidious *Pasteuria* spp. using the wide knowledgebase of well studied *Bacillus* endospores. The primary aim was to characterize the immunogeneic determinants that are possibly responsible for the attachment of *Pasteuria* endospores to the host nematode cuticle by a combination of computational and lab-based approaches. To approve the suggested phylogenetic closeness of *Pasteuria* to *Bacillus*, the first part of the study focused on phylogeny reconstruction of *Pasteuria* spp. amongst *Bacillus* spp. and other members of the phylum Firmicutes. This was followed by *in silico* studies to identify candidate collagen-like genes in *P. penetrans*; the putative functional proteins encoded by these candidate genes were then comparatively characterized with

collagens from other organisms including the members of the genus *Bacillus.* The surface associated collagen-like proteins and other possible immunogens on the endospores of *Pasteuria* were characterized by protein immunoblotting, lectin blotting and immunofluorescence microscopy and comparisons were made with *B. thuringiensis* endospores. Lastly, endospore attachment assays were done to test the hypothesis that collagens and carbohydrates play a role in *Pasteuria* endospore attachment.

The results of the computational analyses suggest a family of collagen coding putative genes in the *Pasteuria* genome, all of which are predicted to have varied biochemical properties and are seemingly of diverse evolutionary origin. The Western blot and microscopic analyses show that the endospores of *P. penetrans* and *B. thuringiensis* share some common immunodominant surface epitopes. The attachment assays confirm the involvement of collagens and at least one carbohydrate (N-acetylglucosamine) in the endospore attachment. However, the results also indicate possible involvement of other adhesins in the process; to support this, at the end of the thesis, I propose a new 'Multitype Adhesin Model' for initial interaction of *Pasteuria* endospores with the cuticle of their host nematodes.

The outcomes of this project will help in identifying the molecular basis of the complex *Pasteuria*-nematode interaction. This will provide a basis to develop environmentally benign nematode bio-management strategies.

List of Figures

List of Tables

List of Appendices

Appendix XII: List of Abbreviations

ANOVA	Analysis of Variance
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BLAST	Basic Local Alignment Search Tool
CL	Collagen-like
CLP	Collagen-like sequences
CTCF	Corrected Total Cell Fluorescence
CTD	C-terminal domain
DEA	Diethanolamine
DTT	Dithiothreitol
FITC	Fluorescein isothiocyanate
GSS	Genome Survey Sequences
HGT	Horizontal Gene Transfer
J2	Second stage juvenile
J2Anti-PpWS	Second stage juvenile treated with Anti-PpWS antibody
J2Col1981	Second stage juvenile treated with Col1981 antibody
J2Col1982	Second stage juvenile treated with Col1982 antibody
J2Collagenase	Second stage juvenile treated with collagenase
J2NAGase	Second stage juvenile treated with NAGase
J2WGA	Second stage juvenile treated with WGA
kDa	Kilo dalton
LCR	Low complexity region
MSCRAMMs	Microbial Surface Components Recognizing Adhesive Matrix Molecules
NAG	N-acetylglucosamine
NAGase	N-acetylglucosaminidase
NCBI	National Center for Biotechnology Information
NTD	N-terminal domain
PAB	Polyclonal Antibody
PasAnti-PpWs	*Pasteuria* endospores treated with Anti-PpWS antibody
PasCol1981	*Pasteuria* endospores treated with Col1981 antibody
PasCol1982	*Pasteuria* endospores treated with Col1982 antibody
PasCollagenase	*Pasteuria* endospores treated with collagenase
PasNAGse	*Pasteuria* endospores treated with NAGase
PasWGA	*Pasteuria* endospores treated with WGA

PBS	Phosphate bufferd Saline
PBST	Phosphate bufferd Saline with Tween
PPN	Plant Parasitic Nematode
PVDF	Polyvinylidene fluoride
RAST	Rapid Annotations using Subsystems
RKN	Root knot nematodes
rpm	Rotations per minute
SASPs	Small Acid Soluble Proteins
SDS-PAGE	Sodium Dodecyl Sulphate
spp.	Species (plural)
TTBS	Tris Tween Buffered Saline
WGA	Wheat germ agglutinin

Chapter 1

Introduction

1.1 FOOD SECURITY

With the enormously increasing human population, predicted to be around 9.7 billion by 2050, a major global challenge for the scientific community is to ensure food security (United Nations, 2015). The latest FAO report on food security suggests that every ninth person in the world suffers from chronic hunger, and around 12.9% of the global population is unable to meet their dietary energy requirements (FAO, 2015). Attempts to meet the dietary needs of today have led to agricultural intensification which in turn is predicted to alter biotic stresses due to pests, pathogens, and weeds. Moreover, climate change is very likely to influence the distribution and biology of plant pests, pathogens and weeds in unpredictable ways (Coakley *et al.*, 1999, Dukes *et al.*, 2009, Gregory *et al.*, 2009, War *et al.*, 2016).

1.2   NEMATODES: OFTEN UNNOTICED AGRICULTURAL PESTS

Nematodes (Phylum: *Nematoda*) are the most ancient, abundant and diverse multicellular animals on the earth. The oldest fossil nematodes indicate that these ubiquitous organisms are at least more than 135 million years old (Poinar Jr *et al.*, 1994, Manum *et al.*, 1994). Since, nematodes are soft-bodied, only a fraction of them have been fossilized, and not much can be certainly said about the origin and ancestry of nematodes. However, nematodes have always been historically associated with several human diseases (Levine, 1968, Garcia and Bruckner, 1997, Yorke and Maplestone, 1926, Rodger, 1895). The earliest found description of the human parasitic nematodes in ancient Chinese literature dates back to 2700 B.C. (Maggenti, 1981a). Plant parasitic nematodes are considered to be even more ancient (253 B.C.) based on another ancient Chinese documented usage of a symbol for a soybean root infesting parasite that resembles an adult female of a soybean cyst nematode (Noel *et al.*, 1992, Morse *et al.*, 1949). In general, nematodes can thrive in all conceivable habitats from freshwater to marine water, from alkaline soils to acidic soils, from tropics to poles, from deserts to marshlands (Norton, 1978, Ullberg and Ólafsson, 2003, Bird and Koltai, 2000, Heip *et al.*, 1985, Goodey, 1963, Hauquier *et al.*, 2016, Gambi and Danovaro, 2016, Suzuki *et al.*, 2017, Rocka, 2017). They can be free-living

(herbivores, microbivores or predators) or can live in or on other organisms as commensals or parasites (Yeates *et al.*, 1993, Jensen, 1987, Freckman and Caswell, 1985, Bongers and Bongers, 1998). Over 4,100 species of nematodes, belonging to the orders *Tylenchida* and *Dorylaimida,* are known to parasitize plants (Decraemer *et al.*, 2006). Phytonematodes or plant parasitic nematodes (PPN), being devastating parasites of crops, represent a global threat to food security parasitizing all domesticated crop plants and under conducive circumstances causing devastation of crops with major socio-economic impacts worldwide. They are polyphagous i.e. there is hardly any flora including cereals, fibers, ornamentals, oilseeds, pulses, vegetables, fruits, etc. that remains unaffected by PPN.

Of the total crop losses incurred due to pests and diseases, one third is attributed to PPN; each year PPN infestation causes between US $80 -100 billion global crop losses (Nicol *et al.*, 2011), while in India the losses are estimated to be US $40 million (ICAR 2014). However, these are only conservative estimates and may be higher as in many of the developing nations the farming community is totally unaware of PPNs. Despite the huge losses, they incur to world agriculture, they are often overlooked, even by experienced and advanced growers, due to a lack of distinct symptoms. They disrupt water relations, and nutrient uptake ability of plants is compromised (Elkins *et al.*, 1979). Hence their symptoms are often confused with nutrient deficiency. PPNs adversely affect the physiology and growth of the infested plants in different ways. Apart from causing direct damage to plants, nematodes predispose the infested plants to bacterial, fungal and viral diseases. Some nematodes often act as vectors to some viruses (Powell, 1971, Taylor and Brown, 1997). The nematode acquires the virus when it feeds on a virus–infected plant and transmits it when it feeds on a healthy plant. Indeed, nematodes are major threats to agriculture but because they are small microscopic organisms present mostly below ground their importance is underestimated. In the global contexts of climate change and food security, plant parasitic nematodes need to be recognized as important agricultural pests equally noteworthy as any other pests or pathogens.

1.3 SEDENTARY ENDOSPARASITIC PHYTONEMATODES

Nematodes that parasitize plant roots can be ectoparasitic which remain outside the plant host and feed on the plant cells using their stylet; semi-endoparasitic which

partially penetrate the plant cell and feed through it at least once in their life cycle; migratory endoparasitic which feed by destructively migrating through root tissues; sedentary endoparasitic that remain sedentary once they establish a feeding site inside the root tissues (Maggenti, 1981b). Sedentary endoparasites belong to the family Heteroderidae (Wouts, 1972). These consist of the root knot nematodes (*Meloidogyne* spp.) and the cyst nematodes (*Heterodera* spp. and *Globodera* spp.). Characteristically, the adult females of these species are saccate and sedentary while the adult males are vermiform and mobile.

Meloidogyne spp., commonly known as root knot nematodes (RKN), cause extensive economic damage worldwide. First reported on cucumbers in an English glasshouse by Berkeley (1855), these pests have been associated with a wide range of host plants and are considered one of the world's most globally distributed crop pests (Hockland *et al.*, 2006). The genus *Meloidogyne* includes more than 90 described species (Hunt and Handoo, 2009). Of these, four species viz. *M. incognita*, *M. javanica*, *M. hapla* and *M. arenaria* are pests of global importance and collectively represent about 95% of the RKN species encountered in agricultural fields (Eisenback and Triantaphyllou, 1991).

The cyst nematodes are the nematodes from the two genera *Heterodera* spp. and *Globodera* spp. which include the most devastating soybean cyst nematodes (*H. glycines*), the cereal cyst nematodes (*H. avenae*) and the potato cyst nematodes (*G. pallida* and *G. rostochiensis*). Cyst nematodes are particularly problematic because of their ability to survive for prolonged periods in the soil when the infective juvenile of the nematode can remain within the hard cysts. Hatching of the juveniles occurs in response to host-derived chemical cues (Perry, 1996); this means that the juvenile lies in a dormant form within the cyst unless it receives some triggering signals from the host roots. Thus, the cyst nematodes can remain dormant in the soil for several years, reportedly up to 20 years (Grainger, 1964).

1.3.1 Life Cycle and Development

The life cycle of the sedentary endoparasitic nematodes has been extensively studied by various researchers across the world. It consists of morphogenetically distinct stages: an embryonic stage, four juvenile stages, and an adult stage. The embryo of a female develops into a first-stage juvenile (J1) which moults inside the egg and emerges as a second-stage juvenile (J2) as the egg hatches. *Figure 1.1* shows a

pictorial representation of the life cycles of the cyst nematodes and the root knot nematodes (Williamson and Gleason, 2003, De Guiran and Ritter, 1979, Chitwood and Perry, 2009, Turner *et al.,* 2006, Williamson and Hussey, 1996).

The infective juvenile, J2, leaves the plant root and moves freely through the soil in search of a new host root meanwhile using the lipids stored in its gut (Eisenback and Triantaphyllou, 1991). It invades a susceptible root tip with the help of cell-wall degrading enzymes and secretions from its subventral and oesophageal glands (Chitwood and Perry, 2009, Smant *et al.,* 1998, Bohlmann and Sobczak, 2014, Caillaud *et al.,* 2008). Once the juvenile reaches the vascular cylinder and establishes a feeding site, it becomes sedentary and starts feeding (Wyss *et al.,* 1992). The juvenile of a root knot nematode induces the formation of giant cells resulting by repeated nuclear divisions in the root tissue (Bird, 1961, Bird, 1962, Rodiuc *et al.,* 2014). The juvenile of a cyst nematode forms syncytia induced by the coordinated enzymatic breakdown of neighbouring root cell walls which coalesces the cell cytoplasm of adjacent cells together (Jones and Northcote, 1972, Endo, 1971, Rodiuc *et al.,* 2014). This results in the formation of galls/cysts, the typical symptoms of root knot/cyst nematode infestation (Rodiuc *et al.,* 2014).

While remaining embedded within the galls/ cysts, the J2 undergoes morphological changes and becomes flask-shaped (Taylor and Sasser, 1978). It stops feeding and moults into a third-stage juvenile (J3) and then into a fourth-stage juvenile (J4) and finally into an adult. In RKN, adult males are rare as they mostly reproduce by parthenogenesis/ apomixes (Trudgill and Blok, 2001), whereas the cyst nematodes follow an obligate sexual life cycle/ amphimixis (Turner *et al.,* 2006). The vermiform males migrate out of the roots. The females remain sedentary, start feeding again, enlarge considerably and become globose/saccate. The female body of cyst nematodes breaks through the plant root and become visible on the root surface. The RKN females lay several eggs in a gelatinous matrix that protrudes outside the nematode body. In contrast, the eggs of the cyst nematode females are retained inside the female body. The length of the life-cycle is variable depending on the soil temperature, host susceptibility, soil type and climate (Eisenback and Triantaphyllou, 1991, Madulu and Trudgill, 1994).

1.3.2 Host Range

Meloidogyne spp. are polyphagous and are known to parasitize more than 3,000 plant species including monocotyledonous, dicotyledonous, woody and herbaceous plants (Hussey and Janssen, 2002). Most vegetable crops including tomato, cucumber, melon, carrot, aubergine, okra, lettuce, peas and peppers are highly susceptible to these pests (Sasser *et al.*, 1983, Goodey *et al.*, 1965). However, some crops such as garlic, cassava, maize, sesame, sorghum, sudan grass are naturally resistant to root knot nematodes (Dobson *et al.*, 2002).

The host range of soybean cyst nematodes (SCNs) ranges from legumes to some ornamental plants such as poppies, *Delphinium*, snapdragon and *Verbena* to some weed species like wild mustard and hop clovers (Davis and Mitchum, 2005, Venkatesh *et al.*, 2009).The potato cyst nematodes (PCNs) are known to parasitize mostly solanaceous plants like potato, sweet potato, eggplant (aubergine), and tomato (Trudgill *et al.*, 1975). The cereal cyst nematodes (CCNs) are serious pests of graminaceous crops worldwide (Cook, 1982).
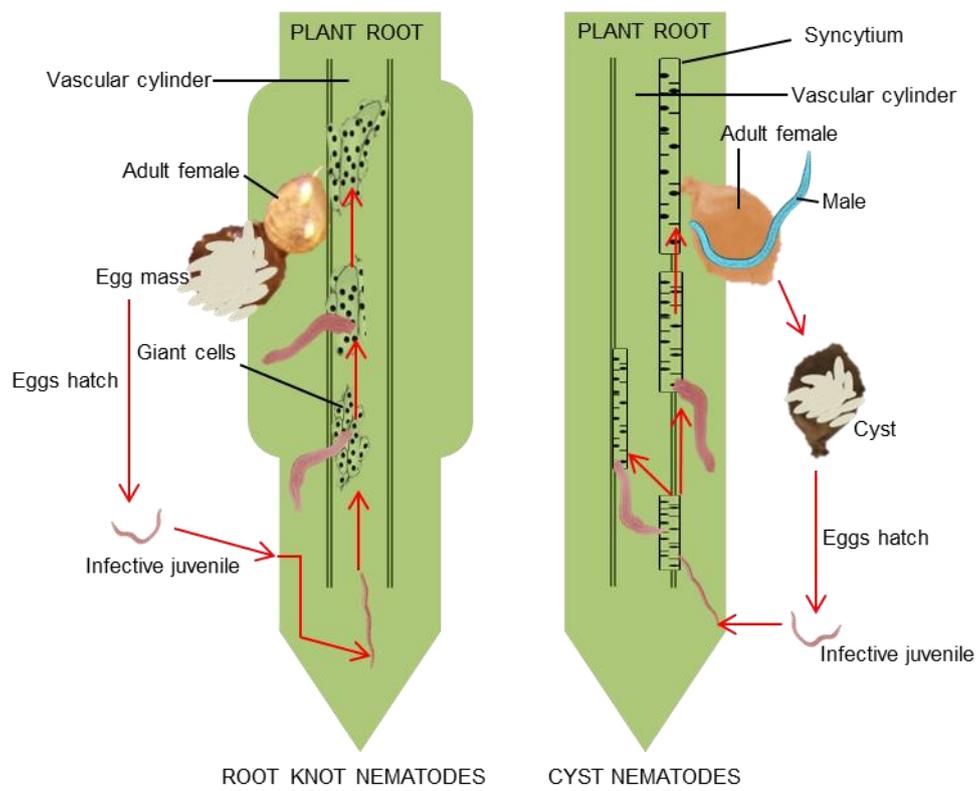
Figure 1. 1: Life cycles of root knot nematodes in comparison with cyst nematodes. The figure is not to scale.

1.4   NEMATODE MANAGEMENT

1.4.1 Traditional methods of nematode control

1) **Crop Rotation**: Traditionally PPNs, when recognized, were controlled by crop rotation where host crops were grown in rotation with non-host crops (Nusbaum and Ferris, 1973, Rodriguez-Kabana and Canullo, 1992). This decreases the nematode population densities below damage thresholds. It still remains the most widely used natural method of PPN management but cannot be solely relied upon where there is an incidence of polyphagous species of nematodes(Kratochvil *et al.*, 2004). Additionally, the rotation crop may result in a facilitated increase in the population density of an alternative nematode species present within the soil (McSorley *et al.*, 1994). However practicing crop rotation as a part of an integrated system of PPN management is a good idea indeed.

2) **Chemical Nematicides**: Since the middle of the last century these pests, in developed agriculture, have been controlled by nematicides which paralyze or kill nematodes (Sikora and Marczok 2005). These chemicals do not guarantee a hundred percent control of nematodes but reduce their number to a level that reduces plant damage and increases crop productivity (Tobin *et al.*, 2008). Many of these chemicals are highly toxic and currently being legislated against and withdrawn from use in Europe, the USA and elsewhere (Noling and Becker, 1994).

3) **Antagonistic Plants**: An ecologically safe method of PPN control is the use of antagonistic plants such as marigold, chrysanthemum and mustard (Hackney and Dickerson, 1975). These plants release allelopathic bioactive compounds known to have nematicidal effects. Their field efficacy is still questionable as it has been reported that their effect is restricted to upper layers of soil while the nematodes may survive deep into the soil (Ploeg, 2002).

4) **Fallowing and flooding**: Some suggest a combination of fallowing and flooding as an effective control measure (Sarah, 1989, Newhall, 1955, Brown, 1933) but both the options are uneconomical and impractical in most situations. Additionally they have deleterious effects on soil organic matter and soil structure (Ponnamperuma, 1984).

5) **Soil solarization**: This is another simple and economical alternative available where radiant heat from the sun is trapped in soil using clear plastic sheets (Oka *et al.*, 2007, Stapleton and DeVay, 1995). Over a period of time the temperature of

7

the soil beneath the plastic sheath rises to a level lethal to PPN along with several other pests and pathogens. The process, however, may also adversely affect the beneficial microflora of soil including the plant growth promoting rhizobacteria that are sensitive to high heats (Scopa and Dumontet, 2007).

1.4.2 Nematode Resistant GM Crops

Genetically modified plants made by manipulating, inserting or silencing specific genes and, thus, making the plants resistant to nematode pests offer another approach for nematode management. Transgenic plants are being developed across the world for root knot and cyst nematodes (Huang *et al.*, 2006, Dutta *et al.*, 2015, Lourenço-Tessutti *et al.*, 2015, Atkinson *et al.*, 2012, Lilley *et al.*, 2012, Green *et al.*, 2012, Roderick *et al.*, 2012, Lin *et al.*, 2013, Liu *et al.*, 2012, Mitchum *et al.*, 2012, Mashela *et al.*, 2016). However, the effect of GM crops on the sustainability of agriculture and environment is a debatable issue and their ethical use is controversial (Horlick-Jones *et al.*, 2007, Tencalla, 2006, Tester, 2001, Reichhardt, 1999, Scoones, 2008, Macnaghten, 2016, Moses, 2016, Abbott, 2016).

1.4.3 Nematode Suppressive Soils and Biocontrol of PPN

Soil being a natural reservoir of varied groups of organisms serves as a niche where several multitrophic interactions occur. Plant parasitic nematodes are often involved in such interactions (Powell, 1971, Powell, 2012, Taylor, 1990, Hussey *et al.*, 1987, Sayre *et al.*, 1988, LaMondia and Timper, 2016, Biere and Goverse, 2016). Phytonematodes are ubiquitous. The damage they incur on their host plants become more severe when the below ground interactions involved are synergistic (Back *et al.*, 2002, Francl and Wheeler, 1993, Garcia and Mitchell, 1975). These positive interactions could be biotic or abiotic. On the contrary, sometimes the soil may act to be nematode suppressive (Weller *et al.*, 2002, Sánchez-Moreno and Ferris, 2007, Westphal, 2005, Pyrowolakis *et al.*, 2002, Hamid *et al.*, 2017, Eberlein *et al.*, 2013). This occurs when the soil has a considerably good number of native antagonists against nematodes as a result of which there is no obvious incidence of nematode infestation in spite of the conditions being favourable (Jaffee and Muldoon, 1989, McSorley *et al.*, 2008). The natural antagonists to phytonematodes could serve as potential candidates for their biocontrol. With the increasing public concern and strict regulations over the use of pesticides, there has been an increased interest in the use

of biological control agents for nematode management. There are a wide variety of life forms including bacteria, fungi and actinomycetes that are often associated with naturally occurring nematode suppressive soils. These organisms can be exploited as a benign alternative to the harmful chemicals (Stirling *et al.*, 1988, Mankau, 1980b, Stirling, 2014, Kaya, 2002).

The first use of a nematode biocontrol agent dates back to more than eight decades ago when Linford (1937) used nematode trapping fungi for nematode management. In 1951, Duddington reported *Dactylella lobata* as a predacious nematode trapping fungus. Since then there have been several reports of nematode antagonistic fungi including the ubiquitous soil fungi such as *Trichoderma* spp., *Fusarium* spp., the nematode trapping fungi *Arthrobotrytis* spp., the nematode egg parasitizing fungi *Pochonia chlamydosporia* and *Purpureocilium lilacinus*, the endoparasitic fungi *Hirsutella* spp., and the arbuscular mycorrhizal fungus *Glomus mosseae* just to name a few (Mankau, 1980b, Mankau, 1980a, Kerry, 1988, Olivares-Bernabeu and López-Llorca, 2002, Jansson and Lopez-Llorca, 2004, Stirling and Kerry, 1983, Bengtsson, 2015). Among actinomycetes, *Streptomyces avermitilis* is known to produce antibiotic compounds called avermectins which are effective nematicides (Wright *et al.*, 1983, El-Nagdi and Youssef, 2004, Jayakumar, 2010). The bacterial enemies of PPN include *Agrobacterium* spp., *Alcaligenes*, *Bacillus firmus*, *Bacillus sphaericus*, *Bacillus subtilis*, *Paenibacillus* spp., *Pasteuria* spp., *Phyllobacillus* spp., *Pseudomonas fluorescens, Telluria* spp., *Xanthomonas* spp. and many more (Kerry, 2000, Keren-Zur *et al.*, 2000, Jonathan *et al.*, 2000, Hashem and Abo-Elyousr, 2011, Aalten *et al.*, 1998).

In a recent report on currently marketed bionematicides, Wilson and Jackson (2013) suggest that amongst the excess of bioagents available in the market the key products are VOTiVO (*B. firmus*), DiTera (*Myrothecium verrucaria*) and BioAct (*P. lilacinus*); some other commercially available products include KlamiC (*P. chlamydosporium*), Econem (*Pasteuria penetrans*), Blue Circle (*Burkholderia cepacia*), Biostart (*Bacillus* spp. mixture). Due to their inherent limitations, bionematicides may seem less effective as compared to chemical nematicides, but there is a growing evidence that they are invaluable for sustainable and safe agriculture (Coombs and Hall, 1998, Wilson and Jackson, 2013). The recent ability to mass produce the obligate bacterium *Pasteuria penetrans* has stimulated increased interest in this organism (Hewlett *et al.*, 2006).

1.4.4 *Pasteuria* spp. – potential biocontrol agents against nematodes

The *Pasteuria* group of Gram positive endospore-forming bacteria represent potential biocontrol agents for a wide range of economically important nematode pests (Starr and Sayre, 1988, Ciancio *et al.*, 1994). They are naturally occurring obligate endoparasites of PPN and are often associated with nematode suppressive soils (Pyrowolakis *et al.*, 2002, Chen *et al.*, 1996). Their non-motile, highly robust endospores are resistant to heat, dessication and mechanical shear and are found in soil in their free forms. *Pasteuria* endospores adhere to the cuticle of freely moving juveniles of PPN as they move in the soil matrix. This, in the first instance, mechanically reduces the ability of a nematode juvenile to penetrate a susceptible plant root (Davies *et al.*, 1988). In the event of root penetration by the juvenile being successful, the nematodes establish a feeding site and just before the first moult, the endospores germinate by forming an infection peg followed by a germ tube. The vegetative cells of *Pasteuria* undergo various developmental stages within the nematode pseudocoelom. They multiply on the developing gonads of infected nematodes, thus, preventing reproduction. The bacteria undergo exponential growth and development, ultimately, leading to the release of as many as 2.5 million mature endospores in the soil from the carcass of a single saccate nematode (Davies *et al.*, 2011).

1.5 THE MEMBERS OF THE GENUS *PASTEURIA*

The bacteria of the genus *Pasteuria* belong to the family *Pasteuriaceae* (Kingdom: Bacteria; division: Firmicutes; class: Bacilli; order: Bacillales). These bacteria are obligate hyperparasites to organisms belonging to two different phyla, the Arthropoda and the Nematoda; they parasitize water fleas (crustaceans) and plant parasitic nematodes and suppress their fecundity (Ebert *et al.*, 1996, Stirnadel and Ebert, 1997, Chen and Dickson, 1998, Giblin-Davis *et al.*, 2003, Preston *et al.*, 2003, Davies *et al.*, 1988). There are seven species of *Pasteuria* known. The classical species *Pasteuria ramosa* infects water fleas of the genera *Daphnia* (Metchnikoff, 1888). The nematode parasitic species described so far include *Pasteuria penetrans* (Thorne, 1940), on *Meloidogyne* spp. (Starr and Sayre, 1988); *Pasteuria thornei*, on *Pratylenchus brachyurus* (Starr and Sayre, 1988); *Pasteuria nishizawae*, on cyst nematodes of *Heterodera* genera (Noel *et al.*, 2005); 'Candidatus P. usage', on *Belonolaimus*

longicaudatus (Giblin-Davis *et al.*, 2003); and *Pasteuria hartismeri* on *Meloidogyne ardenensis* (Bishop *et al.*, 2007); and 'Candidatus *Pasteuria aldrichii*' sp. nov., on *Bursilla* spp. (Giblin-Davis *et al.*, 2011). But their infectivity is not limited to the species of nematodes on which they were first described. At least 200 species belonging to 90 different genera of nematodes have been described to be susceptible to *Pasteuria* spp. (Chen and Dickson, 1998).

The taxonomic classification of *Pasteuria* has been ambiguous (*Table 1.1*). Bacteria of the genus *Pasteuria* were first described by Metchnikoff in 1888 as endospore forming bacterial parasites of waterfleas (*Daphnia* spp.) and were named as *Pasteuria ramosa* in the honour of Louis Pasteur. A similar species was reported for the first time on a nematode *Dorylaimus bulbiferous* but based on the morphology, this species was declared as a protozoan (Cobb, 1906) and in 1940, the species was placed in the genus *Duboscqia* of group Protozoa and renamed as *Duboscquia penetrans.* The evidences provided by electron microscopic studies revealed that the bacteria were more *Bacillus*-like rather than a protozoan. They were therefore, renamed as *Bacillus penetrans* (Mankau, 1975). Sayre and Starr (1989) classified them as a member of *Actinomycetales* due to the presence of mycelia structures. They also highlighted their similarities to Metchnikoff's *P. ramosa* and hence reverted the genus back to *Pasteuria*, hence the name *Pasteuria penetrans.* Years later the bacteria were placed in the family *Alicyclobacillaceae* (Garrity *et al.*, 2005) and finally in 2007 they became the member of the family *Pasteuriaceae* (Ludwig *et al.*, 2007).

Table 1. 1: Timeline of taxonomical classification of *Pasteuria* spp. and the historical ambiguities

1888	First described on *Daphnia* (a waterflea) Named as *Pasteuria ramosa*, in honour of Louis Pasteur Metchnikoff, 1888
1906	First described and studied on a nematode, *Dorylaimus bulbiferous* Declared as a protozoan due to its morphology Cobb, 1906
1940	The nematode parasite accepted as a protozoan Named as *Duboscquia penetrans* Thorne, 1940
1975	Electron microscopic studies revealed *Bacillus*-like characteristics of the bacteria rather than a protozoan *Duboscquia penetrans* renamed as *Bacillus penetrans* Mankau, 1975
1977	Rediscovery of Metchnikoff's work Discovery of similarities of the nematode parasite *Bacillus penetrans* to the crustacean parasite *Pasteuria ramosa* Sayre and Wergin, 1977
1985	The name of the genus was reverted to *Pasteuria* *Bacillus penetrans* renamed as *Pasteuria penetrans* Sayre and Starr, 1985
1989	Classified as *Actinomycetales* based on its morphology (gram-positive, mycelial formation, endospore formation) Sayre and Starr, 1989
2005	*Pasteuria* spp. placed in the family *Alicyclobacillaceae* of the order Bacillales Bacteria>Firmicutes>Bacilli>Bacillales>Alicyclobacillaceae>*Pasteuria* Garrity et al, 2005
2007	The bacteria reported to be more closely related to *Thermoactinomyces* but due to their distinctive phenotypic and genotypic traits, the two genera were not placed in the same family. Placed in a new family called Pasteuriaceae Bacteria>Firmicutes>Bacilli>Bacillales>*Pasteuriaceae*>*Pasteuria* Ludwig et al, 2007

1.5.1 Life Cycle and morphological stages

The life-cycle of *Pasteuria* (*Figure 1.2*) proceeds in synchronization with its nematode host. When in soil, it remains in a dormant form called an 'endospore'. As an infective nematode juvenile migrates in search of a host root, *Pasteuria* endospores adhere to its cuticle by what has been hypothesized to be a 'Velcro-like' attachment system (Davies, 2009) . Most spores germinate only after the juvenile establishes a feeding site in the host root tissues (Chen and Dickson, 1998). In some species, e.g. *Heterodera avanae*, they readily germinate before the J2 enters the root (Davies *et al.*, 1990). During germination, the endospore forms a germ tube that develops into rhizoid-like mycelial structures which proliferate throughout the developing nematode. The rhizoid-like structures fragment and develop into rod-shaped Gram positive vegetative cells in both root-knot (Davies *et al.*, 2011) and cyst nematodes (Mohan *et al.*, 2012). The vegetative cells start dividing exponentially eventually forming granular masses within the female. When the host nutrient supply is depleted, the growth becomes stationary and sporulation is triggered producing microcolony-like structures which start the process of sporogenesis. This process involves the breakup of the microcolonies and the production of quartets followed by doublets which eventually gives rise to single endospores. These endospores are highly resistant to various environmental stresses and can remain viable for many years in soil. The process of sporulation in *Pasteuria* is similar to that in most of the endospore forming bacteria and is described later in *Section 1.7.2*.

Germm tube   **Nematode cuticle**
Nematode hypodermis

Pasteuria endospores in soil

In soil, endospores come in contact with nematode J2 migrating in search of host root

Attachment of endospores to the cuticle of J2 and Germination (formation of an infection peg / germ tube) usually after the nematode J2 establishes a feeding site within host root tissue

Formation of rhizoids or exploratory filaments from the germ tube. These filaments penetrate deep into the pseudocoelom

Stage IV: Mature endospores in sporangium

Stage III: Immature endospores

Stage II: Tetrads and diads

Stage I: Cauliflower-like microcolonies

Fragmentation of rhizoids into *Bacillus*-like rods followed by their exponential growth

Sporulation

Figure 1. 2: Life cycle stages of *Pasteuria* showing different stages including attachment to nematode juveniles; germination and infection peg formation; rhizoid formation; fragmentation of rhizoids into rod-like cells and their exponential growth; sporulation stages. The figure is not to scale.

14

1.5.2 *Pasteuria* endospores and their attachment to the nematode cuticle

Endospore formation is a survival strategy exhibited by certain members of the phylum Firmicutes (Fritze, 2004). Endospores are dormant structures surrounded by proteinaceous layers that form a spore coat. In the pathogenic endospore formers an additional layer called the exosporium further surrounds the spore coat that are thought to give the endospores their adhesive property (Kozuka and Tochikubo, 1985). Like all pathogenic members of the phylum, the endospores of *Pasteuria* spp. also possess an exosporium. As shown by electron micrographic studies of *Pasteuria* endospore, beneath its exosporium lies a skirt-like structure made up of parasporal fibres thought to be involved in endospore attachment to the nematode cuticle (Davies, 2009). Enzyme based assays and antigen based attachment tests by various researchers suggest that attachment of *Pasteuria* endospores to the nematode cuticle involve protein-carbohydrate interactions (Davies and Danks, 1993, Davies and Redden, 1997, Mohan *et al.*, 2001, Vaid *et al.*, 2002).

Analysis of sequenced genomes of *Bacillus* spp. have identified collagen-like fibres associated with the surface of the endospores (Sylvestre *et al.*, 2002, Sylvestre *et al.*, 2003, Steichen *et al.*, 2003). More recently, collagen-like fibres have been identified in *Pasteuria* (Davies and Opperman, 2006, Mouton *et al.*, 2009). The characterisation of these collagens in *Pasteuria ramosa*, an obligate parasite of freshwater crustaceans *Daphnia* spp., provides valuable insights into the structure and diversity of these molecules (Mouton *et al.*, 2009, McElroy *et al.*, 2011). In *P. penetrans*, it has been suggested that these collagen-like fibres interact with a cuticle receptor on the body surface of the nematode possibly involving mucin-like molecules, through a suggested *Velcro*-like attachment process (Davies 2009; Davies & Curtis 2011) (*Figure 1.3*). However, the exact molecular mechanism of endospore attachment and nature of the host specificity is still not known.

Collagen-like fibres
parasporal fibres of
the endospore

Central body of endospore

Spore Coat

Skirt-like exosporium

Muclin-like molecules
on nematode cuticle

Nematode epicuticle/
glycocalyx
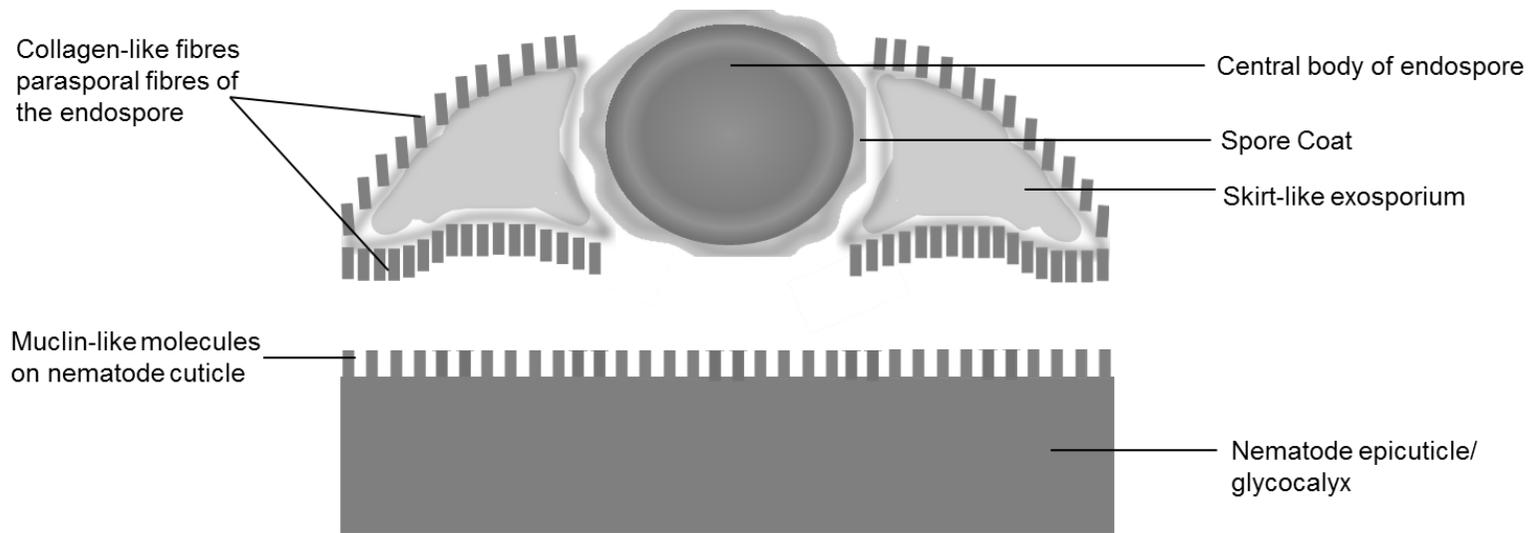
Figure 1. 3: Velcro-like attachment of *Pasteuria* endospore to the host nematode cuticle involving collagen-like parasporal fibres and mucin-like molecules. (Adapted from Davies, 2009)

1.6 CONSTRAINTS TO THE SUCCESSFUL DEPLOYMENT OF PASTEURIA SPP. AS A BIO-NEMATICIDE

1.6.1 Mass production *in vivo* and *in vitro*

Due to their obligate nature and their highly-synchronized life cycle to that of the PPN host, the mass production of *Pasteuria* endospore is one of the primary technical challenges to the successful exploitation of biocontrol potential of *Pasteuria*. *Pasteuria* endospores can be isolated from root knot nematodes by dissecting infected nematode females from plant root material and crushing them in water. This gives a suspension full of endospores. For routine *in vivo* mass culture of *Pasteuria*, the endospore suspension is allowed to attach to freshly hatched nematode juveniles which are then inoculated onto a susceptible plant, generally, in a glass house. The plants are uprooted after about 8 weeks and the roots are air-died and ground. The root powder containing *Pasteuria* endospores can be used as an inoculant for further culture or stored indefinitely (Stirling and Wachtel, 1980). The *in vitro* large-scale production of *Pasteuria* endospores for commercial agriculture has been a subject of intensive study since Stirling (1984) reported its potential to control *Meloidogyne javanica*. In a recent breakthrough, an American company *Pasteuria Biosciences* LLC now acquired by Syngenta has overcome this obstacle by developing a proprietary *in vitro* culturing system (Kojetin *et al.*, 2005, Hewlett *et al.*, 2006), and produces *Clariva*^TM a product based on *Pasteuria* endospores. However, whether or not during the *in vitro* culture process or scale-up and formulation, the *Pasteuria* endospores retain their viability and virulence is still questionable (Crow *et al.* 2011). Based on this *in vitro* Pasteuria technology, the launch of *Clariva*™ as a seed treatment bionematicide in the US in 2013 represents a milestone in this area.

1.6.2 Host specificity

Another major constraint, in line with mass production of *Pasteuria* endospores, is the varying degree of attachment profiles exhibited by *Pasteuria* endospores. It is believed that endospores from individual isolates of the bacterium do not adhere to or recognize all populations of nematodes (Sharma and Davies, 1997, Wishart *et al.*, 2004). However, more recent research has identified an isolate of *Pasteuria sp.* which appears to have a wider host range (Mohan *et al.*, 2012); research on this isolate originating on *Heterodera cajani* (pigeon pea cyst nematode), a major nematode pest

of legumes in India, has shown it to be promiscuous in that it has a wider host range and exhibit cross-generic attachment thereby infecting *Globodera pallida* (potato cyst nematodes) which are globally important pests of potatoes. A suggested hypothetical reason for their host specificity is the immunological heterogeneity of collagen-like fibres present on the exosporium of the endospores (Davies & Curtis 2011; Davies 2009). However the exact molecular mechanism is yet to be discovered.

1.6.3 DNA Isolation of *Pasteuria* – a major constraint in genomic studies

Despite the availability of highly efficient classical sequencing approaches and the continually advancing next generation sequencing technologies, the genome of *Pasteuria* has not been successfully sequenced and published. This is because obtaining sufficient amount of good quality *Pasteuria* DNA, free from host nematode and plant DNA, is difficult due to its obligate nature. However, in the sequence databases of National Centre for Biotechnology Information, a total of 4031 nucleotide sequences of *Pasteuria penetrans* are publicly available. Of these 3903 sequences are currently present in the Genome Survey Sequence (GSS) database. The GSS are the shotgun sequences for *Pasteuria penetrans* population Res147 (Opperman *et al.*, 2003). The genome size of *Pasteuria* spp. has been estimated to be 4.2- 5.0 Mb based on the genome size of the members of the closely related genus *Bacillus* (Bird *et al.*, 2003a, Davies and Spiegel, 2011)*.*

1.7 *BACILLUS* SPP. – A MODEL TO STUDY *PASTEURIA* BIOLOGY

Considering the numerous constraints in exploiting the potential of *Pasteuria* (described in *Section 1.6*)*,* using a model organism that has been studied extensively could provide important insights into *Pasteuria* biology. *Bacillus* spp. are closely related to *Pasteuria* and are contemporary organisms as far as their taxonomic positioning is concerned (*Table 1.2*). This genus of Gram positive endospore forming bacteria has been extensively studied at cultural, physiological, biochemical and genomic levels (Völker and Hecker, 2005, Rasko *et al.*, 2005, Sonenshein *et al.*, 1993, Lai *et al.*, 2003). New *Bacillus* strains are continually being identified and studied and newly sequenced *Bacillus* genomes being uploaded in public databases and made available to be studied. Using *Bacillus* spp. as a model could open fields of comparative biology, proteomics and genomics and could lead us to an understanding of the *Pasteuria* biology and its mechanisms of interactions with the nematode host.

1.7.1 Phylogenetic closeness of *Bacillus* and *Pasteuria*

Prior to the sequencing era, study of evolutionary relationships mainly relied upon morphology and life style of organisms. The initial taxonomic misplacement of *Pasteuria penetrans* amongst protozoans was based on morphology (as discussed in *Section 1.5*, *Table 1.1*). Since the advancement of the science of molecular phylogenetics, numerous subsequent attempts have been made to resolve the placement of *Pasteuria* spp. amongst bacteria. Through various molecular phylogenetic studies *Pasteuria* spp. have been clearly positioned as a member of the phylum Firmicutes. Based on 16S rRNA automated sequencing studies Anderson *et al.* (1999) showed that *Pasteuria* spp. were the members of *Clostridium-Bacillus-Streptococcus* clade of gram-positive eubacteria. This claim was further supported by the 16S rDNA studies by Preston *et al.* (2003). Trotter and Bishop (2003) verified a similar phylogeny by examining the *spo0A* gene. A multilocus study using 40 concatenated housekeeping genes from 33 bacterial species conclusively placed *Pasteuria penetrans* in the *Bacillus-Clostridium* clade of Gram-positive bacteria (Charles *et al.*, 2005).

Table 1. 2: List of attributes shared by *Bacillus* and *Pasteuria*

Attribute	*Bacillus* vs *Pasteuria*	References
Taxonomic classification	*Bacillus* spp. Bacteria>Firmicutes>Bacilli>Bacillales>Bacillaceae *Pasteuria* spp. Bacteria>Firmicutes>Bacilli>Bacillales>Pasteuriaceae	Cohn, 1872 Ludwig et al, 2007
16S rRNA gene/ 16S rDNA	Several studies place *Pasteuria* in *Bacillus-Clostridium-Streptococcus* clade	Atibalentja, 2000; Anderson, 1999
spoOA gene	A study clearly shows that *P.penetrans* is a close relative of *Bacillus* spp.	Trotter and Bishop, 2003
Housekeeping genes	A study of 40 concatenated housekeeping genes place *Pasteuria* in *Clostridium- Bacillus* clade	Charles et al, 2005
Cellular Morphology	*Bacillus* spp. are rod-shaped bacteria Bacillus-like rods observed in life cycle of *Pasteuria*	Cohn, 1872 Davies *et al.*, 2011
Endospore-formation	Both are endospore-forming	Winslow et al, 1920; Mankau, 1975
Exosporium	Endospores of both pathogenic *Bacillus* spp. and *Pasteuria* are associated with a loose-fitting exosporium In the pathogenic species of *Bacillus* as well as in *Pasteuria*, the first contact and interaction of the dormant endospore with the host body occurs via the exosporium	Flugge, 1886; Imbriani and Mankau, 1977; Sayre and Wergin, 1977
Collagen in exosporium	The hairy nap of the exosporium is known to be made up of a collagen-like protein called BclA in pathogenic *Bacillus* spp. Similar collagen-like proteins are postulated in *Pasteuria* spp.	Sylvestre, 2002 Mouton, 2009; McElroy, 2011; Davies, 2006; Davies, 1993; Mohan, 2001

1.7.2 *Bacillus* endospores and attachment mechanisms

Endospores are dormant, non-reproductive, survival stage structures formed by some Gram positive bacteria of phylum Firmicutes. During unfavourable conditions, the vegetative cells of endospore forming bacteria undergo an asymmetric septation which divides the cell into two sister cells, both bearing a copy of the bacterial genome (Vinter, 1959, Piggot and Hilbert, 2004). The larger cell called the 'mother cell' engulfs the smaller cell called the 'pre-spore' and in the process the pre-spore gets insulated within several concentric layers (spore coat and cortex) composed of proteins and peptidoglycan (Bath *et al.*, 2000, Errington, 2003, Popham *et al.*, 1996, Kim and Schumann, 2009). The mother cell, thereafter, undergoes autolysis releasing the mature endospore to survive in the harsh environment (Smith and Foster, 1995). A mature endospore is, thus, made up of a dehydrated central core containing the genetic material (DNA) with a number of and stabilizing factors, like Small Acid Soluble Proteins (SASPs) and Ca^{2+}- dipicolinic acid, that provide stability to the DNA (Setlow, 2007); a peptidoglycan cortex; and a tough multi-layered proteinaceous spore coat that shields the core and cortex (Driks, 2002). Some pathogenic species including *B. subtilis*, *B. cereus* and *B. thuringiensis* possess an additional outermost loose fitting layer called the exosporium.

DesRosier and Lara (1984) define the exosporium as a "loose-fitting, balloon-like layer" surrounding the spore. The exosporium, essentially made up of proteins, lipids and carbohydrates has a basal layer that supports a filamentous "hairy nap" (Sylvestre *et al.*, 2002, Kailas *et al.*, 2011, Driks, 2002, Gerhardt and Ribi, 1964, Charlton *et al.*, 1999). The hair-like filaments of the nap are known to be made up of an immuno-dominant protein BclA (Steichen *et al.*, 2003, Sylvestre *et al.*, 2002). A number of carbohydrates components, including rhamnose, galactosamine and anthrose, have been shown to be associated with the exosporium and endospores of different bacilli (Matz *et al.*, 1970, Daubenspeck *et al.*, 2004b, Maes *et al.*, 2016, Fox *et al.*, 1993, Wunschel *et al.*, 1995, Waller *et al.*, 2004). In many of the endospore-forming Gram positive bacteria bacterial adhesins called MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) bound to ligands like fibronectin, collagen, fibrinogen have been shown to be involved in adhesion to extracellular matrix molecules (Patti *et al.*, 1994, Patti and Höök, 1994, Xu *et al.*, 2004, Tulli *et al.*, 2013, Hitsumoto *et al.*, 2014).

1.8 AIMS AND OBJECTIVES

The primary aim of this project is to exploit the knowledge available for well-studied *Bacillus* spp. to help characterize the surface of *Pasteuria* endospores. Surface characterization of *Pasteuria* endospores will eventually assist in defining the molecular basis of the attachment specificity of *Pasteuria* to its nematode host.

1.8.1 Hypothesis

The overarching hypothesis for this project is that the molecular genetics involved in endospore structure and function will be similar between *Bacillus* spp. and *Pasteuria penetrans.*

1.8.2 Aims and Objectives by Chapter

Chapter 2: Phylogenetic placement of *Pasteuria* spp. within Firmicutes

1) To construct a phylogeny of *Pasteuria* in relation to *Bacillus* spp. and other closely related Firmicutes using 16S rRNA gene.
2) To construct a phylogeny of *Pasteuria* in relation to *Bacillus* spp. and other closely related Firmicutes using *groEL*, *gyrB* and *spo0A* genes and their gene products.

Chapter 3: Putative collagens in *P. penetrans* and their in-silico characterization

1) To investigate the presence of candidate genes coding for collagen-like proteins in a set of unpublished contigs from an isolate of *P. penetrans.*
2) To characterize the putative CLPs of *P. penetrans* using different bioinformatics tools and prediction algorithms.
3) To compare putative CLPs of *P. penetrans* with other bacterial CLPs based on sequence homology and conserved motifs.
4) To derive the cladistics of putative CLPs of *Pasteuria* with other related CLPs based on the amino acid composition of their low-complex GXY repeat regions.

Chapter 4: Comparative endospore protein characterization using Western Blotting

1) To study the endospore protein composition of *P. penetrans* and *B. thuringiensis* using standard SDS-PAGE technique.
2) To identify immunodominant proteins with shared epitopes in *P. penetrans* and *B. thuringiensis* endospores using *Pasteuria*-specific polyclonal antibody for immunodetection.
3) To identify immunodominant collagen-like proteins with shared epitopes in *P. penetrans* and *B. thuringiensis* endospores using collagen specific antibodies raised to synthetic collagen-like peptides from *P. penetrans* contigs.
4) To validate the presence of collagens and glycoproteins by means of collagenase digestion, glycoprotein staining and lectin blotting.

Chapter 5: Comparative surface characterization of *P. penetrans* endospores

1) To investigate surface localization of the epitopes, previously identified in Chapter 4, by means of immunolocalization studies using the *Pasteuria*-specific polyclonal antibodies raised to whole endospore and to collagen-like peptides.
2) To investigate the surface localization of N-acetylglucosamine (NAG) on *Pasteuria* endospores.

Chapter 6: *In vitro* nematode attachment assays

1) To investigate the involvement of collagens in the attachment of *Pasteuria* endospores to the cuticle of nematode juveniles.
2) To investigate the involvement of NAG in the attachment of *Pasteuria* endospores to the cuticle of nematode juveniles.

Chapter 2

Phylogenetic placement of *Pasteuria* spp. within Firmicutes

2.1 INTRODUCTION

Host specificity is regulated by many complex molecular interactions between the pathogen and the host. Obligate organisms have limited scope of laboratory experimentation. Genomics and proteomics, that involve studying an organism through its DNA and protein sequences, offer an invaluable approach to study the overall biology of such organisms. When applied to host-pathogen interaction studies, these modern techniques could help in resolving the molecular basis of pathogenicity. In the *Pasteuria*-nematode system, the complex tritrophic interaction of the host plant, the nematode parasite and the bacterial hyperparasite makes it difficult to obtain the pure DNA of *Pasteuria*. Even the most advanced and vigorous genomic extraction procedures yield DNA contaminated with nematode and plant genetic material. This makes the genomic studies of *Pasteuria* challenging and currently there is no complete genome sequence for *Pasteuria* available. One of the approaches used to identify a new gene in an unsequenced microbial genome is to find a close homolog in another organism. Comparative genomics offers a method by which the genomic organization of *Pasteuria* and the gene functions governing its pathogenicity and host preferences may be determined. As discussed in the previous chapter (*Section 1.7.1*), early phylogenetic studies have tended to focus on using 16S rRNA genes to differentiate species and populations (Anderson *et al.*, 1999, Duan *et al.*, 2003, Sturhan, 1988); however, later work using multilocus housekeeping genes suggests that *Pasteuria* spp. are related to the members of the *Bacillus* genus (Charles *et al.*, 2005). More recent work using single nucleotide polymorphisms has suggested that protein-encoding genes provided increased discrimination than using 16S rRNA sequences (Mauchline *et al.*, 2011).

2.1.1 *groEL* and *gyrB* as emerging candidates for phylogeny reconstruction

Since the gene for the small subunit of ribosomal RNA (16S rRNA) was used for the reconstruction of the 'tree of life' by Woese (Woese, 1987), it has been universally relied upon as a tool for the identification of prokaryotes and their phylogenetic reconstruction and classification. The 16S rRNA gene sequences from environmental

samples have also been used for the evaluation of microbial diversity in the field of metagenomics. Apart from their known conservation at the level of nucleotide sequences and secondary structures, the classical concept presumed that these genes do not undergo horizontal gene transfer events. However, several studies suggest the horizontal transfer of segments of this gene and advocate the need for alternatives to avoid misidentification and interpretation of discordant phylogenies (Eardly *et al.*, 1996, Yap *et al.*, 1999, Schouls *et al.*, 2003, Rajendhran and Gunasekaran, 2011). Protein-coding housekeeping genes have advantages over ribosomal RNA genes. Firstly, they are supposedly less prone to horizontal gene transfer. Secondly, they are known to be functionally conserved, thus, making them more suitable to study the evolution of closely-related species. Two such housekeeping genes are *gyrB* and *groEL.* The *gyrB* and *groEL* genes code for DNA gyrase subunit B and for GroEL chaperone proteins, respectively. Several studies suggest the use of *gyrB* gene as a phylogenetic marker for bacteria due to their ubiquitous conservation amongst eubacteria (Yanez *et al.*, 2003, Dauga, 2002, Bavykin *et al.*, 2004, Yamamoto *et al.*, 2000). Similarly, the *groEL* gene sequences have been previously used for making inferences of eubacterial phylogenies where 16S phylogenies were not so informative due to the high conserved 16S sequences (Karlin and Brocchieri, 2000, Viale *et al.*, 1994, Marston *et al.*, 1999, Goh *et al.*, 1996, Kwok *et al.*, 1999, Yu *et al.*, 2001).

2.1.2 The gene *spo0A* as an endospore-specific candidate for phylogeny reconstruction

The *spo0A* gene codes for a transcription factor, Spo0A, that is found exclusively in the endospore-forming bacterial species (Hoch, 1993). Spo0A is a signal transduction protein that gets activated in response to stress, by a multicomponent phosphorelay system (Burbulys *et al.*, 1991). Once activated by phosphorylation, the protein binds to DNA resulting in repression of several vegetatively expressed genes and activation of sporulation related genes and this leads to the onset of sporulation (Strauch *et al.*, 1990). Due to its exclusive association with endospore formers, the *spo0A* gene could be deployed to study how sporulation evolves. There have been at least two independent studies that used the *spo0A* gene as a genetic marker for phylogenetic

analyses of two genera of endospore formers, *Pasteuria* spp. and *Geobacillus* spp. respectively (Trotter and Bishop, 2003, Kuisiene *et al.*, 2009).

2.1.3 Aims and objectives

Major Aim: To validate the proposition that members of the *Bacillus* genus are suitable candidates for genomic and proteomic comparative studies with nematode parasitic *Pasteuria* species. This is based on the hypothesis that *Pasteuria* are sufficiently related to *Bacillus* spp. as to be good model organisms.

Specific Objectives:

1) To construct a phylogeny of *Pasteuria* in relation to *Bacillus* spp. and other closely related Firmicutes using 16S rRNA gene.
2) To construct a phylogeny of *Pasteuria* in relation to *Bacillus* spp. and other closely related Firmicutes using *groEL*, *gyrB* and *spo0A* genes and their gene products.

2.2   MATERIALS AND METHODS

This section describes the methodologies used to infer the evolutionary history of *Pasteuria* spp. amongst selected members of Firmicutes based on gene sequences coding for 16S rRNA, and nucleic acid and protein sequences of three other genes: *gyrB*, *groEL* and *spo0A*.

2.2.1   Sequences for phylogenetic analyses

The publicly available nucleotide/protein sequences of the cladoceran parasitic *P. ramosa* and nematode parasitic *Pasteuria* spp. were compared with selected nucleotide/protein sequences of *Clostridium* spp., *Paenibacillus* spp. and *Pelosinus* spp. and three major pathogenic *Bacillus* spp. (*B. anthracis*, *B. thuringiensis* and *B. cereus*). Cyanobacterial species including *Microcystis elabens*, *Arthrospira platensis*, *Cyanobium gracile* and/or *Gloeothece* sp. were used as outgroups for 16S, *gyrB* and *groEL* analyses. No outgroup was used in the case of *spo0A* gene and its product as this gene is specifically associated only with endospore formers. See *Table 2.1* for the complete list of sequences used for all phylogenetic analyses.

2.2.2   Molecular phylogenetics based on different genetic markers

Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016) using the Maximum Likelihood method based on Tamura-Nei model (for nucleic acid sequences) and JTT matrix-based model (protein sequences) were used for nucleic acid and protein sequences respectively (Tamura and Nei, 1993, Jones *et al.*, 1992). The initial tree was made using BioNJ algorithm (Gascuel, 1997). The bootstrap consensus tree was inferred from 500 replicates (Felsenstein, 1985).

Table 2. 1: **List of gene sequences used for phylogenetic analyses**

Gene	Sequence code	Accesion number (database in parenthesis)	Organism
16S rRNA	P.gottin	AF515699.1 (ENA)	*Pasteuria goettingianae*
	P.hart	AJ878853.1 (ENA)	*Pasteuria hartismeri*
	P.nishi	AF134868.2 (ENA)	*Pasteuria nishizawae*
	P. ramosa	U34688.1.1435 (ENA)	*Pasteuria ramosa*
	P.pen.HcP	JN592479.1 (ENA)	*Pasteuria penetrans* HcP
	Ba.Ames	Ba16SA (KEGG)	*Bacillus anthracis* Ames
	BA.AmesAnc	GBAA_5958 (KEGG)	*Bacillus anthracis* Ames Ancestor
	BA.CDC684	BAMEG_0007 (KEGG)	*Bacillus anthracis* CDC 684
	BA.A0248	BAA_0007 (KEGG)	*Bacillus anthracis* A0248
	BC.ATCC14579	BC0007 (KEGG)	*Bacillus cereus* ATCC 14579
	BC.ATCC10987	BCE_5738 (KEGG)	*Bacillus cereus* ATCC 10987
	BC.AH187	BCAH187_A0007 (KEGG)	*Bacillus cereus* AH187
	BC.G9842	bcg:BCG9842_B5313 (KEGG)	*Bacillus cereus* G9842
	BT.AlHakam	(BALH_r16S01) (KEGG)	*Bacillus thuringiensis* Al Hakam
	BT.kurstakiHD1	BTK_r29390 (KEGG)	*Bacillus thuringiensis* serovar kurstaki HD-1
	BT.MC28	MC28_r03 (KEGG)	*Bacillus thuringiensis* MC28
	BT.IS5056	H175_rrna01 (KEGG)	*Bacillus thuringiensis* serovar thuringiensis IS5056
	Bs.168	BSU_rRNA_1 (KEGG)	*Bacillus subtilis* subsp. subtilis 168
	Bs.spiziW23	BSUW23_r20632 (KEGG)	*Bacillus subtilis* subsp. spizizenii W23
	Bs.natto.EST195	BSNT_06293 (KEGG)	*Bacillus subtilis* subsp. natto BEST195
	Bs.BSn5	BSn5_r21080 (KEGG)	*Bacillus subtilis* BSn5
	Paeni.poly.E681	PPE_00010 (KEGG)	*Paenibacillus polymyxa* E681
	Paeni.muci.K02	B2K_r38186 (KEGG)	*Paenibacillus mucilaginosus* K02
	Paeni.ter	HPL003_r28210 (KEGG)	*Paenibacillus terrae*
	Paeni.larv	ERIC2_c00090 (KEGG)	*Paenibacillus larvae*
	Pelo.UF01	UFO1_R0122 (KEGG)	*Pelosinus* sp. UF01
	Pelo.ferm	JBW_RNA0104 (KEGG)	*Pelosinus fermentans*
	Clos.acet	CACET_c00110 (KEGG)	*Clostridium aceticum*
	Clos.botu	CBOr001 (KEGG)	*Clostridium botulinum* A ATCC 3502
	Clos.tetE88	CTC_0r03 (KEGG)	*Clostridium tetani* E88
	Micro.ela	AB001724.1 (GenBank)	*Microcystis elabens* NIES42
	Arthr.pla	AB074508.1 (GenBank)	*Arthrospira platensis* IAM M-135
	Gloe.HSC34	EF150783.1 (GenBank)	*Gloeothece* sp. HSC34
groEL	*Pasteuria* sp.	K9L117_9BACL (UniProtKB)	uncultured *Pasteuria* sp.
	Pelo.UF01	UFO1_1284 (KEGG)	*Pelosinus* sp. UF01
	Pelo.ferm	JBW_02996 (KEGG)	*Pelosinus fermentans*
	BC.ATCC14579	BC0295 (KEGG)	*Bacillus cereus* ATCC 14579
	BC.AH187	BCAH187_A0317 (KEGG)	*Bacillus cereus* AH187
	BC.G9842	BCG9842_B5026 (KEGG)	*Bacillus cereus* G9842
	BC.ATCC10987	BCE_0289 (KEGG)	*Bacillus cereus* ATCC 10987
	BA.Ames	BA_0267 (KEGG)	*Bacillus anthracis* Ames
	BA.AmesAnc	GBAA_0267 (KEGG)	*Bacillus anthracis* Ames Ancestor
	BA.A0248	BAA_0309 (KEGG)	*Bacillus anthracis* A0248
	ba.cdc684	BAMEG_0312 (KEGG)	*Bacillus anthracis* CDC 684
	BT.AlHakam	BALH_0251 (KEGG)	*Bacillus thuringiensis* Al Hakam
	BT.kurstaki	BTK_01330 (KEGG)	*Bacillus thuringiensis* serovar kurstaki HD-1
	BT.MC28	MC28_4955 (KEGG)	*Bacillus thuringiensis* MC28
	BT.IS5056	H175_ch0242 (KEGG)	*Bacillus thuringiensis* serovar thuringiensis IS5056
	BS.168	BSU06030 (KEGG)	*Bacillus subtilis* subsp. subtilis 168
	BS.spiziW23	BSUW23_03085 (KEGG)	*Bacillus subtilis* subsp. spizizenii W23
	Clos.acet	CACET_c03570 (KEGG)	*Clostridium aceticum*
	Clos.botu	AJF28459.1 (ENA)	*Clostridium botulinum*
	Clos.tet	CTC_02413 (KEGG)	*Clostridium tetani* E88
	Paeni.poly	PPE_01188 (KEGG)	*Paenibacillus polymyxa* E681
	Paeni.muc	B2K_33950 (KEGG)	*Paenibacillus mucilaginosus* K02
	Paeni.ter	HPL003_14225 (KEGG)	*Paenibacillus terrae*
	Paeni.larv	ERIC2_c03660 (KEGG)	*Paenibacillus larvae*
	Micro.aeru	MAE_03410 (KEGG)	*Microcystis aeruginosa* NIES-843
	Arthro.plat	NIES39_A01110 (KEGG)	*Arthrospira platensis* NIES-39
	Cyano.gra	Cyagr_3154 (KEGG)	*Cyanobium gracile* PCC 6307
gyrB	P.pen	G9C4T5_PASPE (UniprotKB)	*Pasteuria penetrans*
	P.pen	contig11607 (unpublished)	*Pasteuria penetrans*
	P.hart	G9C4T6_9BACL (UniprotKB)	*Pasteuria hartismeri*

	Pelo.UF01	UFO1_4475 (UniprotKB)	*Pelosinus* sp. UFO1
	Pelo.ferm	JBW_00007 (UniprotKB)	*Pelosinus fermentans* JBW45
	BC.ATCC10987	Q73FK1_BACC1 (UniprotKB)	*Bacillus cereus* ATCC 10987
	BC.ATCC14579	Q81JD1_BACCR (UniprotKB)	*Bacillus cereus* ATCC 14579
	BC.G9842	B7IS24_BACC2 (UniprotKB)	*Bacillus cereus* G9842)
	BC.AH187	B7HPS1_BACC7 (UniprotKB)	*Bacillus cereus* AH187
	BA.Ames	BA_0005 (KEGG)	*Bacillus anthracis* Ames
	BA.AmesAnc	GBAA_0005 (KEGG)	*Bacillus anthracis* Ames Ancestor
	BA.A0248	BAA_0005 (KEGG)	*Bacillus anthracis* A0248
	BA.CDC684	BAMEG_0005 (KEGG)	*Bacillus anthracis* CDC 684
	BT.AlHakam	BALH_0005 (KEGG)	*Bacillus thuringiensis* Al Hakam
	BT.kurstaki	BTK_00025 (KEGG)	*Bacillus thuringiensis* serovar kurstaki HD-1
	BT.MC28	MC28_4725 (KEGG)	*Bacillus thuringiensis* MC28
	BT.IS5056	H175_ch0005 (KEGG)	*Bacillus thuringiensis* serovar thuringiensis IS5056
	BS.168	BSU00060 (KEGG)	*Bacillus subtilis* subsp. subtilis 168
	BS.spiziW23	BSUW23_00035 (KEGG)	*Bacillus subtilis* subsp. spizizenii W23
	Clos.acet	CACET_c00060 (KEGG)	*Clostridium aceticum*
	Clos.botu	AJF28149.1 (ENA)	*Clostridium botulinum*
	Clos.tet	AAO34744.1 (ENA)	*Clostridium tetani* E88
	Paeni.poly	PPE_00006 (KEGG)	*Paenibacillus polymyxa* E681
	Paeni.muc	B2K_00030 (KEGG)	*Paenibacillus mucilaginosus* K02
	Paeni.ter	HPL003_07520 (KEGG)	*Paenibacillus terrae*
	Paeni.larv	ERIC2_c00060 (KEGG)	*Paenibacillus larvae*
	Micro.aeru	BAG04645.1 (KEGG)	*Microcystis aeruginosa* NIES-843
	Arthro.plat	BAI92164.1 (KEGG)	*Arthrospira platensis* NIES-39
	Cyano.gra	AFY29354.1 (KEGG)	*Cyanobium gracile* PCC 6307
spo0A	P.pen	Q7X533_PASPE (UniprotKB)	*Pasteuria penetrans*
	P.ram	Q7X532_9BACL (UniprotKB)	*Pasteuria ramosa*
	Pelo.UF01	UFO1_1966 (KEGG)	*Pelosinus* sp. UFO1
	Pelo.ferm	JBW_02424 (KEGG)	*Pelosinus fermentans*
	BC.ATCC14579	BC4170 (KEGG)	*Bacillus cereus* ATCC 14579
	BC.AH187	BCAH187_ A4301 (KEGG)	*Bacillus cereus* AH187
	BC.G9842	BCG9842_B0952 (KEGG)	*Bacillus cereus* G9842
	BC.ATCC10987	BCE_4243 (KEGG)	*Bacillus cereus* ATCC 10987
	BA.Ames	BA_4394 (KEGG)	*Bacillus anthracis* Ames
	BA.AmesAnc	GBAA_4394 (KEGG)	*Bacillus anthracis* Ames Ancestor
	BA.A0248	BAA_4413 (KEGG)	*Bacillus anthracis* A0248
	BA.CDC684	BAMEG_4431 (KEGG)	*Bacillus anthracis* CDC 684
	BT.AlHakam	BALH_3780 (KEGG)	*Bacillus thuringiensis* Al Hakam
	BT.kurstaki	BTK_22035 (KEGG)	*Bacillus thuringiensis* serovar kurstaki HD-1
	BT.MC28	MC28_3464 (KEGG)	*Bacillus thuringiensis* MC28
	BT.IS5056	BT.H175_ch4254 (KEGG)	*Bacillus thuringiensis* serovar thuringiensis IS5056
	BS.168	BSU24220 (KEGG)	*Bacillus subtilis* subsp. subtilis 168
	BS.spiziW23	BSUW23_11985 (KEGG)	*Bacillus subtilis* subsp. spizizenii W23
	Clos.acet	CACET_c18630 (KEGG)	*Clostridium aceticum*
	Clos.botu	AJF28293.1 (ENA)	*Clostridium botulinum*
	Clos.tet	CTC_01569 (ENA)	*Clostridium tetani* E88
	Paeni.poly	PPE_02831 (KEGG)	*Paenibacillus polymyxa* E681
	Paeni.muc	B2K_14880 (KEGG)	*Paenibacillus mucilaginosus* K02
	Paeni.terr	HPL003_24020 (KEGG)	*Paenibacillus terrae*
	Paeni.larv	ERIC2_c32740 (KEGG)	*Paenibacillus larvae*

2.3   RESULTS

2.3.1   Molecular phylogenetics of *Pasteuria* based on the 16S rRNA gene

The phylogenetic tree with the highest log likelihood (-11556.6217) is shown in *Figure 2.1*. The 16S rRNA gene sequences from all the five analyzed species of the genus *Pasteuria*, including the cladoceran parasite, *P. ramosa* clustered together with high bootstrap support (100%). *Pasteuria* spp. were observed to be more closely related to *Clostridium* spp. than *Bacillus* species. However, their close-relatedness with clostridia was supported with a low bootstrap value of only 58%. All the *Bacillus* spp. and *Paenibacillus* spp. were grouped together supported by a high bootstrap value. The pathogenic *Bacillus* spp. i.e. *B. cereus*, *B. thuringiensis* and *B. anthracis* were separated from the non-pathogen *B. subtilis* with 100% support. In 95% of the tree replicates *Pelosinus* spp. was separately grouped as an outgroup with the cyanobacteria.

2.3.2   Molecular phylogenetics of *Pasteuria* based on the *groEL* gene and its product

Based on the *groEL* gene sequences *Pasteuria*, *Clostridum* spp. and *Bacillus* spp. formed a super-clade with a high bootstrap support of 99% (*Figure 2.2 a*). This superclade also contained the *Pelosinus* sequences. *Pasteuria* was placed close to *Clostridium* spp. (100% bootstrap support). Similar to phylogenetic tree based on 16S rRNA gene, the pathogenic *Bacillus* spp. clustered as a separated sub-clade from *B. subtilis* (79% support). *Paenibacillus* spp. were outgrouped with cyanobacteria. The tree based on protein sequences (*Figure 2.2 b*) clustered *Pasteuria* in the '*Bacillus - Clostridium – Paenibacillus*' super-clade (73% bootstrap support). However, *Pasteuria* were observed to be closer to *Bacillus* spp. (75% support). *Pelosinus* spp. were clustered distinctly while the cyanobacterial sequences formed the outgroup.

2.3.3   Molecular phylogenetics of *Pasteuria* based on the *gyrB* gene and its product

The two nucleotide sequences from *P. penetrans* were placed close to *Bacillus* spp. in the *Bacillus-Clostridium* superclade. In 58% of the tree replicates, *P. penetrans* clustered together with bacilli (*Figure 2.3 a*). Unexpectedly, *P. hartismeri* clustered together with the cyanobacterium *Cyanobium gracile* in a superclade that also

contained *Paenibacillus* species. The other two cyanobacterial sequences from *M. elabens and A. platensis* appeared as outgroups but were still placed close to '*Paenibacillus – Cynobium - P. hartismeri*' superclade (91% support). Based on the GyrB protein sequences (*Figure 2.3 b*) the three *Pasteuria* sequences were grouped as a subclade close to *Paenibacillus* spp. (92% support). This '*Pasteuria - Paenibacillus*' clade formed part of a superclade that contained *Bacillus* species (99% support). The *Pelosinus* spp. and two of the clostridia formed separate clades. One of the clostridia appeared as an outgroup with the cynobacteria species.

2.3.4 Molecular phylogenetics of *Pasteuria* based on the *spo0A* gene and its product

Based on the *spo0A* gene and protein sequences, both *P. penetrans* and *P. ramosa* appeared to be more closely related to *Clostridium* than *Bacillus* (*Figure 2.4 a*) and b)*. The nucleotide-based tree had two superclades: the '*Bacillus - Paenibacillus*' superclade (84% bootstrap support) and the '*Pasteuria – Clostridium - Pelosinus*' superclade (65% support). The other tree based on the protein sequences of Spo0A had a superclade containing all the *Bacillus* spp. (82% support) and a '*Pasteuria - Clostridium – Paenibacillus - Pelosinus*' superclade (low bootstrap support of 28%).
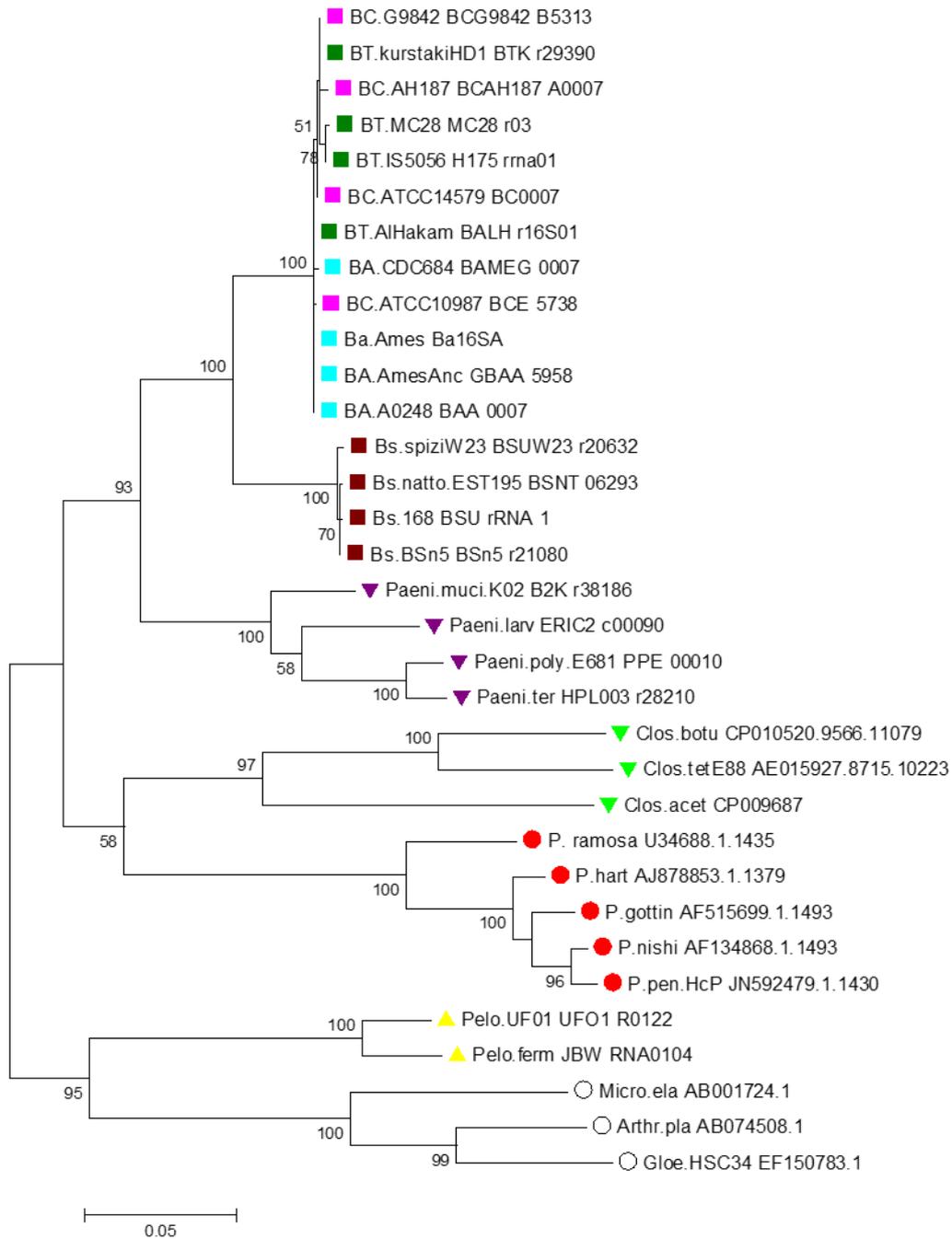
Figure 2. 1: Molecular Phylogeny based on 16S rRNA gene sequences by Maximum Likelihood method. The tree with the highest log likelihood (-11556.6217) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Numbers shown next to the branches are the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates). Only bootstrap values >50% are shown. The analysis involved 33 nucleotide sequences. There were a total of 1821 positions in the final dataset.
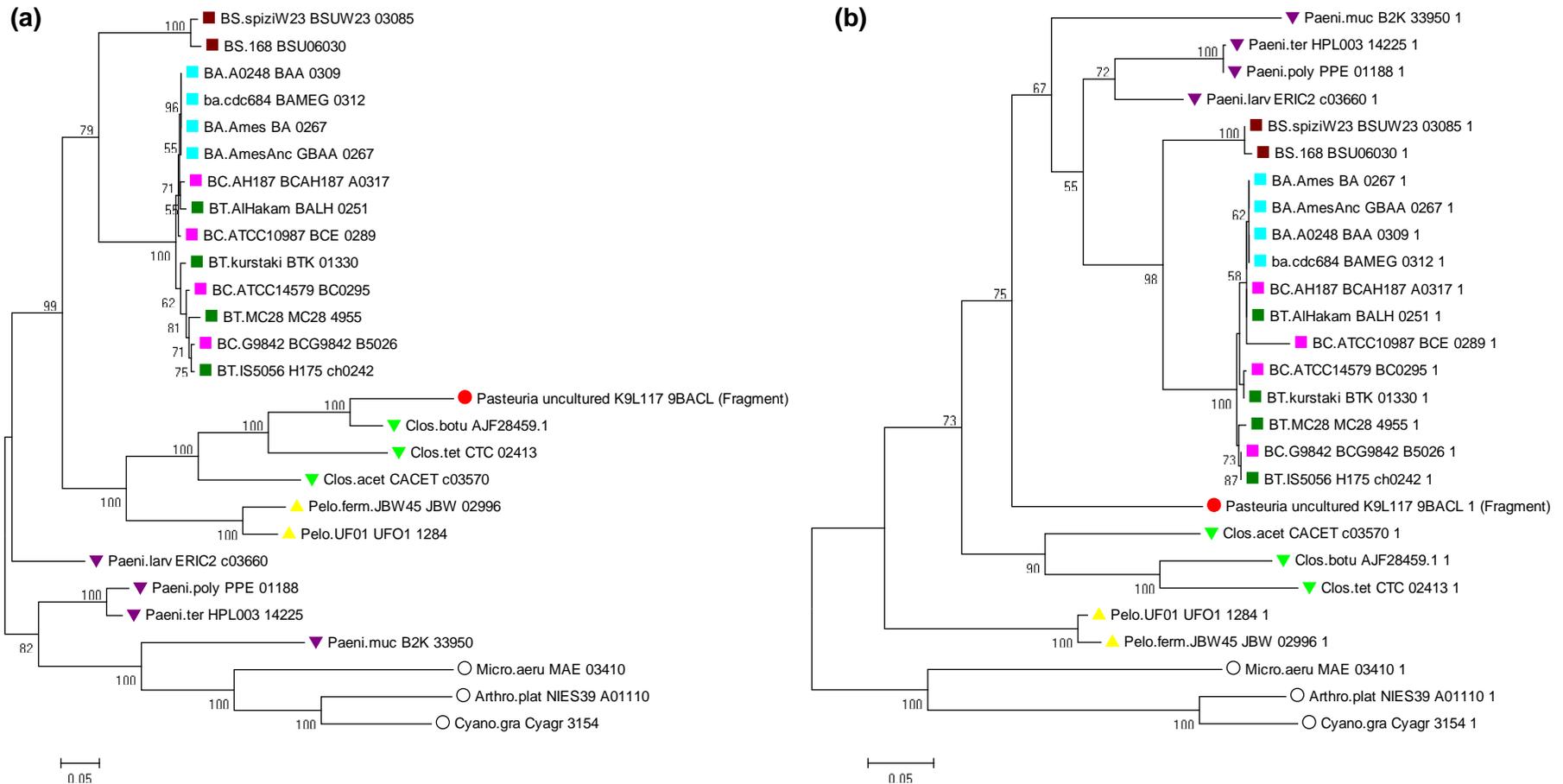
Figure 2. 2: Molecular Phylogeny based on (a) groEL gene and (b) GroEL protein sequences by Maximum Likelihood method (Bootstrap consensus tree) The trees with the highest log likelihood are shown. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Numbers shown next to the branches are the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates). Only bootstrap values >50% are shown. The analysis involved 27 nucleotide/protein sequences. There were a total of 1735 and 561 positions in the final datasets for (a) and (b) respectively. The log likelihood of the trees are: (a) -18978.6702 and (b) -7225.7735.
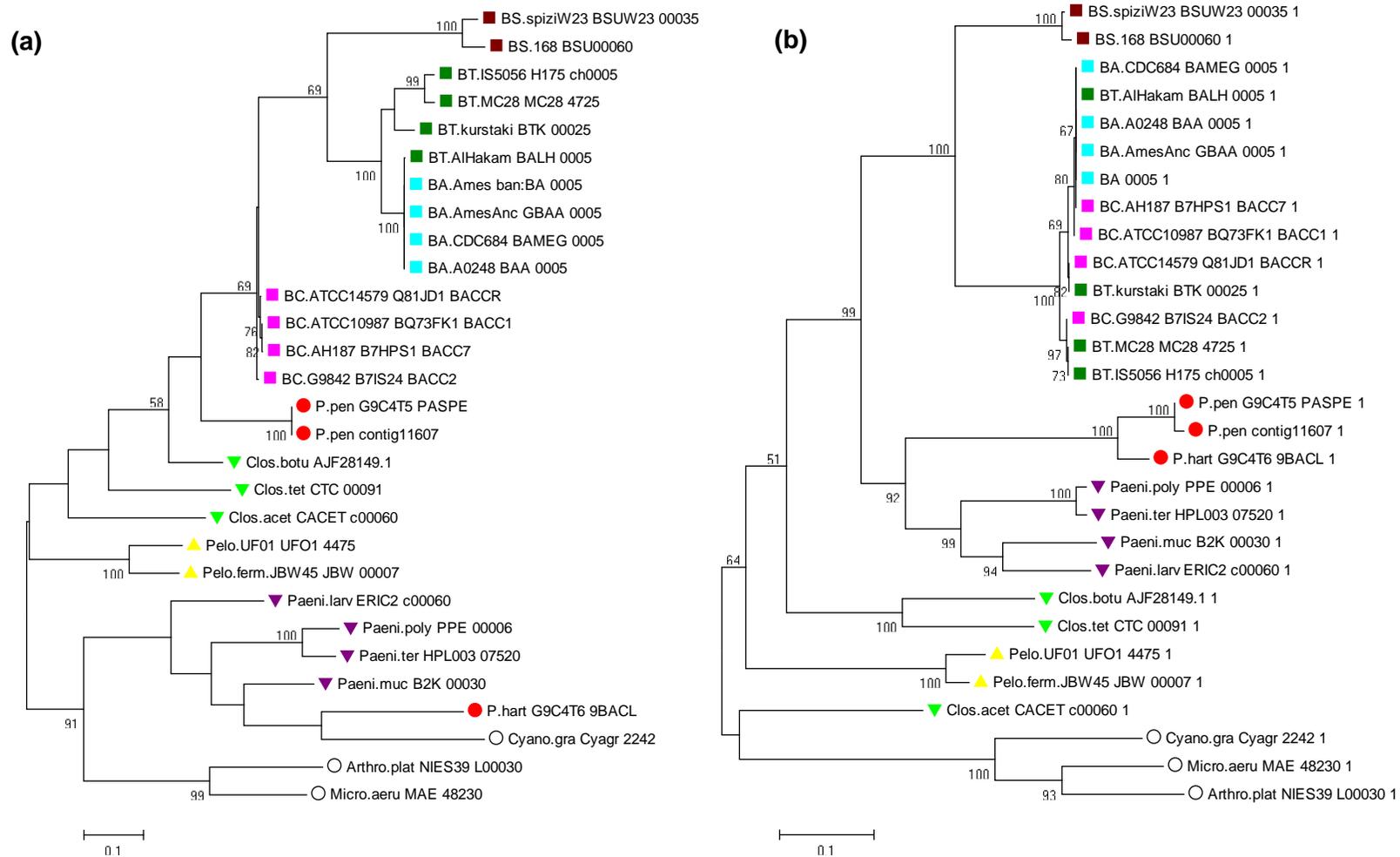
Figure 2. 3: Molecular Phylogeny based on (a) gyrB gene and (b) GyrB protein sequences by Maximum Likelihood method (Bootstrap consensus tree). The trees with the highest log likelihood are shown. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Numbers shown next to the branches are the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates). Only bootstrap values >50% are shown. The analysis involved 28 nucleotide sequences. There were a total of positions in the final dataset. There were a total of 4051 and 679 positions in the final datasets for (a) and (b) respectively. The log likelihood of the trees are: (a) -42453.5218 (b) -10535.1492.
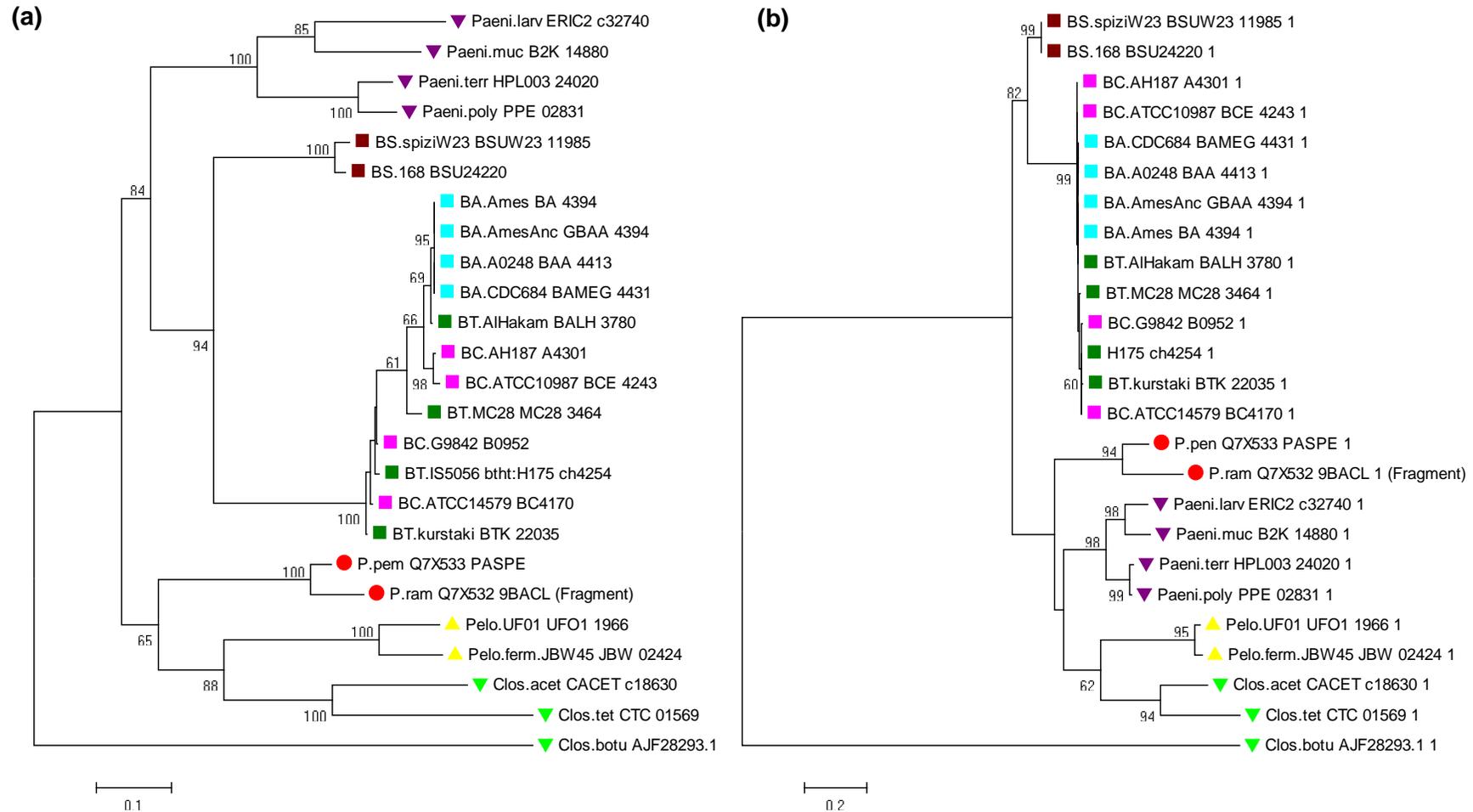
Figure 2. 4: Molecular Phylogeny based on (a) spo0A gene and (b) Spo0A protein sequences by Maximum Likelihood method (Bootstrap consensus tree). The trees with the highest log likelihood are shown. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Numbers shown next to the branches are the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates). Only bootstrap values >50% are shown. The analysis involved 25 nucleotide sequences. There were a total of 993 and 330 positions positions in the final datasets for (a) and (b) respectively. The log likelihood of the trees are: (a) -9721.7711and (b) -4390.0887.

2.4   SUMMARY OF RESULTS

The results of the phylogenetic reconstruction studies presented in this chapter suggest the close-relatedness of *Pasteuria* spp. to the members of *Bacillus* genus and therefore, it is speculated that *Bacillus* spp. can be used as comparative models for the genomic and proteomic studies of *Pasteuria* species.

2.5   DISCUSSION

The objective of the phylogenetic studies reported in this chapter was to support the idea that *Bacillus* spp. are suitable candidates for comparative genomic, proteomic and biochemical studies with *Pasteuria.* The study involved the analysis of important members of Firmicutes based on the classic 16S rRNA gene compared with the nucleic acid and amino acid sequences of three protein encoding genes viz., *gyrB*, *groEL* and *spo0A*. The 16S rRNA gene analysis placed *Pasteuria* closer to *Clostridium* than *Bacillus.* Each of the phylogenetic trees based on the three protein coding genes (*groEL*, *gyrB*, *spo0A*) slightly differed from each other in terms of positioning *Pasteuria* spp. amongst selected Firmicutes. However, the placement of *Pasteuria* in the '*Bacillus – Clostridium*' super-clade emerges as a consensus for most of the genes. This is in line with previous studies on *Pasteuria* phylogeny (Anderson *et al.*, 1999, Preston *et al.*, 2003, Trotter and Bishop, 2003, Charles *et al.*, 2005). Apart from the slight incongruence in the phylogenies estimated from different set of genes, there was also an observed phylogenetic biasedness between the trees based on the nucleic acid and amino acid sequences. The *groEL* gene and the *spo0A* gene and protein sequences placed *Pasteuria* closer to *Clostridium* while the GroEL protein and the *gyrB* gene sequences placed them closer to *Bacillus.* On the other hand, the GyrB protein sequence of *Pasteuria* were most closely related to *Paenibacillus* followed by *Bacillus* and *Clostridium.*

A possible explanation for the differences in the DNA-based and protein-based phylogenies is that proteins are under selective constraint due to their functional roles and are fairly conserved over geological timescale (Huynen and Bork, 1998, Romero and Arnold, 2009) as compared to their corresponding nucleotide coding sequence; degeneracy of codon usage plays a vital role in this conservation. Biasedness in the phylogenetic resolution of an organism using different genes could be attributed to

several factors; perhaps this biasedness hints for some important biological phenomena more than just being an artefact resulting from erroneous computational estimation of phylogeny and tree building. One such phenomenon is incomplete lineage sorting which is widespread across the kingdoms of life and reflects upon the evolutionary time-based selection pressures on the genomes (Maddison and Knowles, 2006, Degnan and Rosenberg, 2006). This selection pressure does not allow the convergence of ancestries of individual genes of a species to its overall observed phylogeny. Another important biological event that contributes to gene tree discordance is horizontal gene transfer (Maddison and Knowles, 2006, Philippe and Douady, 2003). Horizontal gene transfer (HGT) is common amongst bacteria, facilitates bacterial diversity and plays a major role in bacterial evolution (Philippe and Douady, 2003, Dutta and Pan, 2002). Thus, organismal phylogeny and taxonomy must be inferred in the light of such phenomena as incomplete lineage sorting and HGT to avoid any misinterpretation of incongruent trees. HGT events are also known to be an important source of ecological variances between closely related taxa (Cohan and Koeppel, 2008, Wiedenbeck and Cohan, 2011). Since soil is a dominantly shared habitat for the nematode pathogenic *Pasteuria* spp. and other Firmicutes, it is speculative that these organisms share a common horizontal gene pool.

Based on the molecular phylogenetic analyses presented in this chapter, it can be concluded that members of either of the two genera, *Clostridium* or *Bacillus*, could be used for comparative genomics with *Pasteuria. Bacillus* spp. are widely accepted as model organisms to study the process of sporulation in endospore forming bacteria. They are also easy to culture and work with. Thus, in the studies that follow, *Bacillus* spp. were preferred over *Clostridium* spp. as comparative model organisms to study *Pasteuria.* Considering collagens as important pathogenicity determinants thought to be important in *Pasteuria*-nematode interactions, in the next chapter, an attempt has been made to identify and characterize putative collagens of *P. penetrans* and compare them with collagens from other bacteria including closely-related *Bacillus* and *Clostridium* species.

Chapter 3

Putative collagens in *P. penetrans* and their *in-silico* characterization

3.1 INTRODUCTION

The previous chapter described the biological systematics of *Pasteuria* spp. in the context of its phylogenetic relationships within the Firmicutes using *groEL*, *gyrB* and *spo0A*. Genes that are responsible for host parasitic interactions are subject to the Red Queen hypothesis, i.e. they are involved in arms races and evolve at different rates (Van Valen, 1974). Previous studies have hypothesized that the collagen-like fibers on the surface of the *Pasteuria* endospore are involved in the *Pasteuria*-nematode interaction. Exploring the potential of the genome of *P. penetrans* to code for collagen-like proteins (CLPs) that could localize extracellularly on the endospore surface will be an important step important step in determining their role as adhesins and possible virulence factors. The current chapter describes the computational studies that were done to identify and characterize potential CLPs in *P. penetrans* that could be involved in *Pasteuria*-nematode interactions.

3.1.1 Collagen superfamily– the most ubiquitous protein family

Collagen is the most abundant protein present across the animal kingdom from the anatomically simplest sponges to the complex and advanced humans. It is a fibrous protein constituent of a number of our tissues and vital body parts like skin, corneas, bones, cartilage, ligament, tendons, the dentine in our teeth and the blood vessels (Lehninger, 1975). The word collagen comes from a Greek word 'Kolla' meaning 'glue' and hence 'collagen' means 'glue producing'. In ancient times, the skin, tendons and cartilage of animals were boiled to make protein colloidal glues resulting from hydrolysis of collagen. In multicellular organisms, a complex of collagens with some other specialized macromaolecules like proteoglycans form an extracellular matrix that acts like a glue or cement which binds cells and tissues together and are, thus, involved in providing structural strength and integrity. To date, at least 28 different types of collagens are known (Ricard-Blum, 2011). Type I, II and III are most abundantly found in humans (Kühn, 1987). Type IV collagens are found with an almost conserved homology from humans to the fruit fly, *Drosophila melanogaster* through to the free-living nematode, *Caenorhabditis elegans,* and the freshwater polyp *Hydra*

(Fowler *et al.*, 2000, Yasothornsrikul *et al.*, 1997, Johnstone, 2000). Interestingly, the *C. elegans* genome is known to possess 150 distinct collagen genes (Johnstone, 2000).

Initially, collagens were thought to be found only in animals. In the 1960s the first evidence of an extracorporeal existence of a 'collagen-like' substance was reported in the cocoons of a hymenopteran insect gooseberry sawfly (Rudall, 1967, Rudall and Kenching.W, 1971). Later studies proved their existence in bacterial kingdom (Engel and Bächinger, 1999, Charalambous *et al.*, 1988, Erickson and Herzberg, 1990). Genomic and proteomic studies have led to the identification of many collagen-like proteins (CLPs) in bacteria. Some examples of such proteins include the *Streptococcus* Scl proteins, the *Bacillus* proteins BclA, BclB, ExsH and ExsJ and the CL sequences in the endospore appendages of *Clostridium taeniosporum* (Sylvestre *et al.*, 2002, Todd *et al.*, 2003, Lukomski *et al.*, 2000, Rasmussen *et al.*, 2000, Walker *et al.*, 2007). Studies suggest that most of these bacterial CLPs are surface-associated. In the case of bacteria belonging to the phylum Firmicutes, such proteins have been known to be associated with the surface of endospores. Several studies suggest that surface associated bacterial CLPs act as virulence factors that allow a bacterial pathogen to interact with specific host surface receptors and help in invasion and pathogenicity.

3.1.2 The collagen superhelix and the low complexity regions (LCRs)

Mammalian collagen proteins are characterized by a unique 'supercoiled triple helix' (Ramachandran and Kartha, 1954) consisting of tightly packed amino acid chain rich in Proline or Hydroxyproline, essentially having a long stretch of amino acids with every third residue as Glycine (repetitive G-X-Y, where X and Y can be any amino acid), flanked by non-repetitive N- and C-terminal (Brodsky and Ramshaw, 1997, Ramachandran, 1956). However, a consensus amino acid composition of collagens is not evident across the animal kingdom. To quote as an example, Hydroxyproline which is consistently found in most of the animal collagens, is not present in the cocoons or the silk fibres drawn from the salivary glands of sawflies (Sutherland *et al.*, 2013, Rudall and Kenching, 1971). The low complexity G-X-Y regions can sometimes be interrupted with some non-repetitive amino acids. Amongst bacteria, formation of collagen triple helix has been demonstrated in the SclA protein of *Streptococcus* spp.

and the BclA protein of *Bacillus* spp. (Xu *et al.*, 2002, Boydston *et al.*, 2005, Rasmussen *et al.*, 2000). Like their mammalian equivalents, bacterial CLPs are rich in Proline and are characterized by the presence of the signature G-X-Y repeats with Pro residues mostly in the X position (Rasmussen *et al.*, 2003). However, bacterial CLPs are not known to undergo post-translational hydroxylation of proline.

3.1.3 Evolution of collagens and the role of LCRs

Low complexity regions, e.g. G-X-Y repeat regions in collagens, play major roles in the context of evolution (Zilversmit *et al.*, 2010, Toll-Riera *et al.*, 2011, Radó-Trilla and Albà, 2012). Due to the degeneracy of genetic code LCRs in protein sequences do not always mean low complex coding regions. For example, since the amino acid Serine (S) can be coded for by six different triplet codons (TCT, TCC, TCA, TCG, AGT, AGC), a low complex repetitive region of polypeptide chain rich in serine (such as S-S-S-S-S-S) can either be coded by a low complex nucleotide region having a single codon repeated six times in series or by a combination series of two or more codons making it comparatively complex at nucleotide level. One of the major evolutionary events thought to be involved in the generation of low complexity at nucleotide level is replication slippage (DePristo *et al.*, 2006). A replication slippage is when, at the sites of repetitive nucleotides, mispairing occurs between the newly synthesized strand of DNA and the template during the replication process (Levinson and Gutman, 1987). This leads to either extension or reduction in the number of repetitive units in the new strand. In non-repetitive/complex sequences, replication slippage events could lead to deleterious frameshift mutations (Streisinger *et al.*, 1966) and, therefore, the chances of such an event are eliminated by natural selection and evolutionary bias. However, when replication slippage occurs in highly repetitive nucleotide sequences, the reading frame of the protein remains the same. Thus, LCRs are an important source of genetic variability. The existence of collagens or collagen-like proteins across the different kingdoms of life can, therefore, be attributed by the presence of LCRs in them which make them prone to rapid evolution.

3.1.4 BclA - an endospore associated glycosylated CLP in *Bacillus* spp.

The hair-like exosporial nap projected out from the exosporium of most pathogenic *Bacillus* spp. (*B. anthracis*, *B. cereus*, *B. thuringiensis*) is composed of a fibrous collagen-like glycoprotein called BclA for '*Bacillus collagen-like* protein of *anthracis*' as it was first characterized in *Bacillus anthracis* endospores (Sylvestre *et al.* 2002). Like other collagens, BclA essentially consists of three domains: the N-terminal (NTD), the central GXY repeat region, and the C-terminal (CTD). BclA is known to be synthesized temporally as an immunodominant protein during early sporulation stages. The first few amino acids of the translated protein get proteolytically cleaved off before the NTD is anchored to the paracrystalline basal layer of the exosporium (Thompson and Stewart, 2008, Tan and Turnbough, 2010a). The globular CTD is thought to facilitate a stable trimerization of the mature BclA protein monomers (Boydston *et al.*, 2005). The protein has a predicted molecular weight of 37 kDa but has an apparent molecular weight of >250 kDa due to heavy glycosylation and trimerization of the collagen-like helices. The length of the exosporial fibres depends on the number of GXY triplet repeats, most of which are GPT repeats (Sylvestre *et al.*, 2003). The GXY region is heavily O-glycosylated at specific serine/threonine residues (Daubenspeck *et al.*, 2004b, Maes *et al.*, 2016). The endospore associated glycans are thought to play important roles in the interaction of endospores with host surface receptors (Lequette *et al.*, 2011, Bozue *et al.*, 2005). BclA is the most characterized endospore associated glycoprotein to date.

The gene *bclA* is located within a cluster of genes which includes the genes of the highly conserved rhamnose cluster operon (Todd *et al.*, 2003, Steichen *et al.*, 2003). The genes in the rhamnose operon i.e. *rfbA*, *rfbB*, *rfbC* and *rfbD* code for glucose-1-phosphate thymidylyltransferase (EC 2.7.7.24), dTDP-glucose 4,6-dehydratase (EC 4.2.1.46), dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13) and dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) respectively. These are the enzymes involved in the synthesis of rhamnose, one of the sugars known to glycosylate the BclA protein. Mutation in *rfbA* gene has shown a decrease in the binding of endospores to macrophages, suggesting a role of rhamnose in host interaction, possibly as a glyco-conjugate of BclA (Bozue *et al.*, 2005).

3.1.5 Parasporal 'BclA-like' fibres of *Pasteuria*

The parasporal fibres associated with the skirt-like exosporium of *Pasteuria* spp. (Davies 2009), appear to be corresponding to the hair-like nap of the exosporium of closely related bacilli (Gerhardt and Ribi, 1964). Persidis *et al.*, (1991) subjected endospores to various chemical disruptive techniques and negative staining showed an extensive fibrillar network that when disrupted greatly reduced endospore adhesion to nematode cuticle. *Figure 3.1* shows a comparison of electron micrographs of endospores of *Bacillus anthracis* and *P. penetrans* showing the similarities in their exosporial fibres. Interestingly, collagen-like (CL) domains have also been reported in the cladoceran hyperparasites *Pasteuria ramosa* and a triple helical structure has been proposed (Mouton *et al.*, 2009, McElroy *et al.*, 2011). Schaff et al (2011) reported genes analogous to some 12 genes from the Rhamnose Cluster Operon to be present in *P. penetrans* based on BLASTp searches using *P. penetrans* sequences against *Bacillus anthracis*, *Bacillus thuringiensis* and *Bacillus subtilis* yielding very significant hits of E-values <e^{-14}. A genome survey sequencing project of *Pasteuria penetrans* had also revealed the presence of similar CL sequences in *P. penetrans* strain RES147 (Davies and Opperman, 2006). Previous studies have shown an inhibited attachment of *Pasteuria* endospores to their host nematodes when pre-treated with collagenase or with a collagen-binding domain of fibronectin (Davies and Danks, 1993, Mohan *et al.*, 2001). Such studies indicate an involvement of CLPs in the attachment of *Pasteuria* endospores to the cuticle of the host nematode. CLPs in the nematode parasitic *Pasteuria* spp. might be involved not just in the initial interactions of *Pasteuria* endospores to the nematode cuticle but also in the structural integrity of the endospores.

3.1.6 Synteny analysis

A pivotal strategy in comparative genomics is studying the conservation in synteny between organisms (Zody, 2007). Synteny can be defined as the physical localization of genes on the complete genome (macro-synteny) or on a local area within the genome (micro-synteny) of an organism. The concept of synteny analysis relies on genetic linkage and on the fact that genes which are functionally related tend to be located closely on the chromosome of an organism (Wei *et al.*, 2002). The co-localization of genes on the chromosomes of different species, known as conserved

or shared synteny, is an important clue for evolutionary and speciation studies (Larkin *et al.*, 2009). Events like gene translocations and transversions disrupt the syntenic conservation between organisms of different taxa during the course of evolution. Thus, shared synteny can be used to establish functional relationships between genes. Synteny conservation maps derived for closely related species can help in identifying orthologous genes or in inferring the genome organization in a partly sequenced or unsequenced genome (Yelton *et al.*, 2011). A synteny-based approach could be used for the genome assembly of *Pasteuria* spp. and for the functional prediction of their genes.

3.1.7 Aims and objectives

Major Aim: To characterize the putative collagen-like proteins that are thought to be associated with the endospore and possibly involved in the adhesion of the endospore of *P. penetrans* to the nematode cuticle.

Specific Objectives:

1) To investigate the presence of candidate genes coding for collagen-like proteins in a set of unpublished contigs from an isolate of *P. penetrans.*

2) To characterize the putative CLPs of *P. penetrans* using different bioinformatics tools and prediction algorithms.

3) To compare putative CLPs of *P. penetrans* with other bacterial CLPs based on sequence homology and conserved motifs.

4) To derive the cladistics of putative CLPs of *Pasteuria* with other related CLPs based on the amino acid composition of their low-complex GXY repeat regions.
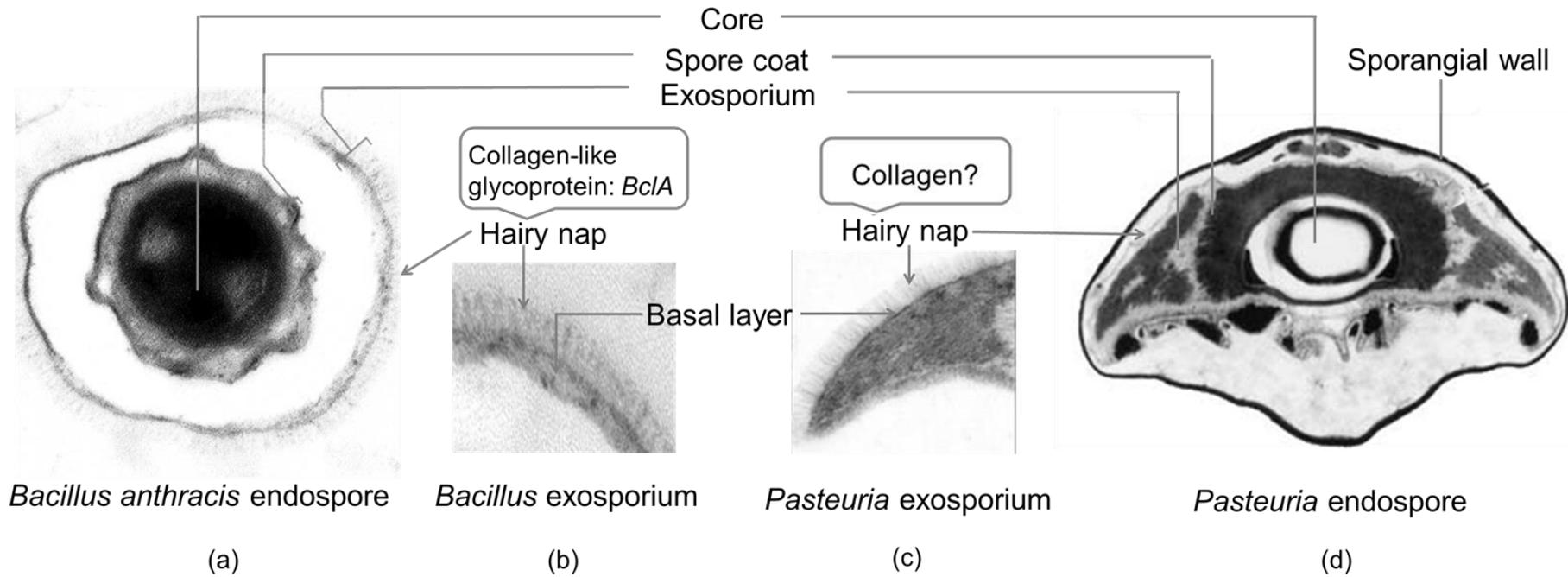
Figure 3. 1: Comparative micrographs of B. anthracis and P. penetrans endospores. (a) & (b): Adapted from Driks (2002); (c) From Davies (2009); (d) From Giblin-Davis *et al.* (2001). The figures are not to relative scales.

3.2 MATERIALS AND METHODS

3.2.1 *Pasteuria* sequences

All the *in silico* analyses were conducted on either the publicly available genome survey sequences (GSS) for *Pasteuria penetrans* Res147 or on the unpublished contigs for *Pasteuria* Res148 isolate, a related but highly passaged sub-population, of RES147 (Mauchline *et al.*, 2011). The *Pasteuria penetrans* Res148 contigs were kindly provided by Dr. Timothy Mauchline (Rothamsted Research, Harpenden).

3.2.2 Search for protein sequences coding for putative collagens in *Pasteuria*

Protein sequences of the BclA orthologs in six, thirteen and twelve different strains of *B. anthracis*, *B. cereus* and *B. thuringiensis*, respectively, were retrieved from the NCBI database. These sequences were used as a query to look for similar sequences in *Pasteuria penetrans* Res148 in the GSS database using the tBLASTn tool from the NCBI BLAST package (Altschul *et al.*, 1990). The scoring parameters were set as default (Matrix: BLOSUM 62; Gap costs: existence 11, extension 1; Conditional compositional score matrix adjustment). The low complexity region filter was turned off. Once the presence of similar CLPs was indicated from the tBLASTn results, the next objective was to identify and retrieve the complete collagen-like sequences. To this end, an unpublished set of sequence contigs from *P. penetrans* Res148 was used. The gene predictions on the *P. penetrans* Res148 contigs were done using the RAST annotation web server (Aziz *et al.*, 2008) and the annotations were searched for any predicted collagen-like sequences based on the comparison of contig annotations with sequenced genome of a closely related firmicute, *Bacillus thuringiensis* strain Al Hakam. To look for any collagen-like sequences not predicted by RAST, the contigs were uploaded on to Artemis genome browser and annotation tool (Carver *et al.*, 2012) and were manually searched for open reading frames containing 'G-X-Y' triplet amino acid repeats using the 'navigator' feature of Artemis (Rutherford *et al.*, 2000). The sequences were further tested for the presence of collagen motifs using MOTIF search tool of GenomeNet web server (Kanehisa *et al.*, 2002).

3.2.3 Comparative synteny analysis of *Pasteuria* CLP-coding genes with *bclA*

The local conservation of microsyteny in the upstream and downstream regions of *bclA* across different strains of Bacillus spp. was studied in the web-based multi-genome browser of the Biocyc database collection (Caspi *et al.*, 2016). The nucleotide sequences of the syntenic genes were retrieved for further analyses. The contigs for *P. penetrans* Res148 were searched for the orthologs of these genes using BLASTp tool on the RAST annotation server (Aziz *et al.*, 2008). The synteny around *Pasteuria* CLP-coding genes was studied based on the RAST annotations.

3.2.4 Characterization of putative *Pasteuria* collagen-like proteins using membrane-spanning, hydrophobicity and glycosylation prediction tools

A. Prediction of membrane-bound/spanning collagens based on hydrophobicity

The proteins that function outside of the bacterial cell cytosol are either membrane-bound or membrane-spanning proteins that are translocated across bacterial cytoplasmic membranes after their synthesis in the cytosol. The proteins that are to be targeted to the outside of cytoplasmic membrane are synthesized with an extended N-terminal containing a hydrophobic signal peptide which gets cleaved away from the protein when the protein is being translocated to the membrane. Thus, the presence of signal peptides in the amino acid sequence of a protein serves as an indication of the membrane localization of a bacterial protein. Signal peptide cleavage sites in the selected putative collagen-like sequences identified in *Pasteuria* were predicted using SignalP 4.1 server (Petersen *et al.*, 2011) and TatP 1.0 server (Bendtsen *et al.*, 2005). Only the NTD was used as a query for cleavage site prediction. The TMHMM 2.0 server (Krogh *et al.*, 2001) was used to predict the transmembrane regions in the sequences. The hydrophobicity profiles were predicted using the inbuilt Kyte and Doolittle program in Bioedit software (Hall, 2007, Kyte and Doolittle, 1982).

B. Prediction of glycosylation sites in putative *Pasteuria* collagens

Glycosylation is an important post-translational modification of a protein that involves the covalent attachment of a carbohydrate to the protein at a specific site. In bacterial

glycoproteins, glycosylation is either N-linked or O-linked. N-glycosylation occurs at a consensus amino acid site Asn-Xaa-Ser/Thr (where Xaa is any amino acid, except Pro). O-glycosylation which is much more common in bacterial glycoproteins is known to occur at Ser and Thr residues and some at Tyr residues. Apart from effecting the chemical properties, glycosylation has major effects on the biological activity of a protein including its antigenicity and its cell to cell interactions (Lequette *et al.*, 2011, Schmidt *et al.*, 2003). Thus, predicting the potential of glycosylation on a putative protein coding sequence is of importance in the prediction of its biological function. To this end, both N- and O-glycosylation sites in the putative collagen-like sequences in *Pasteuria* were predicted using the NetNGlyc 1.0 (Gupta *et al.*, 2004) and NetOGlyc 40 servers (Steentoft *et al.*, 2013).

3.2.5 Comparison of CL sequences in *Pasteuria* and other collagens

Several strategies were used to compare the putative collagen-like sequences identified in *Pasteuria* with other bacterial CLPs. First, the putative *Pasteuria* collagens were used as query sequences for BLASTp searches targeting non-redundant (nr) protein database. The low complexity filter was turned on for these searches to avoid any random hits to low complex G-X-Y repeats of collagens. The idea was to look for sequences similar to the non-collagenous NTD and/or CTD of the putative *Pasteuria* collagens. BLASTp hits with significantly low E-values were pooled together with the putative *Pasteuria* collagens and multiple sequence alignments were performed in Geneious 10.0.8 (Kearse *et al.*, 2012) using MUSCLE program (Edgar, 2004) with default settings. The alignments were observed for the presence of any conserved motifs in the NTD and the CTD. The web-based version of MEME (Multiple Expectation Maximization for Motif Elicitation) was used for the prediction of statistically significant conserved motifs (Bailey *et al.*, 2006).

The G-X-Y repeat regions of selected CLPs were analyzed for the diversity in their percentage amino composition. The G-X-Y repeat regions were extracted from the sequences (i.e. without NTD and CTD). A customized script was written in *R* (R Development Core Team, 2010) to do the following analyses. The Manhattan distance between the percentage amino acid compositions for all possible pairs of sequences was computed. The resulting distance matrix was subjected to an agglomerative hierarchical clustering method using hclust function (Murtagh and Legendre, 2011).

The heatmap.2 function of gplots package in R (**http://cran.r-project.org/web/packages/gplots/index.html**) was used to generate a heatmap representation of the Manhattan distances between the percentage G-X-Y composition of different collagen-like sequences. The dendrogram was exported in Newick format and was edited and visualized in FigTree (**http://tree.bio.ed.ac.uk/software/figtree/**). See *Appendix Ia* for the R-script.

3.3   RESULTS

Figure 3.2 shows the workflow and summary of the prediction of putative CLPs in *P. penetrans* and their in-silico characterization.

3.3.1 Putative collagen-like proteins in *Pasteuria*

The initial tBLASTn searches suggested that there are *bclA*-like genes present in the *P. penetrans* Res147 genome. As high as 200 hits were obtained using the BclA-like sequences from different *Bacillus* species. The topmost significant hits had an E-value range of between 5e-05 to 7e-16. Only some of the query sequences hit a fewer number of target sequences; all these hits were to non-collagen-like sequences. *Table 3.1* lists the topmost hits for all the queried sequences. All the significant hits were for the G-X-Y repeat region of the sequences (*Figure 3.3*). None of the NTD or CTD regions of the sequences were found to have hits to BclA. Using sequence-based comparison tool of RAST, 17 putative collagen coding genes were identified in different contigs obtained from the sequencing of *Pasteuria* Res148. Through manual search a further of 16 open reading frames containing G-X-Y repeat regions were found. These sequences (33 in total) were named as Ppcl for '**P**asteuria **p**enetrans **c**ollagen **l**ike sequences' and they were numbered as Ppcl1 to Ppcl33. Out of these, 23 sequences were unique and intact (i.e. started with a start codon and ended with a stop codon). To confirm that these sequences are related to the collagen superfamily, these sequences were searched for collagen motifs in MOTIF search. Only 17 of these sequences hit the Pfam: collagen family and were considered as the putative CLPs in *P. penetrans* Res148 (*Table 3.2* and *3.3*). For all further analyses, only these 17 Ppcl sequences were used.
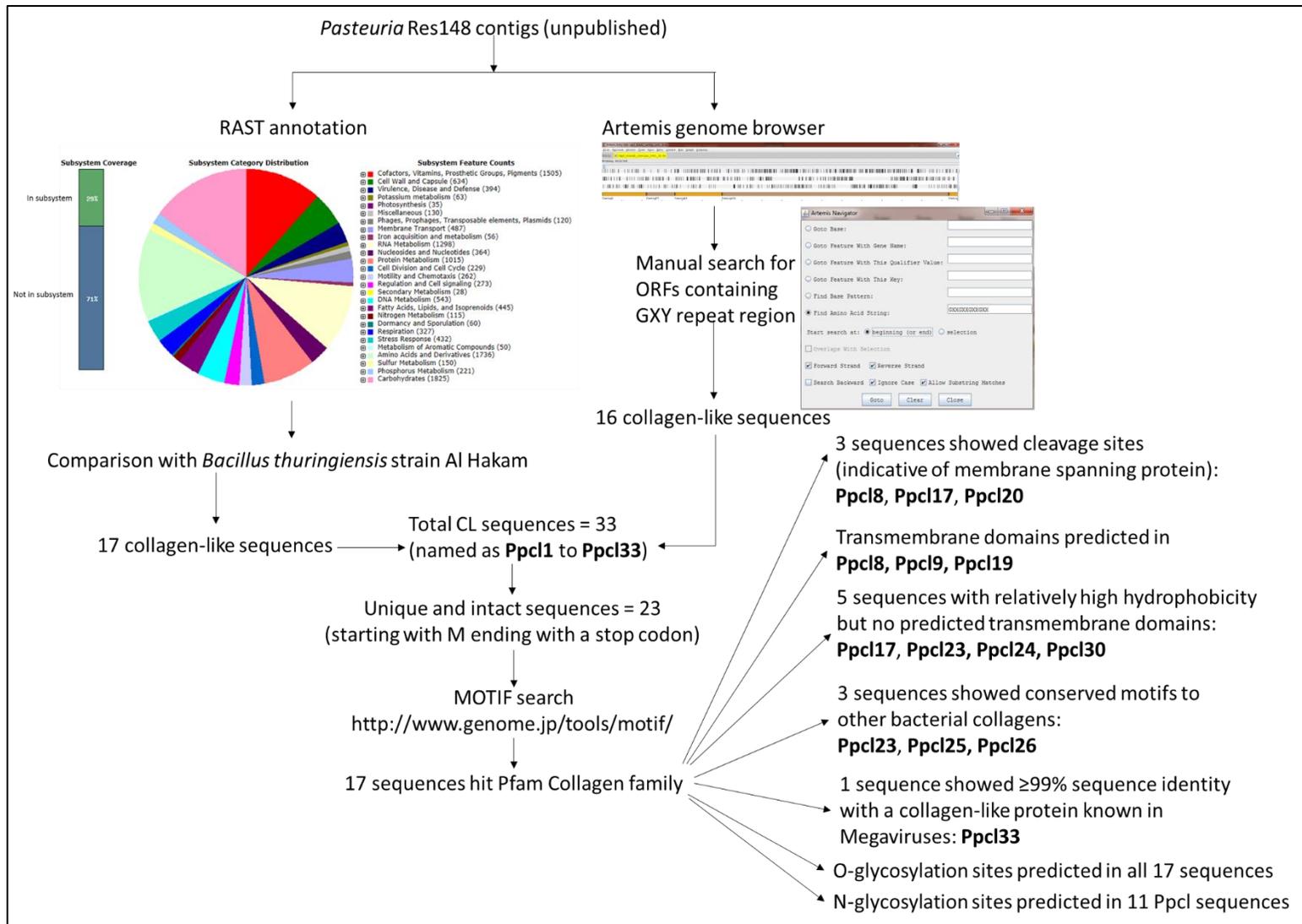
Figure 3. 2: Workflow of identification and characterization of putative collagens in Pasteuria penetrans Res148

Table 3. 1: Summary of tBLASTn hits to *P. penetrans* Res147 using BclA protein sequence from pathogenic Bacillus spp. as query. The table shows the topmost hits with their corresponding E-values and percent identity scores. The low number of hits are to non-collagen-like sequences and hence have non-significant E-values.

Query *Bacillus* strains	tBLASTn hits to *P. penetrans* Res147 (Genome survey sequences)			
	Total number of hits	E-Value (topmost hit)	Percentage Identity	Accession number of *Pasteuria* hit
B. anthracis Ames Ancestor	200 hits	3e-14	73%	CG896951.1
B. anthracis A0248	6 hits	4.9	22%	CG898186.1
B. anthracis Ames	200 hits	3e-14	73%	CG896951.1
B. anthracis CDC 684	200 hits	6e-13	68%	CG896951.1
B. anthracis H9401	200 hits	3e-14	65%	CG896951.1
B. anthracis Sterne	200 hits	4e-14	73%	CG896951.1
B. cereus 03BB102	200 hits	5e-15	34%	CG897377.1
B. cereus AH187	5 hits	3.1	35%	CG897701.1
B. cereus AH820	200 hits	5e-15	56%	CG897377.1
B. cereus ATCC 10987	200 hits	4e-14	56%	CG896951.1
B. cereus ATCC 14579	3 hits	6.5	33%	CG897701.1
B. cereus B4264	200 hits	1e-15	59%	CG897377.1
B. cereus biovar anthracis CI	200 hits	5e-05	67%	CG896951.1
B. cereus E33L	200 hits	1e-15	56%	CG897377.1
B. cereus F837-76	200 hits	1e-15	72%	CG896951.1
B. cereus FRI35	200 hits	7e-13	60%	CG896123.1
B. cereus G9842	200 hits	8e-14	54%	CG896951.1
B. cereus NC7401	200 hits	2e-15	71%	CG896951.1
B. cereus Q1	200 hits	5e-14	56%	CG897377.1
B. thuringiensis Al Hakam	200 hits	1e-15	72%	CG896951.1
B. thuringiensis BMB171	200 hits	1e-15	59%	CG897377.1
B. thuringiensis Bt407	200 hits	7e-16	59%	CG897377.1
B. thuringiensis HD771	200 hits	3e-12	52%	CG896951.1
B. thuringiensis HD789	200 hits	1e-12	58%	CG896951.1
B. thuringiensis MC28	3 hits	8.0	42%	CG898253.1
B. thuringiensis chinensis CT43	3 hits	6.8	33%	CG897701.1
B. thuringiensis finitimus YBT020	200 hits	1e-14	59%	CG897377.1
B. thuringiensis IS5056	200 hits	1e-14	60%	CG896951.1
B. thuringiensis konkukian 97-27	200 hits	1e-15	70%	CG896951.1
B. thuringiensis kurstaki HD73	200 hits	4e-13	70%	CG896951.1
B. thuringiensis YBT 1518	200 hits	5e-12	66%	CG896951.1

Figure 3. 3: *bclA*-like (collagen-like) protein sequences in *Pasteuria*. The alignments of the tBLASTn hits to *Pasteuria* Res147 with the bclA genes from *B. cereus* strain G9842 and *B. thuringiensis* strain HD771, have been shown here. The most commonly occurring amino acids (G, P, T) have been color-coded to make the identities visually clear (G=red, P=Green, T=yellow). In between the query and the subject sequences the identities are shown.

Table 3. 2: List of 17 putative CLPs in *P. penetrans* Res148 (See Appendix II for complete sequences)

S.No.	Name	Contig	Contig size	Length	5' terminal	3' terminal	Location within the contig	Frame within the contig
1.	Ppcl1	2047	4238 bp	1215 bp/ 405 aa	MSNLELLHRLCC	RQVVVIELPSGN	221..1435	+2
2.	Ppcl8	21351	5054 bp	753 bp/ 251 aa	MPNHSGLRGSPL	GFVGLVENRGGL	4236..4988	-1
3.	Ppcl9	2415796	4210 bp	669 bp/ 223 aa	MISVVVTMTSPL	SRSPHAEMDYLP	1971..2639	-1
4.	Ppcl16	176	4092 bp	606 bp/ 202 aa	MYHNDYQGKMSD	PCPPPPYPHREY	1231..1836	+1
5.	Ppcl17	2415632	9401 bp	1242 bp/ 414 aa	MKRSTKYPFLAM	GQAANLIIRRVF	3647..2406	-2
6.	Ppcl18	Contig4909	4701 bp	1170 bp/390 aa	MKIKTLLLFILG	TTSISMYVRQIA	2726..1557	-3
7.	Ppcl19	Contig4909	4701 bp	1143 bp/ 381 aa	MIMKAILNIYLI	TAASLLIKRIAS	3882..2740	-3
8.	Ppcl20	Contig4909	4701 bp	780 bp/ 260 aa	MRGNARIGGNLI	RATASVMIRQIF	4669..3890	-1
9.	Ppcl21	2415917	14096 bp	1938 bp/ 646 aa	MLEFHLPESSYI	SSGASFTIRRVA	96..2033	+3
10.	Ppcl23	Contig14026	6064 bp	1158 bp/ 386 aa	MLAVLLSAPLCA	SISASVLVRRIA	204..1361	+3
11.	Ppcl24	Contig14026	6064 bp	1203 bp/ 401 aa	MNEVTQLSQADY	GTAFSLMIRRLN	4030..5232	+1
12.	Ppcl25	Contig14026	6064 bp	1179 bp/ 393 aa	MKKIIIYLLLIS	SINASILIRQIS	1375..2553	+1
13.	Ppcl26	Contig20147	16310 bp	837 bp/ 279 aa	MASLNKVRVQLL	TATQANLFFKLV	9144..9980	+3
14.	Ppcl28	Contig2047	4238 bp	843 bp/ 281 aa	MILNLFPPCGFP	VTITKYSDSICS	3118..3960	+1
15.	Ppcl29	2415796	4210 bp	837 bp/ 279 aa	MILNLFPPCGFP	VFQYSTNICISQ	2197..3033	+1
16.	Ppcl30	Contig20796	1784 bp	774 bp/ 258 aa	MLIGGNLFVNGT	GTAFSLTIIRLN	1644..871	-1
17.	Ppcl33	Contig21350	6571 bp	1770 bp/ 590 aa	MSRSQNNIINYV	SQKTWILIEQIY	4612..6381	+1

Table 3. 3: CL motifs in putative collagen-like sequences in *P. penetrans* Res148. CL motifs were predicted by GenomeNet MOTIF using the Pfam database. The location of each CL motif within the sequence is shown along with the E-value for each motif.

Sequence	Location of CL motif	E-values	CL motifs along putative CL sequences in *Pasteuria* RES 147
Ppcl1	21..57 62..119 110..166 181..224 220..278	0.00021 1.3e-09 9.3e-07 0.0041 5.2e-09	
Ppcl8	128..176 156..210	2e-06 1.7e-07	
Ppcl9	87..121	0.88	
Ppcl16	67..121 97..154	1.1e-09 6.3e-10	
Ppcl17	198..253	1.3e-05	
Ppcl18	166..210 180..234	2.2e-06 3.1e-08	
Ppcl19	154..207 176..228	2.2e-08 1.5e-07	
Ppcl20	28..75 53..111	3.4e-07 1.6e-09	
Ppcl21	387..444 432..490	3.4e-07 1.6e-09	
Ppcl23	159..209 191..246	1e-05 1.3e-08	
Ppcl24	145..197 190..246	2.8e-08 6.5e-08	
Ppcl25	169..217 190..245	8e-07 4.4e-09	
Ppcl26	124..169 232..256	1.2e-08 0.00019	
Ppcl28	22..70 46..97	2.1e-07 7.4e-07	
Ppcl29	22..73 51..97	5.5e-08 1.3e-07	
Ppcl30	25..80 59..112	1e-08 8.8e-08	
Ppcl33	34..89 79..130 130..181	7.6e-09 7.5e-08 4.1e-07	

3.3.2 Comparative synteny analysis

The in-silico studies suggest that across pathogenic *Bacillus* spp. the synteny around the gene *bclA* is fairly conserved apart from a few exceptional insertions (*Figure 3.4*). Broadly, *bclA* was always found in the close vicinity of the rhamnose operon. Two genes coding for endospore coat related proteins, spore coat protein Z1 and spore coat protein Z2, were found in the syntenic regions upstream to *bclA*. An exosporium related gene *exsF* was found sandwiched between the genes for spore coat proteins Z1 and Z2.

The synteny around the CLP-coding genes in *Pasteuria* could not be assessed because of the small size of the available contigs. However, BLASTp results indicated that genes syntenic to *bclA* have orthologs in *P. penetrans* Res148. At least 21 out of 34 genes syntenic to *bclA* gave considerable hits to *P. penetrans* Res148 with significantly low E-values.
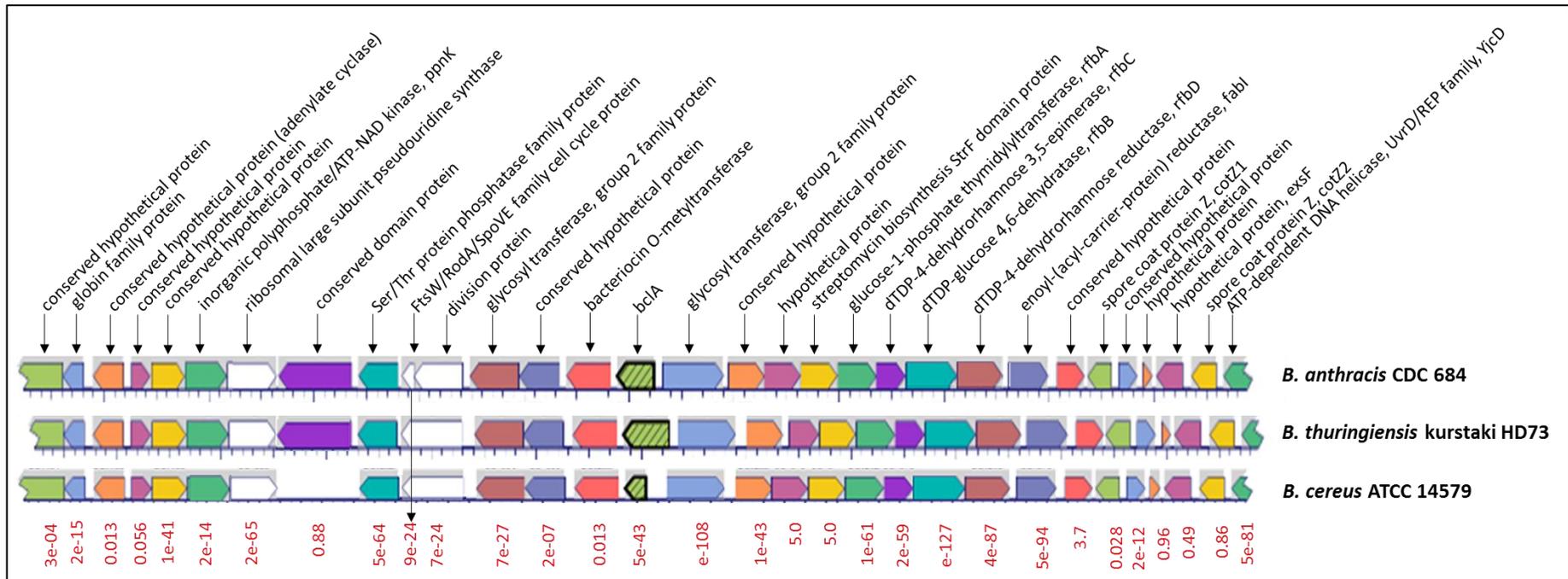
Figure 3.4: Conservation of synteny around the *bclA* gene in *B. anthracis* CDC 684 and *B. thuringiensis* kurstaki HD73). The numbers in red below the synteny maps are the E-values of the topmost BLASTp hits to *P. penetrans* Res148 contigs (irrespective of the query *Bacillus* species). The synteny map was obtained using the Biocyc comparative genome tool and was edited to show the names of encoded proteins and names of bacterial species and strains (on the right).

3.3.3 Characterization of Ppcl proteins

A. Membrane-bound/spanning Ppcl and their hydrophobicity profiles

To further characterize the Ppcl sequences, they were run through the TMHMM 2.0 server, that predicts transmembrane helices; three sequences, namely Ppcl8, Ppcl9 and Ppcl19, were predicted to contain transmembrane helices. Figure 3.5 and 3.6 show the hydrophobicity Kyte and Doolittle hydrophobicity profiles and TMHMM predictions of selected Ppcl proteins, Pcl1a and Pcl2 in comparison with ExsJ and BclA proteins. It is notable that the predicted transmembrane domains of Ppcl8 are similar to those of the ExsJ protein of *B. cereus*, localized at the CTD in both. Two of the *P. ramosa* collagens Pcl1a and Pcl2 were predicted to contain transmembrane domains, one domain near the N-terminal in Pcl1a, and one domain each on N- and C-terminals in Pcl2. No transmembrane domains were predicted for BclA, a collagen-like protein that was initially identified in *B. anthracis* and was used as a comparator; however, it was predicted to be hydrophobic at its CTD. Similar properties were predicted for four Ppcl sequences (Ppcl17, Ppcl23, Ppcl24 and Ppcl30) which had no predicted transmembrane regions but reasonably hydrophobic CTD. The remaining putative Ppcl proteins were predicted to be on the outside of the membrane. Since the TMHMM tool is particularly for the prediction of transmembrane helices and not the localization of a protein, the exact prediction about the localization of proteins (outside/inside) cannot be completely relied upon. When putative collagens from *P. penetrans* Res148 were run through SignalP 4.1 and TatP 1.0 prediction algorithms, only Ppcl17 was found to contain a SignalP cleavage site between amino acid positions 28 and 29; TatP sites were predicted for Ppcl8 (a.a. position 55-56: GFI-GP) and Ppcl20 (a.a. position 26-27: GQA-A). *P. ramosa* collagen Pcl2 and *Bacillus* spp. collagen BclA were predicted to possess SignalP and TatP sites respectively (*Figure 3.7*). No cleavage sites were predicted for ExsJ and Pcl1a.

B. Glycosylation sites in putative Ppcls

Several O-glycosylation sites were predicted for all Ppcl sequences ranging from 2 sites in Ppcl16 to 114 sites in Ppcl21 (*Table 3.4*). BclA, ExsJ, Pcl1a and Pcl2 were predicted to have 83, 57, 9 and 30 putative sites for O-glycosylation respectively. Apart from O-glycosylation sites, putative sites for N-glycosylation were predicted for Ppcl8,

Ppcl17, Ppcl18, Ppcl21, Ppcl23, Ppcl25, Ppcl26, Ppcl28, Ppcl29, Ppcl30, Ppcl33 (*Figure 3.8*). One site each was predicted in BclA, ExsJ and Pcl2. No N-glycosylation sites were predicted for Pcl1a.
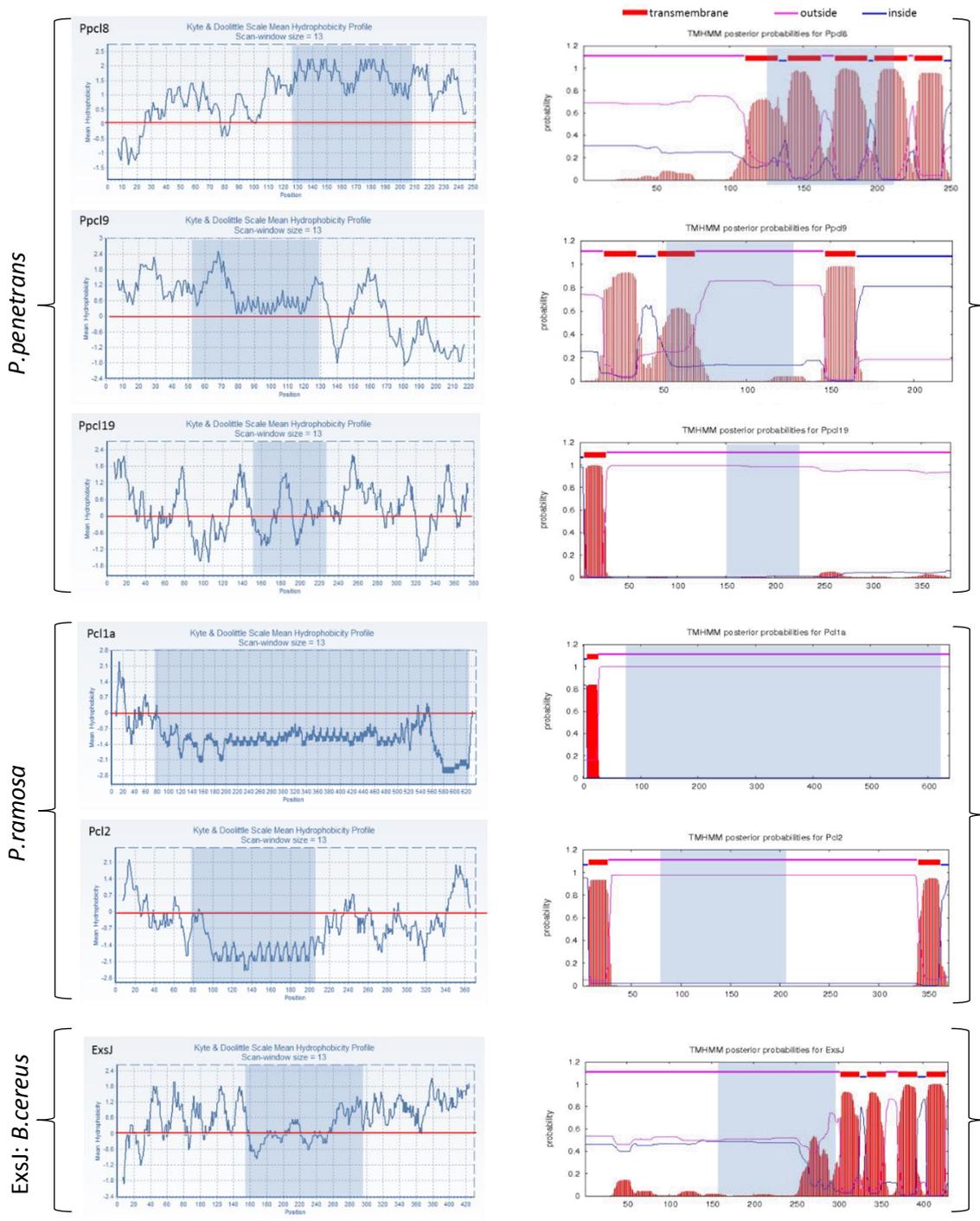
Figure 3. 5: Hydrophobicity and transmembrane domain predictions in selected CLPs from *P. penetrans*, *P. ramosa* and *B. cereus*. On the left are the Kyte and Doolittle hydrophobicity profiles; on the right are the TMHMM predictions. The G-X-Y repeat regions within each sequence are shaded in blue.

59
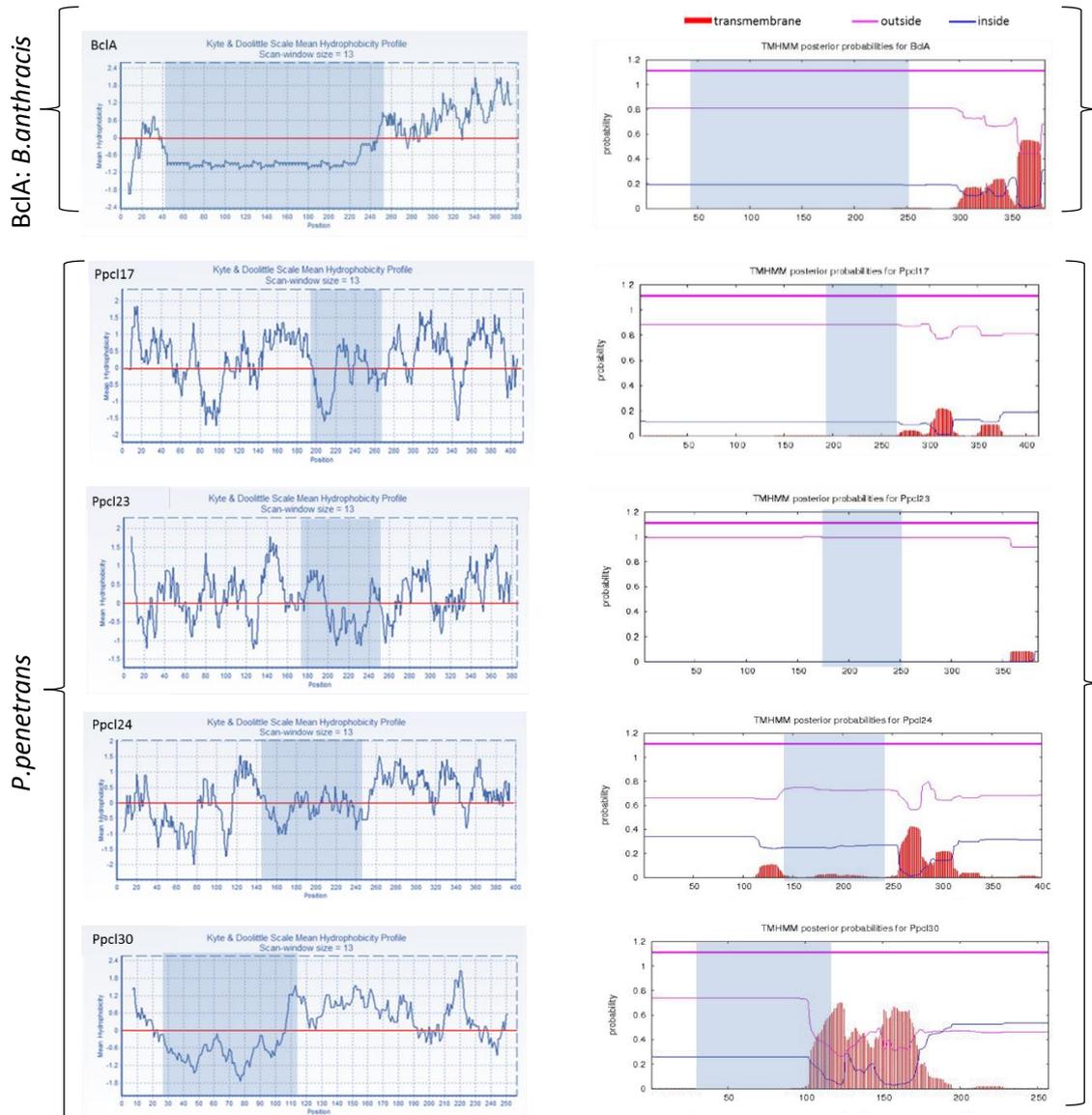
Figure 3. 6: Kyte and Doolittle Hydrophobicity (on the left) profiles for BclA in contrast with Ppcl sequences showing high hydrophobicity at CTD but no transmembrane domains (TMHMM predictions on the left). The GXY repeat regions within each sequence are shaded in blue.
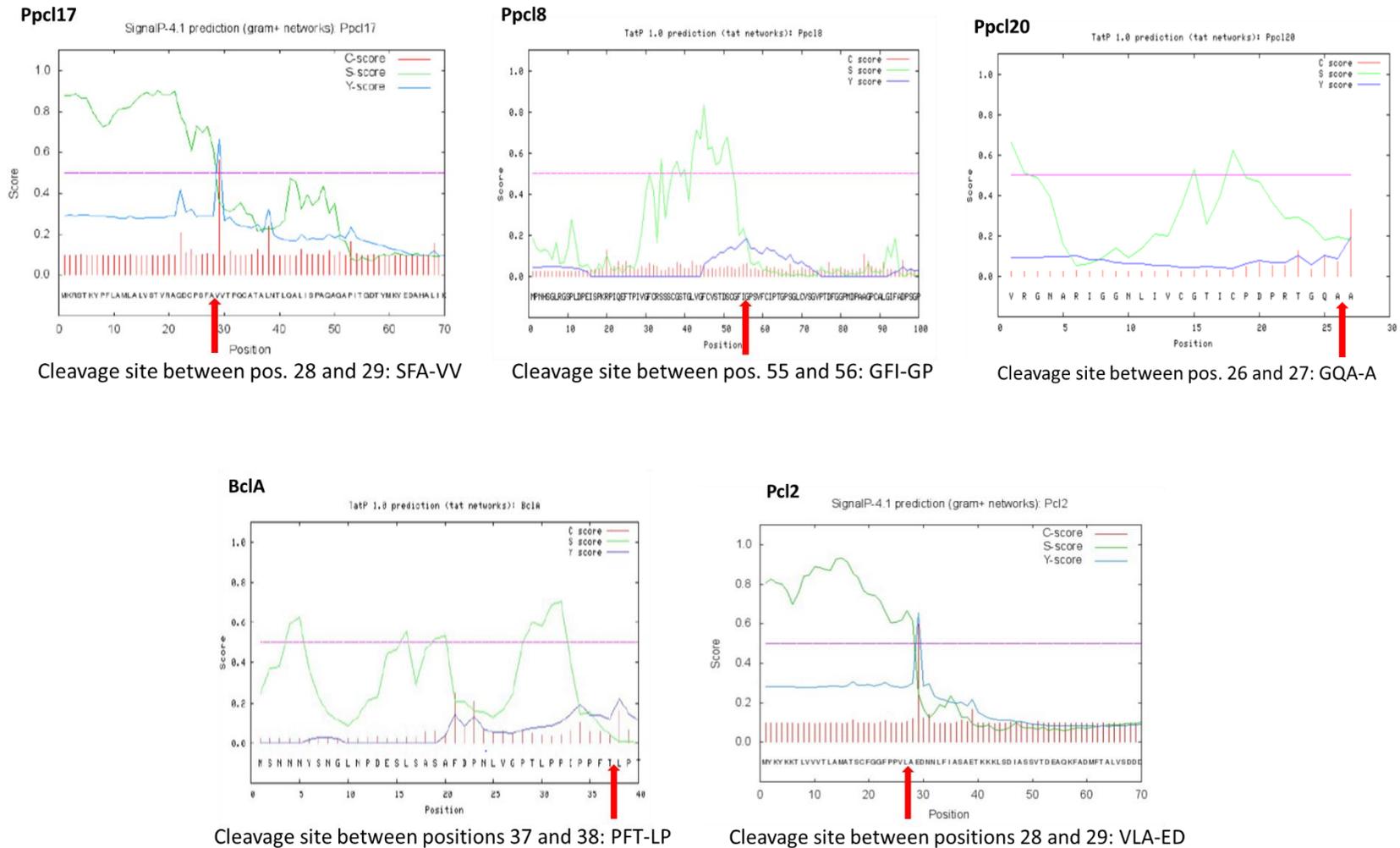
Figure 3. 7: Predicted signal peptide cleavage sites in the N-terminal domain of selected Ppcl sequences as compared with the BclA protein of *Bacillus* spp. and Pcl2 of *P. ramosa.* The predictions were done on SignalP and TatP servers. The cleavage sites are shown as red arrows.

Table 3.4: Predicted O-glycosylation sites in Ppcl sequences as compared with BclA and ExsJ of *Bacillus* spp. and Pcl1a and Pcl2 of *P. ramosa*. The predictions were done on NetOGlyc 4.0 server. The number of predicted sites are shown based on their location within the collagen sequence.

	Sequence	Number of predicted O-glycosylation sites			
		Total	NTD	G-X-Y	CTD
P. penetrans	Ppcl1	24	24	0	0
	Ppcl8	22	21	0	1
	Ppcl9	14	10	3	1
	Ppcl16	2	0	2	0
	Ppcl17	36	14	14	8
	Ppcl18	53	10	21	22
	Ppcl19	41	8	11	22
	Ppcl20	45	2	21	22
	Ppcl21	114	6	100	8
	Ppcl23	39	13	19	7
	Ppcl24	63	16	26	21
	Ppcl25	46	13	16	17
	Ppcl26	60	25	6	29
	Ppcl28	51	2	28	21
	Ppcl29	64	2	26	36
	Ppcl30	37	4	23	10
	Ppcl33	23	2	9	12
P. ramosa	Pcl1a	9	4	5	0
	Pcl2	30	9	0	21
Bacillus spp.	BclA	83	6	75	2
	ExsJ	57	6	33	18
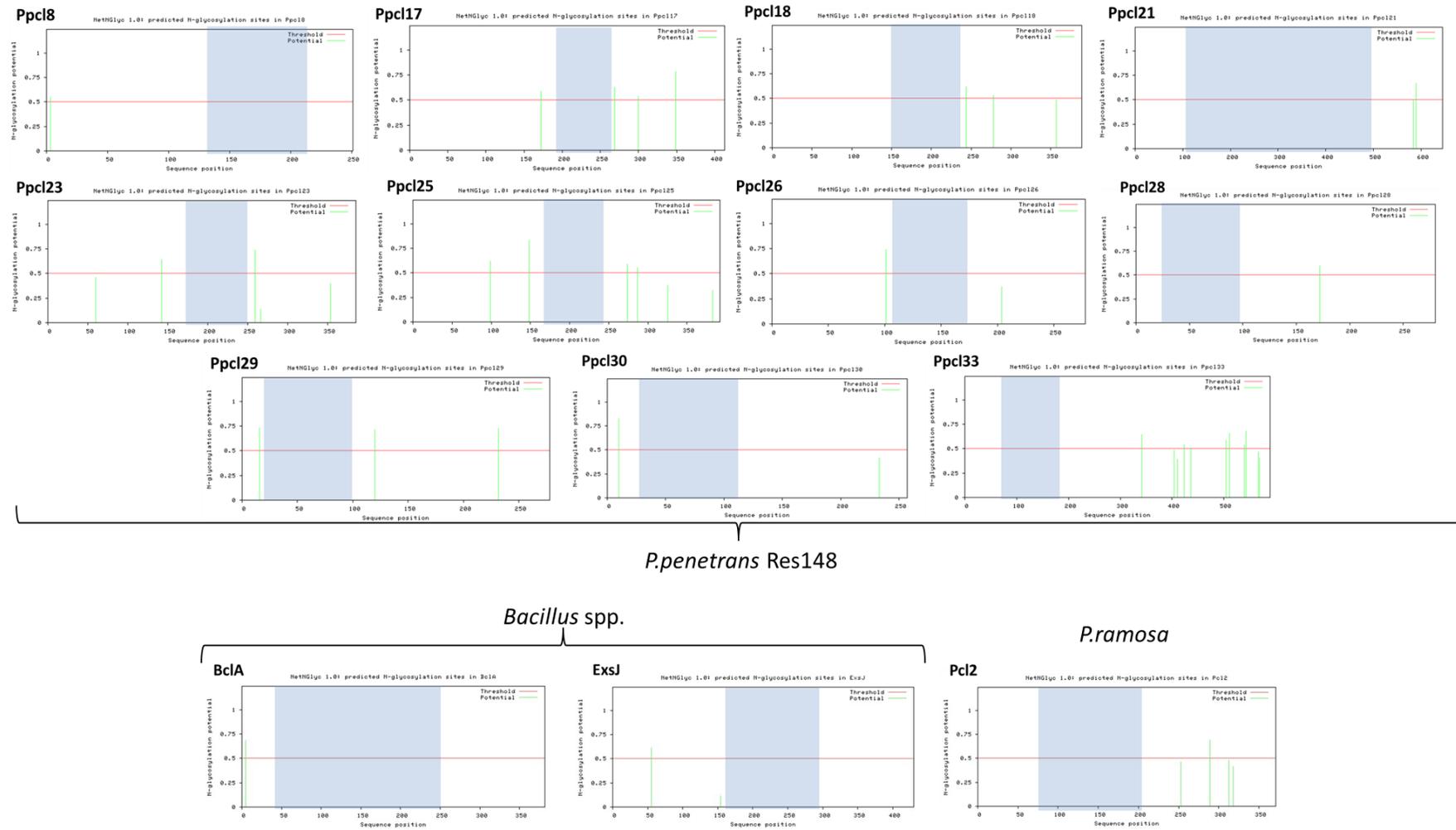
Figure 3. 8: Predicted N-glycosylation sites in Ppcl sequences as compared with BclA and ExsJ of *Bacillus* spp. and Pcl2 of *P. ramosa*. The predictions were done on NetNGlyc 1.0 server. The GXY repeat regions within each sequence are shaded in blue.

3.3.4 Sequences similar to Ppcls

Putative CL sequences from *P. penetrans* (Ppcls, *Table 3.2*) were used to search (BLASTp) non-redundant protein databases; significant matches were found to Ppcl23, Ppcl25, Ppcl26 and Ppcl33. Ppcl23 hit CL sequences from *Bacillus* spp. (98 hits) *Clostridium arbusti* (1 hit) and *Sediminibacillus albus* (1 hit); all hits were in the E-value range of 0.008 to 3e-08. Ppcl25 had 108 BLASTp hits (E-value ≤1e-07) from *Bacillus* spp. (96 hits), *Paenibacillus* spp. (7 hits), *Clostridium* spp. (4 hits), *Pithovirus sibericum* (1 hit). All the hits to Ppcl23 and Ppcl25 were in the CTDs of the sequences (query cover: ≤43% for Ppcl23; ≤49% for Ppcl25). Ppcl26 hit a CL proteins from a range of bacterial genera including *Clostridium* spp., *Ruminococcus torques*, *Fusicatenibacter* spp., *Eisenbergiella* spp., *Desulfotomaculum* spp., *Lachnospiraceae* spp., *Eubacterium dolichum*, *Blautia producta*, *Bifidobacterium* spp., *Lactonifactor longoviformis*, *Epulopiscium* sp., *Bacteroides gallinarum*, *Parabacteroides* spp., *Methanobrevibacter* spp., *Lactobacillus* spp., *Flavonifractor plautii, Leuconostoc* spp., *Veillonella dispar*, *Prevotella* sp. tc2-28. The topmost hits were for proteins from *Clostridium* spp. that had ≥90% query cover and very low E-values ≤1e-62. Ppcl33 showed sequence similarity with Megaviruses and *Ruminococcus* spp. (E-value: 0 to 2e-128; as high as 99% identity for 77% query cover). See *Appendix III* for the list of all CLP sequences that were selected for further analyses.

3.3.5 Identification of conserved motifs in sequences similar to Ppcl

The conserved motifs in the CTD and NTD of selected CL sequences are shown in *Figure 3.9*. At least two unique motifs in Ppcl23 (23_CTD_motif1, 23_CTD_motif2), two other motifs in Ppcl25 (25_CTD_motif1, 25_CTD_motif2), and four motifs in Ppcl26 (26_NTD_motif1, 26_NTD_motif1, 26_CTD_motif1, 26_CTD_motif2) were identified. Up to 77% of Ppcl33 sequence showed identities with CLPs from *Rumincoccus* spp. and Megaviruses. *Figure 3.10* shows graphical views of the multiple sequence alignments and the location of the conserved motifs. *Table 3.5* displays the amino acid sequence of the motifs and their percentage sequence similarity (BLOSUM45) to the corresponding Ppcl sequences (see *Appendix IV* for pairwise sequence similarity BLOSUM45 matrix generated in Geneious).
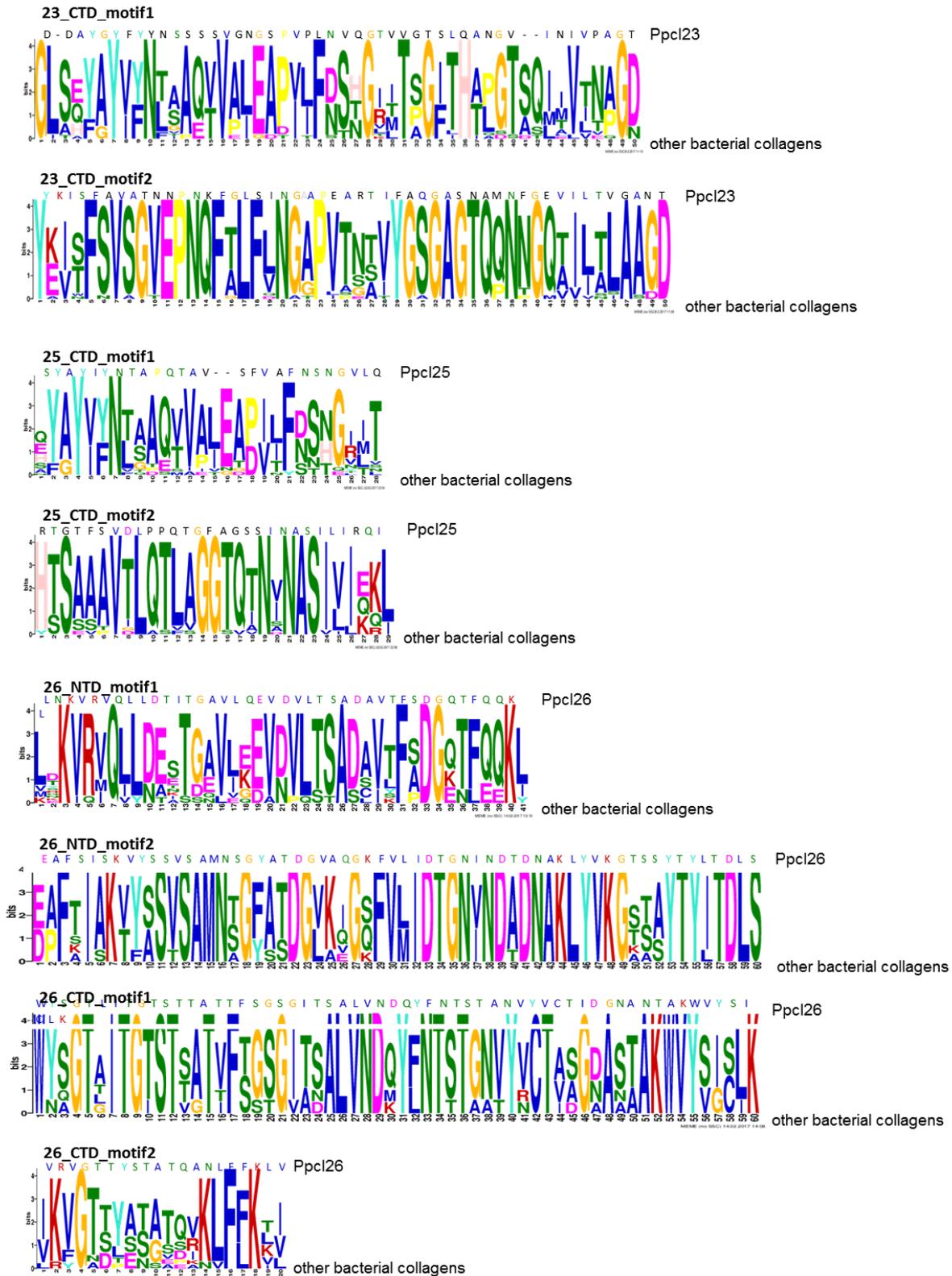
Figure 3. 9: Conserved motifs at the N-terminal and C-terminal domains of selected Ppcl sequences and other bacterial CLPs. The motifs were predicted by MEME and the motif graph so produced was edited to highlight the Ppcl sequences. The stack of letters show the amino acids at each position of the motif in different species. The colour of the letters represents distinct amino acid properties (MEME colour scheme). The smaller letters on the top are for the motif amino acids in Ppcl sequences. The horizontal axes represent the amino acid position in corresponding motif. The vertical axis shows the MEME bit score.

(a) Ppcl23

Figure 3. 10: Graphical representation of the multiple sequence alignments of a selection of sequences similar to (a) Ppcl23, (b) Ppcl25, (c) Ppcl26 and (d) Ppcl33 and the location of the conserved motifs with respect to the GXY repeat region. The respective Ppcl sequence were used as reference for the colour-coding of the alignments. Amino acid residues same as in the Ppcl sequence have been shown in colours similar to Ppcl residues while differences are shown in grey. The conserved motifs are shown as black arrowheads. No motifs have been marked up in the Ppcl33 alignment as the whole length of sequence was found to be considerably conserved. The E-values derived from BLASTp results are shown on the right of each alignment. See Appendix III for organism codes.
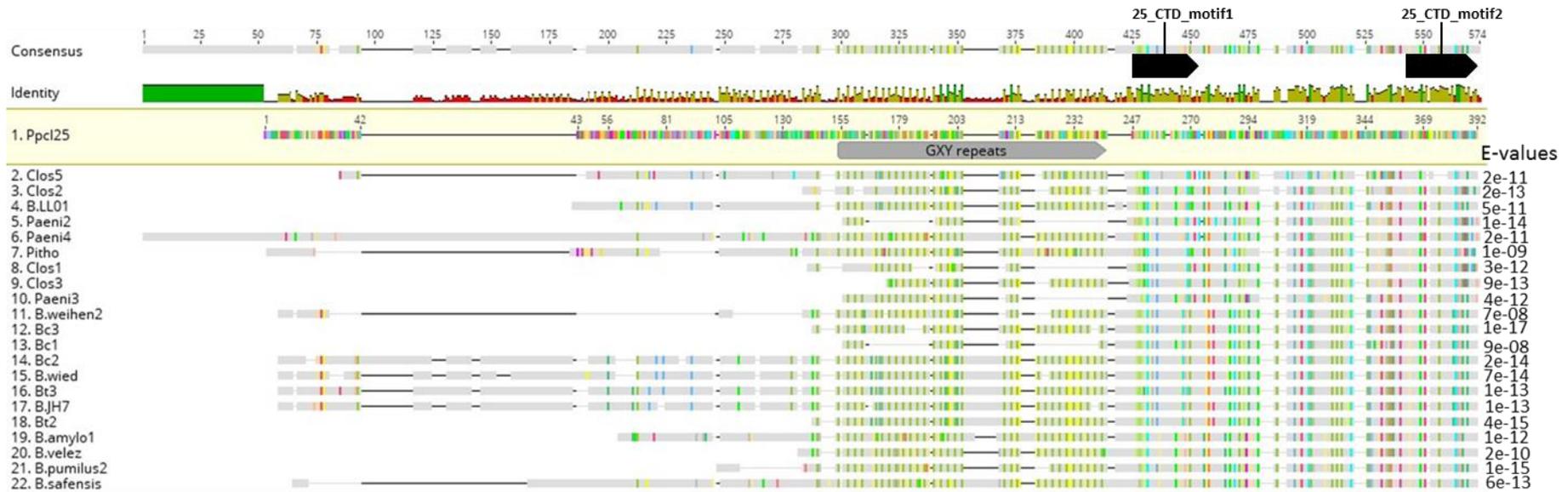
(b) Ppcl25

Figure 3. 10:*(continued)*

(c) Ppcl26
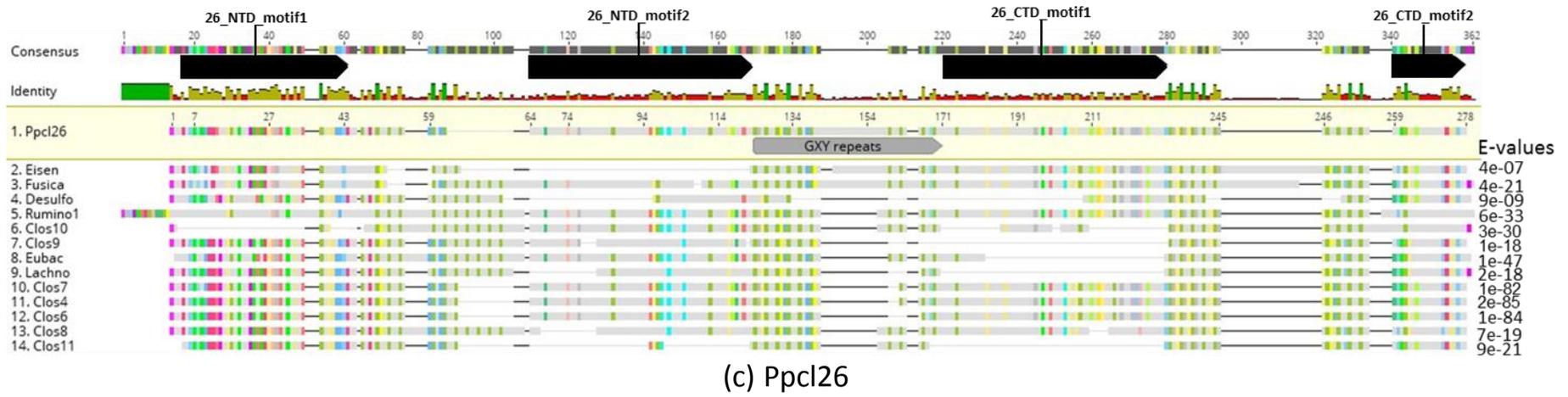
Figure 3. 10:(continued)
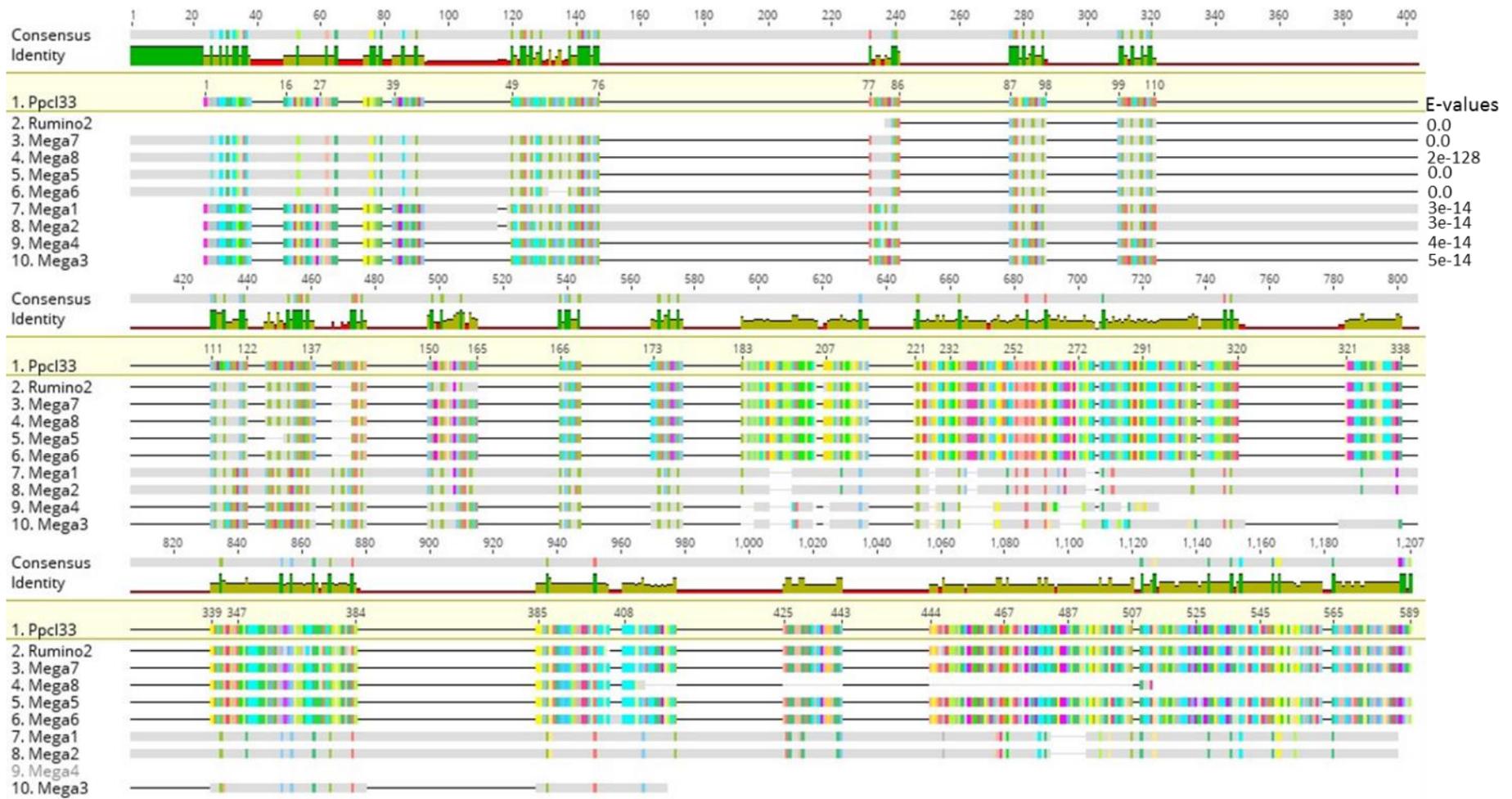
(d) Ppcl33

Figure 3. 10:(continued)

Table 3.5: Amino acid sequence of the identified conserved motifs and their percentage sequence similarity (BLOSUM45) to the corresponding Ppcl sequences. See Appendix III for organism codes.

Ppcl23: CTD_motif1		Percentage similarity to Ppcl23
1. Ppcl23	D-DAYGYFYYNSSSSVGNGSPVPLNVQGTVVGTSLQANGV--INIVPAGT	
2. Bc1	GLSQYAYVFNTAAQVVALEAPVLFNSHGTITSGFTHTLGTSQMTVINAGD	68.00%
3. B.wied	GLSQYAYVFNTAAQVVALEAPILFNSHGRITSGFTHTLGTSQMTVINAGD	66.00%
4. Bc2	GLSQYAYVFNTAAQVVALEAPILFNSHGRITSGFTHTLGTSQMTVINAGD	66.00%
5. B.LL01	GLAQYGYIYNFTPQTVEIGDPVLFDSNGVITSGILHIASTSTIEVVEQGD	64.58%
6. Bt4	GLSHYAYVFNTAAQVVALEAPILFNSHGRMTSGFTHTLGTSQLMVLNAGD	64.00%
7. Bt2	GLSHYAYVFNTAAQVVALEAPILFNSHGRMTSGFTHTLGTSQLMVLNAGD	64.00%
8. Bc5	GLSHYAYVFNTAAQVVALEAPILFNSHGRMTSGFTHTLGTSQLMVLNAGD	64.00%
9. Bc6	GLSHYAYVFNTAAQVVALEAPILFNSHGRMTSGFTHTLGTSQLMVLNAGD	63.83%
10. B.weihen1	GLSQYAYVFNTVAQVVALEAPILFNSHGKITSGFTHTLGTSQMTVINAGD	63.83%
11. B.safensis	GLSEFAYVYNLSAETVALEAPVIFDSTGIITSGITHAPGTSQIIVTNPGN	62.50%
12. Bt3	GLSQYAYIFNTAAQVVALEAPILFNSNGLMTPGFTHTPGTSQIMVINAGD	62.00%
13. Clos1	GLSAYAYIYNLAAQVVPLEADIIFSTNGILTAGITHAQDTASIALVNAGD	61.22%
14. B.acidicola	GLSQYAYIYNLIAQTVPIEATVTFDSNGVITPGIIHAPGTDSILVTVAGD	60.42%
15. B.pumilus2	GLAEFAYVYNLSAQTVALEAPVIFDSTGITTPGITHAPGTSQIIVTNPGN	59.57%
16. B.velez	GLSEYAYVYNLSAETVAIEAPVIFDSTGITTAGITHAPGTAQIIVTTPGD	59.57%
17. B.amylo2	GLSEYAYVYNLSAETVAIEAPVIFDSTGITTAGITHAPGTAQIIVTTPGD	59.57%
18. B.amylo1	GLSEYAYVYNLSAETVAIEAPVIFDSTGITTAGITHAPGTAQIIVTTPGD	59.57%
19. B.pumilus1	GLAEFAYVYNLSAQTVALEAPVIFDSTGITTPGITHAPGTSQIIVTNPGN	57.14%
20. Sedimini	GLSEFGYIYNFGAQVVPLEAPILFDSNGILTPGITHAPGTSQIIVTNPGD	56.25%
21. Paeni1	GITEFGYIYNLGAQVVPIEADVIFDTNGLLTPGITHAPGSAQIAVTNAGD	54.17%

Ppcl23: CTD_motif2		Percentage similarity to Ppcl23
1. Ppcl23	YKISFAVATNNPNKFGLSINGAAPEARTIFAQGASNAMNFGEVILTVGANT	
2. B.wied	YKISFSVSGVEPNQFALFLNG-APVTNSIYGSGAGTQQNNGQTILTLAAGD	84.31%
3. Bc2	YKISFSVSGVEPNQFALFLNG-APVTNSIYGSGAGTQQNNGQTILTLAAGD	84.31%
4. Bt3	YKISFSVSGVEPNQFTLFLNG-APVTNAVYGSGAGTQQNNGQTILSLAAGD	82.35%
5. Bc1	YKISFSVSGVEPNQFALFLNG-APVTNSVYGSGAGTQQNNGQTVLTLAADD	84.31%
6. Bt4	YKISFSVSGVEPNQFTLFLNG-APVTSAVYGSGAGTQQNNGQTILALAAGD	80.39%
7. B.weihen1	YKISFSVSGVEPNQFALFLNG-APVTNSVYGSGAGTQQNNGQTVLNLAAGD	84.31%
8. Bc5	YKISFSVSGVEPNQFTLFLNG-APVTSAVYGSGAGTQPNNGQTILTLAAGD	78.43%
9. Bc6	YKISFSVSGVEPNQFTLFLNG-APVTSAVYGSGAGTQPNNGQTILTLAAGD	78.43%
10. Bt2	YKISFSVSGVEPNQFTLFLNG-APVTSAVYGSGAGTQPNNGQTILALAAGD	78.43%
11. B.pumilus2	YEVTFSVSGVEPNQFTLFVNG-APVTNTVYGSGAGTQQNNGQAILALAAGD	82.35%
12. B.pumilus1	YEVTFSVSGVEPNQFTLFVNG-APVTNTVYGSGAGTQQNNGQAILALAAGD	82.35%
13. B.safensis	YEVTFSVSGVEPNQFTLFVNG-GPVTNTVYGSGAGTQQNNGQAILALAAGD	82.35%
14. B.amylo2	YEVTFSVSGVEPNQFTLFVNG-GPVTNTVYGSGAGTQQNNGQAIITLAAGD	82.35%
15. B.amylo1	YEVTFSVSGVEPNQFTLFVNG-GPVTNTVYGSGAGTQQNNGQAIITLAAGD	82.35%
16. B.velez	YEVTFSVSGVEPNQFTLFVNG-GPVTNTVYGSGAGTQQNNGQAIITLAAGD	82.35%
17. Sedimini	YEVTFSVSGVEPNQFTLFLNG-TPVTNTVYGSGAGTQQNNGQAIIAIAAGD	82.35%
18. B.LL01	YEVTFSVSGTEPNQFALFLNA-APLPETVYGSGAGTQQNNGQVILTLAAGD	84.31%
19. B.acidicola	YEITFSVSGVEPNQFALFDNG-APIAGTVYGSGAGTQQNTGQVIVTLAAGD	80.39%
20. Paeni1	YEINFSVSGVEPNQFGLFLNG-APVAGTIYGSGAGTQQNTGQAIIALASGD	82.35%
21. Clos1	YAIWFNVAGVEPNQFTLFQNG-APVSGATYGSGAGTQPNPGMVIITSAAGD	64.71%

70

Table 3.5 (continued)

Ppcl25_CTD_motif1		Percentage similarity to Ppcl25
1. Ppcl25	SYAYIYNTAPQTAV—SFVAFNSNGVLQ	
2. Clos1	AYAYIYNLAAQVVPLEADIIFSTNGILT	64.29%
3. Paeni2	SYAYVYNTSAQTVAIEADVTFDSNQNLT	64.29%
4. Clos3	NYAYIYNLAAQVVAIEADINFSNNGIIV	64.29%
5. Bt3	QYAYIFNTAAQVVALEAPILFNSNGLMT	64.29%
6. B.safensis	EFAYVYNLSAETVALEAPVIFDSTGIIT	60.71%
7. Paeni4	SYGYIFNTSTQSVATETDITFDSNENLT	60.71%
8. Clos2	EYAYIYNVGAQTIPLETDVIFSNNGVIS	60.71%
9. Pitho	SYAYIYNLGAQVVPLEADITYDTNGVIL	60.71%
10. Bc1	QYAYVFNTAAQVVALEAPVLFNSHGTIT	60.71%
11. Bt2	HYAYVFNTAAQVVALEAPILFNSHGRMT	60.71%
12. B.JH7	HYAYVFNTAAQVVALEAPILFNSHGRMT	60.71%
13. Bc3	HYAYVFNTAAQVVALEAPILFNSHGRMT	60.71%
14. B.weihen2	HYAYVFNTAAQVVALEAPILFNSHGRMT	60.71%
15. B.wied	QYAYVFNTAAQVVALEAPILFNSHGRIT	60.71%
16. Bc2	QYAYVFNTAAQVVALEAPILFNSHGRIT	60.71%
17. B.pumilus2	EFAYVYNLSAQTVALEAPVIFDSTGITT	57.14%
18. B.velez	EYAYVYNLSAETVAIEAPVIFDSTGITT	57.14%
19. B.amylo1	EYAYVYNLSAETVAIEAPVIFDSTGITT	57.14%
20. B.LL01	QYGYIYNFTPQTVEIGDPVLFDSNGVIT	57.14%
21. Clos5	AYGYIYNLQDSMAIVNGDILFSNNGYLN	55.56%
22. Paeni3	QFGYVYNLGAQVVPIEADVTFDSNGILT	53.57%

Ppcl25_CTD_motif2		Percentage similarity to Ppcl25
1. Ppcl25	RTGTFSVDLPPQTGFAGSSINASILIRQI	
2. B.LL01	YSSSSAVILASVVGGTQANVNASIVIKKL	61.29%
3. Paeni4	HSSASTVDLQTLAGGSQINANASILIQQL	58.06%
4. B.velez	HTSAAAVTLQTLAGGTQTNVNASIVLKKL	58.06%
5. B.amylo1	HTSAAAVTLQTLAGGTQTNVNASIVLKKL	58.06%
6. Pitho	HTSSAAVTLQTLAGGTVANSNASVLIQQI	56.67%
7. B.pumilus2	HTSAAAVTLQTLAGGTQTNVNASIVIKKL	54.84%
8. Bt2	HTSAAAVTLQTLAGGTQTNINASIVIEKL	54.84%
9. B.JH7	HTSAAAVTLQTLAGGTQTNINASIVIEKL	54.84%
10. Bt3	HTSAAAVTLQTLAGGTQTNINASIVIEKL	54.84%
11. Bc2	HTSAAAVTLQTLAGGTQTNINASIVIEKL	54.84%
12. Bc1	HTSAAAVTLQTLAGGTQTNINASIVIEKL	54.84%
13. Bc3	HTSAAAVTLQTLAGGTQTNINASIVIEKL	54.84%
14. B.safensis	HTSAAAVTLQTLAGGTQTNVNASIIIKKL	54.84%
15. Paeni3	HSSSAAVTLQTLAGGTQTNVNASIVIKRL	54.84%
16. Clos3	HSSAAAVTLQTLSGGTQINSNASILIQKL	51.61%
17. Clos1	HTSAAAVTLQTLVGGTQINDNASILIQQI	51.61%
18. B.wied	HTSAAAVTLQTLAGGTQTNVNASIVIEKL	51.61%
19. B.weihen2	HTSAAAVTLQTLAGGTQTNINASIVIEKL	51.61%
20. Paeni2	HSSASAVGLQTLAGGTQINANASILIQQL	51.61%
21. Clos2	HTSEVPVILQTFAGGTQVNVNASILIQRI	48.39%

Table 3.5 (continued)

Ppcl26: NTD_motif1		**Percentage similarity to Ppcl26**
1. Ppcl26	LNKVRVQLLDTITGAVLQEVDVLTSADAVTFSDGQTFQQKL	
2. Clos4	LDKVRVQLLDESTGAVLKEVNVLTSADAVTFADGQTFQQKL	95.12%
3. Clos7	LDKVRIQLLDESTGAVLKEVNVLTSADAVTFADGQTFQQKL	95.12%
4. Clos6	LDKVRVQLLDESTGAVLKEVNVLTSADAVTFADGQTFQQKL	95.12%
5. Clos8	LSKVRVQLLDEATGNVIEEVDVLTSADAVKFADGQTFQQKL	95.12%
6. Lachno	LKKVRVQLLDENTSAVIEEVDVLTSADAVTFSDGETFQQKL	95.12%
7. Clos11	MEKVRVQLLDEKTGEVVKEVDVLTSADAVTFSDGQTFQQKL	92.68%
8. Clos9	LRKVRVQLLNESTGAVEEEVNVLTSADCVTFADGETFEQKL	90.24%
9. Eisen	LTKVRMQILDAETDAVLEEVDVLSSASSILFPDGKNLEEKI	90.24%
10. Fusica	LTKIRMQILNAETDEVLEEVDVLSSASSILFPDGKNLEEKI	87.80%
11. Eubac	KQKVQIQLLNESSGEVIGDVDPLTSADCVSFSDGKTFQQKY	82.93%
12. Desulfo	VYKVRVQVYDETTGQLLGDADVQTTADLVYFSDGETFQQKL	80.49%

Ppcl26: NTD_motif2		**Percentage similarity to Ppcl26**
1. Ppcl26	EAFSISKVYSSVSAMNSGYATDGVAQGKFVLIDTGNINDTDNAKLYVKGTSSYTYLTDLS	
2. Rumino	DAFAIAKTFASVSAMNSGFSTDGVKEGQFVMIDTGNVNDADNAKLYVKGKSAYTYITDLS	96.67%
3. Clos10	EAFTIAKTYASTSAMNSGYASDGVKVGQFVMIDTGNVNDADNAKLYVKGTAAYTYITDLS	96.67%
4. Eubac	EAFKIAKTYASVSAMNAGFASDGVKQGQFVMIDTGNVNDADNAKLYVKGASSYTYITDLS	96.67%
5. Clos7	EPFTIAKVYSSVSAMNTGFATDGLKIGSFVLIDTGNVNDADNAKLYVKGSTAYTYITDLS	93.33%
6. Clos6	DPFTIAKVYSSVSAMNTGFATDGLKIGSFVLIDTGNINDADNAKLYVKGSTAYTYITDLS	93.33%
7. Clos4	DPFTIAKVYSSVSAMNTGFATDGLKIGSFVLIDTGNINDADNAKLYVKGSTAYTYITDLS	93.33%

Ppcl26: CTD_motif1		**Percentage similarity to Ppcl26**
1. Ppcl26	WYSGTLITGTSTTATTFSGSGITSALVNDQYFNTSTANVYVCTIDGNANTAKWVYSICLK	
2. Clos4	WYSGTAITGTSTSATVFTGSGITSALVNDQYFNTSTGNVYVCTASGDASTAKWVYSICLK	96.67%
3. Clos6	WYSGTTITGTSTSATVFTGSGIASALVNDQYFNTSTGNVYVCTASGDASTAKWVYSICLK	96.67%
4. Clos7	WYSGTAITGISTSATVFTGSGITSALVNDQYFNTSTGNVYVCTASGDASTAKWVYSISLK	93.33%
5. Fusica	WYQGTGITGTSTVGTVFSSSGVANALVNDKYLNTTTGAVYNCTVAGNASAAKWVYVGSIK	88.33%
6. Eisen	WYQGTTITGTSTTAAVFSNSGIAAALVNDKYLNTSTGAVYNCTVAGAASVAKWAYAGSIK	88.33%
7. Rumino	WNAGTAITGTSTTATIFSGTGITDALVNDMYLNTSTGNTYRCTVAGAAAAAKWVYVGSLK	83.33%

Ppcl26: CTD_motif2		**Percentage similarity to Ppcl26**
1. Ppcl26	VRVGTTYSTATQANLFFKLV	
2. Clos11	VKVGATYANSTQVKLFFKLV	100%
3. Clos8	VKFGTTLATAADVKVFFKKI	95.00%
4. Clos9	VKVGTSLSSATQKKLFFKIV	95.00%
5. Clos6	IKVGTDYASGTQVKLFLKTI	90.00%
6. Clos4	IKVGTDYASGTQVKLFLKTI	90.00%
7. Clos7	IKVGNDYASGTQVKLFLKTI	90.00%
8. Eubac	IKVGTTPESATSRKLFFKVI	90.00%
9. Desulfo	IKVGTSYETGESRKLFFKLL	90.00%
10. Fusica	IKFGTSYATASDIKLFFKKV	90.00%
11. Lachno	IKVGTTTSNAVPRKLFFKVI	85.00%
12. Eisen	LKYGTSYETASEIKLFLKKL	85.00%

3.3.6 Characterization of low-complex G-X-Y repeat region

The cluster analysis of the low complex G-X-Y repeat regions showed that after Glycine (G), Threonine (T) is the second most commonly occurring amino acid in these low complexity regions, followed by Proline (P). However other amino acids like Alanine (A) and Valine (V) also occur.

The hierarchical clustering dendrogram extracted from the heatmap (Appendix Ib) and edited in FigTree (**http://tree.bio.ed.ac.uk/software/figtree/**) is shown in *Figure 3.11*. The tree splits the sequences categorically into 7 clusters forming 2 superclades. Cluster 3 members form a superclade (Superclade 2) that is separated from a much bigger superclade (Superclade 1) from which all the remaining clusters branch out. The reason for the clustering of Cluster 3 members as a separate superclade becomes clear on observing the heatmap for the amino acid composition of its member sequences. Cluster 3 members are unique because they have Lysine (K) and Aspartic acid (D) as the most commonly occurring amino acid after Glycine.

The Ppcl sequences from *P. penetrans* are spread across four clusters. Ppcl1, Ppcl8, Ppcl9 and Ppcl26 cluster together (Cluster 2) with other collagen-like sequences from *Bacillus* sp., *Clostridium* spp., *P. ramosa*, *Paenibacillus* sp., *Protochlamydia* sp. and *Pithovirus.* Ppcl33 along with *Megavirus* sequences and a sequence from *Ruminococcus torques* constitute Cluster 3. Cluster 5 consists of the maximum number of Ppcl sequences (11 sequences) including Ppcl26 and Ppcl29 which are placed close to BclA. Ppcl16 groups with 5 *Bacillus* sequences, 1 *Clostridium* sequence and 1 *P. ramosa* sequence. Most of the *P. ramosa* sequences (25 out of 37) clustered together in Cluster 1 with one *Clostridium* sequence. Cluster 4 has *Bacillus* and *Paenibacillus* sequences. Cluster 6 consists of *Ruminococcus* sp., *Fusicatenibacter* sp., *Eisenbergiella tayi*, *Desulfotomaculum guttoideum.*, *Lachnospiraceae* sp., *Protochlamydia naegleriophila*, *Eubacterium dolichum* along with six *Clostridium* sequences.
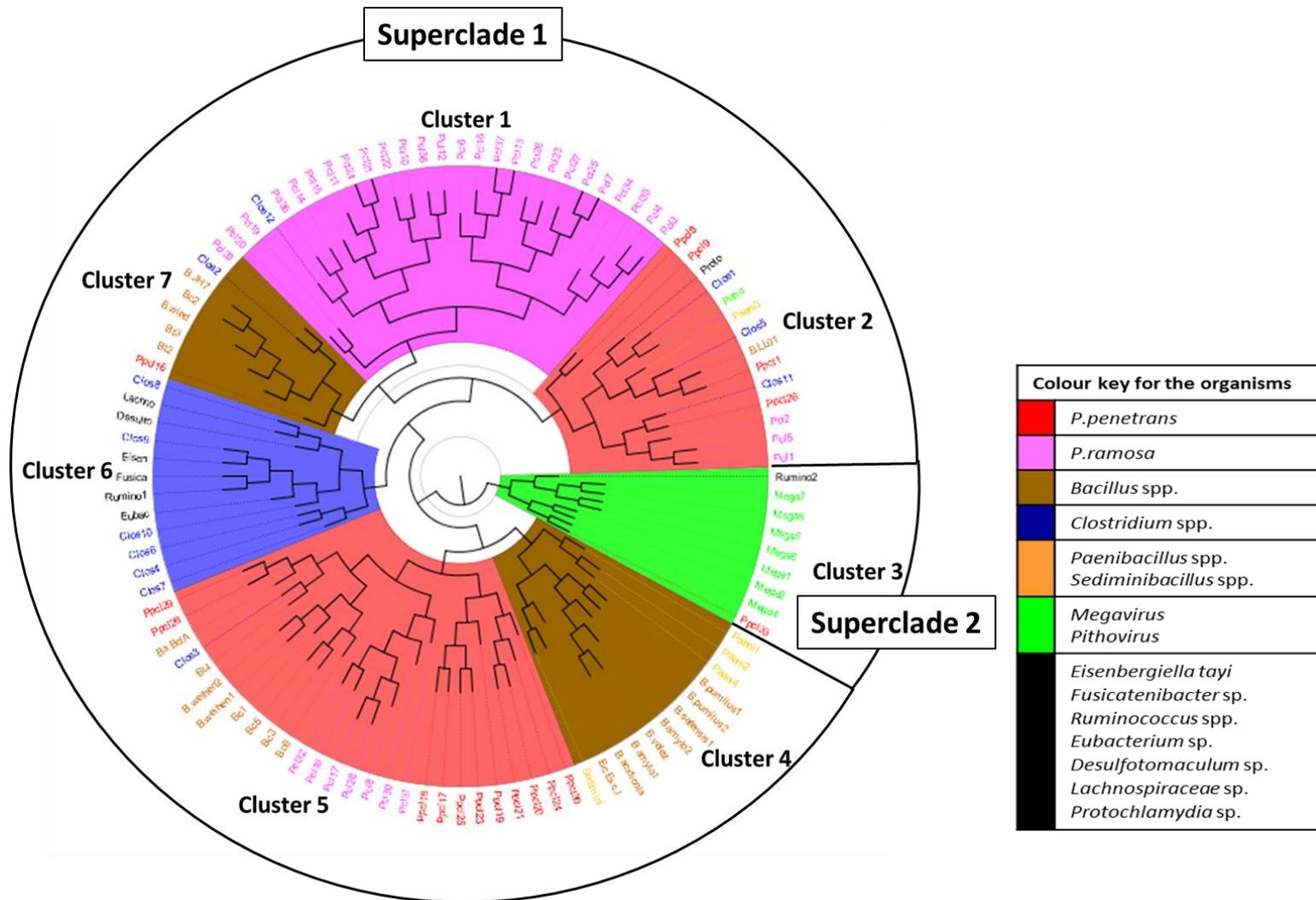
Figure 3.11: Hierarchical clustering dendrogram based on percentage amino acid composition of G-X-Y repeat regions of Ppcl sequences with those in other closely related collagen-like proteins. See Appendix III for list of CLP sequences used for this analysis

3.4   SUMMARY OF RESULTS

In silico studies suggest the presence of an array of genes coding for CLPs in *P. penetrans*. These putative proteins are predicted to be very varied in their biochemical characteristics like hydrophobicity, membrane spanning regions, glycosylation sites. The sequence comparisons and cluster analysis of putative *Pasteuria* CLP with those from other organisms suggest that all the CLPs are very diverse, are evolutionary linked and are possibly involved in adhesion.

3.5   DISCUSSION

This study was set out with the aim to identify and characterize putative CLPs in *P. penetrans.* Throughout this research project, it is hypothesized that the parasporal fibres of the *Pasteuria* exosporium are composed of some BclA-like proteins, that determine the binding properties of the *Pasteuria* endospores to the cuticle of their host nematodes. Addressing this hypothesis, the initial tBLASTn searches indicated the presence of *bclA*-like low complexity regions containing repetitive G-X-Y motifs in *P. penetrans* Res147 genome. The results of synteny analysis across different species of *Bacillus* spp. suggested considerable syntenic conservation around the *bclA* gene indicating the functional relationship of *bclA* with its neighbouring genes. BLASTp searches targeting the unpublished sequences of *Pasteuria penetrans* Res148, a single spore isolate derived from RES147, indicated the presence of putative orthologs of several genes syntenic to *bclA*. It is likely that the CL genes in *Pasteuria* follow a synteny pattern similar to that around the *bclA* gene in *Bacillus*. However, studying the synteny around the identified putative Ppcl genes in *Pasteuria* was beyond the scope of this project as the contigs available for the analyses were of short length.

Computational characterization of the Ppcl sequences in *P. penetrans* Res148 suggest their diverse biochemical characteristics. According to the TMHMM predictions, at least three Ppcl sequences (Ppcl8, Ppcl9, Ppcl19) were projected as membrane-spanning proteins. Amongst these the transmembrane domain of Ppcl8 was observed to be similar to the ExsJ protein of *B. cereus*. Both possessed a series of transmembrane domains nearer to the C-terminal of the proteins. However, four out of five transmembrane domains in Ppcl8 spanned across G-X-Y repeat region, while none of the four transmembrane regions in ExsJ occurred in the G-X-Y region. Four

Ppcl sequences (Ppcl17, Ppcl23, Ppcl24, Ppcl30) were unique in possessing hydrophobic C-terminal domains but no transmembrane regions. Similar traits (hydrophobic CTD, no TMM domain) were predicted for the BclA protein. Pcl1a of *P.ramosa* appeared functionally similar to Ppcl19 of *P. penetrans* as both were predicted to have one transmembrane domain each on their N-terminals. Ppcl8, Ppcl17, Ppcl20 were predicted to have SignalP/TatP cleavage sites at their NTD. Proteins bearing such sites are known to be exported into the extracellular environment. Similar cleavage sites were observed in BclA and *P. ramosa* Pcl2, the former of which is known to occur as a membrane bound protein. Thus, these three putative proteins, viz. Ppcl8, Ppcl17 and Ppcl20, are likely to be localized either on the outer surface of the vegetative cell membrane of *Pasteuria* or on the exosporial wall of their endospore. The presence of O-glycosylation on all the 17 putative Ppcl sequences and N-glycosylation sites on 11 of the sequences suggest that it is highly likely that the investigated CLPs in *Pasteuria* are all post-translationally modified by glycosylation.

When compared with the set of all proteins present in the NCBI non-redundant database, the amino acid sequences of Ppcl23 and Ppcl25 showed sequence similarities with CLPs in *Bacillus* spp. while Ppcl26 was more like CLPs in Clostridia. However, all three of these hit CLP sequences from other organisms as well. This suggests an evolutionary link between the CLPs from different organisms. Most surprisingly, one of the sequences, Ppcl33 showed up to 99% identity with CLPs in Megaviruses.

A heuristic approach was used to compare Ppcl sequences of *Pasteuria* with the CLPs showing sequence homology to them. Considering collagens as 'three-domain systems', the NTD, the G-X-Y repeat region and the CTD of the sequences were separated for this analysis. Sequence conservation motifs were identified at their NTDs and CTDs. Since low complexity regions in proteins are more prone to replication slippage and are thereby susceptible to rapid evolution, the low complex GXY repeat regions of selected CLPs were analyzed separately to get insights into the evolution of these proteins. Cluster analysis based on the percentage amino acid composition of the GXY repeat regions suggest that the CLPs in *P. penetrans* are extensively diverse as the Ppcl sequences were spread across four clusters with CLPs

of varied origin ranging across CLPs from *P. ramosa*, *Bacillus* spp., *Clostridium* spp., *Paenibacillus* spp., *Ruminococcus* spp., *Protochlamydia* spp. to *Pithovirus* and *Megavirus.*

The sequence similarity of Ppcl33 with CLPs from Megaviridae is an unanticipated, and interesting finding. The members of the Megaviridae family are giant viruses (0.7 µm) that are known to infect protozoans in aquatic ecosystems. The most fascinating fact about these viruses is their unusually big genome (≥1.2 Mbp); the genome size of conventional viruses ranges from 8 to 250 Kbp (Colson *et al.*, 2012, Legendre *et al.*, 2012, Arslan *et al.*, 2011). The megavirus genome is predicted to encode more than a thousand protein coding genes including metabolic genes not found in any other viruses. It has been postulated that Megaviruses acquired a large set of genes from bacteria during the course of their evolution (Filée *et al.*, 2007). The fact that *P. ramosa* is a bacterial endosymbiont of water fleas means that both *P. ramosa* and Megaviruses possibly share the same ecological niche. This would possibly make horizontal gene transfer easier between the two organisms. It is possible that the Ppcl33 gene in *P. penetrans* was acquired from Megaviruses via *P. ramosa* while the two *Pasteuria* species were evolving as parasites of their respective hosts. What is even more interesting is that Megaviruses are known to possess fine hair-like structures on their capsids which are thought to be involved in host-parasite interaction. One could easily compare these hair-like structures to the parasporal fibres of *Pasteuria* and BclA nap of *Bacillus.* However, the presence of Ppcl33 in the sequence contigs of *P. penetrans* does not guarantee that this protein is expressed in *Pasteuria.*

All the findings of computational analyses are open to various interpretations. To test the predictions, extensive experiments are needed. The results discussed in this chapter and the previous chapter formed the basis of the subsequent wet laboratory experiments, discussed in the next chapters, wherein endospore proteins of *P. penetrans* were biochemically characterized in comparison with endospore proteins from *B. thuringiensis.*

Chapter 4

Comparative Endospore Protein Characterization using Western Blotting

4.1 INTRODUCTION

Based on previous research findings, collagen emerges as a candidate exosporium protein of *Pasteuria* endospores involved in their attachment to the nematode cuticle. The previous two chapters of this thesis were based on *in-silico* based studies, the results of which suggested, firstly, that *Bacillus* spp. are suitable for comparative studies with *Pasteuria* spp. and, secondly, that the *P. penetrans* genome potentially contain a range of collagen-like proteins with diverse physicochemical properties and sub-cellular localization. Studying the protein biochemistry of *Pasteuria* endospores and their surface proteins could lead to the identification of the immunogenic factors responsible for their host attachment and specificity.

4.1.1 Bacterial endospore proteins and their roles

Proteins make an important component of the bacterial endospores surface coat and exosporium along with some lipids and carbohydrates. The multi-layered proteinaceous layers of the endospore coat essentially act as a shield surrounding the core, provide mechanical integrity and prevent the spore from extreme external environments (Driks, 2002, Henriques and Moran, 2000). Several of the endospore coat proteins serve as enzymes and regulate germination of the endospores (Foster and Johnstone, 1990, Setlow, 2003, Moir *et al.*, 2002, Paidhungat and Setlow, 2000). Some proteins within the core of the endospore like the Small Acid-Soluble Proteins (SASPs) are associated with condensing and stabilizing the DNA by tightly binding to it (Johnson and Tipper, 1981, Setlow, 1988, Fairhead *et al.*, 1993). Surface-associated proteins of bacterial endospores have been shown to be involved in adhesion to host surfaces (Sánchez *et al.*, 2009).

4.1.2 Previous studies on proteins in *Bacillus* endospores

The coat of *Bacillus* endospores is known to be composed of at least 30 unique protein species (Driks, 2002). These proteins are either structural proteins that are assembled together as tough proteinaceous concentric layers surrounding the core or

morphogenetic proteins like SpoIVA and CotE that control the spore coat assembly (Roels *et al.*, 1992, Little and Driks, 2001, Zheng *et al.*, 1988).

A study using 2-dimensional electrophoresis protein signatures of *B. anthracis*, *B. cereus* and *B. thuringiensis* revealed that the overall protein pattern for each species is different with multiple unique spots (DelVecchio *et al.*, 2006). The first spore surface protein was discovered in *B. anthracis* and was named as BclA (*Bacillus* collagen-like protein of anthracis) for being a collagen-like glycoprotein (Steichen *et al.*, 2003, Sylvestre *et al.*, 2003, Sylvestre *et al.*, 2002). Localized to the exosporial nap of the *Bacillus* endospores, this glycoprotein has a peptide backbone of ~39 kDa and a heavy glycosylation makes the intact protein to have an apparent mass of >250 kDa (Sylvestre *et al.*, 2002).

Two other glycoproteins with characteristic collagen-like (CL) repeats have been reported as ExsJ and ExsH in *B. thuringiensis* and *B. cereus* respectively (Garcia-Patrone and Tandecarz, 1995, Charlton *et al.*, 1999, Todd *et al.*, 2003). Both ExsJ and ExsH have been shown to migrate as 205-kDa species during electrophoresis (Todd *et al.*, 2003, Garcia-Patrone and Tandecarz, 1995). However, Todd *et al.* (2003) suggest the existence of the protein ExsJ in two forms: a 205-kDa multimer and a 70-kDa monomer.

A 58 kDa protein similar to GroEL, a molecule chaperone protein involved in protein folding in a number of bacteria, has been found to be present in large quantities in the exosporium of *Bacillus* spp. (Charlton *et al.*, 1999, Redmond *et al.*, 2004). Some other proteins found to be tightly associated with exosporia in *B. anthracis* include alanine racemase (43 kDa), inosine-preferring nucleoside hydrolase (33 kDa), ExsF/BxpB (17 kDa), Spore coat protein Z (14 kDa), a CotB homologue (14 kDa), ExsK (10 kDa) and many more (Redmond *et al.*, 2004). Several other exosporium proteins, e.g. ExsB (26.5 kDa), ExsC (30 kDa), ExsD (66 kDa), ExsE (34 kDa), ExsF (13 kDa and ExsG (5.4 kDa), have been identified and sequenced in *B. cereus* (Todd *et al.*, 2003).

4.1.3 Previous studies on proteins in *Pasteuria* endospores

The *Pasteuria* endospores are predominantly composed of proteins and carbohydrates. A study on the polypeptide components of the parasporal fibres of *P. penetrans* using polyacrylamide gel electrophoresis revealed the presence of at least 15 polypeptide components (Vaid *et al.*, 2002). Another such study involving different populations of *P. penetrans* showed different banding profiles with more than one unique band for each population (Davies and Redden, 1997, Davies *et al.*, 1992). All such studies suggest that a wide range of epitopes are present on the surface of *Pasteuria* endospores and are potentially responsible for the host preferences of these bacteria (Davies and Danks, 1992). Adhesins, such as N-acetylglucosamine, have been reported to be associated with the endospore surface (Persidis *et al.*, 1991, Davies and Danks, 1993) which are thought to interact with the cuticle surface by electrostatic (Afolabi *et al.*, 1995) or hydrophobic forces (Davies *et al.*, 1994).

Some earlier studies linked the endospore heterogeneity of *Pasteuria* to the parasporal fibres associated with the exosporium of their endospores (Davies *et al.*, 1994, Mohan *et al.*, 2001). The presence of CLPs appearing to be closely related to the BclA protein of *Bacillus* spp., have been indicated in the genome of *P. penetrans* Res147 (Davies and Opperman, 2006). Our results of computational analyses (Chapter 2 of this thesis) support the presence of CLPs in another isolate, *P. penetrans* Res148. It has been proposed that such BclA-like CLPs constitute the parasporal fibres of the exosporium and interact with mucin-like peptides present on the nematode cuticle via a velcro-like attachment mechanism (Davies, 2009). In the cladoceran parasitic species, *P. ramosa*, CLPs have previously been identified and characterized (Mouton *et al.*, 2009, McElroy *et al.*, 2011). An immunological comparative study between *P. penetrans* and *P. ramosa* demonstrated that the two species share a "universally conserved epitope", although the nematode parasite possess "a more diverse molecular array of antigenically reactive peptides" (Schmidt *et al.*, 2008).

4.1.4 Aims and Objectives

Major Aim: To explore and characterize the protein profiles of the endospores of *P. penetrans* and closely related *Bacillus thuringiensis* and search for proteins that are putatively conserved between the two bacteria.

Specific Objectives:

1) To study the endospore protein composition of *P. penetrans* and *B. thuringiensis* using standard SDS-PAGE technique.
2) To identify immunodominant proteins with shared epitopes in *P. penetrans* and *B. thuringiensis* endospores using *Pasteuria*-specific polyclonal antibody for immunodetection.
3) To identify immunodominant collagen-like proteins with shared epitopes in *P. penetrans* and *B. thuringiensis* endospores using collagen specific antibodies raised to synthetic collagen-like peptides from *P. penetrans* contigs.
4) To validate the presence of collagens and glycoproteins by means of collagenase digestion, glycoprotein staining and lectin blotting.

4.2 MATERIALS AND METHODS

4.2.1 *Pasteuria* endospores

The *Pasteuria penetrans* endospores (isolate Res148) used for characterization experiments were produced *in vivo* on *Meloidogyne incognita* as per standard technique (Stirling and Wachtel, 1980) and were kindly provided by Dr. K. G. Davies. The spores were counted using a haemocytometer and the count was adjusted to 10^8 spores/ml in PBS. The stock endospore suspension was stored at -20°C in aliquots.

4.2.2 *Bacillus thuringiensis* cultures

Three *Bacillus thuringiensis* strains, viz. *B. thuringiensis* str. Al Hakam, *B. thuringiensis* Berliner (ATCC 10792) and a cry- mutant of *B. thuringiensis* subspecies *kurstaki* were kindly provided by Dr A Bishop. The strains were maintained at UH and preserved in 40% glycerol and stored at -80°C in cryovials. The cultures were sub-cultured routinely on nutrient agar (Sigma-Aldrich, N4019) in 9 cm disposable petri dishes and incubating them at 30°C for 24-36 hours. Liquid cultures were made in Luria-Bertani Broth (Sigma-Aldrich, L3522) or Nutrient broth (Sigma-Aldrich, N7519) and incubating the cultures on a shaking incubator (120 rpm) at 30°C. (See *Appendix V* for media composition)

4.2.3 *Bacillus* endospore purification

For maximum sporulation *Bacillus* strains were cultured in Sporulation Broth (*Appendix V*) in Nunc™ 6-well, 3 ml cell culture plates (Thermo Fisher Scientific) incubated at 37°C at 120 rpm. Sporulation was monitored by endospore Schaeffer-Fulton staining (see section 4.2.4) until most of the cells had sporulated in 6-8 days. The broth culture was then centrifuged at 6000 rpm for 10 minutes. The pellet was washed twice with ice-cold PBS to remove the chemical media and finally re-suspended in 1 ml of sterile distilled water. The suspension was heated to 100°C to kill any of the vegetative cells, if present. The heat-treated suspension was centrifuged at 6000 rpm for 5 minutes to remove any vegetative debris as supernatant. The purity of the endospores in the pellet, thus obtained, was again determined using the standard Schaeffer-Fulton endospore staining technique as described below. The purified endospores were finally re-suspended in sterile distilled water. Endospores were counted using a haemocytometer and the count was adjusted to 10^8 spores/ml. The endospore stock was stored at -20°C in aliquots.

4.2.4   Schaeffer-Fulton staining to estimate *Bacillus* endospore purity

Bacillus endospores were stained using malachite green and safranin by the differential endospore staining technique as described by (Schaeffer and Fulton, 1933). This technique exploits the fact that the endospore wall is thick and resistant and does not allow a stain to permeate easily. Malachite green is used as a primary stain to stain endospores, it enters the spore wall while the endospores are being gently heated on a slide. Once the heating is stopped the green coloured stain fixes is locked inside the spore wall integuments. At this stage, malachite green also enters the vegetative cells but is easily washed away with water. Counterstaining with safranin stains vegetative cells in red. For the staining, *Bacillus* cultures were smeared on clean glass slides, heat fixed and the slides were kept over a steaming water bath on a stand. The smears were drenched in malachite green while there was continuous steaming from underneath. This was done for 10 minutes and the stain was not allowed to dry. The green stained smear was washed with tap water and counter-stained with safranin for a minute. The slides were washed thoroughly with tap water till the running off water became clear. The slides were observed under 100X objective of a light microscope (Leitz laborlux, Leica Mikroskope und Systeme GmbH, Wetzlar, Germany) using immersion oil.

4.2.5   Protein extraction from endospores

The *Pasteuria* and *Bacillus* endospore suspension (10^8 spores/ml) stored at -20°C were thawed at room temperature and centrifuged at 9000g for 2 minutes. The pellet was re-suspended in equal volume of 2X SDS-PAGE sample buffer (*Appendix V*), boiled for 5 minutes and then re-centrifuged at 9000g for 5 minutes. The supernatant was used as a protein sample for the subsequent protein characterization studies. The extracts were always made in small volumes so that they may be used as fresh as possible. The prepared extracts were stored in small aliquots of 100 µl at -20°C to avoid repetitive freeze-thawing.

4.2.6   Protein Quantification

The proteins extracted from *Bacillus* and *Pasteuria* endospores were quantified by measuring the absorbance in the Ultraviolet (UV) range at 280 nm using the Eppendorf Biophotometer Plus. The final concentration of proteins to be used for SDS-PAGE was at least 2 mg/ml for each sample.

4.2.7 Sodium dodecyl Sulphate –Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate (SDS-PAGE) was done using 12% or 7% (w/v) resolving gel and 4% (w/v) stacking gel (Laemmli, 1970). The protein extracts from *Pasteuria* and *Bacillus* endospores (20 µl each) were loaded into the wells of the stacking gel. Precision Plus Protein™ Standards (Biorad, UK) were used as standard markers of which 5 µl were loaded in each well. The proteins were allowed to fractionate at a constant voltage of 200V for about 1.5- 2 hours. The duration of electrophoresis was altered as per the molecular weight of the protein under consideration, for e.g., for proteins larger than 200 kDa, the electrophoresis was run until the 75 kDa protein band in the ladder reached near the bottom of the resolving gel.

4.2.8 Total Protein Staining of SDS-PAGE gels

A. Coomassie staining

The gels were stained with Coomassie stain (*Appendix V*) for 45 minutes on a rocking shaker. The stain was poured away down the sink and replaced with the de-staining solution. A sponge bung was kept in the staining tray to accelerate the de-staining process which took about 30 minutes or more.

B. Silver staining

To visualize the bands separated by SDS-PAGE, the gels were silver stained using standard protocol (Wray *et al.*, 1981). The gels were successively placed in fixing solution, washing solution and 10% glutaraldehyde, each for 30 minutes, followed by several subsequent washes in distilled water at least for 4 hours. The sensitized gels were then treated with dithiothreitol (DTT) for 30 minutes. Without pouring off DTT, silver nitrate solution was added to the gels. After 30 minutes, the gels were rinsed with distilled water and then with developer. The stain was allowed to develop keeping the gels immersed completely in developer. When the bands were developed as desired the reaction was stopped by stopping solution (For the composition of all the reagents see *Appendix V*).

4.2.9 Western Blotting and Immunodetection

For further characterization of the endospore proteins separated by SDS-PAGE, the proteins were electro-blotted onto PVDF (Polyvinylidene fluoride) membrane using a Biorad Western blot wet transfer unit at 200mA constant current for 2 hours (Towbin *et al.*, 1979). The tank was kept cool by placing an icepack in it during the transfer. Prior to setting up the transfer the PVDF membranes had to be activated by soaking them in methanol for 15 seconds, further soaking in distilled water for 2 minutes, and in blotting buffer for 15 minutes. The success of the transfer was determined visually by the complete transfer of the pre-stained ladder. Proteins on the PVDF membrane were immune-detected as described below.

Following the Western Blotting transfer, proteins bound to the PVDF membrane were immuno-detected (Davies *et al.*, 1992).

A. Antibodies and Pre-immune sera used for Immunodetection

The primary polyclonal antibodies were either raised to whole endospores and produced at Rothamsted Research according to Davies *et al.* (1992) or were obtained from Sigma-Genosys Ltd., Haverhill, U.K as listed below. The antibody solutions were made according to Davies *et al.,* (1992) to the desired concentration, between 1:500 to 1:2000, depending on the stock antibody concentration in PBST (0.05% v/v Tween in PBS).

i. **PAB to Whole Spore of *Pasteuria* (Anti-PpWS)**: A polyclonal antibody to whole *Pasteuria* endospore, raised in rabbit (Davies *et al.*, 1992).

ii. **PABs to CL sequence of *Pasteuria* (Col1981/Col1982)**: The two polyclonal antibodies Col1981 and Col1982 were raised to the two custom-synthetized collagen-like peptides GTPGTPGPAGPAGPA, and GPQGPQGTQGIQGIQ respectively (Sigma-Genosys Ltd., UK). These peptides were synthesized from common CL-motifs (G–X–Y repeat sequences) identified from *Pasteuria* genome survey sequences (Davies and Opperman, 2006). Both the antibodies were raised in rabbit following the Sigma-Genosys standard protocol.

iii. **Pre-immune serum**: Pre-bleeds were undertaken prior to any immunizations and used as pre-immune sera.

iv. **Secondary Antibody**: The secondary antibody used in all experiments was anti-rabbit IgG produced in goat (A3687, Sigma-Genosys Ltd., UK), conjugated to alkaline phosphatase.

B. Steps for Immunodetection

All the incubation and washing steps were performed in small staining trays kept on a rocker at room temperature.

i. **Blocking**: The non-specific binding sites of the membrane were blocked by placing the membrane in a blocking solution (PBST with 1% skimmed milk powder) for 30 minutes.

ii. **Incubation in Primary Antibody**: The blocked membrane was incubated overnight in primary antibody solution (Anti-PpWS,1:2000 or Col1981, 1:500 or Col1982, 1:500) at room temperature, followed by subsequent washes in washing solution (PBST). The rabbit pre-immune antiserum was used in place of the primary antibody, as a negative control, to rule out any non-specific cross-reactivity of endospore proteins with rabbit sera.

iii. **Incubation in Secondary Antibody**: The membrane was then incubated in anti-rabbit IgG (1:1000) for 1 hour at room temperature.

iv. **Incubation in substrate**: The proteins specific to the primary antibodies were detected by a chromogenic substrate BCIP (5-bromo-4-chloro-3-indolyl-phosphate). A solution of disodium salt of BCIP (2 mg/ml) was freshly prepared in Diethanolamine (DEA) buffer. The membrane was placed in the substrate solution until blue coloured bands developed. Once the desired colour developed, the membrane was washed in distilled water for about 10 minutes and allowed to air-dry.

v. **Observation**: The blue bands on the membrane were observed. These bands corresponded to the endospore proteins recognized by the primary antibodies used.

4.2.10 Detection of glycoproteins on PVDF
A. Glycoprotein Staining

The proteins immobilized on PVDF membranes, following the Western transfer of SDS-PAGE separated endospore proteins, were stained using the Pierce™ Glycoprotein Staining Kit (Thermo Fisher Scientific, USA) that detects any carbohydrate moieties conjugated to glycoproteins. The following procedure was used for the staining. All the incubation and washing steps were performed in small staining trays kept on a rocker at room temperature.

i. **Washing in 3% Acetic Acid**: The membrane was gently washed in 20 ml of 3% acetic acid for 10 minutes. The step was repeated once.

ii. **Incubation in Oxidizing Solution**: The membrane was transferred to 10 ml Oxidizing Solution and incubated for 15 minutes.

iii. **Washing in 3% Acetic Acid**: The membrane was washed in 10 ml of 3% acetic acid for 5 minutes. The step was repeated twice.

iv. **Incubation in Glycoprotein Staining Reagent**: The membrane was transferred to the Glycoprotein staining reagent and gently agitated for 15 minutes.

v. **Incubation in Reducing Solution**: The membrane was transferred to 10 ml of Reducing Solution and incubated for 5 minutes

vi. **Final Washing**: The membrane was washed extensively with 3% acetic acid followed by another washing with ultrapure water.

vii. **Observation**: The membrane was observed for the presence of magenta coloured bands of glycoproteins.

B. Lectin blotting

The endospore protein extracts were separated by SDS-PAGE and blotted on to PVDF membrane using standard Western blotting technique as described previously. Biotinylated wheat germ agglutinin (WGA) (5 mg/ml biotinylated wheat germ agglutinin in 10 mM HEPES, 0.15 M NaCl, pH 7.5, 0.08% sodium azide, 0.1 mM Ca++ obtained from Vector Labs, Peterborough, England) was then used to detect any glycoproteins with N-acetyl glucosamine (NAG) as their glyco-conjugates. The steps followed are described below. All the washing and incubation steps were performed in small staining trays kept on a rocker at room temperature.

i. **Blocking**: The non-specific binding sites of the PVDF membrane were blocked by placing the membrane in 1% bovine serum albumin in TTBS (Tris buffered saline with 0.1% Tween) for 30 minutes.

ii. **Incubation in WGA**: The blocked membrane was transferred to 5 µg /ml biotinylated WGA solution in TTBS and incubated for 1 hour.

iii. **Washing away unbound WGA**: The membrane was washed thrice in TTBS over 15 minutes.

iv. **Incubation in Vectastain® ABC-AP reagent (AK-5000, Vector Laboratories Ltd., UK)**: The ABC-AP reagent was prepared as per manufacturer's instructions. The membrane was incubated in ABC-AP reagent for 30 minutes.

v. **Washing away excess Vectastain® ABC-AP reagent:** The membrane was washed thrice in TTBS over 15 minutes.

vi. **Final Washing step**: The membrane was washed once in 100 mM Tris, pH 9.5 buffer.

vii. **Incubation in Substrate**: The glycoproteins with NAG residues as their glyco-conjugate were finally detected using freshly prepared substrate solution containing 2 mg/ml BCIP in 10% DEA buffer. The membrane was placed in the substrate solution until blue coloured bands developed. Once the desired colour developed, the membrane was washed in distilled water for about 10 minutes and allowed to air-dry.

viii. **Observation**: The blue bands on the membrane were observed. These bands corresponded to the endospore glycoproteins with NAG as their glycoconjugate.

4.2.11 Detection of collagens in endospore protein extracts

The collagens, present in the heterogeneous protein mixture extracted from *Pasteuria* and *Bacillus* endospores, were digested by treating the extract with collagenase enzyme in PBS-Ca^{2+} buffer. Prior to the enzyme treatment, the protein extracts were desalted and precipitated using acetone to remove any salts or buffer components that could interfere with the enzyme activity.

i. **Desalting and buffer exchange**: Prior to collagenase treatment, the protein extracts were desalted to remove all the molecules from the protein extract that could interfere with collagenase activity. The protein extracts were diluted with sterile H$_2$O, loaded on to Vivaspin® 2 centrifugal concentrators (Vivaproducts Inc., USA) and centrifuged at 4000 g for 15 minutes at 4°C in a Heraeus Labofuge 400R centrifuge (ThermoFisher, UK). The flow through was discarded and the retentate was topped up to 1 ml with sterile H$_2$O. The steps were repeated twice. The final retentate free from most of the buffer components was subjected to acetone precipitation to remove any SDS micelles still present.

ii. **Acetone precipitation**: Ice-cold acetone was added to the Vivaspin®-desalted extracts in 4:1 ratio. The acetone-protein mixture was incubated at -20°C for 1 hour and then centrifuged at 13,000 g for 10 minutes. The supernatant was removed carefully, leaving a volume equivalent to about half of the initial volume of the Vivaspin®-desalted extract used for precipitation. Any traces of acetone in the precipitated protein were allowed to evaporate at room temperature. The protein precipitate was re-dissolved in sterile PBS.

iii. **Collagenase treatment**: 6 µl of each protein extract was mixed with 4 µl of collagenase enzyme stock (1500 units/ml) in a 25 µl-tube and incubated at 37°C for 2 hours.

4.2.12 SDS-PAGE, Western blot and immuno-detection of collagenase treated protein extracts

The collagenase treated protein extracts (10 µl of each) were diluted to equal volume of 2X SDS-PAGE sample buffer. The SDS-PAGE and Western Blots were performed as described previously in 4.2.7 and 4.2.9. The untreated protein samples were run simultaneously as controls. The two anti-collagen antibodies Col1981 and Col1982 (see *Section 4.2.9 A* above) were used to detect the presence of collagens as described in *Section 4.2.9 B*. Collagens with NAG as their glycoconjugate were detected by a lectin blot as described in *Section 4.2.10 A*. The bands obtained for untreated protein extracts were compared to those obtained for collagenase treated extracts. A shift in the location of a band to a lower molecular weight marker or a missing band in collagenase treated extracts would indicate a collagen or CL peptide. Gelatin (2 mg/ml stock) was mixed with an equal volume of 2X sample buffer and used as a positive control.

4.2.13 Estimation of molecular weight of proteins from gels and Western blots

To determine the molecular weight of protein/polypeptide components in the endospore protein extracts, after SDS-PAGE/ Western blotting, the images of the stained gels/blots were digitally captured and analyzed using the image processing and analysis software tool Fiji (Schindelin *et al.*, 2012). The migration distances of each visible band of the protein samples and the molecular markers was measured from the top of each resolving gel. The relative migration distance (R_f) of each band of

molecular marker and of the endospore proteins, was determined using the following equation:

$$R_f = \frac{Migration\ distance\ of\ protein}{Length\ of\ the\ resolving\ gel}$$

<div align="right">(Equation 4. 1)</div>

Graphs were plotted using the values of R_f obtained for the molecular markers on the x-axis and the logarithmic values of the known molecular weights of the molecular markers on the y-axis. Lines of best fit were drawn and the consequent straight line equations were derived. The molecular weights of the unknown proteins were determined by interpolating their R_f values into the straight-line equation.

4.3 RESULTS

4.3.1 Purification of endospores of *Bacillus* spp.

Maximum sporulation was observed on the 8th day after inoculation. The endospores obtained and purified from the sporulation broth were at least 98% pure as observed by endospore staining (*Appendix VI*).

4.3.2 SDS-PAGE of protein extracts from *Bacillus* and *Pasteuria* endospores

Coomassie staining revealed at least nine bands in *B. thuringiensis* strain subspecies Al Hakam, three bands in *B. thuringiensis* kurstaki and ten bands in *B. thuringiensis* Berliner (*Figure 4.1*). Only one extremely faint band (Band 9) of approximately 36 kDa was observed in *P. penetrans* endospore extract, which was not seen in any of the *B. thuringiensis* strains. Silver stain, being more sensitive, detected more bands than Coomassie stain (*Figure 4.2*). At least four bands were clearly visible in *Pasteuria* endospore extracts. Band 10 (~58 kDa) was commonly observed in all the bacterial strains while Band 16 (~19 kDa) was common in *P. penetrans* and *B. thuringiensis* kurstaki cry- strain. Band 4 (~112 kDa) and Band 11 (~42 kDa) were peculiar to *P. penetrans. Figures 4.1 (B) and 4.2 (B)* show the occurrence of the protein bands in decreasing order of their molecular masses as visible after coomassie and silver staining, respectively. See *Appendices VIIa* and *VIIb* for graphs and calculations for the estimation of molecular weights.

Figure 4. 1: Protein profiles of endospore protein extracts separated by SDS-PAGE and stained by Coomassie stain (12% resolving, 4% stacking gels; electrophoresis at 200 V for 45 minutes). (A) Image of Coomassie stained gel (prominent bands marked as arrowheads, numbered in decreasing order of their molecular weights); (B) Tabulated representation of occurrence of the bands in different bacterial endospore samples (band numbers are same as numbered in the image). See Appendix VIIb (table 1) for calculations to estimate molecular weight.
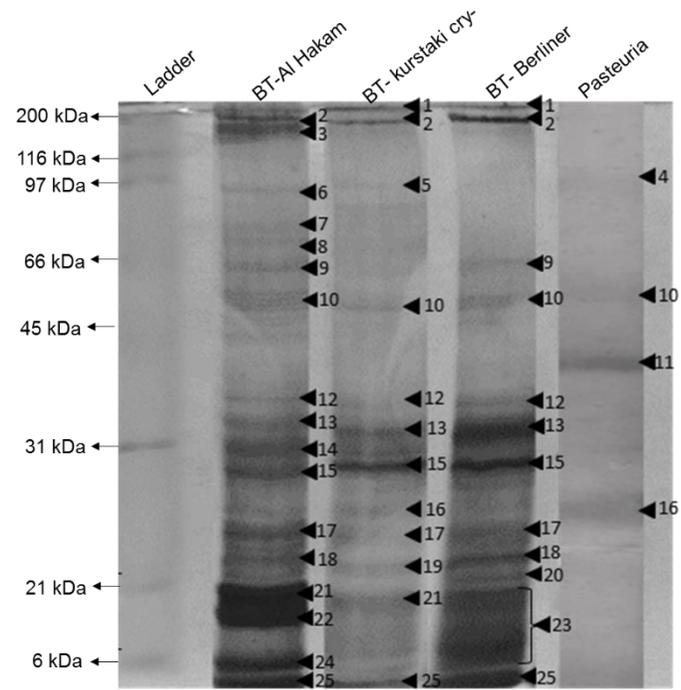
A

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Band 8	Band 9	Band 10	Band 11	Band 12	Band 13	Band 14	Band 15	Band 16	Band 17	Band 18	Band 19	Band 20	Band 21	Band 22	Band 23	Band 24	Band 25
BT-Al Hakam		✓	✓			✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓			✓	✓		✓	✓
BT-kurstaki cry-	✓	✓			✓					✓		✓	✓		✓	✓	✓		✓		✓				✓
BT-berliner	✓	✓							✓	✓		✓	✓		✓		✓	✓		✓			✓		✓
Pasteuria				✓						✓	✓					✓									

B

Figure 4. 2: Protein profiles of endospore protein extracts separated by SDS-PAGE and stained by Silver stain (12% resolving, 4% stacking gels; electrophoresis at 200 V for 45 minutes). (A) Image of Silver stained gel (prominent bands marked as arrowheads, numbered in decreasing order of their molecular weights); (B) Tabulated representation of occurrence of the bands in different bacterial endospore samples (band numbers are same as numbered in the image). See Appendix VIIb (table 2) for calculations to estimate molecular weight.

4.3.3 Western blots of electrophoresed proteins from *Bacillus* endospores

The proteins on SDS-PAGE gels were successfully transferred to PVDF membranes by Western blotting. The success of the transfer was determined as the complete transfer of pre-stained protein markers from the gel onto the membrane.

4.3.4 Immuno-detection of endospore proteins using Anti-PpWS

Probing the Western blot of endospore proteins with Anti-PpWS revealed two bands in *B. thuringiensis* subspecies Al Hakam, three bands each in *B. thuringiensis* subspecies berliner and *B. thuringiensis* kurstaki, while just one band in *P. penetrans* Res148 (*Figure 4.3*). Band 1 (≥250 kDa) was common to all *B. thuringiensis* strains but was not visible in *Pasteuria.* Band 2 (~99 kDa) was observed in kurstaki and Berliner strains of *B. thuringiensis* but not in Al Hakam, while band 3 (~75 kDa) was only observed in Al Hakam. Band 4 (~72 kDa) was common to *Pasteuria* and *B. thuringiensis* kurstaki cry- and Berliner.

4.3.5 Immuno-detection of endospore collagens using Col1981/Col1982

Endospore extracts when probed with the Col1981 revealed one band of >250 kDa (bands 1 and 2) in each of the bacterial samples (*Figure 4.4 A, B)*. Band 1 in *Pasteuria, B. thuringiensis* subspecies berliner and *B. thuringiensis* subspecies kurstaki cry-migrated slightly less than band 2 in *B. thuringiensis* Al Hakam. No other band was visible in *Pasteuria* extract, while there were two additional bands in *B. thuringiensis* berliner (band 5, ~103 kDa; band 6, ~94 kDa) and one band each in *B. thuringiensis* Al Hakam (band 4, ~139 kDa) and *B. thuringiensis* kurstaki (Band 3, 155 kDa).

Col1982 antibody recognized a ≥250 kDa protein (band 1) and three other protein bands in *Pasteuria* (band 3, ~142 kDa; band 6, ~91 kDa; band 7, ~72 kDa), one protein band (band 2, ~156 kDa) in *B. thuringiensis* Al Hakam, two bands in *B. thuringiensis* kurstaki cry- (band 4, ~111kDa; band 5, ~103 kDa) and no bands in *B. thuringiensis* subspecies berliner (*Figure 4.4 C, D*).

	BT-Al Hakam	BT-kurstaki cry-	BT-berliner	Pasteuria
Band 1 (>250 kDa)	✓	✓	✓	
Band 2 (~100 kDa)		✓	✓	
Band 3 (~75 kDa)	✓			
Band 4 (~72 kDa)		✓	✓	✓

(A)   (B)
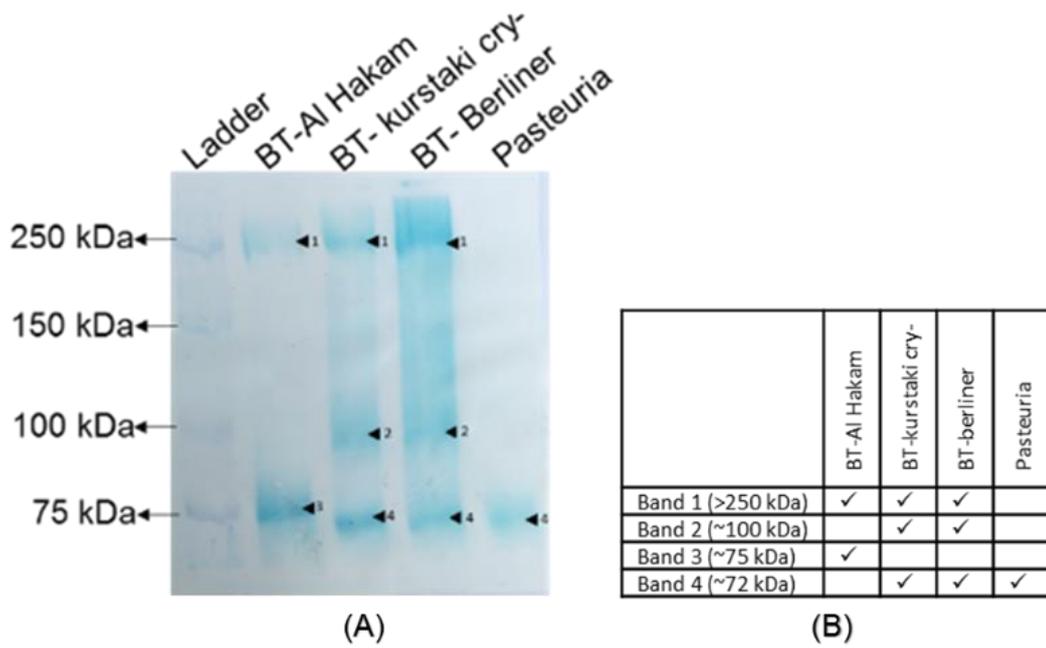
Figure 4. 3: Immuno-detection of endospore protein extracts from three strains of *B. thuringiensis* and *P. penetrans* Res148 using Anti-PpWS (A) Western blot image (B) Estimated molecular masses (kDa) of protein bands recognized by Anti-PpWS. See Appendix VIIb (table 3) for calculations.
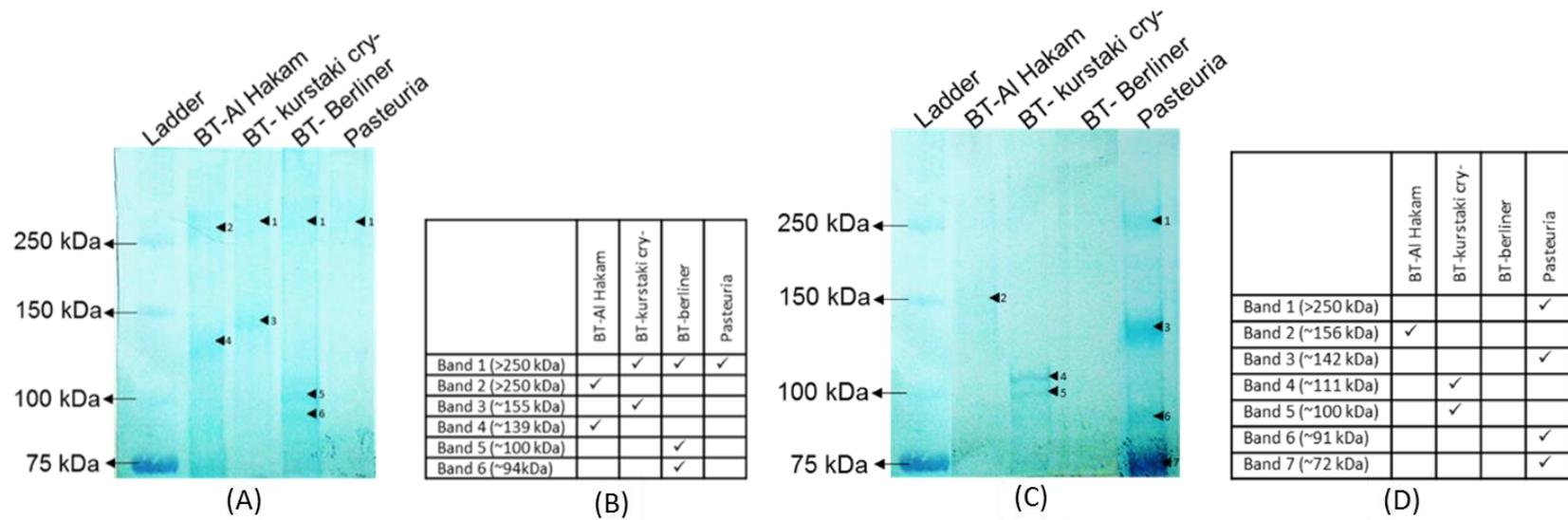
Figure 4. 4: Immuno-detection of collagen-like proteins in the endospore protein extracts from three strains of *B. thuringiensis* and *P. penetrans* Res148 using Col1981. (A) Western blot probed with Col1981 (B) Estimated molecular masses (kDa) of protein bands recognized by Col1981 (C) Western blot probed with Col1982 (D) Estimated molecular masses (kDa) of protein bands recognized by Col1982. See Appendix VIIb (table 4, 5) for calculations.

4.3.6 Detection of glycoproteins

Staining of SDS-PAGE gels with the Thermo Scientific Pierce^{TM} Glycoprotein Staining Kit revealed a band more than 250 kDa (band 1) in *Pasteuria* endospore protein extract (*Figure 4.5*). Similar higher molecular weight bands were observed in two of the *B. thuringiensis* strains, namely, *B. thuringiensis* Al Hakam and *B. thuringiensis* kurstaki (band 2). Additionally, a smear of about 43 to 66 kDa (band 3) was also visible in *B. thuringiensis* kurstaki. No glycoprotein was detected in *B. thuringiensis* Berliner.

Probing the Western blots of endospore extracts with wheat germ agglutinin (a lectin that detects N-acetylglucosamine) revealed a single band of about 139 kDa (band 4) in *B. thuringiensis* Al Hakam, while no bands in the other *B. thuringiensis* strains (*Figure 4.6*). In *P. penetrans* Res148 endospore extracts, at least 6 bands were distinctly visible (band 1, >250 kDa; band 2, ~224 kDa; band 3, ~176 kDa; band 5, ~126 kDa; band 6, ~83 kDa; band 7, ~69 kDa). These results indicated that the some of the polypeptides resolved from *B. thuringiensis* and *P. penetrans* endospores are glycosylated with NAG.

4.3.7 Digestion of collagen proteins Collagenase

The success of collagenase treatment was confirmed by a visual shift of higher molecular bands to lower molecular weights in gelatin sample (*Figure 4.7*). In untreated gelatin sample separated by SDS-PAGE and stained by silver stain, two bands of about 149 kDa (band 1) and 142 kDa (band 2) were observed. While the number of bands remained the same in the gelatin sample treated with collagenase for two hours, a shift in mobility of the bands to about 104 kDa (band 3) and 68 kDa (band 4) was observed.

4.3.8 Detection of collagens after collagenase treatment of protein extracts

In an experiment to study the effect of a two hours' collagenase treatment on the Western blot recognition by wheat germ agglutinin, the WGA was still able to recognize proteins/peptides in the treated samples (*Figure 4.8*). In the case of *B. thuringiensis* Al Hakam, a 141 kDa band (band 3) was observed in both the treated and untreated protein samples. However, there was a marked reduction in the intensity of the band in the treated sample. In the case of *Pasteuria*, three distinct bands (band 1, >250 kDa; band 2, ~194 kDa; band 4, ~130 kDa) were observed in both treated and

untreated protein samples. An additional weaker band at about 106 kDa (band 5) was seen in *P. penetrans* protein sample treated with collagenase for 2 hours.

In another experiment where *Pasteuria* endospore proteins were treated with collagenase enzyme for 6 hours, the number of bands recognized by WGA decreased from five bands in untreated protein extract (band 1, >250 kDa; band 2, ~199 kDa; band 3, ~147 kDa; band 6, ~90 kDa, band 8, 69 kDa) to three bands in treated extract (band 4, ~131 kDa; band 5, ~115 kDa; band 7, ~73 kDa) (*Figure 4.9 A, B*). This indicates that some of the proteins/peptides were digested, partially or completely, and the proteins/peptides or peptide fragments detected on the blot were the ones that had NAG attached to them as a glyco-conjugate. When the Western blot, of the same (6 hours' collagenase treated) samples, was probed with Col1982, it produced a single band of about 151 kDa (band 1) in untreated protein sample whereas two distinct bands at about 122 kDa (band 2) and 79 kDa (band 3) in collagenase treated protein sample of *Pasteuria* endospores (*Figure 4.9 C, D*).

	BT-Al Hakam	BT-kurstaki cry-	BT-berliner	Pasteuria
Band 1 (>250 kDa)				✓
Band 2 (>250 kDa)	✓	✓		
Band 3 (Smear: 66 – 43 kDa)		✓		

(A)　　　　　(B)
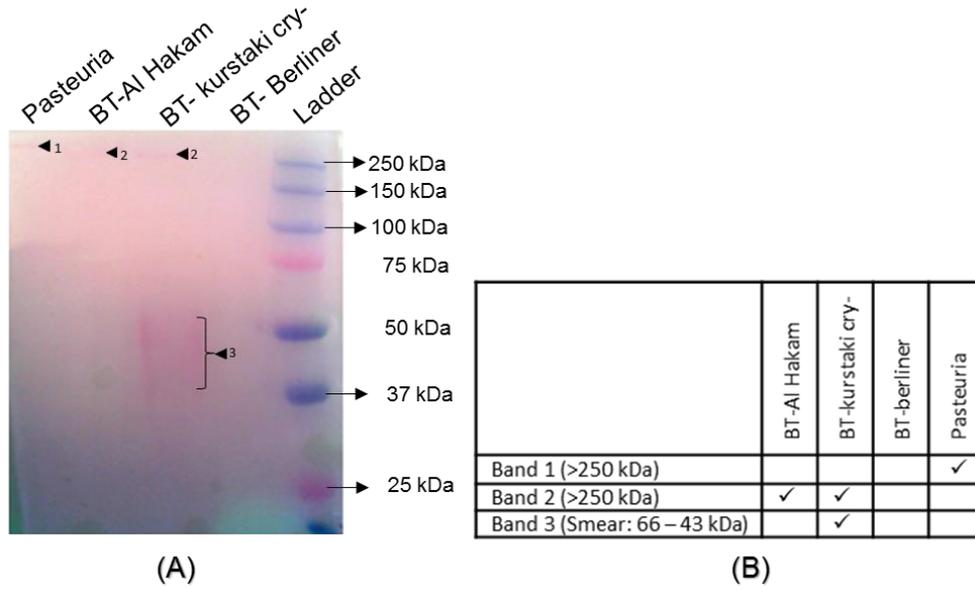
Figure 4. 5: Detection of glycoproteins in endospore protein extracts. (A) SDS-PAGE gel stained with Thermo Scientific Pierce™ Glycoprotein Staining Kit (B) Estimated molecular masses (kDa) of protein bands recognized by Glycoprotein stain. See Appendix VIIb (table 6) for calculations.

	BT-Al Hakam	BT-kurstaki cry-	BT-berliner	Pasteuria
Band 1 (>250 kDa)				✓
Band 2 (~224 kDa)				✓
Band 3 (~176 kDa)				✓
Band 4 (~139 kDa)	✓			
Band 5 (~126 kDa)				✓
Band 6 (~83 kDa)				✓
Band 7 (~72 kDa)				✓

(A)　　　　　(B)
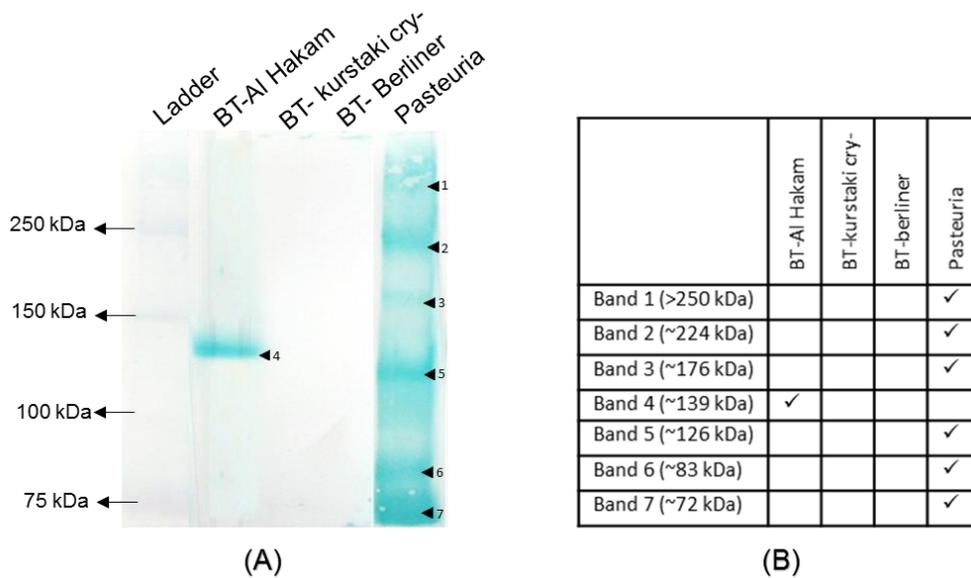
Figure 4. 6: (A) Detection of glycoproteins having NAG as glyco-conjugate. (A) Western blot probed with WGA (B) Estimated molecular masses (kDa) of protein bands recognized by WGA. See Appendix VIIb (table 7) for calculations.
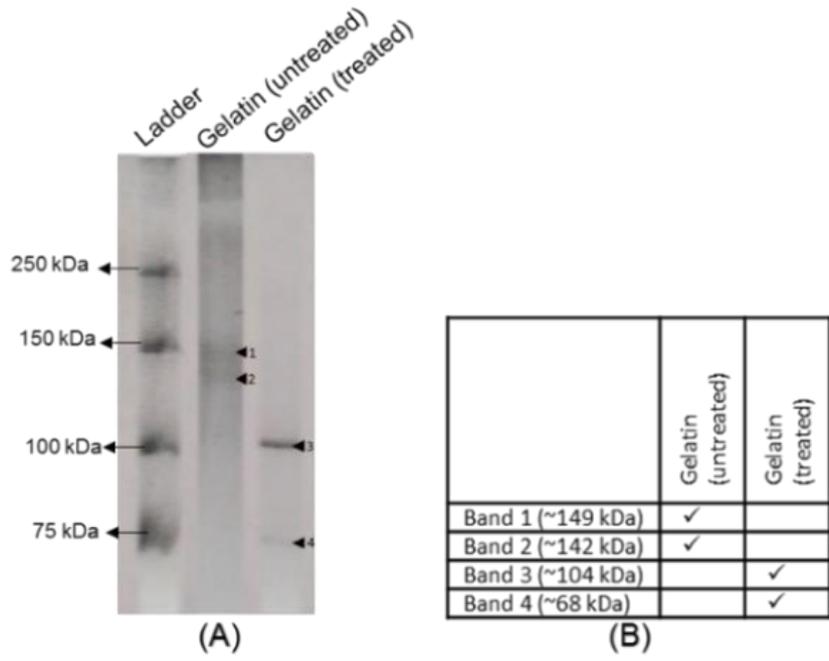
Figure 4. 7: Gelatin as positive control for Collagenase treatment. (A) Silver stained gel of collagenase treated and untreated gelatin extract separated by SDS-PAGE (B) Estimated molecular masses (kDa). The collagenase treatment was done at 37°C for 2 hours. See *Appendix VIIb* (table 8) for calculations.

	Gelatin (untreated)	Gelatin (treated)
Band 1 (~149 kDa)	✓	
Band 2 (~142 kDa)	✓	
Band 3 (~104 kDa)		✓
Band 4 (~68 kDa)		✓
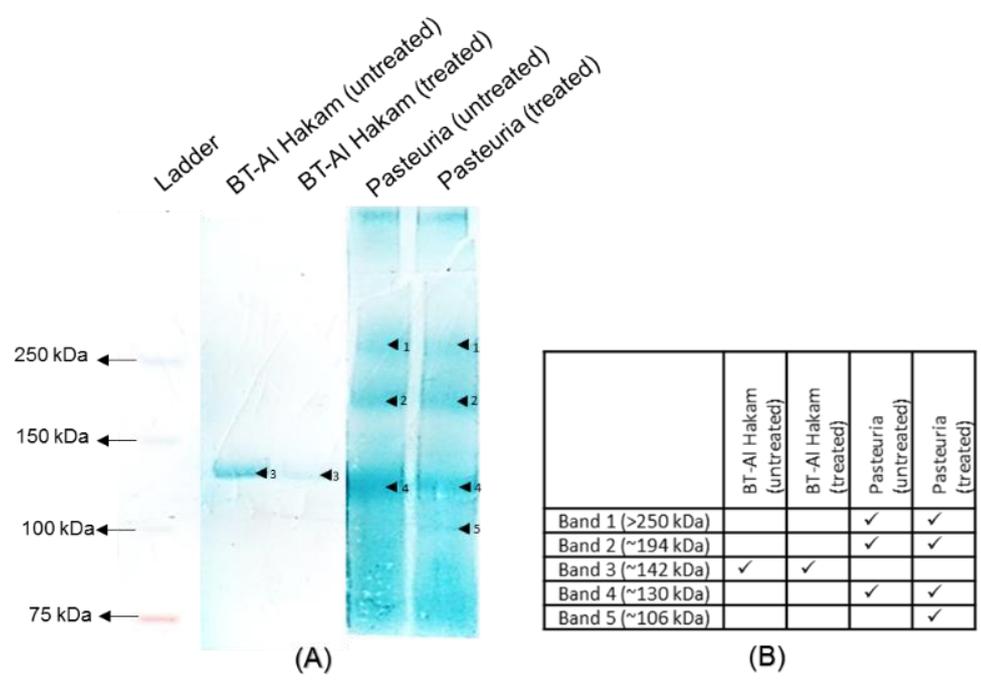
Figure 4. 8: Detection of collagens with NAG as glyco-conjugate (Two hours' collagenase treatment). The endospores protein extracts were treated with collagenase at 37°C for 2 hours prior to SDS-PAGE and Western blotting. (A) Western blot probed with WGA (B) Estimated molecular masses (kDa) of protein bands recognized by WGA. See *Appendix VIIb* (table 9) for calculations.

	BT-Al Hakam (untreated)	BT-Al Hakam (treated)	Pasteuria (untreated)	Pasteuria (treated)
Band 1 (>250 kDa)			✓	✓
Band 2 (~194 kDa)			✓	✓
Band 3 (~142 kDa)	✓	✓		
Band 4 (~130 kDa)			✓	✓
Band 5 (~106 kDa)				✓

100
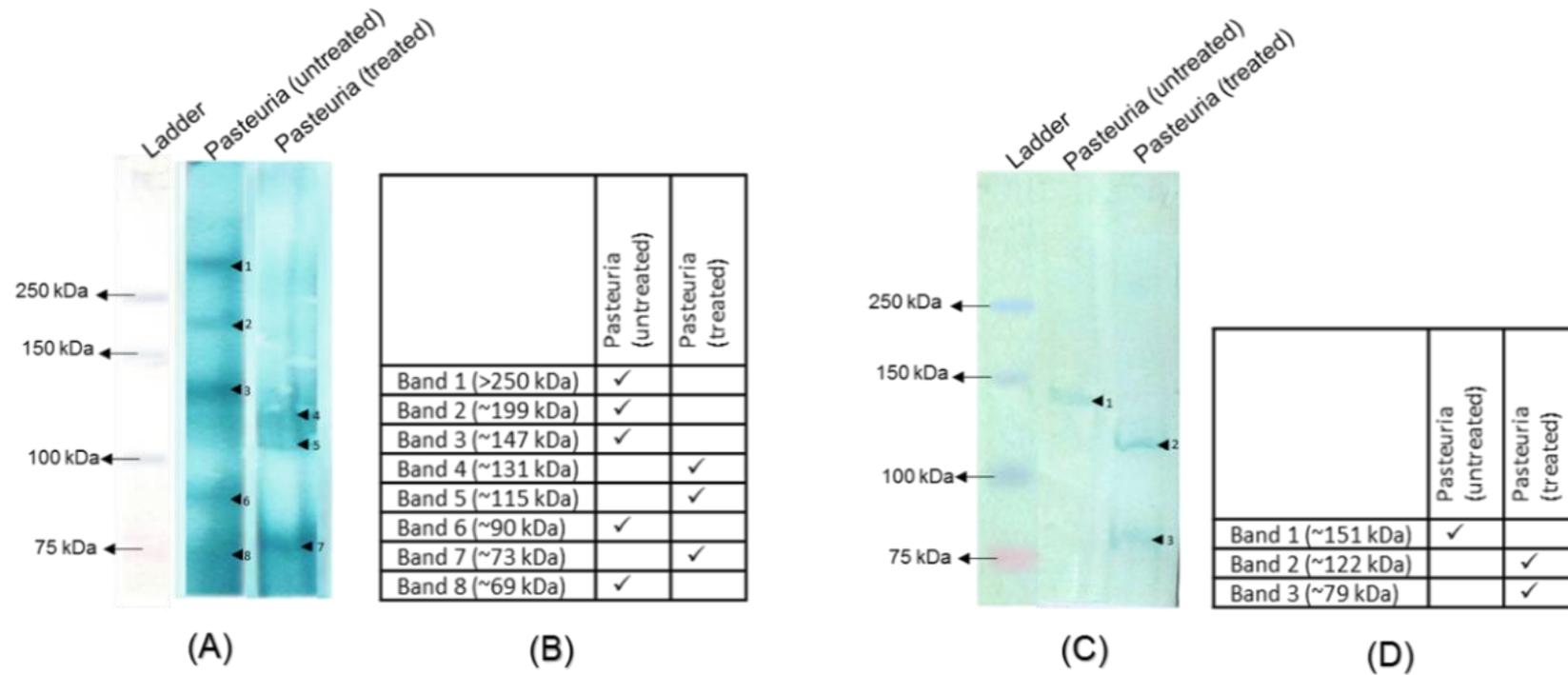
Figure 4. 9: Detection of collagens with NAG as glyco-conjugate (Six hours' collagenase treatment). The endospores protein extracts were treated with collagenase at 37°C for 6 hours prior to SDS-PAGE and Western blotting. (A) Western blot probed with WGA (B) Estimated molecular masses (kDa) of protein bands recognized by WGA (C) Western blot probed with Col1982 (D) Estimated molecular masses (kDa) of protein bands recognized by Col1982. See *Appendix VIIb* (table 10, 11) for calculations.

4.4   SUMMARY OF RESULTS

The results of the protein characterization studies suggest that the endospores of *P. penetrans* and *B. thuringiensis* share proteins with common epitopes, including some CLPs that are glycosylated with N-acetylglucosamine.

4.5   DISCUSSION

The aim of the experiments discussed in this chapter was to comparatively characterize the endospore proteins of *P. penetrans* with those of *B. thuringiensis.* The results from Coomassie staining and silver staining of the endospore extracts suggest some considerable similarities in the endospore protein composition between different strains of *B. thuringiensis* and *P. penetrans.* As expected, silver staining being more sensitive in detecting lower concentration of proteins revealed more bands than Coomassie stain. More number of bands were revealed in the endospore extracts from the three *B. thuringiensis* strains than *P. penetrans.* This could be, possibly, due to tougher endospore coat of *P. penetrans* which could not be readily dissolved by the sample buffer resulting in substantially lesser protein in the extract.

As evident from the total protein profile, there was at least one protein (~58 kDa) that was commonly found in *P. penetrans* and all three *B. thuringiensis* strains. In an earlier study a 58 kDa protein, not revealed by silver staining, appeared on the Western blot of endospore extracts of three different populations of *P. penetrans* when probed with a polyclonal antibody raised to *P. penetrans* (Davies, Robinson, Laird, 1992)*.* Strikingly, the molecular chaperone protein GroEL, which is known to be abundant in bacterial exosporia, also migrates with an apparent molecular weight of 58 kDa (Charlton *et al.*, 1999, Redmond *et al.*, 2004). GroEL is present throughout the tree of life and it is interesting that it is also present in the endospores although we don't know of their definitive function in the endospores. A ~19 kDa protein was apparently common between *P. penetrans* and *B. thuringiensis* kurstaki cry-, but was not observed in the other two *B. thuringiensis* strains. It is possible that this ~19 kDa protein is either absent in the two strains or is present in lesser concentration not enough to be detected by Coomassie or silver staining. Notably, a 17 kDa protein ExsF/BxpB is known to be associated with the exosporium of *Bacillus* species

(Redmond *et al.*, 2004). Also, a 17 kDa polypeptide has been shown in *P. penetrans* endospore extract in a previous study (Davies, Robinson, Laird, 1992). Both of these *Bacillus* and *Pasteuria* proteins could be comparable to this ~19 kDa protein band visible in the silver stained gel. Two bands of ~112 kDa and ~42 kDa respectively were unique to *P. penetrans* endospores.

The Western blot analyses show the presence of proteins with common epitopes in the endospores of *P. penetrans* and all the three tested strains of *B. thuringiensis.* The *Pasteuria*-specific polyclonal antibody Anti-PpWS recognized a >250 kDa protein in the *B. thuringiensis* strains but no >250 kDa band in *P. penetrans* itself*.* A similar >250 kDa protein band was seen in all the bacteria, including *P. penetrans*, when the blot was probed with Col1981 and in *P. penetrans* when probed with Col1982. Both Col1981 and Col1982 are antibodies raised to two different collagen-like synthetic polypeptides originated from *Pasteuria* GSS database. Therefore, the above results suggest that a high molecular weight collagen or CLP could be common between *P. penetrans* and *B. thuringiensis* endospores. The BclA protein, an immunodominant CLP in the exosporium of *Bacillus* spp. is known to migrate with an apparent molecular weight of >250 kDa due to heavy glycosylation (Sylvestre *et al.*, 2002). Interestingly, glycoprotein staining revealed bands of >250 kDa in *P. penetrans* and *B. thuringiensis* Al Hakam and kurstaki cry- strains, suggesting their possible similarity with BclA.

Another protein band that stands out amongst others is a ~72-75 kDa band that was commonly recognized in all the protein samples by Anti-PpWS. Clearly, Anti-PpWS detected a common epitope in these endospore proteins that have comparable molecular weight in *P. penetrans* and *B. thuringiensis*. Col1982 antibody detected the same band in *P. penetrans.* This suggests that the ~72 kDa band in *P. penetrans* is a collagen-like protein. It is possible that the bands recognized by Anti-PpWS represent orthologous proteins differing in a specific epitope recognized by Col1982 in *P. penetrans* that is either not present in the *B. thuringiensis* orthologue or is not conformationally accessible for the binding of Col1982. Another possibility is that these proteins are not the same but have epitopes that are commonly recognized by the Anti-PpWS antibody. If the former is true, it can be inferred that these ~72-75 kDa proteins are all collagen-like proteins common between *B. thuringiensis* and *P. penetrans.* The possible similarity of these protein bands with ExsJ, a collagen-like

protein associated with *B. cereus* exosporium, known to migrate as a 70 kDa monomer (Todd *et al.*, 2003), is a strong possibility .

Apart from these two observed similarities (a ~250 kDa protein and a ~72 kDa protein) in the immunogenic profiles of *P. penetrans* and *B. thuringiensis*, there were some bands that were differentially detected in the endospore proteins of the two genera. The Anti-PpWS antibody recognized a ~100 kDa protein band in two of the *B. thuringiensis* strains (kurstaki cry- and Berliner). Col1981 and Col1982 detected similar sized bands in *B. thuringiensis* Berliner and *B. thuringiensis* kurstaki cry- respectively, thus indicating the presence of a collagen-like epitope associated with this ~100 kDa protein. Additionally, Col1981 detected three other bands of ~139 kDa, ~155 kDa and ~94 kDa in *B. thuringiensis* strains Al Hakam, kurstaki cry- and Berliner respectively. Similarly, Col1982 recognized a ~156 kDa in *B. thuringiensis* Al Hakam and a ~103 kDa protein in *B. thuringiensis* kurstaki cry- strain. In *Pasteuria* endospore proteins, in addition to the peculiar ~250 kDa and ~72 kDa protein bands discussed in the beginning, two other bands of ~142 kDa and ~91 kDa were seen in Col1982-probed blot. Therefore, it can be concluded that *P. penetrans* endospores have at least four different collagen-like polypeptides present in the endospore with Col1982-specific epitopes (approximately of 250 kDa, 72 kDa, 142 kDa and 91 kDa) and at least two of these polypeptides (250 kDa and 72 kDa) are also present in all the tested *B. thuringiensis* strains in addition to other proteins/polypeptides with similar epitopes. Of these, the 250 kDa protein is also revealed to be a glycoprotein by glycoprotein staining.

N-acetylglucosamine (NAG) is a sugar commonly associated with many bacterial glycoproteins. A lectin staining with wheat germ agglutinin (WGA) was done to detect if any of the proteins on the blot had N-acetylglucosamine as a glyco-conjugate. A ~139 kDa protein band in *B. thuringiensis* Al Hakam, previously recognized by Col1981, was detected by WGA. Thus, this suggests the 139 kDa protein is a collagen-like glycoprotein decorated with NAG sugars. No bands were observed in the other two strains of *B. thuringiensis.* WGA staining displayed six bands (>250 kDa, ~224 kDa, ~176 kDa, ~126 kDa, ~83 kDa and ~72 kDa) in *P. penetrans.* Out of these, the ~126 kDa and ~83 kDa protein bands are comparable in size and glycosylation to two previously reported potential adhesins in *P. penetrans* endospores, one of ~126 kDa and the other of ~89 kDa molecular weight, and both bearing terminal N-

acetylglucosamine residues (Persidis, 1991). The current study conveys that *P. penetrans* contains at least six endospore proteins that are glycosylated with NAG, two of which are collagen-like glycoproteins (250 kDa and 72 kDa proteins detected by Col1982).

To further confirm the presence of collagen in the endospore proteins, the extracted proteins were treated with collagenase enzyme prior to running them on a gel for immunodetection. A collagen band was expected to either disappear or shift from its original position. When detected with WGA, the 139 kDA band in *B. thuringiensis* Al Hakam became extremely faint after 2 hours of collagenase treatment, probably due to partial digestion of this protein. After 2 hours of treatment of *P. penetrans* endospore proteins with collagenase and staining the blot with WGA, none of the bands disappeared when compared with untreated sample. However, in the treated sample the ~142 kDa protein band was comparatively fainter and an additional band of ~106 kDa was observed. After 6 hours of collagenase treatment of *P. penetrans* endospore proteins, the number of bands detected by WGA was reduced from six to only three bands. Moreover, the three bands that were observed in the treated sample were not at the same positions as in the untreated samples. In another experiment, when collagenase treated (6 hours) *Pasteuria* protein samples were probed with Col1982, a shift in the size of bands was evident with the detection of an additional band.

All the results discussed in this chapter indicate the presence of collagen-like proteins in the endospores of *P. penetrans* and *B. thuringiensis* and that they are glycosylated with N-acetylglucosamine sharing common epitopes recognized by Anti-PpWS, Col1981 and Col1982. The only negative control used in the experiments was pre-immune serum from the immunized animal. Including the protein extracts from a bacterium with least known similarities with *P. penetrans* could have been beneficial to further validate the results of this study. The proteins or polypeptides immunologically detected in this study are components of endospores, whether any of these proteins are surface associated and are directly involved in attachment of the endospores to the host nematode cuticle is not yet clear. The chapters that follow address this question.

Chapter 5
Surface characterization of *P. penetrans* endospores

5.1   INTRODUCTION

From previous studies on *Pasteuria* endospore heterogeneity and on the role surface proteins play in pathogenicity throughout the bacterial kingdom, it is perceptible that characterizing the endospore surface epitopes of *Pasteuria penetrans* is a pivotal step towards understanding the mechanisms of *Pasteuria*-nematode interaction. The results shown in the previous chapter (Chapter 4) suggest antigenic similarities between the endospore proteins of *P. penetrans* and *B. thuringiensis* and the presence of some collagen-like proteins and proteins glycosylated with *N*-acetylglucosamine (NAG) is evident.

5.1.1   Bacterial surface antigens play key roles in pathogenesis

The adherence of a bacterial pathogen/parasite to the host cell surface is the first and major step before invasion into the host, colonization, and subsequent pathogenicity (Pizarro-Cerdá and Cossart, 2006). The factors of the pathogen and the host involved in the process of adherence or attachment are often called adhesins (pathogen) and receptors (host). Both adhesins and receptors could be any macromolecules which are integral components of the cell surface of the pathogen and the host. Pathogens evolve biochemically diverse adhesins to be able to interact with the equally diverse range of host cell receptors (Kline *et al.*, 2009). Bacterial adhesins allow the pathogens to resist the shearing forces present in the surrounding environment which may make it difficult for the bacteria to stick to the host cells. Adhesins, to a large extent, are also responsible for specifically targeting the bacteria to attach to their host in a 'lock and key' manner (O'Connell, 2003). Several adhesins recognize host extracellular matrix proteins such as collagen, elastin, fibronectin, keratin, laminin, fibrinogen, etc. and are known as 'Microbial surface components recognizing adhesive matrix molecules' or MSCRAMMs in short (Patti *et al.*, 1994). Members of the genus *Staphylococcus* possess specialized fibronectin-binding proteins on their surface (Schwarz-Linek *et al.*, 2003). Enteropathogenic *Yersinia* species produce adhesins that interact with integrins of the host (Isberg and Barnes, 2001). The pathogenic species of many Gram-negative genera, such as, *Neisseria*, *Salmonella*, *Legionella*, *Vibrio* and some

Gram-positive genera, such as, *Streptococcus parasanguis* and *Actinomyces* spp. bear highly specialized hair-like surface protrusions, called pili or fimbriae, that are involved in the attachment of the bacterial cells to the host cell/tissue lining (Craig *et al.*, 2004, Ton-That and Schneewind, 2004). In the evolutionary arms race, some bacteria have even evolved with not just one but an array of adhesins on their cell surface which can recognize and bind to different host cell receptors. *Staphylococcus epidermidis*, for example, is an opportunistic human pathogen that displays at least six different types of cell-wall anchored adhesins that recognize broad category of receptor molecules on the host cells/tissues (Paharik and Horswill, 2016). Although owing to the wide range of pathogens adhesins and host receptors involved, the mechanisms of attachment of the pathogens to the hosts may vary substantially, of all the pathogens rapidly co-evolving with their hosts, the primary purpose is to be able to invade successfully and to establish an infection.

5.1.2 Previous studies on the surface antigens of *Bacillus* endospores

Bacterial endospores can be broadly classified into two: those that contain an exosporium and those that don't. The endospore architecture and the process of its assembly has been studied most extensively using the model organism *Bacillus subtilis* which lacks an exosporium (McKenney *et al.*, 2013). The pathogenic species *B. anthracis, B. thuringiensis* and *B. cereus* possess an exosporium but relatively little is known about its composition and assembly. However, recent advancements in electron microscopic techniques have led to an improved understanding of the endospore architecture and the exosporium layer in particular (Kailas *et al.*, 2011, Giorno *et al.*, 2007, Rodenburg *et al.*, 2014).

A number of proteins associated with the exosporium in *Bacillus* spp. have now been identified, but the exact localization of these proteins on the exosporium has been explored only for a few of them. The most immunodominant exosporial protein characterized to date is the BclA protein (Sylvestre *et al.*, 2002). It is a collagen-like glycoprotein that forms the filaments of the hairy nap that are projected out from the proteinaceous and crystalline basal layer of the exosporium (Sylvestre *et al.*, 2002, Sylvestre *et al.*, 2003, Thompson and Stewart, 2008, Boydston *et al.*, 2005). The glyco-conjugates of BclA include L-rhamnose, N-acetylgalactosamine (NAG), 3-O-methyl rhamnose and anthrose, however, the list is not exhaustive (Daubenspeck *et*

al., 2004a, Tamborrini et al., 2009). Another exosporium associated immunodominant collagen-like glycoprotein is BclB which is supposedly surface exposed above the basal layer of the exosporium but somewhere beneath the BclA filaments (Thompson et al., 2007, Thompson et al., 2011b, Thompson et al., 2012). Yet other collagen-like glycoproteins, namely, BetA, ExsJ and ExsH, have been found to be associated with exosporium of *Bacillus* spp. but not much is known about their polysaccharide composition (Leski et al., 2009, Thompson et al., 2011a, Todd et al., 2003, Garcia-Patrone and Tandecarz, 1995, Lequette et al., 2011).

Several other proteins have been shown to either compose the exosporium or involve in its assembly. ExsY, CotY, ExsFA (BxpB) and ExsFB are the key proteins that have been identified in the basal layer of the exosporium (Redmond et al., 2004, Boydston et al., 2006, Johnson et al., 2006, Thompson et al., 2012, Sylvestre et al., 2005, Steichen et al., 2005, Thompson et al., 2011b). Other proteins, like BxpA, BxpC, ExsB. ExsC, ExsK, GerQ, have been identified in different proteomic analyses of exosporial extracts but their function and surface localization, as exposed antigens or ligands, is yet to be explored (Redmond et al., 2004, Steichen et al., 2003, McPherson et al., 2010, Todd et al., 2003, Severson et al., 2009, Gai et al., 2006).

5.1.3 Previous studies on surface antigens of *Pasteuria* endospores

Endospore heterogeneity of *Pasteuria*, governed by the nature and amount of surface associated proteins, has been, previously, linked to the host specificity of *Pasteuria* spp. (Davies et al., 1992, Davies et al., 1994). Different theories exist in the literature regarding the surface exposed epitopes on the endospore surface of nematode parasitic *Pasteuria* and their role in the attachment of the endospore to the cuticle of host nematodes. Studies suggest the presence of a variety of receptor epitopes or adhesins on the endospore surface of *P. penetrans* (Davies et al., 1996, Persidis et al., 1991). A study using a monoclonal antibody raised to the endospores of a population of *P. penetrans* has shown site-specific localization of surface epitopes on the concave side of the endospores and has indicated endospore surface heterogeneity both within and between different populations of *P. penetrans* (Davies and Redden, 1997). Another immunolocalization study revealed that surface adhesins are almost uniformly distributed over the endospores and they begin to appear only at later stages of sporulation (Brito et al., 2003).

Previous protein characterization studies have indicated the association of *N*-acetylglucosamine (NAG) and *N*-actylmuramic acid (NAM) with the peptides resolved from the endospore protein extracts of *P. penetrans* and the presence of collagens on their endospore surface (Persidis *et al.*, 1991, Davies and Danks, 1993, Charnecki *et al.*, 1998, Davies and Opperman, 2006). The occurrence of carbohydrate binding proteins (lectins) on the cuticle of plant parasitic nematodes has been demonstrated earlier which supports the idea of the presence of glycoproteins or carbohydrates on the endospore surface of *Pasteuria* and their involvement in the attachment of endospores to the nematode cuticle (Bird *et al.*, 1989, Persidis *et al.*, 1991, Davies and Danks, 1993, Davies, 1994). A study suggests the involvement of gelatin-binding domains of fibronectin-like residues present on the nematode cuticle in the interaction with the receptors on the endospores *P. penetrans* (Mohan *et al.*, 2001).

Comparative studies of endospore surface proteins of *P. penetrans* with their relative cladoceran parasite species *P. ramosa* suggest the presence of at least one conserved surface epitope between the two (Schmidt *et al.*, 2008). Collagen-like proteins have been reported previously in *P. ramosa* and one such protein, Pcl1A, has been identified in their endospore surface protein extracts (Mouton *et al.*, 2009, McElroy *et al.*, 2011).

5.1.4 Aims and objectives

Major Aim: To characterize the surface exposed epitopes on the endospores of *P. penetrans* in comparison with *B. thuringiensis.*

Specific Objectives:

1) To investigate surface localization of the epitopes, previously identified in Chapter 4, by means of immunolocalization studies using the *Pasteuria*-specific polyclonal antibodies raised to whole endospore and to collagen-like peptides.
2) To investigate the surface localization of N-acetylglucosamine on *Pasteuria* endospores.

5.2   MATERIALS AND METHODS

5.2.1   *Bacillus* and *Pasteuria* endospores

The endospores from three *B. thuringiensis* strains and *P. penetrans* isolate Res148, as described previously (chapter 4, sections 4.2.1, 4.2.2, 4.2.4), were used in the surface characterization experiments. The endospores were used at a concentration of 10^5 spores/ml.

5.2.2   Antibodies and Pre-immune sera

The primary antibodies used for the localization of antigenic determinants on the endospores of *P. penetrans* and *B. thuringiensis* were either of the three polyclonal antibodies (PABs) Anti-PpWS, Col1981 or Col1982 and the secondary antibody was a FITC-conjugated anti-rabbit IgG as described previously (chapter 4, section 4.2.9 A). In all the experiments the corresponding pre-immune sera were used as negative controls. This was to ensure that the observed fluorescence was due to an interaction of the antibody to a specific epitope on the endospore surface, and not due to some non-targeted cross-reactivity of the other components of antiserum with the endospore surface molecules.

5.2.3   Lectin: Wheat Germ Agglutinin (WGA)

Lectins are proteins that can bind to carbohydrates. One of the commonly used lectins in studying glycoproteins is wheat germ agglutinin (WGA) which specifically binds to *N*-acetylglucosamine. Unconjugated WGA and FITC-labelled WGA, obtained from Vector Labs, Peterborough, England, were used for the detection of surface-exposed NAG residues on *Pasteuria* endospores as described below in 5.2.5 and 5.2.6.

   i.   **Unconjugated WGA:** The unconjugated WGA was obtained in a lyophilized form and was diluted to 5 mg/ml in 10 mM HEPES buffered saline, pH 8.5, 0.1 mM $CaCl_2$.

   ii. **Fluorescein-labelled WGA:** The fluorescein-labelled lectin (5 mg/ml) was used as obtained in a buffer containing 10 mM HEPES, 0.15 M NaCl, pH 7.5, 0.08% sodium azide, 0.1 mM Ca++.

5.2.4   Immuno-localization of surface epitopes - Indirect Immunofluorescence

   i.   **Attaching endospores to the slide surface:** The wells of multi-well slides were coated with 1:10 dilution of 0.1% Poly-L-lysine (P8920, Sigma Aldrich) in

deionized water and left to dry. Once dry, the wells were coated with 10 µl of 10^5 endospores/ml suspension and left to dry again at room temperature.

ii. **Blocking:** The wells of the slide were over-layered with 10 µl of PBST with 2% BSA and incubated for 1 hour at room temperature to block the non-specific binding sites. This step was one of the critical steps in reducing background fluorescence.

iii. **Incubation in Primary Antibody:** The wells of the slide were over-layered with 10 µl of a 1:50 dilution of either of the three PABs, viz., Anti-PpWS, Col1981 or Col1982 in PBST. The slides were incubated in a moist chamber for 2 hours at room temperature, followed by three subsequent gentle washes in PBST to remove the excessive unbound antibodies.

iv. **Incubation in Secondary Antibody:** The wells of the slide were over-layered with 10 µl of a 1:50 dilution of the FITC-labelled anti-rabbit IgG in PBST. The slides were incubated in a dark and moist chamber for 1 hour at room temperature, followed by three subsequent gentle washes in PBST to remove the excessive unbound antibodies.

v. **Mounting in Citifluor:** To prevent photo-bleaching of fluorescent samples, a coverslip with a drop of Citifluor mounting fluid (Citifluor Ltd., London) was mounted on each well. The coverslips were sealed to prevent drying. The prepared slides were observed under microscope as freshly as possible. When needed, the slides were stored at 4°C in moist and dark chamber.

vi. **Observation under Microscope and Imaging:** The immune-stained endospore slides were observed first under the 40X objective and then under the oil immersion objective (100X) using the B1-E bandpass emission filter of Nikon Optiphot-2 Fluorescent microscope.

The schematic representation for the indirect immunofluorescence experiment is shown in *Figure 5.1.*

5.2.5 Localization of surface exposed NAG residues - Direct fluorescence

The endospores were attached to slides coated with poly-L-lysine and the non-specific binding sites were blocked as directed above (Section 5.2.4, i-ii). The endospores in each well of the slide were overlaid with 10 µl of 5 mg/ml Fluorescein-labelled WGA and incubated in dark and moist chamber at room temperature for 2 hours. The unbound WGA was removed by gently washing the wells thrice with PBST. The slides

were mounted in Citifluor mounting fluid and observed under 40X and 100X objectives using the B1-E bandpass emission filter of Nikon Optiphot-2 Fluorescent microscope. The schematic representation for the direct fluorescence using FITC-labelled WGA is shown in *Figure 5.2.*

5.2.6   Interaction of localized epitopes with NAG residues

To identify if any of the epitopes recognized by the PABs in previous experiment (Section 5.2.5) were glycosylated with *N*-acetyl glucosamine (NAG), the endospores were pre-treated with unconjugated WGA, followed by indirect immunofluorescence using one of the three PABs as the primary antibody and FITC-labelled anti-rabbit IgG as the secondary antibody. Endospores attachment onto the Poly-L-Lysine coated slides and blocking of the non-specific binding sites was performed as discussed previously. The schematic representation for the interaction of localized epitopes with NAG residues is shown in *Figure 5.3.*

  i.   **Pre-treatment of spores with unconjugated WGA**

       The endospores adhered to the wells of the slide were coated with unconjugated WGA and incubated for 1 hour in a moist chamber. The logic was to cover all the surface exposed NAG residues. All the unbound WGA was removed by washing the wells thrice gently with PBST.

 ii.   **Indirect Immunofluorescence**

       After treating the endospore with WGA, indirect immunofluorescence was done as described previously (section 5.2.4, iii-vi).

5.2.7   Imaging and quantification of fluorescence

The fluorescent microscopic images along with their corresponding bright field images were captured at 1000X magnification using a Nikon D5200 DSLR camera attached to the microscope. The resulting fluorescent images were then analyzed using Fiji, an image processing and analysis software tool (Schindelin *et al.*, 2012). Images of five different microscopic fields from each treatment were selected and four randomly picked endospores from each image were used for analysis. To measure the fluorescence of each of the selected endospores, all the endospores of interest were selected individually using the 'selection tool' of the software. Using the 'Analyze' menu of the tool bar, the area of each endospore and integrated density were

measured. To measure the background fluorescence, in each image, four different background areas of about the same size as endospores were selected and the mean gray value for each area was measured. The mean of these mean grey values corresponds to the mean background fluorescence. The corrected total cell fluorescence (CTCF) was calculated as follows (McCloy *et al.*, 2014):

$$CTCF = Integrated\ Density$$
$$- (Area\ of\ selected\ spore \times Mean\ background\ fluorescence)$$

(Equation 5. 1)

5.2.8   Statistical Analysis of Data

The Shapiro-Wilk Test for Normality was applied on all data sets. Out of seven, three data sets were significantly normal (*Appendix IX*). The non-normal data failed to get normalized when subjected to transformation. Several studies suggest that ANOVA is robust to a few violations of normality (Glass *et al.*, 1972, Schminder *et al.*, 2010, Goheen *et al.*, 2013, Bachstetter *et al.*, 2015). Hence, the variances of the mean CTCF values obtained (as above) for each treatment of bacterial strains were compared using one-way ANOVA and Tukey's *post hoc* test to find the means that significantly differed from each other.
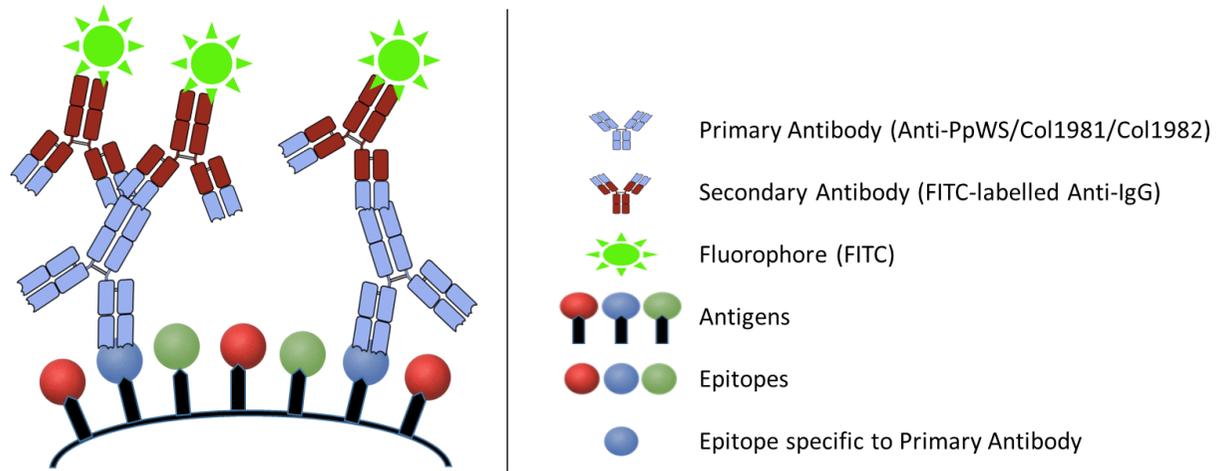
Figure 5. 1: Schematic representation of indirect Immunofluorescence for the immuno-localization of surface epitopes of endospores
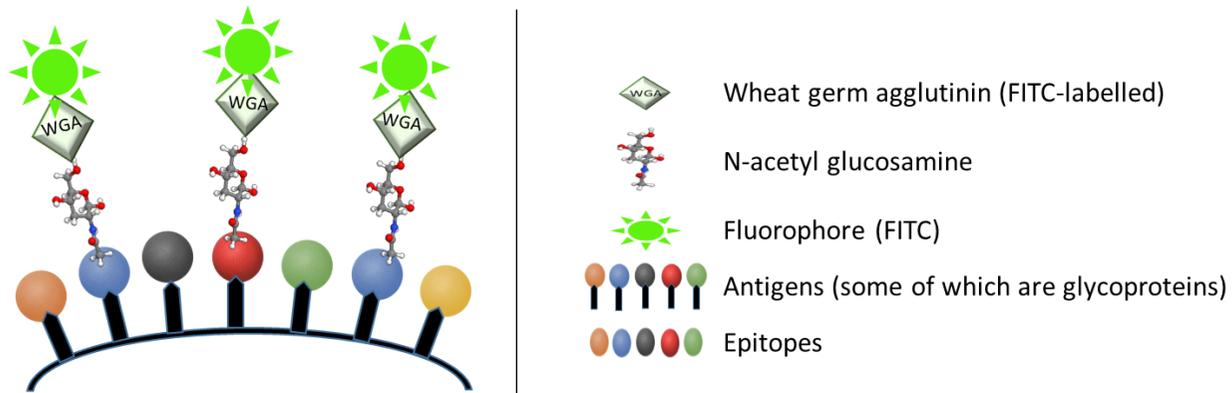
Figure 5. 2: Schematic representation of direct fluorescence for the localization of surface exposed NAG residues associated with the surface epitopes of endospores
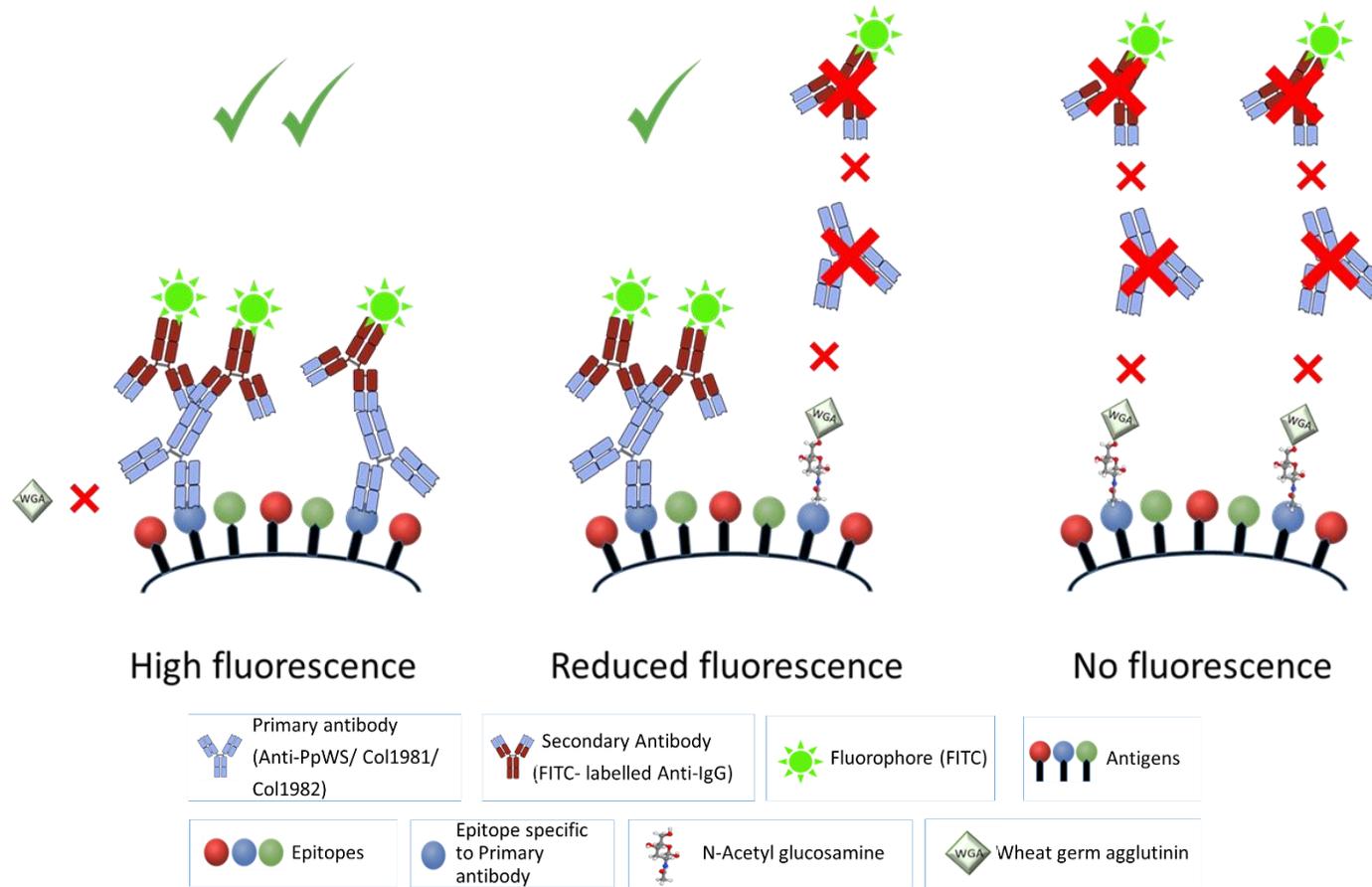
Figure 5. 3: Schematic representation of the interaction of localized epitopes with NAG residues

5.3   RESULTS

Figure 5.4 shows the bright field images of endospores of the three *Bacillus thuringiensis* strains and *Pasteuria penetrans* isolate Res148. Refer to *Appendix VIII* for all the raw data used to calculate the mean CTCF values.

5.3.1   Immuno-localization of surface epitopes – Indirect Immunofluorescence

All the three PABs used in the study, one raised to whole endospore of *P. penetrans* (Anti-PpWS) and the other two (col1981 and col 1982) raised to collagen-like synthetic peptides made from *P. penetrans* genome survey sequences, were able to recognize endospore surface epitopes of *P. penetrans* as anticipated. More interestingly, the three antibodies also recognized the surface epitopes of *B. thuringiensis* endospores.

A.   Anti-PpWS: The antibody to *Pasteuria* endospores, Anti-PpWS, recognized all the *Pasteuria* endospores. At a magnification of 1000X, it seemed that the epitopes recognized by the antibody were almost evenly distributed on individual endospores. However, on a closer observation (1000X microscopic magnification; 2X digital zoom in) it was evident that the antibody did not bind to the central body of the endospore (*Figure 5.5*). Two concentric circles of fluorescence were observed, one just outlining the central body and the other one on the periphery of the endospore. These two concentric layers recognized by Anti-PpWS were separated by a very fine line of non-fluorescent layer. Similarly, in the case of *B. thuringiensis* endospores probed with Anti-PpWS, the surface epitopes on all the three strains were recognized more at the periphery of the endospores than at the center (*Figure 5.6*). An analysis of variance (ANOVA) of the CTCF values (corrected total cell fluorescence) suggests significant variation amongst the endospores ($F_{3, 76}= 74.09; p <2e-16$). A Tukey's *post hoc* analysis showed significant differences ($p<0.005$) between the means of CTCF values of *P. penetrans* endospores and the endospores of *B. thuringiensis* (*Figure 5.11 A*). However, amongst the three strains of *B. thuringiensis*, the differences of means between *B. thuringiensis* subspecies kurstaki and *B. thuringiensis* subspecies berliner were not significant ($p=0.998$). Refer to *Appendix IX* for tables for ANOVA and Tukey's test.

B.   Col1981: The Col1981 antibody recognized the outer side of the endospores in both *P. penetrans* and *B. thuringiensis* (*Figure 5.7* and *5.8*). However, the

recognition pattern of Col1981 appeared to be varying between individual *Pasteuria* endospores. In some of the endospores, the central body could be clearly distinguished as the outer layer of the endospores was recognized more than the center, while in others the antibody was homogenously bound to the entire surface of the endospores. In the case of *B. thuringiensis* endospores, the antibody bound more to the outer layer. One-way ANOVA showed highly significant differences between the CTCF values of the endospores ($F_{3, 76}= 73.12; p <2e-16$). A comparison of all the pairs of means by Tukey's test (*Figure 5.11 B*) revealed that amongst the *B. thuringiensis* strains, the means of *B. thuringiensis* Al Hakam and *B. thuringiensis* var kurstaki were not significantly different ($p=0.955$), the differences in the means of *B. thuringiensis* Al Hakam and *B. thuringiensis* berliner were not so significant ($p=0.02$), while the means of *B. thuringiensis* subspecies kurstaki and *B. thuringiensis* subspecies berliner differed significantly ($p<0.005$). There were highly significant differences in the means of CTCF values values between *Pasteuria* endospores and all the three *B. thuringiensis* strains ($p=0.0000$). Refer to *Appendix IX* for tables for ANOVA and Tukey's test.

C. Col1982: The results of immunolocalization of collagens using Col1982 were similar to that of Col1981, showing variable recognition patterns between individual *Pasteuria* endospores and outer layer of endospores being recognized more in most of the *Pasteuria* endospores and all of the *B. thuringiensis* endospores (*Figure 5.9* and *5.10*). However, statistical analysis (*Figure 5.11 C*) showed that none of the pairs of means of CTCF values amongst the three *B. thuringiensis* strains differed significantly ($p>0.1$). There was high significant difference between the mean values of endospores of the *Pasteuria* isolate and each of the three *B. thuringiensis* strains ($p=0.0000$). Refer to *Appendix IX* for tables for ANOVA and Tukey's test.
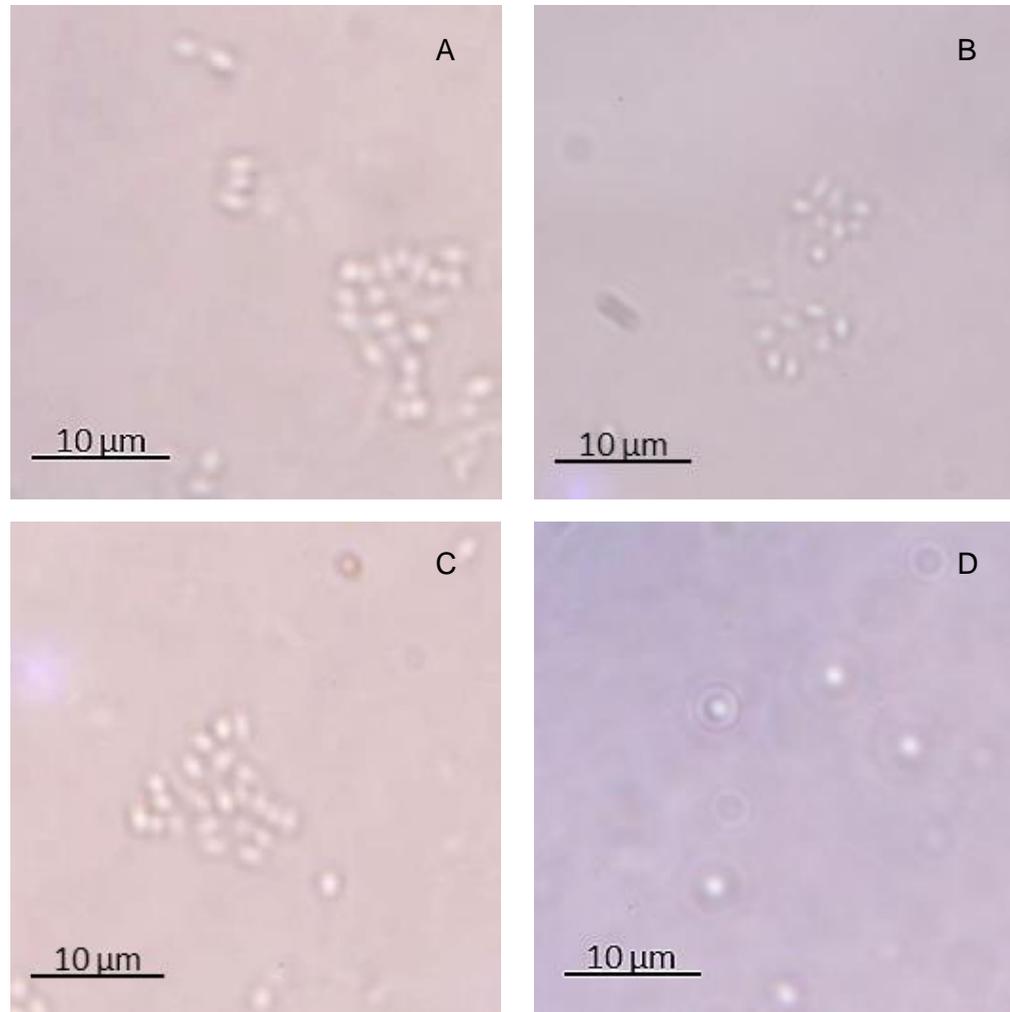
Figure 5. 4: Bright field images of endospores of (A) *Bacillus thuringiensis* strain Al Hakam, (B) *Bacillus thuringiensis* subspecies berliner, (C) *Bacillus thuringiensis* subspecies kurstaki, (D) *Pasteuria penetrans* isolate Res148. The images were captured at 1000X magnification and digitally zoomed to 3X at desired focal point. (Total magnification = 3000X). The scale bar measures 10 μm

(A): Fluorescent images

(B): Bright field images
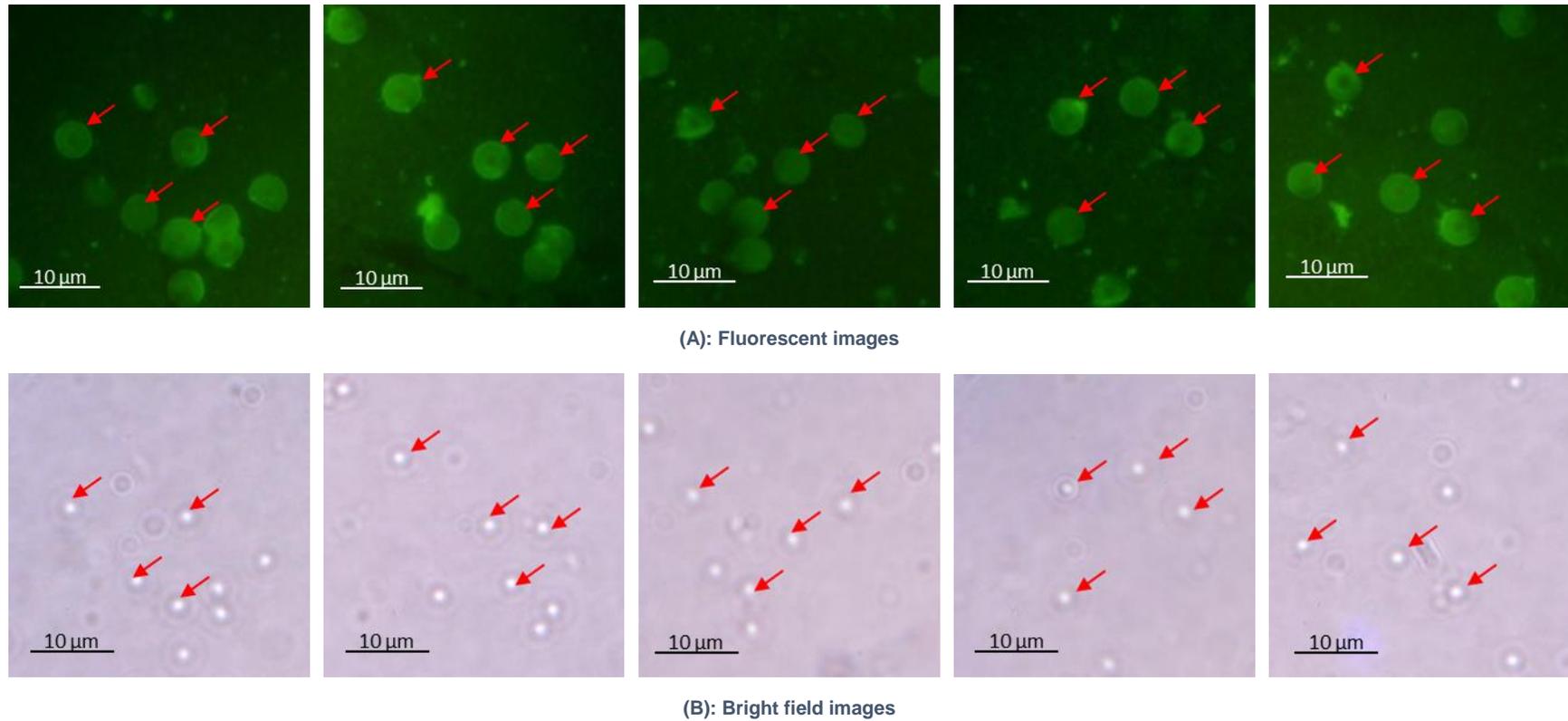
Figure 5. 5: Representative fluorescent micrographs illustrating localization of Anti-PpWS on *P. penetrans* Res148 endospores by indirect immunofluorescence using FITC labelled anti-IgG; and the corresponding bright field images. The red arrows show the endospores that were randomly picked for quantitative analysis of fluorescence. The images were captured at 1000X magnification and digitally zoomed to 2X at desired focal point. (Total magnification = 2000X). The scale bar measures 10 μm.

(A): *Bacillus thuringiensis* **strain Al Hakam**

(B): *Bacillus thuringiensis* **subspecies berliner**

(C): *Bacillus thuringiensis* **subspecies kurstaki**
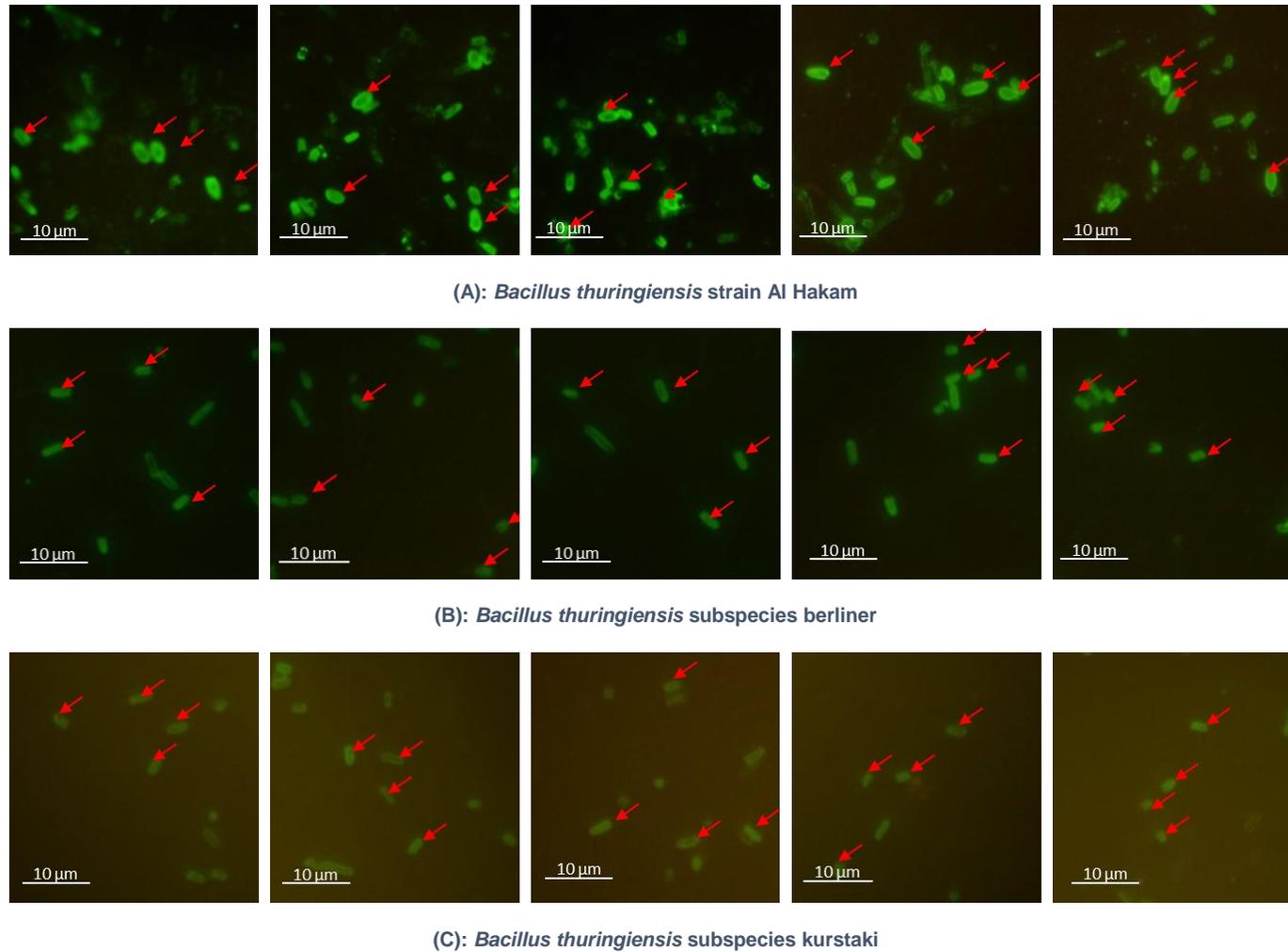
Figure 5. 6: Representative fluorescent micrographs illustrating localization of Anti-PpWS on *B. thuringiensis* endospores by indirect immunofluorescence using FITC labelled anti-IgG. Each image was clicked from a different microscopic field. The red arrows show the endospores that were randomly picked for quantitative analysis of fluorescence. The images were captured at 1000X magnification and digitally zoomed to 2X at desired focal point. (Total magnification = 2000X). The scale bar measures 10 μm.

(A): Fluorescent images

(A): Bright field images
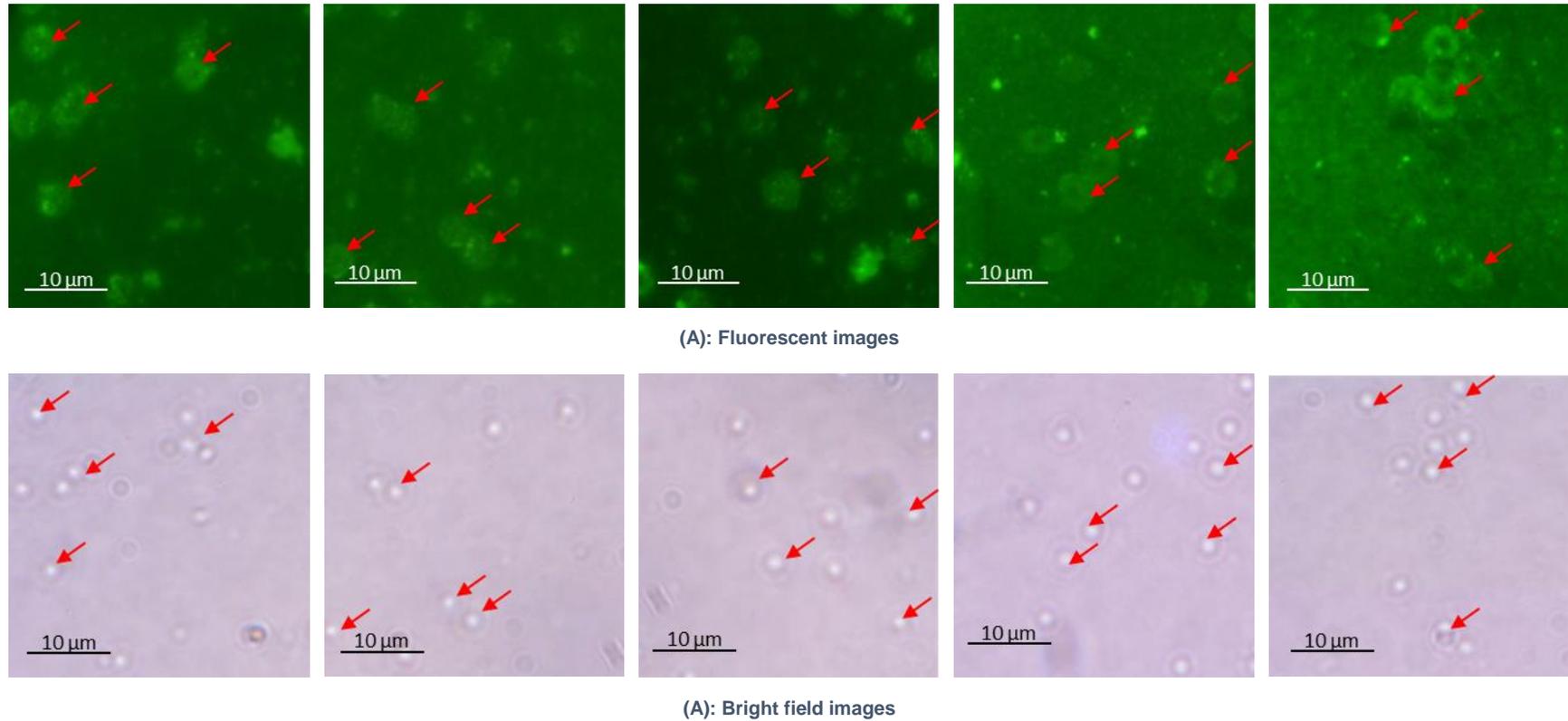
Figure 5. 7: Representative fluorescent micrographs illustrating localization of Col1981 on *P. penetrans* Res148 endospores by indirect immunofluorescence using FITC labelled anti-IgG; and the corresponding bright field images. The images were captured at 1000X magnification and digitally zoomed to 2X at desired focal point. (Total magnification = 2000X). The red arrows show the endospores that were randomly picked for quantitative analysis of fluorescence. The scale bar measures 10 µm.

(A): *Bacillus thuringiensis* **strain Al Hakam**

(B): *Bacillus thuringiensis* **subspecies berliner**

(C): *Bacillus thuringiensis* **subspecies kurstaki**
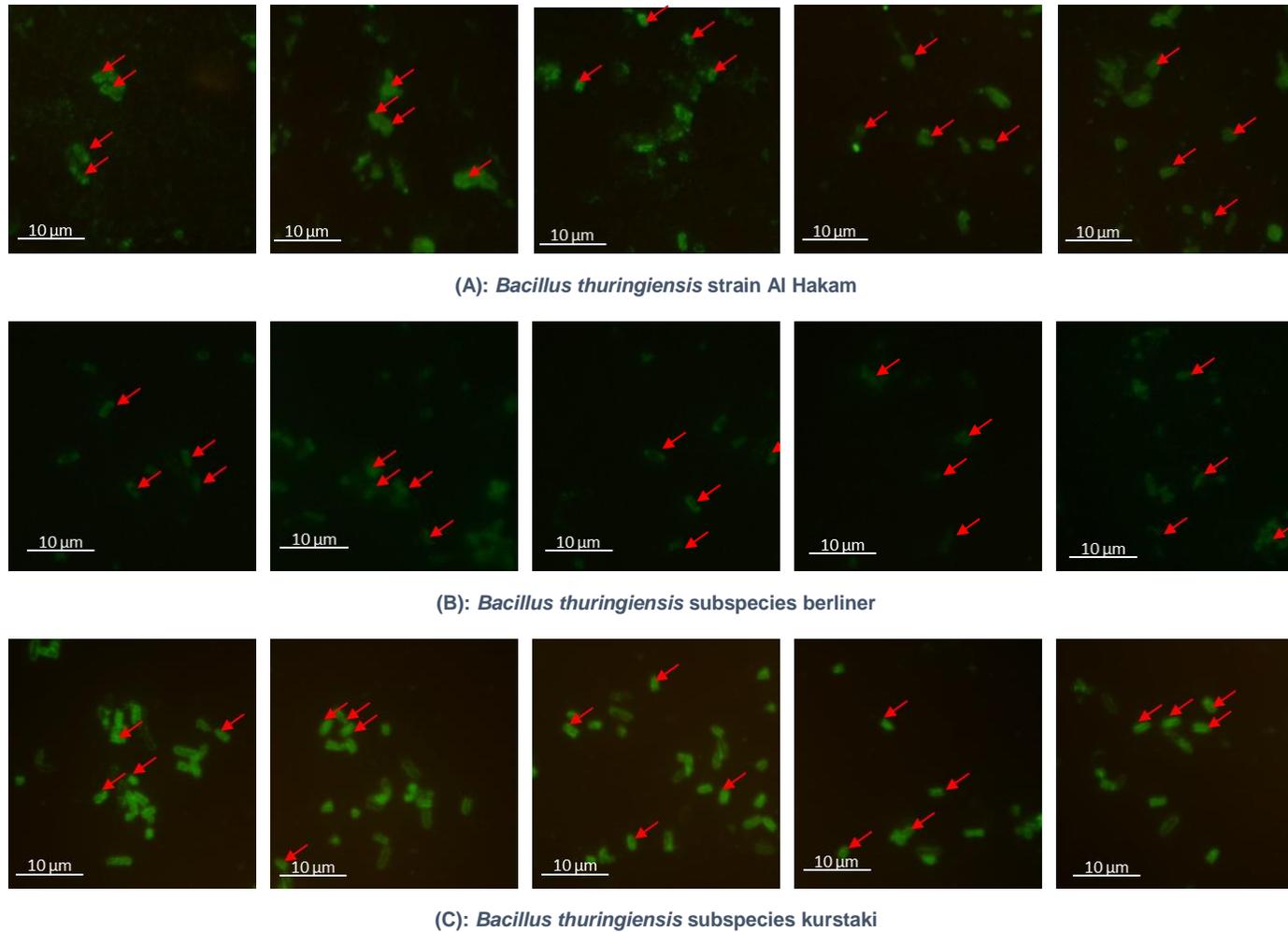
Figure 5. 8: Representative fluorescent micrographs illustrating localization of Col1981 on *B. thuringiensis* endospores by indirect immunofluorescence using FITC labelled anti-IgG. Each image was clicked from a different microscopic field. The red arrows show the endospores that were randomly picked for quantitative analysis of fluorescence. The images were captured at 1000X magnification and digitally zoomed to 2X at desired focal point. (Total magnification = 2000X). The scale bar measures 10 μm.

(A): Fluorescent images

(B): Bright field images
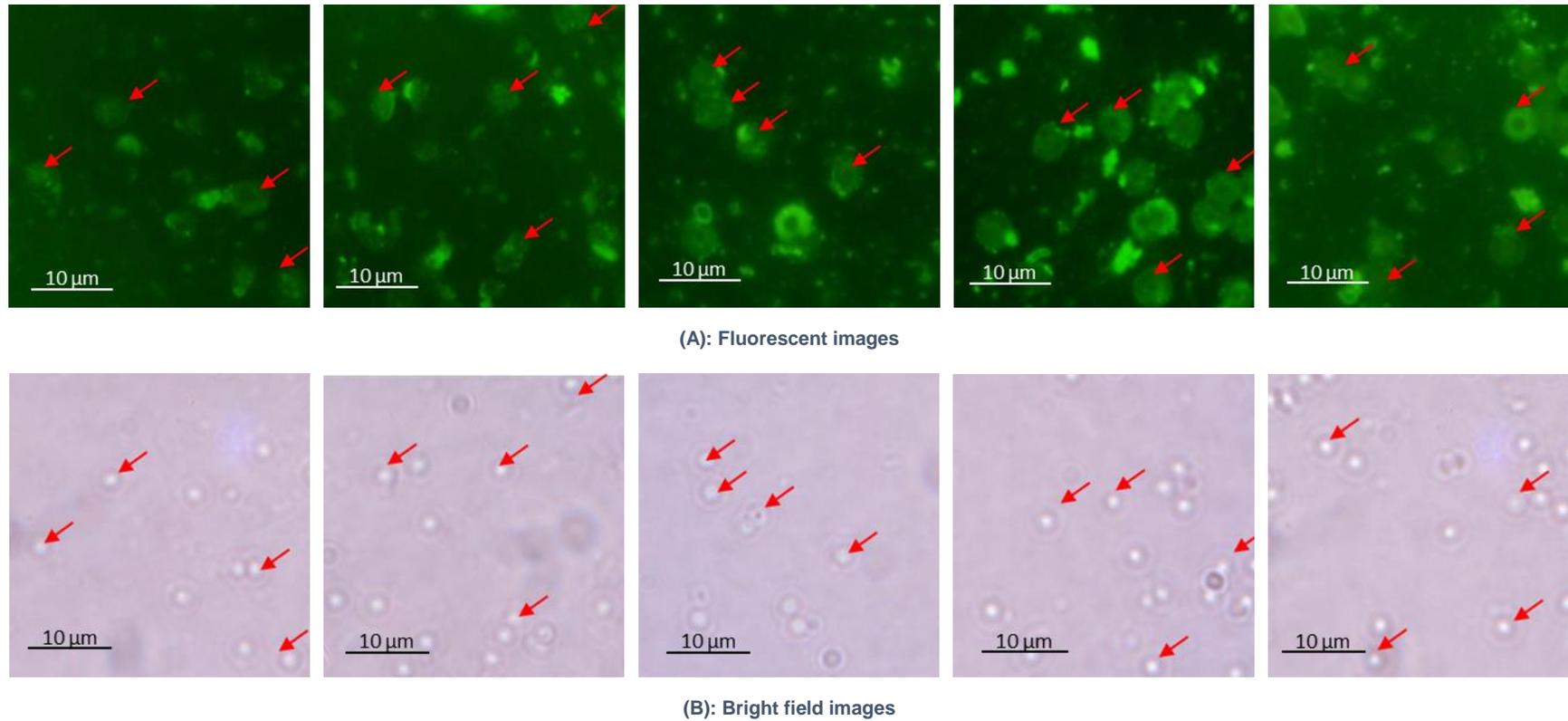
Figure 5. 9: Representative fluorescent micrographs illustrating localization of Col1982 on *P. penetrans* Res148 endospores by indirect immunofluorescence using FITC labelled anti-IgG; and the corresponding bright field images. The images were captured at 1000X magnification and digitally zoomed to 2X at desired focal point. (Total magnification = 2000X). The red arrows show the endospores that were randomly picked for quantitative analysis of fluorescence. The scale bar measures 10 µm.

(A): Bacillus thuringiensis strain Al Hakam

(B): Bacillus thuringiensis subspecies berliner

(C): Bacillus thuringiensis subspecies kurstaki
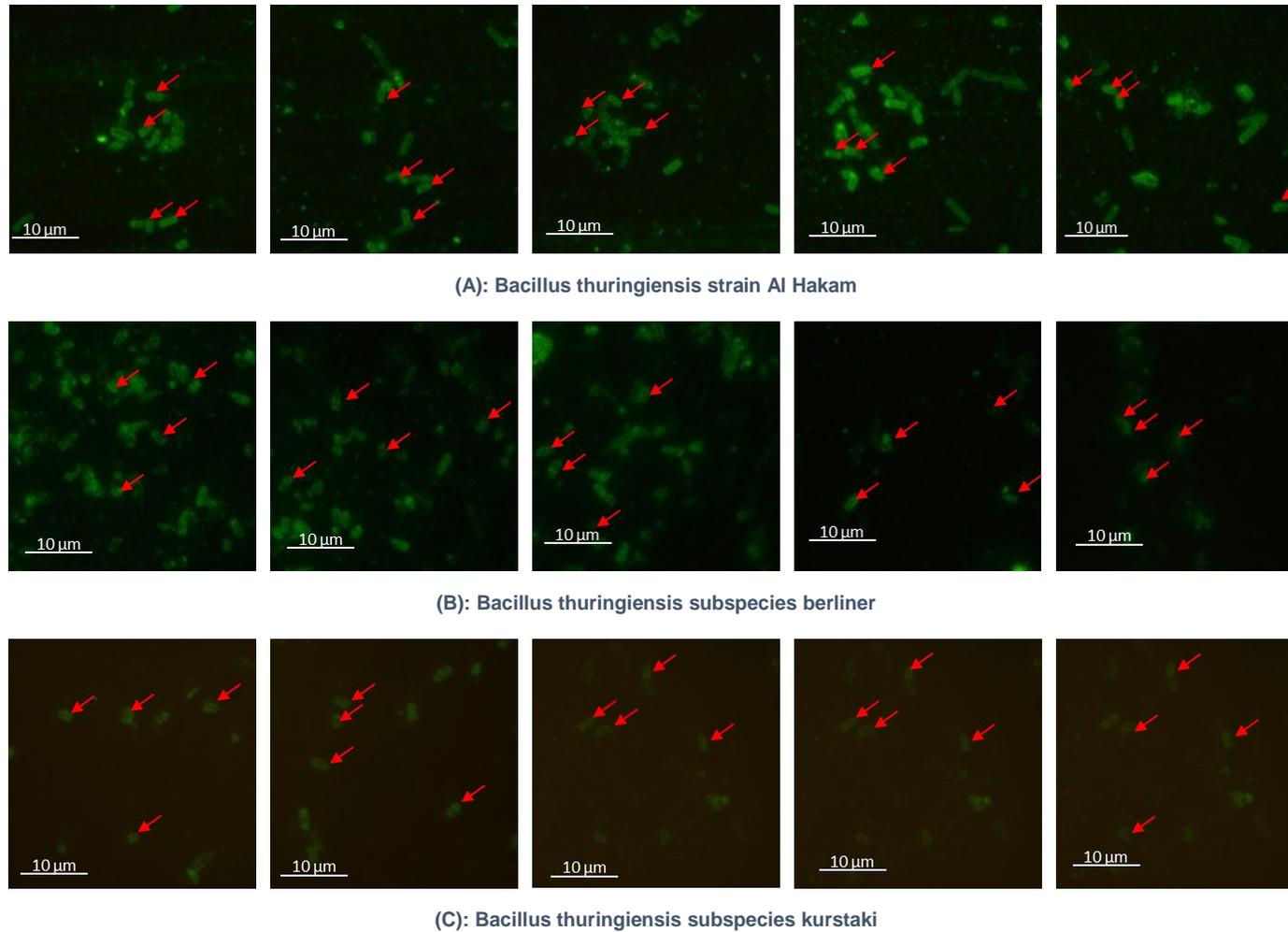
Figure 5. 10: Representative fluorescent micrographs illustrating localization of Col1982 on B. thuringiensis endospores by indirect immunofluorescence using FITC labelled anti-IgG. Each image was clicked from a different microscopic field. The images were captured at 1000X magnification and digitally zoomed to 2X at desired focal point. (Total magnification = 2000X). The red arrows show the endospores that were randomly picked for quantitative analysis of fluorescence. The scale bar measures 10 μm.

124
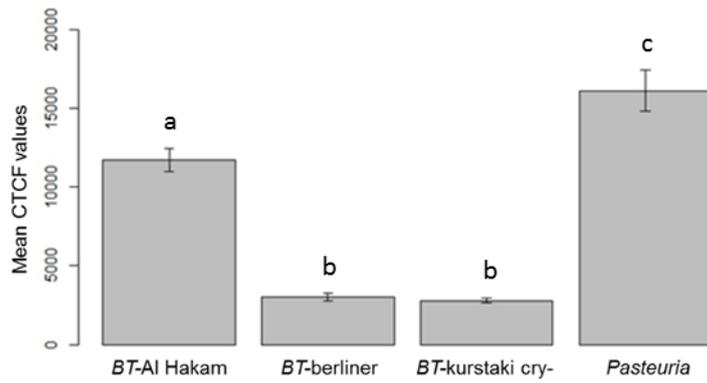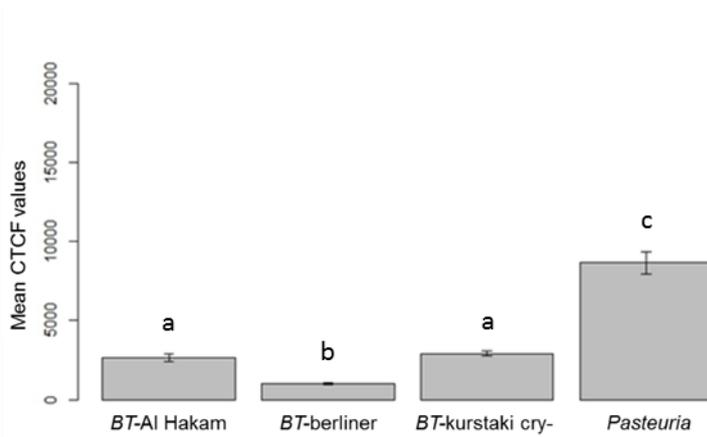
A: Anti-PpWS
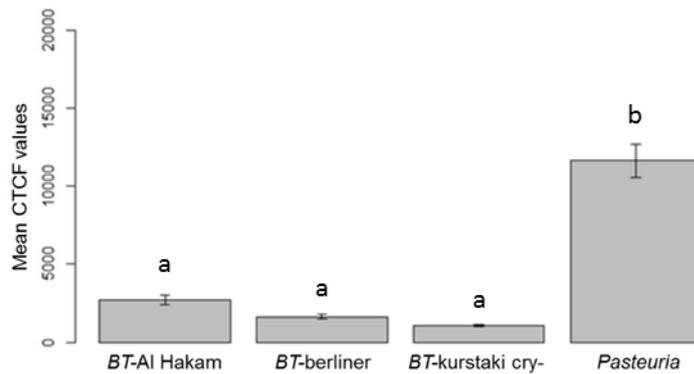
B: Col1981

C: Col1982

Figure 5. 11: Mean CTCF values of endospores of the three strains of *B. thuringiensis* and an isolate of *Pasteuria* when probed with (A) Anti-PpWS (B) Col1981 and (C) Col1982. The error bars show the standard error of the mean values (n=20). The different small letters above the bars denote significant differences between the mean CTCF values of the endospores of the four bacteria (based on Tukey's *post hoc* test). See *Appendix VIII* for calculations of Mean CTCF values and *Appendix IX* for tables for ANOVA and Tukey's test.

5.3.2   Direct fluorescence localization of surface exposed NAG residues

To further characterize the surface epitopes of *Pasteuria*, when the endospores were probed with wheat germ agglutinin (WGA), a lectin that specifically binds to *N*-acetyl glucosamine (NAG), the presence of NAG on the surface of *Pasteuria* endospores was evident by the strong fluorescence (*Figure 5.12* and *5.14*). There was a clear pattern in which WGA recognized the endospores. The lectin bound to the outer side of the endospores and not to the center. As observed in case of Anti-PpWS probing, there were two concentric circles of fluorescence observed, one outlining the central body and the other one on the periphery of the endospore. The inner fluorescent layer was much more intense than the outer layer. Thus, the results showed the presence of surface exposed NAG residues distributed in a definite pattern on the surface of the *P. penetrans.*

5.3.3   Interaction of localized epitopes with NAG residues

Figure 5.13 shows the immunofluorescent micrographs representing the recognition pattern of the three PABs in WGA-treated endospores. When the endospores of *P. penetrans* were pre-treated with unconjugated WGA followed by probing them with Anti-PpWS, there was a highly significant decrease ($p=3.09e^{-10}$) in the recognition of the antibody (*Figure 5.14*). When WGA-treated endospores were probed with Col1982, there was a marked decrease ($p=0.0289$) in the recognition of the endospores as compared to the untreated endospores. Surprisingly in case of Col1981, when pre-treated with WGA, the recognition of the antibody significantly increased ($p=0.000375$) as compared to the untreated endospores. Refer to *Appendix IX* for tables for ANOVA and Tukey's test.

(A): Fluorescent images

(B): Bright field images
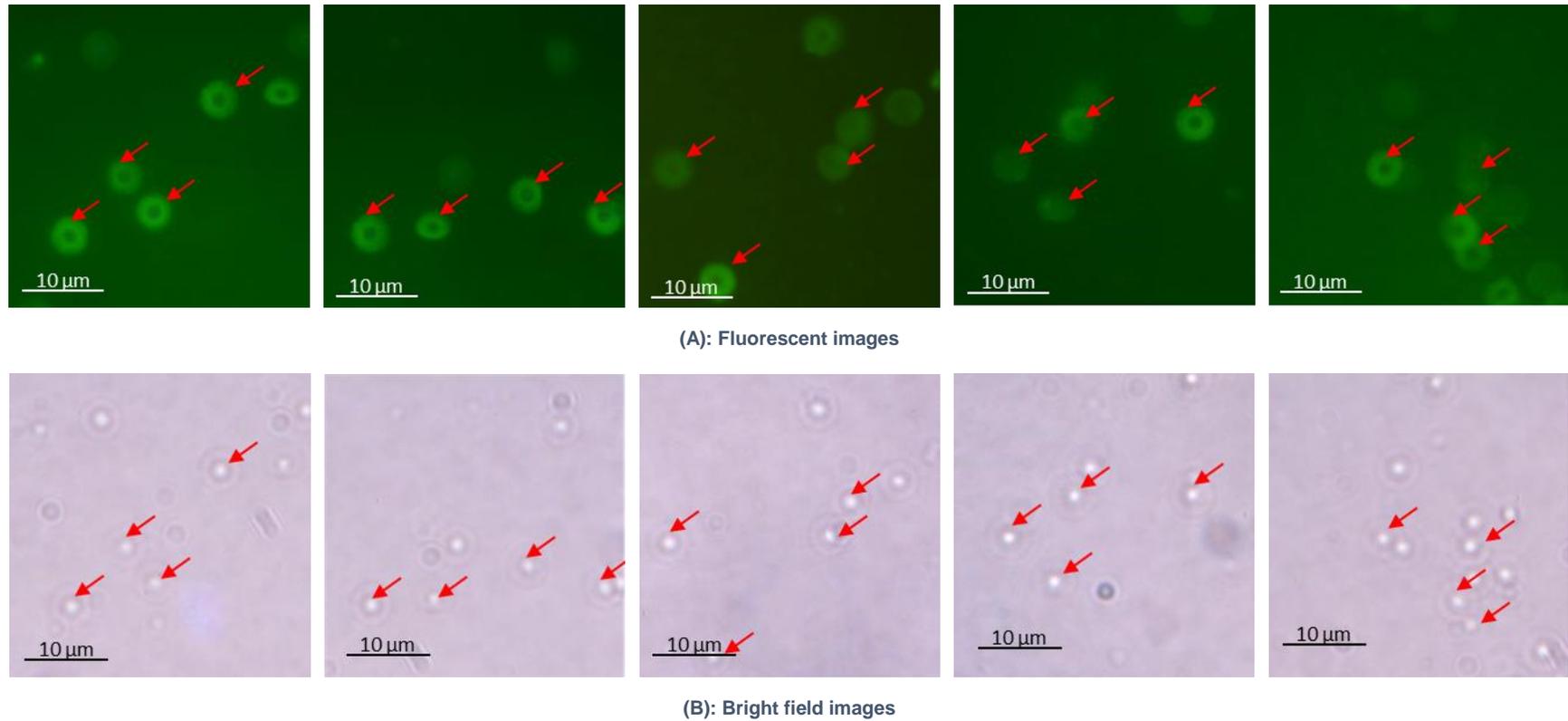
Figure 5. 12: Representative fluorescent micrographs of *P. penetrans* Res148 endospores illustrating localization of NAG residues by direct immunofluorescence using FITC labelled wheat germ agglutinin; and the corresponding bright field images. The red arrows show the endospores that were randomly picked for quantitative analysis of fluorescence. The images were captured at 1000X magnification and digitally zoomed to 2X at desired focal point. (Total magnification = 2000X). The scale bar measures 10 µm.

Untreated endospores   **WGA-treated endospores**

10 µm   10 µm
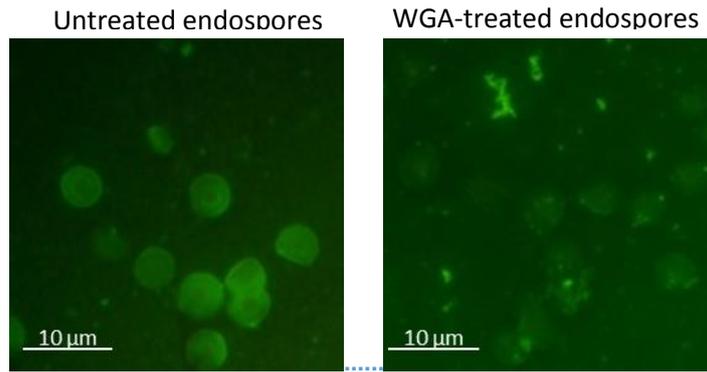
A: Reduction in the recognition of Anti-PpWS antibody by *Pasteuria* endospores when pre-treated WGA

Untreated endospores   **WGA-treated endospores**

10 µm   10 µm
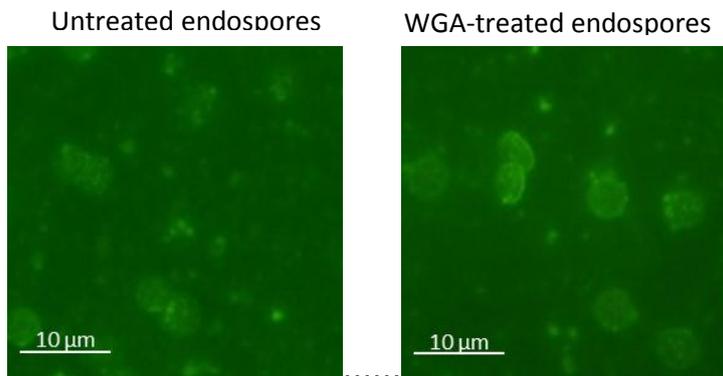
B: Increased recognition of Col1981 antibody by *Pasteuria* endospores when pre-treated WGA

Untreated endospores   **WGA-treated endospores**

10 µm   10 µm
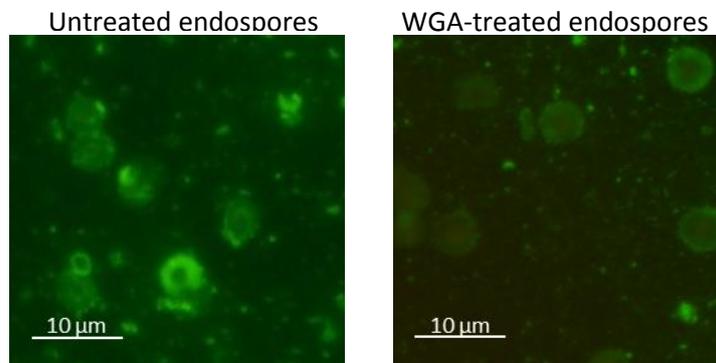
C: Reduction in the recognition of Col1982 antibody by *Pasteuria* endospores when pre-treated WGA

Figure 5. 13: Representative immunofluorescent micrographs depicting changes in the recognition pattern of the three PABs (A, Anti-PpWS; B, Col1981; C, Col1982) after treatment of *P. penetrans* Res148 endospores with WGA. The images were captured at 1000X magnification and digitally zoomed to 2X at desired focal point. (Total magnification = 2000X). The scale bar measures 10 µm.
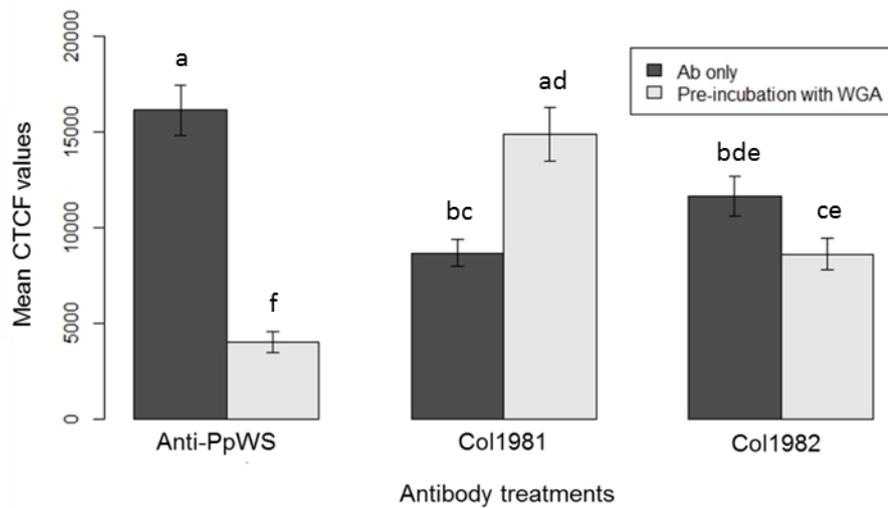
Figure 5. 14: Mean CTCF values of endospores probed with Anti-PpWS, Col1981 and Col1982 in untreated and WGA-treated *Pasteuria* endospores. The different small letters above the bars denote significant differences between the mean CTCF values of the endospores of the four bacteria (based on Tukey's *post hoc* test). See *Appendix VIII* for calculations of Mean CTCF values and *Appendix IX* for tables for ANOVA and Tukey's test.

5.4   SUMMARY OF RESULTS

As suggested by immunolocalization studies described here, there are common collagen-like epitopes on the endospore surface of *P. penetrans* and *B. thuringiensis.* It is also evident from the results that there are surface-exposed NAG residues decorated in a specific pattern on the endospore of *P. penetrans.*

5.5   DISCUSSION

Immunofluorescence microscopy is a powerful technique for immuno-localization studies and has been previously used for quantification of *Pasteuria* endospores and their surface characterization (Fould *et al.*, 2001, Costa *et al.*, 2006, Davies *et al.*, 1992, Duponnois *et al.*, 2000, Davies *et al.*, 1994, Davies and Redden, 1997, Brito *et al.*, 2003). In the current study, polyclonal antibodies were used to characterize the endospore surface of *P. penetrans* in comparison with *B. thuringiensis.* In the immunoblotting experiments discussed in the previous chapter, Anti-PpWS, the polyclonal antibody raised to *Pasteuria* endospores, had shown affinity for various endospore proteins/polypeptides of *P. penetrans* and *B. thuringiensis.* The results described here show that Anti-PpWS recognized surface localized epitopes of both *Pasteuria* and *Bacillus* endospores. Additionally, the endospore surfaces were also recognized by the *Pasteuria*-collagen specific antibodies, Col1981 and Col1982. This suggests that one or more endospore proteins (including CLP), characterized in the previous chapter by immunoblotting, could be surface associated. The affinity of the polyclonal antibodies for the surface epitopes of the endospores was measured in terms of fluorescence intensity (CTCF). It was observed, for all the three antibodies, that the CTCF values for *Pasteuria* endospores were significantly higher than that of *Bacillus* endospores. As these antibodies were either raised to *P. penetrans* endospores or to CL peptides of *P. penetrans*, the observed differences in the fluorescence intensity of *B. thuringiensis* and *Pasteuria* endospores were obvious.

Interestingly, a specific pattern of fluorescence was exhibited by most of the endospores probed with Anti-PpWS. In case of *Pasteuria* endospore, the antibody did bind to the central body and fluorescence was detected in two concentric circles, one just outlining the central body and the other one on the outer side of the exosporium, both separated by a non-fluorescent region. This implies that the epitopes recognized

by Anti-PpWS are apparently decorated on the endospore in two concentric circles surrounding the central body of the endospore. All *B. thuringiensis* endospores were also recognized by Anti-PpWS, prominently on the periphery indicating shared epitopes between *P. penetrans* and *B. thuringiensis* exsoporium. When probed with the two CL-peptide-specific antibodies to *Pasteuria*, Col1981 and Col1982, all the endospores (*Pasteuria* and *Bacillus*) showed similar fluorescence patterns; both the antibodies showed more affinity towards the epitopes present at the outer side of the endospore, seemingly on the exosporium. Based on the antibody recognition pattern displayed by the endospores, it is likely that some common collagen-like epitopes are present on the exosporium of the endospores in both *Pasteuria* and *B. thuringiensis.*

Previously, an immuno-localization based study using a monoclonal antibody raised to whole endospores of *Pasteuria* indicated that an endospore-associated surface epitope was almost uniformly localized on the periphery of *Pasteuria* endospores (Brito *et al.*, 2003). Several other studies using other monoclonal and polyclonal antibodies suggested surface epitope heterogeneity amongst individual endospores of *P. penetrans* within and between populations; this heterogeneity has been linked to their host preferences and specificities (Davies *et al.*, 1994, Davies and Redden, 1997). However, this is the first study where collagen-specific antibodies have been used to characterize *Pasteuria* endospore surface. Formerly, the localization of the BclA protein on the surface of *Bacillus* endospores has been studied using immuno-labelling with anti-BclA antibodies (Sylvestre *et al.*, 2002, Plomp and Malkin, 2008, Thompson and Stewart, 2008). BclA is a collagen-like protein localized in the exosporium of *Bacillus* spp. that has been shown to be involved in host interaction (Bozue *et al.*, 2007, Xue *et al.*, 2011, Gu *et al.*, 2012). Since, Col1981 and Col1982 recognized epitopes on the exosporium of *B. thuringiensis*, it can be speculated that the epitope recognized by the two antibodies could possibly be a part of a BclA-like protein. If this is true, it is expected that this collagen-like protein in *Pasteuria* could play major role in nematode attachment.

Probing *Pasteuria* endospores with FITC-labelled wheat germ agglutinin indicated the presence of N-acetylglucosamine (NAG) in a similar pattern to that exhibited by Anti-PpWS i.e. a patterned fluorescence as two concentric circles. Again, the central body of the endospores were not recognized indicating that the signal was indeed coming

from the exosporia of *Pasteuria* and *Bacillus* endospores. When the endospores pre-treated with WGA were probed with Anti-PpWS, there was a marked decrease in the observed recognition of the epitopes by the antibody. A possible explanation for this could be a masking effect caused by the WGA bound to NAG residues, possibly, present as glyco-conjugates of proteins/collagen-like proteins on the endospore surface. Similar results were observed for Col1981 antibody. However, steric hindrance caused by the large WGA molecule (WGA) cannot be ruled out. Unexpectedly, in case of Col1982 antibody, the observed fluorescence increased after WGA treatment i.e. more Col1982 antibodies were bound to the WGA-treated endospores. This could possibly be an indicative of non-specific binding or cross-reactivity of Col1982 with WGA. Lectins have been earlier deployed to investigate the association of carbohydrate with endospore surface associated proteins of *Pasteuria* (Persidis *et al.*, 1991, Bird *et al.*, 1989, Davies and Danks, 1993, Davies and Redden, 1997). Polypeptides resolved from *Pasteuria* endospore extracts have been found to be linked with NAG and NAM residues (Persidis *et al.*, 1991). Also, carbohydrate binding proteins (lectins) form an integral component of nematode cuticles and carbohydrate-protein interaction has been suggested as a possible mechanism of *Pasteuria*-nematode interaction (Davies and Danks, 1993, Spiegel *et al.*, 1996).

The novel approach of comparing the surface epitopes of *Pasteuria* and *Bacillus* endospores, in this study, could help in improving the understanding of endospore biochemistry and antigenic determinants of *Pasteuria*. The results from the immunolabelling and lectin binding experiments presented in this chapter and the previous chapter are an important observation to the possible roles collagens and carbohydrates play in *Pasteuria*-nematode interaction. This hypothesis, that collagens and protein-associated NAG residues play role in *Pasteuria*-nematode attachment was further tested by in-vitro attachment assays which will be discussed in the next chapter.

Chapter 6
In vitro nematode attachment assays

6.1 INTRODUCTION

Results from protein characterization and immunolocalization studies of *P. penetrans* endospores (Chapter 4 and 5) consistently suggested the presence of surface-associated collagens and/or proteins glycosylated with *N*-acetylglucosamine (NAG) on *Pasteuria* endospores. This chapter focusses on investigating the possible involvement of these characterized proteins in the attachment of *Pasteuria* endospores to the nematode cuticle.

6.1.1 Attachment- the foremost step in establishing infection

Attachment of *Pasteuria* endospore to the cuticle of the second stage juvenile of its host nematode is the first and foremost step towards establishing an infection, though all the endospores that attach do not essentially germinate (Sturhan *et al.*, 1994). A non-motile endospore of *Pasteuria* spp. remains dormant in the soil until a suitable host nematode enters its domain and the endospore attaches to the cuticle of the nematode. Once attached to the host cuticle, a germ tube emerges from the basal side of the germinating endospore that penetrates the pseudocoelom of the nematode (Sayre and Wergin, 1977, Chen and Dickson, 1998) before the bacterium starts multiplying, undergoes various developmental stages, and prevents the infected nematode females from producing eggs. Understanding the mechanisms governing the initial interactions between *Pasteuria* endospore surface and the nematode cuticle is fundamental to gain insights into the host range and specificity of *Pasteuria* spp. and successfully exploit their potential as biocontrol agents of wide range of plant parasitic nematodes.

6.1.2 Previously suggested mechanisms of endospore attachment

Several physiochemical and biochemical mechanisms have been postulated for the *Pasteuria*-nematode initial interactions. According to one postulate (Afolabi *et al.*, 1995), repulsive electrostatic forces exist between the negatively charged *Pasteuria* endospores and nematode cuticle surface, which is counteracted by an attractive hydrophobic force. Whether or not the endospores will successfully adhere to the nematode cuticle depends on which of the two forces dominate. Removal of

hydrophobic forces using KSCN has shown to reduce the ability of endospores to bind to nematode cuticle (Davies *et al.*, 1996). The endospore surface of *Pasteuria* and the cuticle surface of nematodes are both known to exhibit a diverse range of molecular entities that could act as potential antigenic determinants or receptor sites governing the attachment process. Proteins and carbohydrates, that form a variety of molecular receptors and adhesins, compose the outer surface of *Pasteuria* endospores (Persidis *et al.*, 1991, Davies *et al.*, 1994, Davies *et al.*, 1996). The involvement of the parasporal fibres of *Pasteuria* endospores in the interaction to their nematode host has been suggested (Sayre and Wergin, 1977). The nematode cuticle is known to be composed of lipids, carbohydrates and several surface-associated proteins including cuticlins, mucins, fibronectin and collagens (Bird, 1980, Davies *et al.*, 1996, Blaxter *et al.*, 1998, Davies, 2009). More recently, a velcro-like attachment model has been postulated involving collagen-like fibres on *Pasteuria* endospores and mucin-like receptor molecules on nematode cuticle (Davies, 2009).

Several studies based on *in vitro* attachment assays propose carbohydrate-protein interaction as another possible mechanism of adhesion. In one such study, J2s pre-treated with proteolytic enzymes or *N*-acetylglucosaminidase or endospores pre-treated with periodate or lectins or chitinase, resulted in reduced number of *Pasteuria* endospores attached per juvenile (Davies and Danks, 1993). Studies show that pre-treatment of endospores with the lectins concavalin A and wheat germ agglutinin reduce the attachment (Bird *et al.*, 1989, Spiegel *et al.*, 1996). These studies support a view that carbohydrates present on the endospore surface of *Pasteuria* interact with lectins (carbohydrate binding proteins) present on the nematode cuticle. Lectin-carbohydrate interactions are known to be involved in the attachment of nematophagous fungi to the host cuticle (Tunlid *et al.*, 1992, Nordbring-Hertz *et al.*, 2006).

6.1.3 Aims and objectives

Major Aim: To identify the epitopes or molecules, on the surface of *Pasteuria* endospores, that are involved in *Pasteuria*-nematode interaction.

Specific Objectives:

1) To investigate the involvement of collagens in the attachment of *Pasteuria* endospores to the cuticle of nematode juveniles.
2) To investigate the involvement of NAG in the attachment of *Pasteuria* endospores to the cuticle of nematode juveniles.

6.2 MATERIALS AND METHODS

6.2.1 Nematode juveniles and *Pasteuria* endospores

Freshly hatched second stage juveniles (J2) of root knot nematodes (RKN), *Meloidogyne incognita* were used for the attachment assays. To obtain a fresh J2 culture, the following protocol was followed. Aubergine (*Solanum melongena*) plants growing in pots for 28 days were uprooted and the roots were washed gently with tap water; egg masses of root knot nematodes were handpicked using forceps and collected in a small volume of sterile distilled water in a petri dish and incubated at room temperature (25±2°C) overnight. The freshly hatched J2s were concentrated. Excess water was carefully removed using a pipette. The nematode suspension was swirled gently to make it homogenous. The nematode count was adjusted to about 100 J2/ 10μl of suspension.

Pasteuria endospores were obtained by dissecting infected female root knot nematodes obtained from cowpea (*Vigna unguiculata*). The spores were washed repeatedly in sterile distilled water and pelleted by centrifugation (1000 g for 10 minutes). The final concentrated spores were re-suspended in phosphate buffered saline and the spore count was adjusted to 10^5 spores/ml using a haemocytometer.

6.2.2 Pre-treatment of *Pasteuria* endospores and nematode juveniles

Table 6.1 lists all the chemicals used for the treatments and summarizes their dilutions and the optimum temperature at which the treated endospores/ juveniles were incubated. For each endospore treatment, 5 μl of endospore suspension (~ 500 spores) was concentrated at the bottom of a sterile 0.2 ml PCR tube by centrifugation (1000*g* for 10 minutes at room temperature). The supernatant was removed; the pellet was re-suspended in 50 μl of the diluted chemical (antibody/enzyme/lectin) and incubated at optimum temperatures for 1 hour. For the treatment of nematode juveniles, 10 μl of J2 suspension (~ 100 J2) was concentrated at the bottom of a sterile 0.2 ml PCR tubes by centrifugation (1000g for 3 minutes at room temperature). The treatments were done as directed above for endospores. After washing, the treated J2s were re-suspended in 10 μl of PBS.

Control sets of untreated endospores and nematode juveniles were incubated in PBS at room temperature prior to the attachment assay. The treated and untreated

endospores/ juveniles were then washed thrice with PBST (0.05% v/v PBS with Tween-80) by repeated centrifugation and removing the supernatant each time. The final pellet of endospore and juveniles were re-suspended in 5 µl and 10 µl of PBS, respectively.

6.2.3 Attachment assays

To each tube containing 5 µl of treated endospore suspension (~500 spores), 10 µl of untreated J2 suspension (~100 J2) was added; conversely, untreated endospores (5 µl) were added to treated J2 suspension (10 µl). A control assay was done with 5 µl of untreated endospore suspension with 10 µl of J2 suspension. The nematode-*Pasteuria* mixture was centrifuged at 8000 rpm for 5 minutes. The mixture was left undisturbed at 4°C overnight, and attachment was observed the next morning. For each treatment, 20 nematode juveniles were randomly selected and observed under light microscope (400X magnification). The number of spores attached per juvenile was counted.

6.2.4 Statistical analysis of data

The Shapiro-Wilk Test for Normality was applied on all data sets. Most of the data sets were found to be normal (*Appendix X*). The non-normal failed to get normalized when subjected to transformation. As discussed previously in chapter 5, despite the presence of non-normal data, considering ANOVA to be highly robust, the variances between difference data sets were analysed using one-way and two-way ANOVA and Tukey's *post hoc* test to find the significant effect of different treatments on endospore attachment.

Table 6. 1: Treatments given to *Pasteuria* endospores and RKN juveniles prior to attachment assays

Treatments	Dilutions	Diluent	Incubation	Purpose of treatment
Pre-incubation of *Pasteuria* endospores with Col1981	1:50 1:500 1:1000	PBS	25±2°C for 1 h	To block surface-exposed collagens of *Pasteuria* endospores
Pre-incubation of *Pasteuria* endospores with Col1982	1:50 1:500 1:1000	PBS	25±2°C for 1 h	To block surface-exposed collagens of *Pasteuria* endospores
Pre-incubation of *Pasteuria* endospores and RKN J2 with Collagenase	750 u/ml 375 u/ml 250 u/ml	PBS-Ca^{2+}	37°C for 1 h	To digest surface-exposed collagens of *Pasteuria* endospores and on J2 cuticle
Pre-incubation of *Pasteuria* endospores and RKN J2 with N-acetylglucosaminidase (NAGase)	50 u/ml 25 u/ml	PBS	37°C for 1 h	To digest surface-exposed NAG residues of *Pasteuria* endospores and on J2 cuticle
Pre-incubation of *Pasteuria* endospores and RKN J2 with WGA	1:50 1:500 1:1000	PBS	25±2°C for 1 h	To block surface-exposed NAG residues of *Pasteuria* endospores and on J2 cuticle

6.3 RESULTS

Figure 6.1 shows light micrographs showing *Pasteuria* endospores attached to the cuticle of root knot juveniles.

6.3.1 Attachment of untreated endospores to untreated root knot juveniles

As a control to all the chemical treatments done to *Pasteuria* endospores and root knot juveniles, when a set of untreated *Pasteuria* endospores were allowed to interact with a set of freshly hatched second stage juveniles of root knot nematodes, as many as 24 endospores/J2 were observed to attach in the twenty J2s observed. On an average the number of endospores attached per J2 was 15.1±1.05. All the endospores attached in the conventional manner i.e. the concave surface of the skirt-like structure of the exosporium was in contact with the nematode cuticle. In general, the endospores didn't show preference to any region of the nematode body, they attached equally well to the head, neck and tail regions of the juveniles.

6.3.2 Effect of Collagen blocking/digestion on endospore attachment

As compared to the results obtained for the untreated control (*Section 6.3.1*), endospores pre-incubated with the two collagen-specific antibodies to *Pasteuria* showed marked reduction in their ability to attach to healthy root knot juveniles ($F_{1, 38}$= *76.72, p = 1.19e-10* at 1:50 dilution of Col1981; $F_{1, 38}$= *33.08, p = 1.25e-06* at 1:50 dilution of Col1982). A concentration dependent reduction in attachment was observed for both the antibodies (*Figure 6.2a*). Even at the least concentrations of Col1981 and Col1982 (1:1000), the number of endospores attached per J2 reduced by 8.9% and 7.3% respectively. The most effective dilution of both the antibodies (1:50) caused a marked decrease of 73.2% in Col1981 treated endospores and 45.7% in Col1982 treated endospores.

When the endospores were treated with collagenase enzyme, all the three dilutions were very effective in reducing attachment ($F_{1, 38}$= *90.53, p = 1.34e-11* at 250 units/ml collagenase), percentage reduction compared to control being 89.7%, 92.4% and 77.5% for 750 u/ml, 375 u/ml and 250 u/ml collagenase treatments respectively (*Figure 6.2b*). Interestingly, when root knot juveniles, treated with the same dilutions of collagenase, were incubated with untreated *Pasteuria* endospores, similar reduction

rates in attachment were observed (*Figure 6.2b*) ($F_{1, 38}= 118.7, p = 2.99e\text{-}13$ at 250 units/ml collagenase).

6.3.3 Effect of NAG blocking/digestion on endospore attachment

Wheat germ agglutinin (WGA) is a lectin that binds specifically to *N*-acetylglucosamine. When endospores were pre-incubated with WGA, the number of endospores attached per juvenile was highly reduced by 94.7% at 1:50 dilution, the least dilution (1:1000) showing a reduction of 63.6% (*Figure 6.3a*) ($F_{1, 38}= 174.6, p = 8.73e\text{-}16$ at 1:50 dilution of WGA; $F_{1, 38}= 61.79, p = 1.74e\text{-}09$ at 1:1000 dilution of WGA). In case of pre-treated RKN juveniles, the 1:500 dilution of WGA was the most effective (87.8%) in reducing the endospore attachment (*Figure 6.3a*) ($F_{1, 38}= 141.8, p = 2.14e\text{-}14$ at 1:500 dilution of WGA). When treated with the enzyme *N*-acetylglucosaminidase (NAGase), both endospores and J2 treatments resulted in a highly significant reduction in the number of endospores attached per J2, 81.8% and 77.8% respectively at 25 u/ml of NAGase (*Figure 6.3b*) ($F_{1, 38}= 120.4, p = 2.44e\text{-}13$ for endospores; $F_{1, 38}= 92.31, p = 1.03e\text{-}11$ for J2).

6.3.4 Unconventional attachment of *Pasteuria* endospores

An important and rather unique observation was regarding the orientation of the attached *Pasteuria* endospores on the nematode juveniles. *Pasteuria* endospores are most commonly known to attach to the nematode cuticle through the concave side of the skirt-like exosporium, thus, when *Pasteuria*-encumbered juveniles are visualized under microscope the concave side of the endospore faces towards the nematode cuticle. In the current study, while most of the untreated endospores attached to untreated J2 in conventional manner, in case of pre-treated endospores or juveniles, inverted or sideways attachment of endospores on the cuticle surface of juveniles was a common sight. It seemed that treatments of both endospores and nematode juveniles by collagenase, WGA and NAGase caused not just a reduction in the number of endospores attached per juvenile but also triggered the endospores to attach unconventionally (*Figure 6.1b, 6.1c*). When *Pasteuria* endospores were treated with collagenase, up to 80% of the total endospores attached per juvenile were either inverted or attached sideways (*Figure 6.4*). A relatively lesser percentage of unconventional attachment (10% - 20%) was observed when collagenase treated J2 were allowed to interact with untreated endospores. When pre-treated with WGA, both

Pasteuria endospores and nematode juveniles caused 80% of the endospores to attach unconventionally. When RKN juveniles were treated with the enzyme NAGase, up to 40% attachments were unconventional, while in case of NAGase treated *Pasteuria* endospores, 10% unconventional attachments were observed.
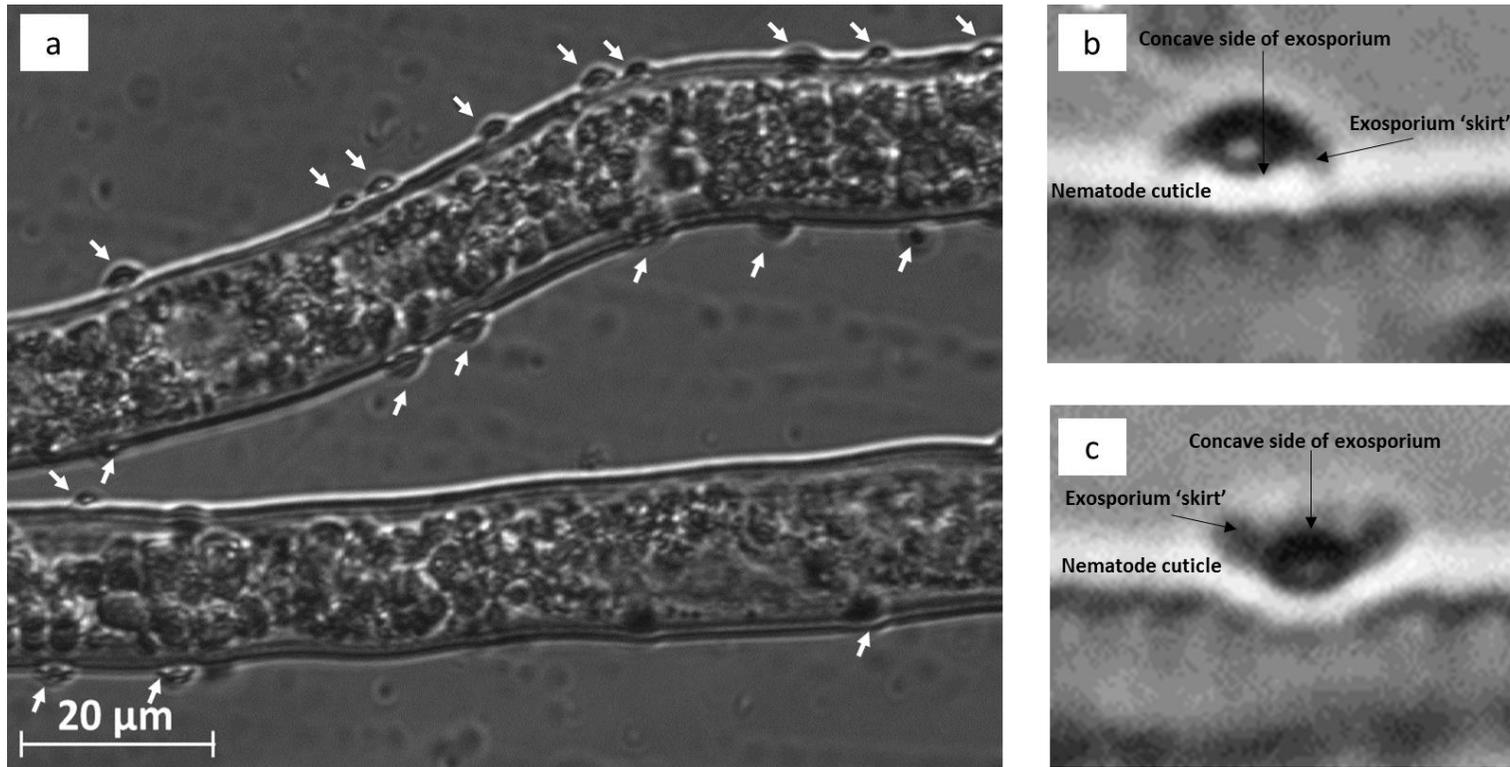
Figure 6. 1: Light micrographs showing *Pasteuria* endospores attached to the cuticle of root knot juveniles. (a) Neck regions of two healthy RKN juveniles; the white arrows show the attached endospores. (b) An endospore conventionally attached to nematode cuticle. Note that the concave side of the skirt-like exosporium is facing towards the nematode cuticle (c) Inverted attachment of an endospore. Note that in this case the endospore did not attach to the nematode cuticle through the concave side of its exosporium. An inverted attachment or a sideways attachment are both considered non-conventional.

Table 6. 2: Number of *Pasteuria* endospores attached per J2 after different treatments. To count the endospores, 20 juveniles were randomly selected from each of the 23 treatments: Spore treatment with Col1981 at 3 different concentrations; Spore treatment with Col1981 at 3 different concentrations; Spore treatment with NAGase (PasNAGase) at 2 different concentrations; J2 treatment with NAGase (J2NAGase) at 2 different concentrations; Spore treatment with WGA (PasWGA) at 3 different concentrations; J2 treatment with WGA (J2WGA) at 3 different concentrations; Spore treatment with Collagenase (PasCollagenase) at 3 different concentrations; J2 treatment with Collagenase (J2Collagenase) at 3 different concentrations; 1 untreated control (spores and J2 in PBS)

J2 (replicates)	Col1981 1/50	Col1981 1/500	Col1981 1/1000	Col1982 1/50	Col1982 1/500	Col1982 1/1000	PasNAGase 50u/ml	PasNAGase 25u/ml	J2NAGase 50u/ml	J2NAGase 25u/ml	PasWGA 1/50	PasWGA 1/500	PasWGA 1/1000	J2WGA 1/50	J2WGA 1/500	J2WGA 1/1000	PasCollagenase 250u/ml	PasCollagenase 375u/ml	PasCollagenase 750u/ml	J2Collagenase 250u/ml	J2Collagenase 375u/ml	J2Collagenase 750u/ml	PBS Control
1.	1	14	15	8	13	16	1	5	4	1	2	2	4	3	2	4	3	0	0	6	2	1	19
2.	5	5	10	5	16	11	1	5	2	3	2	4	2	1	4	6	2	1	0	4	5	2	18
3.	7	16	7	6	11	10	4	1	4	4	0	6	10	6	1	8	2	0	0	1	0	3	22
4.	8	11	11	11	18	9	1	0	4	2	0	2	6	1	3	4	9	3	2	3	2	1	11
5.	2	10	19	12	10	17	1	2	1	5	1	2	8	5	4	9	6	4	2	2	3	1	15
6.	1	8	18	12	8	13	4	5	2	0	0	3	2	7	2	5	4	1	0	3	0	2	18
7.	2	9	10	11	15	16	3	2	2	0	0	2	2	1	1	2	2	3	1	0	2	0	18
8.	1	4	9	10	13	15	1	3	5	3	4	1	5	3	0	4	8	1	0	1	1	1	11
9.	4	15	10	5	12	11	1	0	1	12	0	3	4	4	5	4	5	0	2	4	1	1	9
10.	4	11	15	6	12	16	2	3	2	2	0	1	9	6	1	2	3	0	1	6	0	0	24
11.	7	14	16	7	14	14	2	4	1	0	1	1	6	4	0	4	10	0	2	2	3	1	10
12.	9	6	18	7	9	18	2	1	5	3	0	0	7	2	2	8	2	0	4	3	3	3	18
13.	8	8	11	10	11	19	2	3	4	2	0	2	12	5	0	6	2	3	1	0	1	1	17
14.	3	9	16	10	11	11	0	3	4	5	0	2	4	3	2	3	1	2	3	5	2	1	20
15.	10	12	15	9	10	12	0	5	0	5	0	0	5	0	1	3	1	0	5	4	2	0	11
16.	1	5	19	11	12	10	4	4	4	6	1	0	5	4	0	4	0	1	2	3	0	0	19
17.	1	5	20	8	16	13	2	2	8	6	2	1	7	2	1	8	2	0	3	1	2	1	12
18.	0	8	9	5	14	16	3	0	10	2	2	1	2	4	5	5	0	2	1	3	2	2	11
19.	3	14	18	4	8	16	2	5	10	4	1	2	4	1	1	3	3	2	0	2	4	1	9
20.	4	10	9	7	9	17	1	2	5	2	0	2	6	3	2	6	3	0	2	4	0	0	10
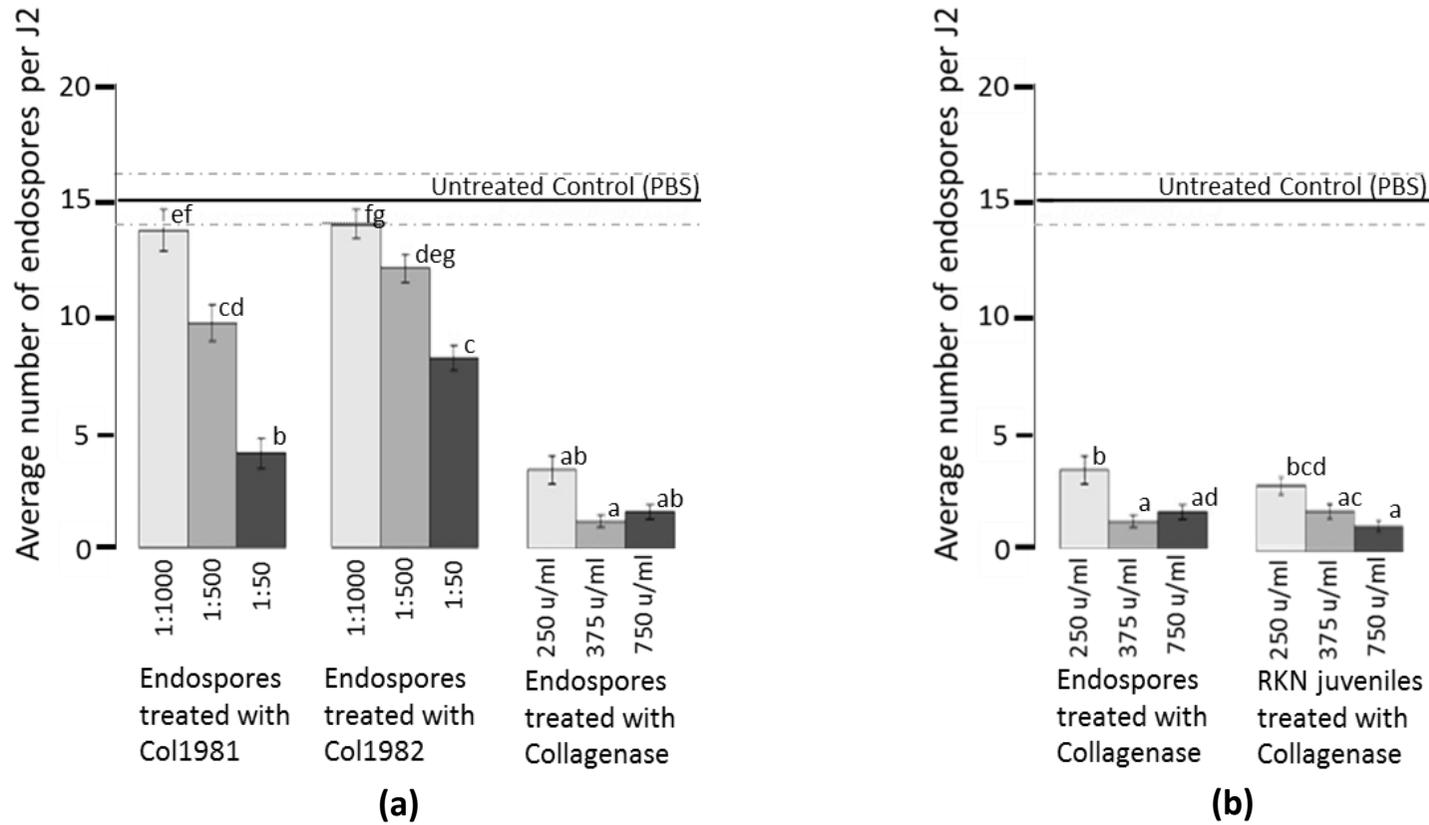
Figure 6. 2: Effect of Col1981, Col1982 and collagenase treatments on endospores attachment. On the X-axis are the chemicals used for pre-incubation of Pasteuria endospores or RKN juveniles at different working dilutions. On the Y-axis is the average number of endospores attached per J2. The error bars denote the standard error of the means. The horizontal line at y=15 represent the average number of endospores attached per juvenile in control treatment (PBS), and the two dotted lines close to the control line denote the standard error of means for the control. The different small letters above the bars denote significant differences between the mean of the number of endospores per J2 (based on Tukey's *post hoc* test). See *Appendix X* for tables for ANOVA and Tukey's test.

Figure 6. 3: Effect of (a) WGA and (b) NAGase treatments on endospores attachment. On the X-axis are the chemicals used for pre-incubation of *Pasteuria* endospores or RKN juveniles at different working dilutions. On the Y-axis is the average number of endospores attached per J2. The error bars denote the standard error of the means. The horizontal line at y=15 represent the average number of endospores attached per juvenile in control treatment (PBS), and the two dotted lines close to the control line denote the standard error of means for the control. The different small letters above the bars denote significant differences between the mean of the number of endospores per J2 (based on Tukey's *post hoc* test). See *Appendix X* for tables for ANOVA and Tukey's test.

Figure 6. 4: Percentage of conventional and unconventional type of endospore attachment observed when *Pasteuria* endospores and RKN juveniles treated with collagenase, WGA and NAGase were allowed to interact with untreated RKN juveniles and untreated *Pasteuria* endospores respectively.

6.4   SUMMARY OF RESULTS

The results presented in this chapter indicate the roles of collagen-like proteins and NAG residues in the initial *Pasteuria*-nematode interaction. It is also suggestive that there are more molecular factors involved in this complex interaction.

6.5   DISCUSSION

As established in the previous chapters (Chapter 4 and 5) collagens, glycoproteins and N-acetylglucosamine form dominant components of *Pasteuria* endospore surface; the results of the attachment assays presented in this chapter suggest their involvement in the *Pasteuria*-nematode interaction.

The putative role of collagen in attachment was tested using two approaches. One approach was by blocking the epitopes of the collagens on *Pasteuria* endospore surface using the anti-collagen antibodies to *Pasteuria* (Col1981 and Col1982); if any of the epitopes recognized by these antibodies are involved in attachment, the bound antibodies would make them unavailable to interact with the receptor nematode. Interestingly, both the antibody treatments caused reduced ability of the endospores to attach to healthy juveniles. Another approach was by enzymatically digesting the collagens. Independent collagenase treatments of endospores as well as juveniles resulted in remarkably low number of endospores attached per juvenile, thus, indicating that the collagens of both *Pasteuria* endospores and nematode cuticle may be involved in the molecular interaction. The pre-treatment of *Pasteuria* endospores has been previously reported to reduce endospore attachment (Davies and Danks, 1993).

In addition to collagens, N-acetylglucosamine (NAG) has been reported previously as a carbohydrate component of *Pasteuria* endospore, supposedly involved in *Pasteuria*-nematode interaction. The Wheat germ agglutinin and N-acetylglucosaminidase (NAGase) treatments prior to attachment assays were aimed at understanding the role of N-acetylglucosamine in attachment; the former was to block any surface-exposed NAG residues and the latter was to enzymatically digest any NAG residues. Both the treatments reduced the attachment greatly, when either of endospore or juveniles were treated. These results support the idea that a carbohydrate-protein molecular interaction makes the *Pasteuria*-nematode interaction possible (Davies and Danks, 1993). Collagens are usually glycosylated; thus, it can be anticipated that the

carbohydrate molecules on *Pasteuria* endospores thought to be involved in *Pasteuria*-nematode interaction could be the glyco-conjugates of collagens or other glycoprotein or could just be present as independent entities.

The most striking observation of the current attachment assays was the high frequency of unconventionally attached endospores when either endospore or J2 was pre-treated with collagenase or NAGase or WGA. A former study reported inverted attachment when endospores of an isolate of *Pasteuria* from pigeon pea cyst nematodes, *Heterodera* cajani, were found to cross-infect potato cyst nematodes, *Globodera pallida* (Mohan *et al.*, 2012) . It was then suggested that the inverted attachment was possibly due to a difference in the molecular receptors of the two genera of nematodes. The current attachment assays were done on a single population of a single genus of nematodes; therefore, the above explanation is not applicable here. The conventional orientation of endospores, where the concave surface of the skirt-like structure formed by parasporal fibres is orientated towards the nematode cuticle, has been previously explained by the presence of denser collagen-like fibres on the concave side than on the convex side (Davies, 2009). It is interesting to note that, in this study, all control treatments (endospores and J2s treated in PBS), involving same set of endospores and J2s as used for various treatments, exhibited conventional attachments. Thus, it is apparent that unconventional attachments are being triggered by collagenase, NAGase or WGA treatments. A possible elucidation could be the presence of different set of adhesins on either surface of the endospores. On the concave surface, when the collagens are digested by collagenase and the NAG residues are digested by NAGase or blocked by WGA, the epitopes/molecules on the convex surface could facultatively become involved in the attachment process.

Thus, the attachment of *Pasteuria* endospores to the nematode cuticle appears to be a complex molecular process which involves more than one biochemical entity, present on the *Pasteuria* endospore surface and the nematode cuticle; to establish an attachment these entities either work together or one by one as and when required. The experiments discussed here confirm at least two of these biochemical entities as collagens and N-acetylglucosamine.

Chapter 7

General Discussion

This chapter intends to evaluate and interpret the major findings of this project which involved a three-tier approach for the identification and characterization of potential adhesins on *Pasteuria* endospores that are involved in the attachment of the endospores to the cuticle of their host nematodes. First, an *in-silico* approach was used to identify putative collagen-like genes and the biochemical properties of their potential functional products were predicted. Secondly, a novel strategy was deployed wherein the endospore proteins components of *P. penetrans* were comparatively immuno-characterized with *B. thuringiensis* endospore proteins; and commonalities between the endospore surface epitopes of the two genera were identified. Lastly, *in vitro* attachment assays were performed to further understand if any of the epitopes identified in the immune-characterization studies could be involved in *Pasteuria*-nematode initial interactions. Following are the principal findings, implications and contributions of this study.

7.1   *Bacillus* spp. are suitable comparative models to study *Pasteuria* spp.

The results from the phylogenetic analyses presented in Chapter 2 of this thesis suggest *Pasteuria* spp. are members of the *Bacillus–Clostridium* clade. Out of the seven analyses performed, *Pasteuria* spp. appeared closer to *Clostridium* in four instances (16S rRNA, *groEL*, *spo0A*, Spo0A), to *Bacillus* in two instances (GroEL, *gyrB*) and to *Paenibacillus* in one instance (GyrB). Some previous studies have placed *Pasteuria* in the *Bacillus–Clostridium* clade (Anderson *et al*., 1999, Preston *et al*., 2003, Trotter and Bishop, 2003, Charles *et al*., 2005). Considering their phylogenetic closeness, both the genera can be used as comparative models for *Pasteuria* studies. There are several reasons why I chose *Bacillus* (particularly *B. thuringiensis*) over *Clostridium* for the lab-based comparative studies described in this thesis. *B. subtilis* has been a traditional and tractable model Gram positive bacterial species used for various genetical, biochemical and physiological studies. However, *B. subtilis* lacks a true exosporium. The pathogenic species *B. anthracis*, *B. cereus* and *B. thuringiensis* each possess an exosporium, much like a simplified version of the exosporium of *Pasteuria*, which is structurally more complicated but also surrounded by a hairy nap. In *Bacillus* spp., the hairy fibres of the nap are made up of a collagen-like protein

(Sylvestre *et al.*, 2002, Steichen *et al.*, 2005); similar chemical structurehas been previously suggested for the *Pasteuria* parasporal fibres (Davies and Opperman, 2006). *Bacillus* spp. are easy to culture; while *Clostridium* spp. need special anaerobic conditions and *B. anthracis* needs containment level III to work with, *B. thuringiensis* by contract are relatively easy to grow and quicker to sporulate. Additionally, *B. thuringiensis* is not known to infect humans, therefore, is safer to work with. The *in silico* characterization of putative collagens of *Pasteuria* (Chapter 3) suggested similarities between several collagens and hypothetical collagen-like proteins of the members of the *Bacillus* genus. Even more interestingly, the comparative investigations on *B. thuringiensis* and *P. penetrans* endospores (Chapter 4 and 5) revealed several biochemical and immunological similarities which will be discussed in more detail later in this chapter.

7.2 The genome of *Pasteuria* contains several extensively diverse putative collagen-like genes

Prior studies have suggested the adhesive role of collagen-like proteins in the *Pasteuria*-nematode interaction (Davies and Danks, 1993, Davies and Opperman, 2006, Mohan *et al.*, 2001, Davies, 2009). A part of this work (Chapter 3) attempts to identify putative collagens in the genome of an isolate of *P. penetrans* and their *in-silico* characterization; in reviewing the literature, no such study for *Pasteuria* spp. was found, except for the cladoceran parasite *P. ramosa* (Mouton *et al.*, 2009, McElroy *et al.*, 2011). In *Bacillus* spp., BclA is the most characterized endospores associated CLP (Collagen-Like Protein), is known to constitute the fibres of the exosporial nap in pathogenic species and is thought to play role in host cell interaction (Sylvestre *et al.*, 2003, Daubenspeck *et al.*, 2004b, Boydston *et al.*, 2005, Sylvestre *et al.*, 2005, Rety *et al.*, 2005, Tan and Turnbough, 2010b, Tan *et al.*, 2011). In this study, from a set of unpublished contigs of *P. penetrans* Res148 genome, at least 17 collagen-like sequences were predicted with the characteristic G-X-Y repeats. Computational characterization of these sequences showed their predicted functional proteins to be biochemically diverse. Seven predicted proteins were shown to possess high hydrophobic regions; out of these, three sequences had predicted membrane spanning regions (Ppcl8, Ppcl9, Ppcl19) and four had no transmembrane regions (Ppcl17, Ppcl23, Ppcl24, Ppcl30).

The comparison of the TMHMM (Transmembrane Helices; Hidden Markov Model) predictions of these putative CLPs of *Pasteuria* spp. with the ExsJ and BclA proteins of *Bacillus* spp. is worth noting. ExsJ is an exosporium associated protein of *B. cereus* but its exact localization is not known; based on the large N-terminal end of ExsJ (~17 kDa) an earlier study had suggested that ExsJ could be contributing to a significant mass of the basal layer of the exosporium (Ball *et al.*, 2008). In the current study, ExsJ was predicted to have four transmembrane domains at the C-terminal end; this was comparable with Ppcl8 of *P. penetrans* Res148 which had five predicted transmembrane domains at the C-terminal and a large N-terminal domain. The high hydrophobic C-terminal domains of Ppcl17, Ppcl23, Ppcl24 and Ppcl30 and the absence of transmembrane domains in these sequences are the two properties that make them relatively similar to BclA.

The prediction of signal peptide cleavage sites in Ppcl8, Ppcl17 and Ppcl20 indicate them to be extracellularly localized. If any of these characterized putative proteins are expressed during sporulation as endospore-associated proteins, their localization within the endospores could be predicted based on this study. In addition to the localization of these putative proteins based on the transmembrane domains and signal peptide cleavage sites, glycosylation sites were predicted in all the seventeen sequences indicating the possibility of post-translational glycosylation of their gene products. However, the presence of these CL sequences in the contigs do not give a conclusive evidence for the presence of these genes; thus, further investigations of their functional expression are recommended via experimental studies.

The comparison of the amino acid sequences of the predicted *Pasteuria* CLPs to other collagens revealed their similarities to known CLPs from different origins. Cluster analysis, based on the percentage amino acid composition of the low complexity G-X-Y repeat regions, placed putative CLPs of *Pasteuria* in four separate clusters along with CLPs from *P. ramosa* and from the members of five different bacterial genera, viz. *Bacillus*, *Clostridium*, *Paenibacillus*, *Ruminococcus*, *Protochlamydia* and two genera of giant viruses, viz. *Megavirus* and *Pithovirus.* These results show that the collagen-like proteins are widely diverse and are linked evolutionarily. Proteins with low complexity regions are prone to rapid evolution due to non-erroneous replication slippage (Zilversmit *et al.*, 2010, DePristo *et al.*, 2006, Toll-Riera *et al.*, 2011, Radó-Trilla and Albà, 2012); since, low complexity G-X-Y repeats are characteristic features

of all collagens; it can be speculated that the putative CLPs of *Pasteuria* and the CLPs with significant sequence similarities with them are evolutionary linked and they possibly evolved to serve similar biological functions.

7.3   *Pasteuria* and *Bacillus* endospores have common surface-associated epitopes

The results of the immuno-blotting and immunofluorescence microscopy examinations presented in this thesis (Chapter 4 and 5) suggest surface epitopes that are common between the endospores *P. penetrans* and *B. thuringiensis.* Probing the Western blots with Col1981 and Col1982 antibodies raised to synthetic CL peptide fragments, two polypeptides of >250 kDa and ~72 kDa were detected commonly in *P. penetrans* and *B. thuringiensis.* In *Pasteuria*, both these polypeptides seemed to be glycosylated by N-acetylglcucosamine as suggested by the results of glycoprotein staining and lectin (WGA) staining. It is interesting that two known endospore-associated glycosylated CLPs of *Bacillus* spp., BclA and ExsJ monomer, are known to migrate at apparent molecular weights of >250 kDa and ~70 kDa respectively. Six other bands (100 kDa, 139 kDa, 155 kDa, 94 kDa, 156 kDa, 103 kDa), detected by Col1981 and Col1982, in endospore extracts of different *B. thuringiensis* strains indicate the presence of other proteins that have CL epitopes common to *Pasteuria.* Of these, the ~139 kDa protein of *B. thuringiensis* strain Al Hakam was detected by WGA staining. Two bands (142 kDa and 91 kDa) detected by Col1982 were unique to *Pasteuria* protein extracts. In addition to the >250 kDa and ~72 kDa protein bands, four more bands (~224 kDa, ~176 kDa, ~126 kDa, ~83 kDa) were detected by WGA staining of *Pasteuria* extracts indicating their association with N-acetylglucosamine. The presence of ~126 kDa and ~83 kDa proteins in endospore extracts of *P. penetrans* and their glycosylation with NAG residues is consistent with a previous report of similar protein bands by Persidis *et al.*, 1991. Thus, it is evident that *Pasteuria* endospores contains at least six NAG-glycosylated proteins (>250 kDa, ~224 kDa, ~176 kDa, ~126 kDa, ~83 kDa, ~72 kDa) of which at least two are CLPs (>250 kDa and ~72 kDa) that share common CL epitopes with *B. thuringiensis.* Immunolocalization studies confirmed that one or more epitopes detected in the Western blotting experiments are associated to endospore surfaces of *P. penetrans* and *B. thuringiensis.* All the three polyclonal antibodies (Anti-PpWS, Col1981, Col1982) recognized epitopes on the surface of endospores. It is noticeable that the epitopes detected on *Pasteuria* endospores were distributed in a

peculiar fashion, localized in two concentric rings within the exosporium. A similar pattern was exhibited when *Pasteuria* endospores were probed with FITC-labelled WGA. These results indicate that the detected surface epitopes of *Pasteuria* endospores are possibly glycosylated with NAG and are decorated on the exosporium of the endospores in a specific pattern.

Whole spore antibodies (polyclonal and monoclonal) have been used previously to characterize endospore protein components and surface epitopes of *Pasteuria* (Brito *et al.*, 2003, Davies *et al.*, 1994, Davies and Redden, 1997). Lectin staining has also been used to indicate the glycoconjugates of these proteins (Persidis *et al.*, 1991, Bird *et al.*, 1989, Davies and Danks, 1993, Davies and Redden, 1997). To the best of my knowledge, this is the first time that *Bacillus* endospore proteins have been used to explore the endospore protein composition of *Pasteuria*. Additionally, this is the first study where CL-specific antibodies to *Pasteuria* have been used to characterize proteins of *Pasteuria* endospores. The study reveals common surface-associated epitopes on the endospores of *Pasteuria* and *B. thuringiensis*; it also indicates the presence of an array of collagens and proteins glycosylated with NAG residues on *Pasteuria* endospores.

7.4 Endospore collagens possibly associated with NAG play important role in the initial *Pasteuria*-nematode interaction

The reduced numbers of *Pasteuria* endospores attached per RKN juvenile *in vitro*, when either the endospores or the juveniles were pre-treated with either collagenase, WGA or NAGase, indicate the role of collagens and NAG in *Pasteuria*-nematode interaction. The results of the attachment assays complement the immuno-characterization studies discussed above which showed the presence of collagens as well as NAG residues on the surface of *Pasteuria* endospores. These results support previous research studies that have indicated the involvement of carbohydrates and proteins in the attachment of *Pasteuria* endospores to the cuticle of nematode (Bird *et al.*, 1989, Spiegel *et al.*, 1996, Davies and Danks, 1993).

Interestingly, apart from reducing the ability of endospores to attach, the pre-treatments of endospores or nematode juveniles with either collagenase or WGA or NAGase caused most of the endospores to adhere to the nematode cuticle in unconventional ways i.e. either inverted or sideways. Conventionally, *Pasteuria*

endospores attach with the concave side of their skirt-like exosporium facing towards the nematode cuticle. Davies (2009) suggested that the presence of denser parasporal fibres on the concave side than on the convex side is responsible for the conventional orientation of *Pasteuria* attachment. Inverted attachment of endospores has been previously reported by Mohan *et al.* (2012) when endospores from an Indian isolate of *Pasteuria* from *Heterodera cajani* (pigeon pea cyst nematode) infected *Globodera pallida* (potato cyst nematode) in the UK. This was an example of cross-generic attachment of *Pasteuria* endospores in two genera of nematodes located in two geographically distant locations with very different agro-climatic conditions. Differences in the molecular receptors of the two nematode genera were then suggested to be possibly governing the way endospores attach to them. In the current study, conventional, inverted and sideways attachments were all observed within a single population of root knot nematodes.

Considering these results and the results from the previous study (Mohan *et al.*, 2012) a possible explanation for the inverted attachment could be the presence of different sets of epitopes on the concave and convex surface of *Pasteuria* endospores. Under optimum conditions (natural host, no harsh chemical treatments), the proteins and their decorating carbohydrates on the concave side of the endospores may be interacting with the molecules on the nematode cuticle for a stable attachment. When the endospores are not able to attach to the nematode cuticle through the concave side of the exosporium, the role of the epitopes on the convex side of the exosporium may come into play.

7.5   The Multitype Adhesion Model

Based on the protein characterization studies and attachment assays presented in this thesis and an extensive review of published literature from the past studies, I propose a 'Multitype Adhesion Model' for the initial *Pasteuria*-nematode interaction (*Figure 7.1*). Earlier an interaction of the glycosylated collagen-like parasporal fibres of *Pasteuria* endospores with the mucin-like fibrous receptors on the nematode cuticle through a 'Velcro-like attachment' process has been suggested (Davies, 2009, Davies and Curtis, 2011). The model proposed here is a modification of the 'Velcro-like attachment' model. According to the 'Multitype Adhesion Model' proposed here, there are multiple types of adhesins present on the surface of *Pasteuria* endospores which

interact with different types of receptor molecules present on the nematode cuticle. It is implicated from the results of the attachment assays that there is a biased distribution of these adhesins on the concave and convex side of the exosporium i.e. different adhesin types are displayed on the either sides. Both collagens and NAG residues are just two parts of this multitype adhesion system; many more adhesins are expected to be involved. Taking into consideration the *in silico* data on the putative CLPs of *Pasteuria* and the Western blots showing several CLPs in *Pasteuria* endospore proteins (presented in Chapter 3 and 4 of this thesis), there may be different types of 'collagen-like adhesins' localized on different parts of the endospores. A protein-protein, a protein-carbohydrate or a carbohydrate-carbohydrate interaction or a combination of all such interactions may be acting between these endospore adhesins and the nematode cuticle receptors possibly through a velcro-like mechanism as suggested previously by Davies (2009). The *Pasteuria*-nematode interaction is complex and needs to be looked at with a broader perspective to fully understand the molecular mechanisms that govern this complex biological process.

7.6   *Pasteuria*-nematode interaction in context of the Red Queen Hypothesis

An arms race is known to exist between rapidly co-evolving species which the evolutionary biologists define as the 'Red Queen Hypothesis' (Van Valen, 1974). The hypothesis describes that an antagonistic 'evolutionary arms race' exists between a host and its pathogen; an increased host fitness triggers an increased pathogen fitness and vice versa. The purpose of the host is to prevent and survive an invasion of a potentially lethal pathogen while the pathogen's motive is to trick any survival strategy adopted by the host. If a pathogen intends to maintain its virulence over evolutionary timescale, it must constantly overcome the selection pressure and enhance its infectivity (Dawkins and Krebs, 1979). The maintenance of the 'Red Queen dynamics' has been observed in the *Daphnia–P. ramosa* system (Decaestecker *et al.*, 2007). It is likely that an evolutionary arms race exists between nematode parasitic species of *Pasteuria* and their host nematodes. This could naturally be leading to the simultaneous progressive evolution of virulence factors in *Pasteuria* and resistance in their nematode hosts. The current study demonstrates in the genome of *P. penetrans* the presence of putative genes that potentially code for biochemically diverse collagen-like proteins; also, the presence of multiple adhesins on the surface of *Pasteuria* endospores is evident. As discussed previously, collagens are proteins with

low complexity regions, and are therefore prone to rapid evolution. Thus, these collagen-like proteins could be an important source to introduce diversity within these bacteria. Possibly *Pasteuria* spp. have evolved these wide range of collagen-like proteins and other adhesin molecules to facilitate their successful invasion of the co-evolving nematode species; subsequently some of the *Pasteuria* species or populations may have evolved mechanisms to cross-infect the members of nematode genera that they were not known to infect earlier as suggested by the reports of cross-generic attachments worldwide. Considering the diversity of putative CLPs in *P. penetrans* and their sequence similarities with CLPs from some other organisms, it is likely that the genes coding for these CLPs may have been acquired via horizontal gene transfer and constitute 'pathogenicity islands' which are known for their role in 'evolution in quantum leaps' (Groisman and Ochman, 1996). The gene pool of *Pasteuria* spp. seems to have evolved to be highly diverse to ensure the biological fitness of these obligate species countering the evolving mechanisms of host resistance. Thus, *Pasteuria* spp. have potential to be developed as robust biocontrol agents targeting broader host range.
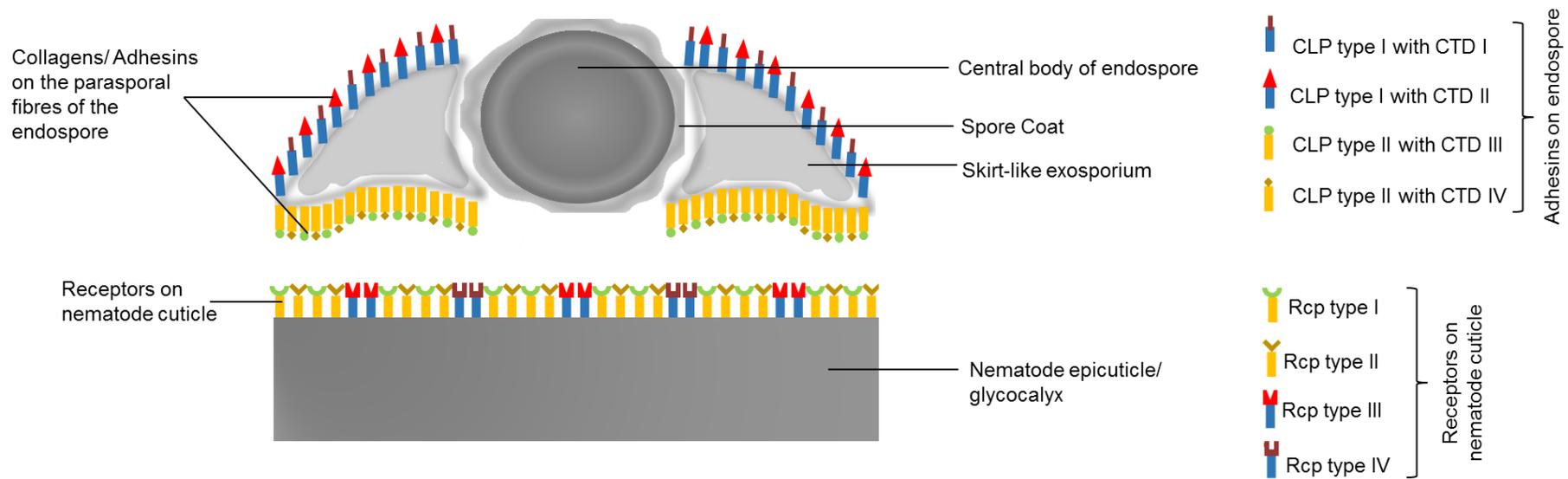
Figure 7. 1: The proposed Multitype Adhesion Model of attachment of Pasteuria endospores to nematode cuticle. This model is a modification of the Velcro-like attachment model proposed by Davies (2009) discussed earlier in Section 1.5.2, Figure 1.3 in Chapter 1 of this thesis. Here it is suggested that there is an array of adhesins present on the surface of Pasteuria endospores; these include different types of collagen-like proteins (CLP I, CLP II, CLP III, CLP IV, etc.). These collagen-like proteins (CLPs) are distributed differently on the concave and convex sides of the endospores. Each CLP has a unique C-terminal domain (CTD I, CTD II, CTD III, CTD IV, etc.) which recognizes a unique type of receptor on the nematode cuticle. Different combinations of CLPs and CTDs are possible and some of the CLPs could even share common CTDs. Additionally, the glyco-conjugates (e.g., NAG) could vary in different CLPs located on either side of the endospore.

7.7   Conclusion

This project was undertaken with the aim of identifying and characterising the virulence factors (epitopes/adhesins) on the surface of *Pasteuria* endospores. The principal implication of the findings presented in this thesis is that they enhance our understanding of the molecular and biochemical basis of the *Pasteuria*-nematode interaction. A novel and promising approach of using *B. thuringiensis* as a comparative tool to understand the endospore biology of *Pasteuria* spp. has been introduced. On one hand, *in silico* identification of several putative collagen-like genes in *P. penetrans* and the predictions of their diverse biological properties opens a new vista for further exploratory investigations about the diverse array of collagens possessed by *Pasteuria*; on the other hand, this study shows collagens and N-acetylglucosamine residues to be key factors in the attachment of *Pasteuria* endospores to the host nematode cuticle. There are many unanswered questions about the other possible factors governing host specificity of nematode parasitizing *Pasteuria*; further studies are needed to comprehended these factors. Such a comprehensive understanding of multiple virulence factors involved in the attachment process will help in exploiting the full potential of *Pasteuria* spp. for the successful management of a wide range of phytonematodes in the soil.

Bibliography

AALTEN, P., VITOUR, D., BLANVILLAIN, D., GOWEN, S. & SUTRA, L. 1998. Effect of rhizosphere fluorescent *Pseudomonas* strains on plant-parasitic nematodes *Radopholus similis* and *Meloidogyne* spp. *Letters in Applied Microbiology,* 27**,** 357-361.

ABBOTT, A. 2016. GM-crop papers spark probe. *Nature,* 529**,** 268-269.

AFOLABI, P., DAVIES, K. & O'SHEA, P. 1995. The electrostatic nature of the spore of *Pasteuria penetrans*, the bacterial parasite of root-knot nematodes. *Journal of Applied Bacteriology,* 79**,** 244-249.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of molecular biology,* 215**,** 403-410.

ANDERSON, J., PRESTON, J., DICKSON, D., HEWLETT, T., WILLIAMS, N. & MARUNIAK, J. 1999. Phylogenetic analysis of *Pasteuria penetrans* by 16S rRNA gene cloning and sequencing. *Journal of Nematology,* 31**,** 319.

ARSLAN, D., LEGENDRE, M., SELTZER, V., ABERGEL, C. & CLAVERIE, J.-M. 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences,* 108**,** 17486-17491.

ATIBALENTJA, N., NOEL, G. & DOMIER, L. 2000. Phylogenetic position of the North American isolate of *Pasteuria* that parasitizes the soybean cyst nematode, Heterodera glycines, as inferred from 16S rDNA sequence analysis. *International Journal of Systematic and Evolutionary Microbiology,* 50**,** 605-613.

ATKINSON, H. J., LILLEY, C. J. & URWIN, P. E. 2012. Strategies for transgenic nematode control in developed and developing world crops. *Current Opinion in Biotechnology,* 23**,** 251-256.

AZIZ, R. K., BARTELS, D., BEST, A. A., DEJONGH, M., DISZ, T., EDWARDS, R. A., FORMSMA, K., GERDES, S., GLASS, E. M. & KUBAL, M. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics,* 9**,** 75.

BACK, M., HAYDOCK, P. & JENKINSON, P. 2002. Disease complexes involving plant parasitic nematodes and soilborne pathogens. *Plant Pathology,* 51**,** 683-697.

BAILEY, T. L., WILLIAMS, N., MISLEH, C. & LI, W. W. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research,* 34**,** W369-W373.

BALL, D. A., TAYLOR, R., TODD, S. J., REDMOND, C., COUTURE-TOSI, E., SYLVESTRE, P., MOIR, A. & BULLOUGH, P. A. 2008. Structure of the exosporium and sublayers of spores of the *Bacillus cereus* family revealed by electron crystallography. *Molecular microbiology,* 68**,** 947-958.

BATH, J., WU, L. J., ERRINGTON, J. & WANG, J. C. 2000. Role of *Bacillus subtilis* SpoIIIE in DNA transport across the mother cell-prespore division septum. *Science,* 290**,** 995-997.

BAVYKIN, S. G., LYSOV, Y. P., ZAKHARIEV, V., KELLY, J. J., JACKMAN, J., STAHL, D. A. & CHERNI, A. 2004. Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *Journal of clinical microbiology,* 42**,** 3711-3730.

BENDTSEN, J. D., NIELSEN, H., WIDDICK, D., PALMER, T. & BRUNAK, S. 2005. Prediction of twin-arginine signal peptides. *BMC Bioinformatics,* 6**,** 167.

BENGTSSON, T. 2015. Biological control of root-knot nematodes (*Meloidogyne* spp.) by the fungus *Pochonia chlamydosporia*.

BERKELEY, M. 1855. Vibro forming excrescence on roots of cucumber plants Grad, Chorn 14: 220. *Jhorne"s Principles of Nematology. McGraw-Hill Book Co; New York***,** 312.

BIERE, A. & GOVERSE, A. 2016. Plant-Mediated Systemic Interactions Between Pathogens, Parasitic Nematodes, and Herbivores Above-and Belowground. *Annual Review of Phytopathology,* 54**,** 499-527.

BIRD, A. F. 1961. The ultrastructure and histochemistry of a nematode-induced giant cell. *The Journal of Biophysical and Biochemical Cytology,* 11**,** 701-715.

BIRD, A. F. 1962. The inducement of giant cells by *Meloidogyne javanica*. *Nematologica,* 8**,** 1-10.

BIRD, A. F. 1980. The nematode cuticle and its surface. *Nematodes as biological models,* 2**,** 213-236.

BIRD, A. F., BONIG, I. & BACIC, A. 1989. Factors affecting the adhesion of micro-organisms to the surfaces of plant-parasitic nematodes. *Parasitology,* 98**,** 155-164.

BIRD, D. M. & KOLTAI, H. 2000. Plant parasitic nematodes: habitats, hormones, and horizontally-acquired genes. *Journal of Plant Growth Regulation,* 19**,** 183-194.

BIRD, D. M., OPPERMAN, C. H. & DAVIES, K. G. 2003a. Interactions between bacteria and plant-parasitic nematodes: now and then. *International Journal for Parasitology,* 33**,** 1269-1276.

BIRD, D. M., OPPERMAN, C. H. & DAVIES, K. G. 2003b. Interactions between bacteria and plant-parasitic nematodes: now and then. *International Journal of Parasitology,* 33**,** 1269-76.

BISHOP, A. H., GOWEN, S. R., PEMBROKE, B. & TROTTER, J. R. 2007. Morphological and molecular characteristics of a new species of *Pasteuria* parasitic on *Meloidogyne ardenensis. Journal of invertebrate pathology,* 96**,** 28-33.

BLAXTER, M. L., ROBERTSON, W. M., PERRY, R. & WRIGHT, D. 1998. The cuticle. *The Physiology and Biochemistry of Free-living and Plant-parasitic Nematodes,* 25-48.

BOHLMANN, H. & SOBCZAK, M. 2014. The plant cell wall in the feeding sites of cyst nematodes. *Frontiers of Plant Science,* 5**,** 89.

BONGERS, T. & BONGERS, M. 1998. Functional diversity of nematodes. *Applied Soil Ecology,* 10**,** 239-251.

BOYDSTON, J. A., CHEN, P., STEICHEN, C. T. & TURNBOUGH, C. L., JR. 2005. Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of *Bacillus anthracis. Journal of Bacteriology,* 187**,** 5310-7.

BOYDSTON, J. A., YUE, L., KEARNEY, J. F. & TURNBOUGH, C. L. 2006. The ExsY protein is required for complete formation of the exosporium of *Bacillus anthracis. Journal of Bacteriology,* 188**,** 7440-7448.

BOZUE, J., MOODY, K. L., COTE, C. K., STILES, B. G., FRIEDLANDER, A. M., WELKOS, S. L. & HALE, M. L. 2007. *Bacillus anthracis* spores of the *bclA* mutant exhibit increased adherence to epithelial cells, fibroblasts, and endothelial cells but not to macrophages. *Infection and Immunity,* 75, 4498-4505.

BOZUE, J. A., PARTHASARATHY, N., PHILLIPS, L. R., COTE, C. K., FELLOWS, P. F., MENDELSON, I., SHAFFERMAN, A. & FRIEDLANDER, A. M. 2005. Construction of a rhamnose mutation in Bacillus anthracis affects adherence to macrophages but not virulence in guinea pigs. *Microbial Pathogenesis,* 38**,** 1-12.

BRITO, J., PRESTON, J., DICKSON, D., GIBLIN-DAVIS, R., WILLIAMS, D., ALDRICH, H. & RICE, J. 2003. Temporal formation and immunolocalization of an endospore surface epitope during *Pasteuria penetrans* sporogenesis. *Journal of Nematology,* 35**,** 278.

BRODSKY, B. & RAMSHAW, J. A. M. 1997. The collagen triple-helix Structure. *Matrix Biology,* 15**,** 545-554.

BROWN, L. N. 1933. Flooding to control root-knot nematodes. *Journal of Agricultural Research,* 47**,** 883-888.

BURBULYS, D., TRACH, K. A. & HOCH, J. A. 1991. Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell,* 64**,** 545-552.

CAILLAUD, M.-C., DUBREUIL, G., QUENTIN, M., PERFUS-BARBEOCH, L., LECOMTE, P., DE ALMEIDA ENGLER, J., ABAD, P., ROSSO, M.-N. & FAVERY, B. 2008. Root-knot nematodes manipulate plant cell functions during a compatible interaction. *Journal of Plant Physiology,* 165**,** 104-113.

CARVER, T., HARRIS, S. R., BERRIMAN, M., PARKHILL, J. & MCQUILLAN, J. A. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics,* 28**,** 464-469.

CASPI, R., BILLINGTON, R., FOERSTER, H., FULCHER, C. A., KESELER, I., KOTHARI, A., KRUMMENACKER, M., LATENDRESSE, M., MUELLER, L. A. & ONG, Q. 2016.

BioCyc: Online Resource for Genome and Metabolic Pathway Analysis. *The FASEB Journal,* 30**,** lb192-lb192.

CHARALAMBOUS, B. M., KEEN, J. N. & MCPHERSON, M. J. 1988. Collagen-like sequences stabilize homotrimers of a bacterial hydrolase. *The EMBO Journal,* 7**,** 2903-2909.

CHARLES, L., CARBONE, I., DAVIES, K. G., BIRD, D., BURKE, M., KERRY, B. R. & OPPERMAN, C. H. 2005. Phylogenetic analysis of *Pasteuria penetrans* by use of multiple genetic loci. *Journal of Bacteriology,* 187**,** 5700-8.

CHARLTON, S., MOIR, A. J. G., BAILLIE, L. & MOIR, A. 1999. Characterization of the exosporium of *Bacillus cereus*. *Journal of Applied Microbiology,* 87**,** 241-245.

CHARNECKI, J., RICE, J., DICKSON, D. & PRESTON, J. Determinants for the attachment of endospores of *Pasteuria penetrans* to phytopathogenic nematodes. Annual Meeting of the American Society for Microbiology, Atlanta, GA (Abstr.), 1998.

CHEN, Z. & DICKSON, D. 1998. Review of *Pasteuria penetrans*: Biology, ecology, and biological control potential. *Journal of Nematology,* 30**,** 313.

CHEN, Z., DICKSON, D., MCSORLEY, R., MITCHELL, D. & HEWLETT, T. 1996. Suppression of *Meloidogyne arenaria* race 1 by soil application of endospores of *Pasteuria penetrans*. *Journal of Nematology,* 28**,** 159.

CHITWOOD, D. J. & PERRY, R. N. 2009. Reproduction, physiology and biochemistry. *Root-knot Nematodes***,** 182-200.

CIANCIO, A., BONSIGNORE, R., VOVLAS, N. & LAMBERTI, F. 1994. Host records and spore morphometrics of *Pasteuria penetrans* group parasites of nematodes. *Journal of Invertebrate Pathology,* 63**,** 260-267.

COAKLEY, S. M., SCHERM, H. & CHAKRABORTY, S. 1999. Climate change and plant disease management. *Annual review of phytopathology,* 37**,** 399-426.

COBB, N. A. 1906. *Fungus maladies of the sugar cane: with notes on associated insects and nematodes*, Hawaiian gazette.

COHAN, F. M. & KOEPPEL, A. F. 2008. The origins of ecological diversity in prokaryotes. *Current Biology,* 18**,** R1024-R1034.

COHN, F. 1872. Untersuchungen über Bakterien. *Beiträge zur Biologie der Pflanzen 1, 1875 (Heft 1), J. U. Kern's Verlag (Max Müller), Breslau.***,** 127-224.

COLSON, P., DE LAMBALLERIE, X., FOURNOUS, G. & RAOULT, D. 2012. Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology,* 55**,** 321-332.

COOK, R. 1982. Cereal and Grass Hosts of Some Gramineous Cyst Nematodes1. *EPPO Bulletin,* 12**,** 399-411.

COOMBS, J. & HALL, K. E. 1998. Dictionary of Biological Control and Integrated Pest Management. *(2nd Edition) CPL Press, Newbury, UK*.

COSTA, S. R., KERRY, B. R., BARDGETT, R. D. & DAVIES, K. G. 2006. Exploitation of immunofluorescence for the quantification and characterization of small numbers of *Pasteuria endospores*. *FEMS Microbiology Ecology,* 58**,** 593-600.

CRAIG, L., PIQUE, M. E. & TAINER, J. A. 2004. Type IV pilus structure and bacterial pathogenicity. *Nature Reviews Microbiology,* 2**,** 363-378.

DAUBENSPECK, J. M., ZENG, H., CHEN, P., DONG, S., STEICHEN, C. T., KRISHNA, N. R., PRITCHARD, D. G. & TURNBOUGH, C. L. 2004a. Novel oligosaccharide side chains of the collagen-like region of BclA, the major glycoprotein of the *Bacillus anthracis* exosporium. *Journal of Biological Chemistry,* 279**,** 30945-30953.

DAUBENSPECK, J. M., ZENG, H., CHEN, P., DONG, S., STEICHEN, C. T., KRISHNA, N. R., PRITCHARD, D. G. & TURNBOUGH, C. L., JR. 2004b. Novel oligosaccharide side chains of the collagen-like region of BclA, the major glycoprotein of the *Bacillus anthracis exosporium*. *Journal of Biological Chemistry,* 279**,** 30945-53.

DAUGA, C. 2002. Evolution of the gyrB gene and the molecular phylogeny of Enterobacteriaceae: a model molecule for molecular systematic studies. *International Journal of Systematic and Evolutionary Microbiology,* 52**,** 531-547.

DAVIES, K. 2009. Understanding the interaction between an obligate hyperparasitic bacterium, *Pasteuria penetrans* and its obligate plant-parasitic nematode host, *Meloidogyne* spp. *Advances in Parasitology,* 68**,** 211-45.

DAVIES, K., AFOLABI, P. & O'SHEA, P. 1996. Adhesion of *Pasteuria penetrans* to the cuticle of root-knot nematodes (Meloidogyne spp.) inhibited by fibronectin: a study of electrostatic and hydrophobic interactions. *Parasitology,* 112**,** 553-559.

DAVIES, K., FLYNN, C., LAIRD, V. & KERRY, B. 1990. The life-cycle, population dynamics and host pecificity of a parasite of *Heterodera avenae*, similar to *Pasteuria penetrans*. *Revue de nématologie,* 13**,** 303-309.

DAVIES, K., REDDEN, M. & PEARSON, T. K. 1994. Endospore heterogeneity in *Pasteuria penetrans* related to adhesion to plant-parasitic nematodes. *Letters in Applied Microbiology,* 19**,** 370-373.

DAVIES, K., ROBINSON, M. & LAIRD, V. 1992. Proteins involved in the attachment of a hyperparasite, *Pasteuria penetrans*, to its plant-parasitic nematode host, *Meloidogyne incognita*. *Journal of Invertebrate Pathology,* 59**,** 18-23.

DAVIES, K. & SPIEGEL, Y. 2011. *Biological Control of Plant-Parasitic Nematodes:: Building Coherence between Microbial Ecology and Molecular Mechanisms*, Springer Science & Business Media.

DAVIES, K. G. 1994. *In vitro* recognition of a 190 kda putative attachment receptor from the cuticle of *Meloidogyne javanica* by *Pasteuria penetrans* spore extract. *Biocontrol Science and Technology,* 4**,** 367-374.

DAVIES, K. G. & CURTIS, R. H. 2011. Cuticle surface coat of plant-parasitic nematodes. *Annual Review of Phytopathology,* 49**,** 135-56.

DAVIES, K. G. & DANKS, C. 1993. Carbohydrate/protein interactions between the cuticle of infective juveniles of *Meloidogyne incognita* and spores of the obligate hyperparasite *Pasteuria peentrans*. *Nematologica,* 39**,** 54-64.

DAVIES, K. G., KERRY, B. R. & FLYNN, C. A. 1988. Observations on the pathogenicity of *Pasteuria penetrans*, a parasite of root-knot nematodes. *Annals of Applied Biology,* 112**,** 491-501.

DAVIES, K. G. & OPPERMAN, C. H. 2006. A potential role for collagen in the attachment of *Pasteuria penetrans* to nematode cuticle. *Multitrophic interactions in soil: IOBC/wprs Bulletin,* 29**,** 11-15.

DAVIES, K. G., OPPERMAN, C. H., MANZANILLA-LÓPEZ, R. & ROWE, J. 2011. Re-evaluation of the life-cycle of the nematode-parasitic bacterium *Pasteuria penetrans* in root-knot nematodes, *Meloidogyne* spp. *Nematology,* 13**,** 825-835.

DAVIES, K. G. & REDDEN, M. 1997. Diversity and partial characterization of putative virulence determinants in *Pasteuria penetrans*, the hyperparasitic bacterium of root-knot nematodes (*Meloidogyne* spp.). *Journal of Applied Microbiology,* 83**,** 227-35.

DAVIS, E. L. & MITCHUM, M. G. 2005. Nematodes: Sophisticated parasites of legumes. *Plant Physiology,* 137**,** 1182-1188.

DAWKINS, R. & KREBS, J. R. 1979. Arms races between and within species. *Proceedings of the Royal Society of London B: Biological Sciences,* 205**,** 489-511.

DE GUIRAN, G. & RITTER, M. 1979. Life cycle of *Meloidogyne* species and factors influencing their development. *Root-knot nematodes***,** 173-191.

DECAESTECKER, E., GABA, S., RAEYMAEKERS, J. A., STOKS, R., VAN KERCKHOVEN, L., EBERT, D. & DE MEESTER, L. 2007. Host–parasite 'Red Queen'dynamics archived in pond sediment. *Nature,* 450**,** 870-873.

DECRAEMER, W., HUNT, D. J., PERRY, R. & MOENS, M. 2006. Structure and classification. *Plant Nematology***,** 3-32.

DEGNAN, J. H. & ROSENBERG, N. A. 2006. Discordance of species trees with their most likely gene trees. *PLoS Genetics,* 2**,** e68.

DELVECCHIO, V. G., CONNOLLY, J. P., ALEFANTIS, T. G., WALZ, A., QUAN, M. A., PATRA, G., ASHTON, J. M., WHITTINGTON, J. T., CHAFIN, R. D. & LIANG, X. 2006. Proteomic profiling and identification of immunodominant spore antigens of *Bacillus*

anthracis, Bacillus cereus, and *Bacillus thuringiensis. Applied and Environmental Microbiology,* 72, 6355-6363.

DEPRISTO, M. A., ZILVERSMIT, M. M. & HARTL, D. L. 2006. On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. *Gene,* 378, 19-30.

DESROSIER, J. P. & LARA, J. C. 1984. Synthesis of the exosporium during sporulation of *Bacillus cereus. Journal of General Microbiology,* 130, 935-940.

DOBSON, H., COOPER, J., MANYANGARIRWA, W. & CHIIMBA, W. 2002. Integrated vegetable pest management: safe and sustainable protection of small-scale brassicas and tomatoes. *Integrated vegetable pest management: safe and sustainable protection of small-scale brassicas and tomatoes.*

DRIKS, A. 2002. Maximum shields: the assembly and function of the bacterial spore coat. *Trends in microbiology,* 10, 251-254.

DUAN, Y., CASTRO, H., HEWLETT, T., WHITE, J. & OGRAM, A. 2003. Detection and characterization of *Pasteuria* 16S rRNA gene sequences from nematodes and soils. *International Journal of Systematic and Evolutionary Microbiology,* 53, 105-112.

DUDDINGTON, C. 1951. Further records of British predacious fungi. II. *Transactions of the British Mycological Society,* 34, 194-209.

DUKES, J. S., PONTIUS, J., ORWIG, D., GARNAS, J. R., RODGERS, V. L., BRAZEE, N., COOKE, B., THEOHARIDES, K. A., STANGE, E. E. & HARRINGTON, R. 2009. Responses of insect pests, pathogens, and invasive plant species to climate change in the forests of northeastern North America: What can we predict? This article is one of a selection of papers from NE Forests 2100: A Synthesis of Climate Change Impacts on Forests of the Northeastern US and Eastern Canada. *Canadian Journal of Forest Research,* 39, 231-248.

DUPONNOIS, R., FARGETTE, M., FOULD, S., THIOULOUSE, J. & DAVIES, K. G. 2000. Diversity of the bacterial hyperparasite *Pasteuria penetrans* in relation to root-knot nematodes (*Meloidogyne* spp.) control on *Acacia holosericea. Nematology,* 2, 435-442.

DUTTA, C. & PAN, A. 2002. Horizontal gene transfer and bacterial diversity. *Journal of Biosciences,* 27, 27-33.

DUTTA, T. K., PAPOLU, P. K., BANAKAR, P., CHOUDHARY, D., SIROHI, A. & RAO, U. 2015. Tomato transgenic plants expressing hairpin construct of a nematode protease gene conferred enhanced resistance to root-knot nematodes. *Frontiers in microbiology,* 6.

EARDLY, B., WANG, F.-S. & VAN BERKUM, P. 1996. Corresponding 16S rRNA gene segments in Rhizobiaceae and *Aeromonas* yield discordant phylogenies. *Current Issues in Symbiotic Nitrogen Fixation.* Springer.

EBERLEIN, C., MOOS, J., PAULSEN, H. M. & WESTPHAL, A. 2013. Soil suppression against the cereal cyst nematode *Heterodera avenae. Berichte aus dem Julius Kühn-Institut,* 30.

EBERT, D., RAINEY, P., EMBLEY, T. M. & SCHOLZ, D. 1996. Development, life cycle, ultrastructure and phylogenetic position of *Pasteuria ramosa* Metchnikoff 1888: rediscovery of an obligate endoparasite of *Daphnia magna* Straus. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 351, 1689-1701.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* 32, 1792-1797.

EISENBACK, J. D. & TRIANTAPHYLLOU, H. H. 1991. Root-knot nematodes: *Meloidogyne* species and races. *Manual of agricultural nematology,* 191-274.

EL-NAGDI, W. & YOUSSEF, M. 2004. Soaking faba bean seed in some bio-agents as prophylactic treatment for controlling *Meloidogyne incognita* root-knot nematode infection. *Journal of Pest Science,* 77, 75-78.

ELKINS, C., HAALAND, R., RODRIGUEZ-KABANA, R. & HOVELAND, C. 1979. Plant-parasitic nematode effects on water use and nutrient uptake of a small-and a large-rooted tall fescue genotype. *Agronomy Journal,* 71, 497-500.

ENDO, B. Y. 1971. Nematode-Induced Syncytia (Giant Cells). Host–Parasite Relationships of Heteroderidae. *In:* ROHDE, B. M. Z. F. M. A. (ed.) *Cytogenetics, Host–Parasite Interactions, and Physiology.* Academic Press.

ENGEL, J. & BÄCHINGER, H. 1999. Collagen-like sequences in phages and bacteria. *Proceedings of the Indian Academy of Sciences - Chemical Sciences,* 111**,** 81-86.

ERICKSON, P. R. & HERZBERG, M. 1990. Purification and partial characterization of a 65-kDa platelet aggregation-associated protein antigen from the surface of *Streptococcus sanguis. Journal of Biological Chemistry,* 265**,** 14080-14087.

ERRINGTON, J. 2003. Regulation of endospore formation in *Bacillus subtilis. Nature Reviews Microbiology,* 1**,** 117-126.

FAIRHEAD, H., SETLOW, B. & SETLOW, P. 1993. Prevention of DNA damage in spores and in vitro by small, acid-soluble proteins from *Bacillus* species. *Journal of Bacteriology,* 175**,** 1367-1374.

FAO, I. A. W. 2015. The State of Food Insecurity in the World 2015. *Meeting the 2015 international hunger targets: taking stock of uneven progress.* FAO, Rome.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution***,** 783-791.

FILÉE, J., SIGUIER, P. & CHANDLER, M. 2007. I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *TRENDS in Genetics,* 23**,** 10-15.

FLUGGE, C. 1886. Die Mikroorganismen. *2nd edn. Leipzig: Verlag F.C.W. Vogel.*

FOSTER, S. & JOHNSTONE, K. 1990. Pulling the trigger: the mechanism of bacterial spore germination. *Molecular microbiology,* 4**,** 137-141.

FOULD, S., DIENG, A., DAVIES, K., NORMAND, P. & MATEILLE, T. 2001. Immunological quantification of the nematode parasitic bacterium *Pasteuria penetrans* in soil. *FEMS Microbiology Ecology,* 37**,** 187-195.

FOWLER, S. J., ‡, JOSE, S., ‡, ZHANG, X., DEUTZMANN, R., SARRAS, M. P., BOOT-HANDFORD, R. P. & ‡ 2000. Characterization of Hydra Type IV Collagen: TYPE IV Collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose. *Journal of Biological Chemistry,* 275**,** 39589-39599.

FOX, A., BLACK, G. E., FOX, K. & ROSTOVTSEVA, S. 1993. Determination of carbohydrate profiles of *Bacillus anthracis* and *Bacillus cereus* including identification of O-methyl methylpentoses by using gas chromatography-mass spectrometry. *Journal of Clinical Microbiology,* 31**,** 887-894.

FRANCL, L. & WHEELER, T. 1993. Interaction of plant-parasitic nematodes with wilt-inducing fungi. *Nematode interactions.* Springer.

FRECKMAN, D. W. & CASWELL, E. P. 1985. The ecology of nematodes in agroecosystems. *Annual Review of Phytopathology,* 23**,** 275-296.

FRITZE, D. 2004. Taxonomy of the genus *Bacillus* and related genera: the aerobic endospore-forming bacteria. *Phytopathology,* 94**,** 1245-1248.

GAI, Y., LIU, G. & TAN, H. 2006. Identification and characterization of a germination operon from *Bacillus thuringiensis. Antonie van Leeuwenhoek,* 89**,** 251-259.

GAMBI, C. & DANOVARO, R. 2016. Biodiversity and life strategies of deep-sea meiofauna and nematode assemblages in the Whittard Canyon (Celtic margin, NE Atlantic Ocean). *Deep Sea Research Part I: Oceanographic Research Papers,* 108**,** 13-22.

GARCIA-PATRONE, M. & TANDECARZ, J. S. 1995. A glycoprotein multimer from *Bacillus thuringiensis* sporangia: dissociation into subunits and sugar composition. *Molecular and Celular Biochemistry,* 145**,** 29-37.

GARCIA, L. S. & BRUCKNER, D. A. 1997. *Diagnostic medical parasitology*, American Society for Microbiology (ASM).

GARCIA, R. & MITCHELL, D. 1975. Synergistic interactions of *Pythium myriotylum* with *Fusarium solani* and *Meloidogyne arenaria* in pod rot of peanut. *Phytopathology,* 65**,** 832-833.

GARRITY, G. M., BELL, J. A. & LILBURN, T. 2005. The revised road map to the manual. *Bergey's Manual® of Systematic Bacteriology.* Springer.

GASCUEL, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution,* 14**,** 685-695.

GERHARDT, P. & RIBI, E. 1964. Ultrastructure of the exosporium enveloping spores of *Bacillus cereus. Journal of Bacteriology,* 88**,** 1774.

GIBLIN-DAVIS, R., WILLIAMS, D., WERGIN, W., DICKSON, D., HEWLETT, T., BEKAL, S. & BECKER, J. 2001. Ultrastructure and development of *Pasteuria* sp.(S-1 strain), an obligate endoparasite of *Belonolaimus longicaudatus* (Nemata: Tylenchida). *Journal of Nematology,* 33**,** 227.

GIBLIN-DAVIS, R. M., NONG, G., PRESTON, J. F., WILLIAMS, D. S., CENTER, B. J., BRITO, J. A. & DICKSON, D. W. 2011. 'Candidatus *Pasteuria aldrichii'*, an obligate endoparasite of the bacterivorous nematode *Bursilla. International Journal of Systematic and Evolutionary Microbiology,* 61**,** 2073-2080.

GIBLIN-DAVIS, R. M., WILLIAMS, D. S., BEKAL, S., DICKSON, D. W., BRITO, J. A., BECKER, J. O. & PRESTON, J. F. 2003. 'Candidatus *Pasteuria* usgae'sp. nov., an obligate endoparasite of the phytoparasitic nematode *Belonolaimus longicaudatus. International Journal of Systematic and Evolutionary Microbiology,* 53**,** 197-200.

GIORNO, R., BOZUE, J., COTE, C., WENZEL, T., MOODY, K.-S., MALLOZZI, M., RYAN, M., WANG, R., ZIELKE, R. & MADDOCK, J. R. 2007. Morphogenesis of the *Bacillus anthracis* spore. *Journal of Bacteriology,* 189**,** 691-705.

GOH, S. H., POTTER, S., WOOD, J. O., HEMMINGSEN, S. M., REYNOLDS, R. P. & CHOW, A. W. 1996. HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *Journal of Clinical Microbiology,* 34**,** 818-823.

GOODEY, J., FRANKLIN, M. & HOOPER, D. 1965. T. Goodey's The nematode parasites of *Meloidogyne arenaria* eggs in an Alabama soil. A mycological survey and greenhouse studies. *Nematropica,* 13**,** 201-213.

GOODEY, T. 1963. Soil and freshwater nematodes. *Soil and freshwater nematodes. A monograph.*

GRAINGER, J. 1964. Factors affecting the control of eelworm diseases 1. *Nematologica,* 10**,** 5-20.

GREEN, J., WANG, D., LILLEY, C. J., URWIN, P. E. & ATKINSON, H. J. 2012. Transgenic potatoes for potato cyst nematode control can replace pesticide use without impact on soil quality. *PloS One,* 7**,** e30973.

GREGORY, P. J., JOHNSON, S. N., NEWTON, A. C. & INGRAM, J. S. 2009. Integrating pests and pathogens into the climate change/food security debate. *Journal of Experimental Botany,* 60**,** 2827-2838.

GROISMAN, E. A. & OCHMAN, H. 1996. Pathogenicity Islands: Bacterial Evolution in Quantum Leaps. *Cell,* 87**,** 791-794.

GU, C., JENKINS, S. A., XUE, Q. & XU, Y. 2012. Activation of the classical complement pathway by *Bacillus anthracis* is the primary mechanism for spore phagocytosis and involves the spore surface protein BclA. *The Journal of Immunology,* 188**,** 4421-4431.

GUPTA, R., JUNG, E. & BRUNAK, S. 2004. NetNGlyc 1.0 Server. *Center for biological sequence analysis, technical university of Denmark available from:* http://www.*cbs.dtudk/services/NetNGlyc*.

HACKNEY, R. & DICKERSON, O. 1975. Marigold, castor bean, and chrysanthemum as controls of *Meloidogyne incognita* and *Pratylenchus alleni. Journal of Nematology,* 7**,** 84.

HALL, T. 2007. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98 (1999). *External Resources Chemical Abstracts Service (CAS).*

HAMID, M. I., HUSSAIN, M., WU, Y., ZHANG, X., XIANG, M. & LIU, X. 2017. Successive soybean-monoculture cropping assembles rhizosphere microbial communities for the soil suppression of soybean cyst nematode. *FEMS Microbiology Ecology,* 93**,** fiw222.

HASHEM, M. & ABO-ELYOUSR, K. A. 2011. Management of the root-knot nematode *Meloidogyne incognita* on tomato with combinations of different biocontrol organisms. *Crop Protection,* 30**,** 285-292.

HAUQUIER, F., BALLESTEROS-REDONDO, L., GUTT, J. & VANREUSEL, A. 2016. Community dynamics of nematodes after Larsen ice-shelf collapse in the eastern Antarctic Peninsula. *Ecology and evolution,* 6**,** 305-317.

HEIP, C., VINCX, M. & VRANKEN, G. 1985. The ecology of marine nematodes. *Oceanography and Marine Biology***,** 399-489.

HENRIQUES, A. O. & MORAN, C. P. 2000. Structure and assembly of the bacterial endospore coat. *Methods,* 20**,** 95-110.

HEWLETT, T. E., GRISWOLD, S. T. & SMITH, K. S. 2006. Biological control of Meloidogyne incognita using in-vitro produced *Pasteuria penetrans* in a microplot study. *Journal of Nematology,* 38**,** 274-274.

HITSUMOTO, Y., MORITA, N., YAMAZOE, R., TAGOMORI, M., YAMASAKI, T. & KATAYAMA, S. 2014. Adhesive properties of *Clostridium perfringens* to extracellular matrix proteins collagens and fibronectin. *Anaerobe,* 25**,** 67-71.

HOCH, J. A. 1993. Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. Annual Reviews of Microbiology. 47:441–465.

HOCKLAND, S., INSERRA, R., MILLAR, L., LEHMAN, P., PERRY, R. & MOENS, M. 2006. International plant health-putting legislation into practice. *Plant Nematology***,** 327-345.

HORLICK-JONES, T., WALLS, J., ROWE, G., PIDGEON, N., POORTINGA, W., MURDOCK, G. & O'RIORDAN, T. 2007. *The GM Debate: Risk, Politics and Public Engagement*, Routledge.

HUANG, G., ALLEN, R., DAVIS, E. L., BAUM, T. J. & HUSSEY, R. S. 2006. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proceedings of the National Academy of Sciences,* 103**,** 14302-14306.

HUNT, D. J. & HANDOO, Z. A. 2009. Taxonomy, identification and principal species. *Root-knot nematodes,* 1**,** 55-88.

HUSSEY, R. & JANSSEN, G. 2002. Root-knot nematodes: *Meloidogyne* species. *Plant resistance to parasitic nematodes***,** 43-70.

HUSSEY, R., MCGUIRE, J., BROWN, R. & KERRY, B. 1987. Interaction with other organisms. *Principles and practice of nematode control in crops.***,** 293-328.

HUYNEN, M. A. & BORK, P. 1998. Measuring genome evolution. *Proceedings of the National Academy of Sciences,* 95**,** 5849-5856.

IMBRIANI, J. & MANKAU, R. 1977. Ultrastructure of the nematode pathogen, *Bacillus penetrans. Journal of Invertebrate Pathology,* 30**,** 337-347.

ISBERG, R. R. & BARNES, P. 2001. Subversion of integrins by enteropathogenic *Yersinia*. *Journal of Cell Science,* 114**,** 21-28.

JAFFEE, B. & MULDOON, A. 1989. Suppression of cyst nematode by natural infestation of a nematophagous fungus. *Journal of Nematology,* 21**,** 505.

JANSSON, H.-B. & LOPEZ-LLORCA, L. V. 2004. Control of nematodes by fungi. *Fungal Biotechnology in Agricultural, Food, and Environmental Applications***,** 205-215.

JAYAKUMAR, J. 2010. Bio-efficacy of *Streptomyces avermitilis* culture filtrates against root knot nematode, *Meloidogyne incognita* and reniform nematodes, *Rotylenchulus reniformis. Karnataka Journal of Agricultural Sciences,* 22.

JENSEN, P. 1987. Feeding ecology of free-living aquatic nematodes. *Marine Ecology Progress Series,* 35**,** 187-196.

JOHNSON, M. J., TODD, S. J., BALL, D. A., SHEPHERD, A. M., SYLVESTRE, P. & MOIR, A. 2006. ExsY and CotY are required for the correct assembly of the exosporium and spore coat of *Bacillus cereus. Journal of Bacteriology,* 188**,** 7905-7913.

JOHNSON, W. C. & TIPPER, D. 1981. Acid-soluble spore proteins of *Bacillus subtilis. Journal of Bacteriology,* 146**,** 972-982.

JOHNSTONE, I. L. 2000. Cuticle collagen genes: expression in *Caenorhabditis elegans*. *Trends in Genetics,* 16**,** 21-27.

JONATHAN, E., BARKER, K., ABDEL ALIM, F., VRAIN, T. & DICKSON, D. 2000. Biological control of Meloidogyne incognita on tomato and banana with rhizobacteria, actinomycetes, and Pasteuria penetrans. *Nematropica,* 30**,** 231-240.

JONES, D. T., TAYLOR, W. R. & THORNTON, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences: CABIOS,* 8**,** 275-282.

JONES, M. & NORTHCOTE, D. 1972. Nematode-induced syncytium-a multinucleate transfer cell. *J. Cell Science,* 10**,** 1.

KAILAS, L., TERRY, C., ABBOTT, N., TAYLOR, R., MULLIN, N., TZOKOV, S. B., TODD, S. J., WALLACE, B. A., HOBBS, J. K. & MOIR, A. 2011. Surface architecture of endospores of the *Bacillus cereus/anthracis/thuringiensis* family at the subnanometer scale. *Proceedings of the National Academy of Sciences,* 108**,** 16014-16019.

KANEHISA, M., GOTO, S., KAWASHIMA, S. & NAKAYA, A. 2002. The KEGG databases at GenomeNet. *Nucleic Acids Research,* 30**,** 42-46.

KARLIN, S. & BROCCHIERI, L. 2000. Heat shock protein 60 sequence comparisons: duplications, lateral transfer, and mitochondrial evolution. *Proceedings of the National Academy of Sciences,* 97**,** 11348-11353.

KAYA, H. K. 2002. Natural Enemies and Other. *Entomopathogenic nematology***,** 189.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S. & DURAN, C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

KEREN-ZUR, M., ANTONOV, J., BERCOVITZ, A., FELDMAN, K., HUSID, A., KENAN, G., MARCOV, N. & REBHUN, M. *Bacillus firmus* formulations for the safe control of root-knot nematodes. *The BCPC Conference: Pests and diseases, Volume 1*. Proceedings of an international conference held at the Brighton Hilton Metropole Hotel, Brighton, UK, 13-16 November 2000., 2000. British Crop Protection Council, 47-52.

KERRY, B. 1988. Fungal parasites of cyst nematodes. *Agriculture, Ecosystems & Environment,* 24**,** 293-305.

KERRY, B. R. 2000. Rhizosphere interactions and the exploitation of microbial agents for the biological control of plant-parasitic nematodes. *Annual Review of Phytopathology,* 38**,** 423-441.

KIM, J. & SCHUMANN, W. 2009. Display of proteins on *Bacillus subtilis* endospores. *Cellular and Molecular Life Sciences,* 66**,** 3127-3136.

KLINE, K. A., FÄLKER, S., DAHLBERG, S., NORMARK, S. & HENRIQUES-NORMARK, B. 2009. Bacterial Adhesins in Host-Microbe Interactions. *Cell Host & Microbe,* 5**,** 580-592.

KOJETIN, D. J., THOMPSON, R. J., BENSON, L. M., NAYLOR, S., WATERMAN, J., DAVIES, K. G., OPPERMAN, C. H., STEPHENSON, K., HOCH, J. A. & CAVANAGH, J. 2005. Structural analysis of divalent metals binding to the Bacillus subtilis response regulator Spo0F: the possibility for in vitro metalloregulation in the initiation of sporulation. *Biometals,* 18**,** 449-66.

KOZUKA, S. & TOCHIKUBO, K. 1985. Properties and origin of filamentous appendages on spores of *Bacillus cereus. Microbiology and Immunology,* 29**,** 21-37.

KRATOCHVIL, R. J., SARDANELLI, S., EVERTS, K. & GALLAGHER, E. 2004. Evaluation of crop rotation and other cultural practices for management of root-knot and lesion nematodes. *Agronomy Journal,* 96**,** 1419-1428.

KROGH, A., LARSSON, B., VON HEIJNE, G. & SONNHAMMER, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology,* 305**,** 567-80.

KÜHN, K. 1987. The Classical Collagens: Types I, II, and III. *In:* BURGESON, R. M. E. (ed.) *Structure and Function of Collagen Types.* Academic Press.

KUISIENE, N., RAUGALAS, J. & CHITAVICHIUS, D. 2009. Phylogenetic, inter, and intraspecific sequence analysis of spo0A gene of the genus *Geobacillus. Current Microbiology,* 58**,** 547.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, msw054.

KWOK, A. Y., SU, S.-C., REYNOLDS, R. P., BAY, S. J., AV-GAY, Y., DOVICHI, N. J. & CHOW, A. W. 1999. Species identification and phylogenetic relationships based on partial HSP60 gene sequences within the genus *Staphylococcus*. *International Journal of Systematic and Evolutionary Microbiology,* 49, 1181-1192.

KYTE, J. & DOOLITTLE, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology,* 157, 105-132.

LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227, 680-5.

LAI, E.-M., PHADKE, N. D., KACHMAN, M. T., GIORNO, R., VAZQUEZ, S., VAZQUEZ, J. A., MADDOCK, J. R. & DRIKS, A. 2003. Proteomic analysis of the spore coats of *Bacillus subtilis* and *Bacillus anthracis*. *Journal of Bacteriology,* 185, 1443-1454.

LAMONDIA, J. & TIMPER, P. 2016. Interactions of Microfungi and Plant-Parasitic Nematodes. *Biology of Microfungi.* Springer.

LARKIN, D. M., PAPE, G., DONTHU, R., AUVIL, L., WELGE, M. & LEWIN, H. A. 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Research,* 19, 770-777.

LEGENDRE, M., ARSLAN, D., ABERGEL, C. & CLAVERIE, J.-M. 2012. Genomics of Megavirus and the elusive fourth domain of Life. *Communicative and Integrative Biology,* 5, 102-106.

LEHNINGER, A. L. 1975. Biochemistry: the molecular basis of cell structure and functions. *Worth, New York,* 2nd Edition, 35.

LEQUETTE, Y., GARÉNAUX, E., TAUVERON, G., DUMEZ, S., PERCHAT, S., SLOMIANNY, C., LERECLUS, D., GUÉRARDEL, Y. & FAILLE, C. 2011. Role played by exosporium glycoproteins in the surface properties of *Bacillus cereus* spores and in their adhesion to stainless steel. *Applied and Environmental Microbiology,* 77, 4905-4911.

LESKI, T. A., CASWELL, C. C., PAWLOWSKI, M., KLINKE, D. J., BUJNICKI, J. M., HART, S. J. & LUKOMSKI, S. 2009. Identification and classification of *bcl* genes and proteins of *Bacillus cereus* group organisms and their application in *Bacillus anthracis* detection and fingerprinting. *Applied and Environmental Microbiology,* 75, 7163-72.

LEVINE, N. D. 1968. Nematode parasites of domestic animals and of man. *Nematode Parasites of Domestic Animals and of Man.*

LEVINSON, G. & GUTMAN, G. A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution,* 4, 203-221.

LILLEY, C., DAVIES, L. & URWIN, P. 2012. RNA interference in plant parasitic nematodes: a summary of the current status. *Parasitology,* 139, 630-640.

LIN, J., MAZAREI, M., ZHAO, N., ZHU, J. J., ZHUANG, X., LIU, W., PANTALONE, V. R., ARELLI, P. R., STEWART, C. N. & CHEN, F. 2013. Overexpression of a soybean salicylic acid methyltransferase gene confers resistance to soybean cyst nematode. *Plant Biotechnology Journal,* 11, 1135-1145.

LINFORD, M. B. 1937. Stimulated activity of natural enemies of nematodes. *Science,* 85, 123-124.

LITTLE, S. & DRIKS, A. 2001. Functional analysis of the *Bacillus subtilis* morphogenetic spore coat protein CotE. *Molecular Microbiology,* 42, 1107-1120.

LIU, S., KANDOTH, P. K., WARREN, S. D., YECKEL, G., HEINZ, R., ALDEN, J., YANG, C., JAMAI, A., EL-MELLOUKI, T. & JUVALE, P. S. 2012. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature,* 492, 256-260.

LOURENÇO-TESSUTTI, I. T., SOUZA JUNIOR, J. D. A., MARTINS-DE-SA, D., VIANA, A. A. B., CARNEIRO, R. M. D. G., TOGAWA, R. C., DE ALMEIDA-ENGLER, J., BATISTA, J. A. N., SILVA, M. C. M. & FRAGOSO, R. R. 2015. Knock-Down of Heat-Shock Protein 90 and Isocitrate Lyase Gene Expression Reduced Root-Knot Nematode Reproduction. *Phytopathology,* 105, 628-637.

LUDWIG, W., SCHLEIFER, K. & WHITMAN, W. 2007. Revised road map to the phylum Firmicutes. *Bergey's Manual Trust website*.

LUKOMSKI, S., NAKASHIMA, K., ABDI, I., CIPRIANO, V. J., IRELAND, R. M., REID, S. D., ADAMS, G. G. & MUSSER, J. M. 2000. Identification and Characterization of the *scl* Gene Encoding a Group A *Streptococcus* Extracellular Protein Virulence Factor with Similarity to Human Collagen. *Infection and Immunity,* 68**,** 6542-6553.

MACNAGHTEN, P. 2016. Responsible innovation and the reshaping of existing technological trajectories: the hard case of genetically modified crops. *Journal of Responsible Innovation,* 3**,** 282-289.

MADDISON, W. P. & KNOWLES, L. L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic biology,* 55**,** 21-30.

MADULU, J. & TRUDGILL, D. 1994. Influence of temperature on the development and survival of *Meloidogyne javanica*. *Nematologica,* 40**,** 230-243.

MAES, E., KRZEWINSKI, F., GARENAUX, E., LEQUETTE, Y., CODDEVILLE, B., TRIVELLI, X., RONSE, A., FAILLE, C. & GUERARDEL, Y. 2016. Glycosylation of BclA glycoprotein from *Bacillus cereus* and *Bacillus anthracis* exosporium is domain-specific. *Journal of Biological Chemistry,* 291**,** 9666-9677.

MAGGENTI, A. 1981a. *General Nematology*, Springer.

MAGGENTI, A. R. 1981b. Nematodes: development as plant parasites. *Annual Reviews in Microbiology,* 35**,** 135-154.

MANKAU, R. 1975. *Bacillus penetrans* n. comb. causing a virulent disease of plant-parasitic nematodes. *Journal of invertebrate Pathology,* 26**,** 333-339.

MANKAU, R. 1980a. Biocontrol: fungi as nematode control agents. *Journal of Nematology,* 12**,** 244.

MANKAU, R. 1980b. Biological control of nematode pests by natural enemies. *Annual Review of Phytopathology,* 18**,** 415-440.

MANUM, S. B., BOSE, M. N., SAYER, R. T. & BOSTRÖM, S. 1994. A nematode (*Captivonema cretacea* gen. et sp. n.) preserved in a clitellate cocoon wall from the Early Cretaceous. *Zoologica Scripta,* 23**,** 27-31.

MARSTON, E. L., SUMNER, J. W. & REGNERY, R. L. 1999. Evaluation of intraspecies genetic variation within the 60 kDa heat-shock protein gene (*groEL*) of *Bartonella* species. *International Journal of Systematic and Evolutionary Microbiology,* 49**,** 1015-1023.

MASHELA, P. W., NDHLALA, A. R., POFU, K. M. & DUBE, Z. P. 2016. Phytochemicals of Nematode-Resistant Transgenic Plants. *Transgenesis and Secondary Metabolism*, 1-16.

MATZ, L. L., BEAMAN, C. & GERHARDT, P. 1970. Chemical Composition of Exosporium from Spores of *Bacillus cereus*. *Journal of Bacteriology,* 101**,** 196-201.

MAUCHLINE, T. H., KNOX, R., MOHAN, S., POWERS, S. J., KERRY, B. R., DAVIES, K. G. & HIRSCH, P. R. 2011. Identification of new single nucleotide polymorphism-based markers for inter- and intraspecies discrimination of obligate bacterial parasites (*Pasteuria* spp.) of invertebrates. *Applied and Environmental Microbiology,* 77**,** 6388-94.

MCCLOY, R. A., ROGERS, S., CALDON, C. E., LORCA, T., CASTRO, A. & BURGESS, A. 2014. Partial inhibition of Cdk1 in G2 phase overrides the SAC and decouples mitotic events. *Cell Cycle,* 13**,** 1400-1412.

MCELROY, K., MOUTON, L., DU PASQUIER, L., QI, W. & EBERT, D. 2011. Characterisation of a large family of polymorphic collagen-like proteins in the endospore-forming bacterium *Pasteuria ramosa*. *Res Microbiol,* 162**,** 701-14.

MCKENNEY, P. T., DRIKS, A. & EICHENBERGER, P. 2013. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. *Nature Reviews Microbiology,* 11**,** 33-44.

MCPHERSON, S. A., LI, M., KEARNEY, J. F. & TURNBOUGH, C. L. 2010. ExsB, an unusually highly phosphorylated protein required for the stable attachment of the exosporium of *Bacillus anthracis*. *Molecular Microbiology,* 76**,** 1527-1538.

MCSORLEY, R., DICKSON, D., DE BRITO, J. & HOCHMUTH, R. 1994. Tropical rotation crops influence nematode densities and vegetable yields. *Journal of Nematology,* 26**,** 308.

MCSORLEY, R., WANG, K.-H. & CHURCH, G. 2008. Suppression of root-knot nematodes in natural and agricultural soils. *Applied Soil Ecology,* 39**,** 291-298.

METCHNIKOFF, E. 1888. *Pasteuria ramosa* un représentant des bactéries à division longitudinale. *Annales de l'Institut Pasteur (Paris),* 2**,** 165-170.

MITCHUM, M. G., KANDOTH, P. K., YECKEL, G. & ITHAL, N. 2012. Genes Implicated In Resistance To Soybean Cyst Nematode Infection And Methods Of Their Use. Google Patents.

MOHAN, S., FOULD, S. & DAVIES, K. G. 2001. The interaction between the gelatin-binding domain of fibronectin and the attachment of *Pasteuria penetrans* endospores to nematode cuticle. *Parasitology,* 123**,** 271-276.

MOHAN, S., MAUCHLINE, T. H., ROWE, J., HIRSCH, P. R. & DAVIES, K. G. 2012. *Pasteuria* endospores from *Heterodera cajani* (Nematoda: Heteroderidae) exhibit inverted attachment and altered germination in cross-infection studies with *Globodera pallida* (Nematoda: Heteroderidae). *FEMS Microbiology Ecology,* 79**,** 675-84.

MOIR, A., CORFE, B. & BEHRAVAN, J. 2002. Spore germination. *Cellular and Molecular Life Sciences,* 59**,** 403-409.

MORSE, W. J., CARTTER, J. & WILLIAMS, L. F. 1949. *Soybeans: Culture and varieties No. 1520. US Department. of Agriculture.*

MOSES, V. 2016. The debate over GM crops is making history. *Nature,* 537**,** 139.

MOUTON, L., TRAUNECKER, E., MCELROY, K., DU PASQUIER, L. & EBERT, D. 2009. Identification of a polymorphic collagen-like protein in the crustacean bacteria *Pasteuria ramosa. Research in Microbiology,* 160**,** 792-9.

MURTAGH, F. & LEGENDRE, P. 2011. Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. *Journal of Classification. 31(3), 274–295.*

NEWHALL, A. 1955. Disinfestation of soil by heat, flooding and fumigation. *The Botanical Review,* 21**,** 189-250.

NICOL, J. M., TURNER, S. J., COYNE, D., DEN NIJS, L., HOCKLAND, S. & MAAFI, Z. T. 2011. Current nematode threats to world agriculture. *Genomics and Molecular Genetics of Plant-Nematode Interactions, 21-43.*

NOEL, G., RIGGS, R. & WRATHER, J. 1992. History, distribution, and economics. *Biology and Management of the Soybean Cyst Nematode***,** 1-13.

NOEL, G. R., ATIBALENTJA, N. & DOMIER, L. L. 2005. Emended description of *Pasteuria nishizawae. International Journal of Systematic and Evolutionary Microbiology,* 55**,** 1681-1685.

NOLING, J. & BECKER, J. 1994. The challenge of research and extension to define and implement alternatives to methyl bromide. *Journal of Nematology,* 26**,** 573.

NORDBRING-HERTZ, B., JANSSON, H. B. & TUNLID, A. 2006. Nematophagous fungi. *Nematophagous fungi. Encyclopedia of life sciences.* Macmillan, Basingstoke.

NORTON, D. C. 1978. *Ecology of plant-parasitic nematodes*, John Wiley & Sons, New York. 268 pp.

NUSBAUM, C. & FERRIS, H. 1973. The role of cropping systems in nematode population management. *Annual Review of Phytopathology,* 11**,** 423-440.

O'CONNELL, D. 2003. Dock, lock and latch. *Nature Reviews Microbiology,* 1**,** 171-171.

OKA, Y., SHAPIRA, N. & FINE, P. 2007. Control of root-knot nematodes in organic farming systems by organic amendments and soil solarization. *Crop Protection,* 26**,** 1556-1565.

OLIVARES-BERNABEU, C. M. & LÓPEZ-LLORCA, L. V. 2002. Fungal egg-parasites of plant-parasitic nematodes from Spanish soils. *Revista iberoamericana de micología,* 19**,** 104-110.

OPPERMAN, C. H., DAVIES, K. G., SOSINSKI, B. R., WATERMAN, J. & BURKE, M. 2003. Genome survey sequences (3903) obtained from *Pasteuria penetrans* (Res-147). .

Centre for the Biology of Nematode Parasitism, University of North Carolina, Raleigh, USA & Nematode Interactions Unit, Rothamsted Research, Harpenden, UK.

PAIDHUNGAT, M. & SETLOW, P. 2000. Role of Ger proteins in nutrient and nonnutrient triggering of spore germination in *Bacillus subtilis. Journal of Bacteriology,* 182**,** 2513-2519.

PATTI, J. M., ALLEN, B. L., MCGAVIN, M. J. & HOOK, M. 1994. MSCRAMM-mediated adherence of microorganisms to host tissues. *Annual Reviews in Microbiology,* 48**,** 585-617.

PATTI, J. M. & HÖÖK, M. 1994. Microbial adhesins recognizing extracellular matrix macromolecules. *Current opinion in cell biology,* 6**,** 752-758.

PERRY, R. N. 1996. Chemoreception in plant parasitic nematodes. *Annual review of phytopathology,* 34**,** 181-199.

PERSIDIS, A., LAY, J. G., MANOUSIS, T., BISHOP, A. H. & ELLAR, D. J. 1991. Characterisation of potential adhesins of the bacterium *Pasteuria penetrans*, and of putative receptors on the cuticle of *Meloidogyne incognita*, a nematode host. *Journal of Cell Science,* 100 (Pt 3)**,** 613-22.

PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods,* 8**,** 785-786.

PHILIPPE, H. & DOUADY, C. J. 2003. Horizontal gene transfer and phylogenetics. *Current Opinion in Microbiology,* 6**,** 498-505.

PIGGOT, P. J. & HILBERT, D. W. 2004. Sporulation of *Bacillus subtilis. Current Opinion in Microbiology,* 7**,** 579-586.

PIZARRO-CERDÁ, J. & COSSART, P. 2006. Bacterial Adhesion and Entry into Host Cells. *Cell,* 124**,** 715-727.

PLOEG, A. T. 2002. Effects of selected marigold varieties on root-knot nematodes and tomato and melon yields. *Plant Disease,* 86**,** 505-508.

PLOMP, M. & MALKIN, A. J. 2008. Mapping of proteomic composition on the surfaces of *Bacillus* spores by atomic force microscopy-based immunolabeling. *Langmuir,* 25**,** 403-409.

POINAR JR, G., ACRA, A. & ACRA, F. 1994. Earliest fossil nematode (Mermithidae) in Cretaceous Lebanese amber. *Fundamental and applied nematology,* 17**,** 475-477.

PONNAMPERUMA, F. 1984. Effects of flooding on soils. *Flooding and Plant Growth***,** 9-45.

POPHAM, D. L., HELIN, J., COSTELLO, C. E. & SETLOW, P. 1996. Analysis of the peptidoglycan structure of *Bacillus subtilis* endospores. *Journal of Bacteriology,* 178**,** 6451-6458.

POWELL, N. 1971. Interactions between nematodes and fungi in disease complexes. *Annual Review of Phytopathology,* 9**,** 253-274.

POWELL, N. 2012. Interaction of plant parasitic nematodes with other disease-causing agents. *Zuckerman, BM, Mai, WF SS Rohde, RA (Eds). Plant Parasitic Nematodes,* 2**,** 119-136.

PRESTON, J., DICKSON, D., MARUNIAK, J., NONG, G., BRITO, J., SCHMIDT, L. & GIBLIN-DAVIS, R. 2003. *Pasteuria* spp.: systematics and phylogeny of these bacterial parasites of phytopathogenic nematodes. *Journal of Nematology,* 35**,** 198.

PYROWOLAKIS, A., WESTPHAL, A., SIKORA, R. A. & BECKER, J. O. 2002. Identification of root-knot nematode suppressive soils. *Applied Soil Ecology,* 19**,** 51-56.

R DEVELOPEMENT CORE TEAM 2010. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

RADÓ-TRILLA, N. & ALBÀ, M. 2012. Dissecting the role of low-complexity regions in the evolution of vertebrate proteins. *BMC Evolutionary Biology,* 12**,** 155.

RAJENDHRAN, J. & GUNASEKARAN, P. 2011. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research,* 166**,** 99-110.

RAMACHANDRAN, G. N. 1956. Structure of Collagen. *Nature,* 177**,** 710-711.

RAMACHANDRAN, G. N. & KARTHA, G. 1954. Structure of Collagen. *Nature,* 174**,** 269-270.

RASKO, D. A., ALTHERR, M. R., HAN, C. S. & RAVEL, J. 2005. Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiology Reviews,* 29, 303-329.

RASMUSSEN, M., EDEN, A. & BJORCK, L. 2000. SclA, a novel collagen-like surface protein of *Streptococcus pyogenes. Infection and Immunity,* 68, 6370-6377.

RASMUSSEN, M., JACOBSSON, M. & BJORCK, L. 2003. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *Journal of Biological Chemistry,* 278, 32313-6.

REDMOND, C., BAILLIE, L. W., HIBBS, S., MOIR, A. J. & MOIR, A. 2004. Identification of proteins in the exosporium of *Bacillus anthracis. Microbiology,* 150, 355-363.

REICHHARDT, T. 1999. US sends mixed message in GM debate. *Nature,* 400, 298-298.

RETY, S., SALAMITOU, S., GARCIA-VERDUGO, I., HULMES, D. J., LE HEGARAT, F., CHABY, R. & LEWIT-BENTLEY, A. 2005. The crystal structure of the *Bacillus anthracis* spore surface protein BclA shows remarkable similarity to mammalian proteins. *Journal of Biological Chemistry,* 280, 43073-8.

RICARD-BLUM, S. 2011. The collagen family. *Cold Spring Harbor Perspectives in Biology,* 3, a004978.

ROCKA, A. 2017. Cestodes and nematodes of Antarctic fishes and birds. *Biodiversity and Evolution of Parasitic Life in the Southern Ocean.* Springer.

RODENBURG, C. M., MCPHERSON, S. A., TURNBOUGH, C. L. & DOKLAND, T. 2014. Cryo-EM analysis of the organization of BclA and BxpB in the *Bacillus anthracis* exosporium. *Journal of Structural Biology,* 186, 181-187.

RODERICK, H., TRIPATHI, L., BABIRYE, A., WANG, D., TRIPATHI, J., URWIN, P. E. & ATKINSON, H. J. 2012. Generation of transgenic plantain (*Musa* spp.) with resistance to plant pathogenic nematodes. *Molecular Plant Pathology,* 13, 842-851.

RODGER, A. 1895. Preliminary Account of Natural History Collections made on a Voyage to the Gulf of St Lawrence and Davis Straits. *Proceedings of the Royal Society of Edinburgh,* 20, 154-163.

RODIUC, N., VIEIRA, P., YOUSSEF BANORA, M. & DE ALMEIDA ENGLER, J. 2014. On the track of transfer cells formation by specialized plant-parasitic nematodes. *Frontiers in Plant Science,* 5.

RODRIGUEZ-KABANA, R. & CANULLO, G. H. 1992. Cropping systems for the management of phytonematodes. *Phytoparasitica,* 20, 211-224.

ROELS, S., DRIKS, A. & LOSICK, R. 1992. Characterization of *spoIVA*, a sporulation gene involved in coat morphogenesis in *Bacillus subtilis. Journal of Bacteriology,* 174, 575-585.

ROMERO, P. A. & ARNOLD, F. H. 2009. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology,* 10, 866-876.

RUDALL, K. 1967. Conformation in chitin–protein complexes. *Conformation of Biopolymers,* 2, 751-756.

RUDALL, K. M. & KENCHING.W 1971. ARTHROPOD SILKS - Problem of fibrous proteins in animal tissues. *Annual Review of Entomology,* 16, 73-&.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M.-A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16, 944-945.

SÁNCHEZ-MORENO, S. & FERRIS, H. 2007. Suppressive service of the soil food web: effects of environmental management. *Agriculture, Ecosystems and Environment,* 119, 75-87.

SÁNCHEZ, B., ARIAS, S., CHAIGNEPAIN, S., DENAYROLLES, M., SCHMITTER, J., BRESSOLLIER, P. & URDACI, M. 2009. Identification of surface proteins involved in the adhesion of a probiotic *Bacillus cereus* strain to mucin and fibronectin. *Microbiology,* 155, 1708-1716.

SARAH, J. L. 1989. Banana nematodes and their control in Africa. *Nematropica,* 19, 199-216.

SASSER, J., EISENBACK, J., CARTER, C. & TRIANTAPHYLLOU, A. 1983. The international *Meloidogyne* project-its goals and accomplishments. *Annual Review of Phytopathology,* 21, 271-288.

SAYRE, R. & STARR, M. 1985. *Pasteuria penetrans* (ex Thome, 1940) nom. rev., comb, n., sp. n., a mycelial and endospore-forming bacterium parasitic in plant-parasitic nematodes. *Proceedings of Helminthological Society of Washington,* 52**,** 149-165.

SAYRE, R. & STARR, M. 1989. Genus *Pasteuria* Metchnikoff, 1888. *Bergey's Manual of Systematic Bacteriology. Baltimore, MD: Williams and Wilkins***,** 2601-2615.

SAYRE, R., STARR, M., POINAR JR, G. & JANSSON, H.-B. 1988. Bacterial diseases and antagonisms of nematodes. *Diseases of Nematodes, vol. 1.***,** 69-101.

SAYRE, R. M. & WERGIN, W. P. 1977. Bacterial parasite of a plant nematode: morphology and ultrastructure. *Journal of Bacteriology,* 129**,** 1091-1101.

SCHAEFFER, A. B. & FULTON, M. D. 1933. A SIMPLIFIED METHOD OF STAINING ENDOSPORES. *Science,* 77**,** 194.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S. & SCHMID, B. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods,* 9**,** 676-682.

SCHMIDT, L. M., MOUTON, L., NONG, G., EBERT, D. & PRESTON, J. F. 2008. Genetic and immunological comparison of the cladoceran parasite *Pasteuria ramosa* with the nematode parasite *Pasteuria penetrans. Applied and Environmental Microbiology,* 74**,** 259-264.

SCHMIDT, M. A., RILEY, L. W. & BENZ, I. 2003. Sweet new world: glycoproteins in bacterial pathogens. *Trends in Microbiology,* 11**,** 554-561.

SCHOULS, L. M., SCHOT, C. S. & JACOBS, J. A. 2003. Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. *Journal of Bacteriology,* 185**,** 7241-7246.

SCHWARZ-LINEK, U., WERNER, J. M., PICKFORD, A. R., GURUSIDDAPPA, S., KIM, J. H., PILKA, E. S., BRIGGS, J. A., GOUGH, T. S., HÖÖK, M. & CAMPBELL, I. D. 2003. Pathogenic bacteria attach to human fibronectin through a tandem β-zipper. *Nature,* 423**,** 177-181.

SCOONES, I. 2008. Mobilizing Against GM Crops in India, South Africa and Brazil. *Journal of Agrarian Change,* 8**,** 315-344.

SCOPA, A. & DUMONTET, S. 2007. Soil solarization: effects on soil microbiological parameters. *Journal of Plant Nutrition,* 30**,** 537-547.

SETLOW, P. 1988. Small, acid-soluble spore proteins of *Bacillus* species: structure, synthesis, genetics, function, and degradation. *Annual Reviews in Microbiology,* 42**,** 319-338.

SETLOW, P. 2003. Spore germination. *Current opinion in microbiology,* 6**,** 550-556.

SETLOW, P. 2007. I will survive: DNA protection in bacterial spores. *Trends in microbiology,* 15**,** 172-180.

SEVERSON, K. M., MALLOZZI, M., BOZUE, J., WELKOS, S. L., COTE, C. K., KNIGHT, K. L. & DRIKS, A. 2009. Roles of the *Bacillus anthracis* spore protein ExsK in exosporium maturation and germination. *Journal of Bacteriology,* 191**,** 7587-7596.

SHARMA, S. & DAVIES, K. 1997. Modulation of spore adhesion of the hyperparasitic bacterium *Pasteuria penetrans* to nematode cuticle. *Letters in Applied Microbiology,* 25**,** 426-430.

SMANT, G., STOKKERMANS, J. P., YAN, Y., DE BOER, J. M., BAUM, T. J., WANG, X., HUSSEY, R. S., GOMMERS, F. J., HENRISSAT, B., DAVIS, E. L., HELDER, J., SCHOTS, A. & BAKKER, J. 1998. Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of National Academy of Sciences USA,* 95**,** 4906-11.

SMITH, T. J. & FOSTER, S. J. 1995. Characterization of the involvement of two compensatory autolysins in mother cell lysis during sporulation of *Bacillus subtilis* 168. *Journal of Bacteriology,* 177**,** 3855-3862.

SONENSHEIN, A. L., HOCH, J. A. & LOSICK, R. 1993. *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics.

SPIEGEL, Y., MOR, M. & SHARON, E. 1996. Attachment of *Pasteuria penetrans* endospores to the surface of *Meloidogyne javanica* second-stage juveniles. *Journal of Nematology,* 28**,** 328.

STAPLETON, J. J. & DEVAY, J. E. 1995. Soil solarization: a natural mechanism of integrated pest management. *Novel Approaches to Integrated Pest Management***,** 309-322.

STARR, M. & SAYRE, R. *Pasteuria thornei* sp. nov. and *Pasteuria penetrans* sensu stricto emend., mycelial and endospore-forming bacteria parasitic, respectively, on plant-parasitic nematodes of the genera *Pratylenchus* and *Meloidogyne.* Annales de l'Institut Pasteur/Microbiologie, 1988. Elsevier, 11-31.

STEENTOFT, C., VAKHRUSHEV, S. Y., JOSHI, H. J., KONG, Y., VESTER-CHRISTENSEN, M. B., SCHJOLDAGER, K. T., LAVRSEN, K., DABELSTEEN, S., PEDERSEN, N. B., MARCOS-SILVA, L., GUPTA, R., BENNETT, E. P., MANDEL, U., BRUNAK, S., WANDALL, H. H., LEVERY, S. B. & CLAUSEN, H. 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO Journal,* 32**,** 1478-88.

STEICHEN, C., CHEN, P., KEARNEY, J. F. & TURNBOUGH, C. L. 2003. Identification of the immunodominant protein and other proteins of the *Bacillus anthracis* exosporium. *Journal of Bacteriology,* 185**,** 1903-1910.

STEICHEN, C. T., KEARNEY, J. F. & TURNBOUGH, C. L. 2005. Characterization of the exosporium basal layer protein BxpB of *Bacillus anthracis. Journal of Bacteriology,* 187**,** 5868-5876.

STIRLING, G. 2014. Integrated soil biology management: the pathway to enhanced natural suppression of plant-parasitic nematodes. *Biological Control of Plant-Parasitic Nematodes: Soil Ecosystem Management in Sustainable Agriculture***,** 304-341.

STIRLING, G. & KERRY, B. 1983. Antagonists of the cereal cyst nematode *Heterodera avenae* Woll. in Australian soils. *Animal Production Science,* 23**,** 318-324.

STIRLING, G. & WACHTEL, M. F. 1980. Mass production of *Bacillus penetrans* for the biological control of root-knot nematodes. *Nematologica,* 26**,** 308-312.

STIRLING, G. R., POINAR JR, G. & JANSSON, H.-B. 1988. *Biological control of plant-parasitic nematodes*, CRC Press Inc.

STIRNADEL, H. A. & EBERT, D. 1997. Prevalence, host specificity and impact on host fecundity of microparasites and epibionts in three sympatric *Daphnia* species. *Journal of Animal Ecology***,** 212-222.

STRAUCH, M., WEBB, V., SPIEGELMAN, G. & HOCH, J. A. 1990. The SpoOA protein of *Bacillus subtilis* is a repressor of the *abrB* gene. *Proceedings of the National Academy of Sciences,* 87**,** 1801-1805.

STREISINGER, G., OKADA, Y., EMRICH, J., NEWTON, J., TSUGITA, A., TERZAGHI, E. & INOUYE, M. 1966. Frameshift mutations and the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology.* Cold Spring Harbor Laboratory Press, 77-84.

STURHAN, D. 1988. New host and geographical records of nematode-parasitic bacteria of the *Pasteuria penetrans* group. *Nematologica,* 34**,** 350-356.

STURHAN, D., WINKELHEIDE, R., SAYRE, R. & WERGIN, W. P. 1994. Light and electron microscopical studies of the life cycle and developmental stages of a *Pasteuria* isolate parasitizing the pea cyst nematode, *Heterodera rodera* goettingiana. *Fundamental and Applied Nematology,* 17**,** 29-42.

SUTHERLAND, T. D., PENG, Y. Y., TRUEMAN, H. E., WEISMAN, S., OKADA, S., WALKER, A. A., SRISKANTHA, A., WHITE, J. F., HUSON, M. G., WERKMEISTER, J. A., GLATTAUER, V., STOICHEVSKA, V., MUDIE, S. T., HARITOS, V. S. & RAMSHAW, J. A. M. 2013. A new class of animal collagen masquerading as an insect silk. *Scientific Reports,* 3.

SUZUKI, A. C., KAGOSHIMA, H., CHILTON, G., GROTHMAN, G. T., JOHANSSON, C. & TSUJIMOTO, M. 2017. Meiofaunal richness in highly acidic hot springs in Unzen-Amakusa National Park, Japan, including the first rediscovery attempt for mesotardigrada. *Zoological Science,* 34**,** 11-17.

SYLVESTRE, P., COUTURE-TOSI, E. & MOCK, M. 2003. Polymorphism in the Collagen-Like Region of the *Bacillus anthracis* BclA Protein Leads to Variation in Exosporium Filament Length. *Journal of Bacteriology,* 185**,** 1555-1563.

SYLVESTRE, P., COUTURE-TOSI, E. & MOCK, M. 2005. Contribution of ExsFA and ExsFB proteins to the localization of BclA on the spore surface and to the stability of the *Bacillus anthracis* exosporium. *Journal of Bacteriology,* 187**,** 5122-5128.

SYLVESTRE, P., MOCK, M. & COUTURE-TOSI, E. 2002. A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. *Molecular Microbiology,* 45**,** 169–178.

TAMBORRINI, M., OBERLI, M., WERZ, D., SCHÜRCH, N., FREY, J., SEEBERGER, P. & PLUSCHKE, G. 2009. Immuno-detection of anthrose containing tetrasaccharide in the exosporium of *Bacillus anthracis* and *Bacillus cereus* strains. *Journal of Applied Microbiology,* 106**,** 1618-1628.

TAMURA, K. & NEI, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution,* 10**,** 512-526.

TAN, L., LI, M. & TURNBOUGH, C. L. 2011. An unusual mechanism of isopeptide bond formation attaches the collagen-like glycoprotein BclA to the exosporium of *Bacillus anthracis. Molecular Biology,* 2**,** e00084-11.

TAN, L. & TURNBOUGH, C. L. 2010a. Sequence motifs and proteolytic cleavage of the collagen-like glycoprotein BclA required for its attachment to the exosporium of *Bacillus anthracis. Journal of Bacteriology,* 192**,** 1259-1268.

TAN, L. & TURNBOUGH, C. L. 2010b. Sequence motifs and proteolytic cleavage of the collagen-like glycoprotein BclA required for its attachment to the exosporium of *Bacillus anthracis. Journal of Bacteriology,* 192.

TAYLOR & SASSER 1978a. Biology, identification and control of root-knot nematodes. *North Carolina State University Graphics.*

TAYLOR, A. & SASSER, J. 1978b. Biology, identification and control of root-knot nematodes. *North Carolina State University Graphics.*

TAYLOR, C. 1990. Nematode interactions with other pathogens. *Annals of Applied Biology,* 116**,** 405-416.

TAYLOR, C. E. & BROWN, D. J. 1997. *Nematode vectors of plant viruses*, CAB INTERNATIONAL.

TENCALLA, F. 2006. Science, politics, and the GM debate in Europe. *Regulatory Toxicology and Pharmacology,* 44**,** 43-48.

TESTER, M. 2001. Depolarizing the GM debate. *New Phytologist,* 149**,** 9-12.

THOMPSON, B. M., HOELSCHER, B. C., DRIKS, A. & STEWART, G. C. 2011a. Localization and assembly of the novel exosporium protein BetA of *Bacillus anthracis. Journal of Bacteriology,* 193**,** 5098-5104.

THOMPSON, B. M., HOELSCHER, B. C., DRIKS, A. & STEWART, G. C. 2012. Assembly of the BclB glycoprotein into the exosporium and evidence for its role in the formation of the exosporium 'cap' structure in *Bacillus anthracis. Molecular Microbiology,* 86**,** 1073-1084.

THOMPSON, B. M., HSIEH, H. Y., SPRENG, K. A. & STEWART, G. C. 2011b. The co-dependence of BxpB/ExsFA and BclA for proper incorporation into the exosporium of *Bacillus anthracis. Molecular Microbiology,* 79**,** 799-813.

THOMPSON, B. M. & STEWART, G. C. 2008. Targeting of the BclA and BclB proteins to the *Bacillus anthracis* spore surface. *Molecular Microbiology,* 70**,** 421-34.

THOMPSON, B. M., WALLER, L. N., FOX, K. F., FOX, A. & STEWART, G. C. 2007. The BclB glycoprotein of *Bacillus anthracis* is involved in exosporium integrity. *Journal of Bacteriology,* 189**,** 6704-13.

THORNE, G. 1940. *Duboscqia penetrans* n. sp.(Sporozoa, Microsporidia, Nosematidae), a parasite of the nematode *Pratylenchus pratensis* (de Man) Filipjev. *Proceedings of the Helminthological Society of Washington,* 7**,** 51-53.

TOBIN, J., HAYDOCK, P., HARE, M., WOODS, S. & CRUMP, D. 2008. Effect of the fungus *Pochonia chlamydosporia* and fosthiazate on the multiplication rate of potato cyst nematodes (*Globodera pallida* and *G. rostochiensis*) in potato crops grown under UK field conditions. *Biological Control, 46*, 194-201.

TODD, S. J., MOIR, A. J. G., JOHNSON, M. J. & MOIR, A. 2003. Genes of *Bacillus cereus* and *Bacillus anthracis* encoding proteins of the exosporium. *Journal of Bacteriology, 185*, 3373-3378.

TOLL-RIERA, M., RADÓ-TRILLA, N., MARTYS, F. & ALBÀ, M. M. 2011. Role of low-complexity sequences in the formation of novel protein coding sequences. *Molecular Biology and Evolution*, msr263.

TON-THAT, H. & SCHNEEWIND, O. 2004. Assembly of pili in Gram-positive bacteria. *Trends in Microbiology, 12*, 228-234.

TOWBIN, H., STAEHELIN, T. & GORDON, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences, 76*, 4350-4354.

TROTTER, J. R. & BISHOP, A. H. 2003. Phylogenetic analysis and confirmation of the endospore-forming nature of *Pasteuria penetrans* based on the spo0A gene. *FEMS Microbiology Letters, 225*, 249-256.

TRUDGILL, D., EVANS, K. & PARROTT, D. M. 1975. Effects of potato cyst-nematodes on potato plants. *Nematologica, 21*, 169-182.

TRUDGILL, D. L. & BLOK, V. C. 2001. Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. *Annual Reviews of Phytopathology, 39*, 53-77.

TULLI, L., MARCHI, S., PETRACCA, R., SHAW, H. A., FAIRWEATHER, N. F., SCARSELLI, M., SORIANI, M. & LEUZZI, R. 2013. CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen. *Cellular microbiology, 15*, 1674-1687.

TUNLID, A., JANSSON, H.-B. & NORDBRING-HERTZ, B. 1992. Fungal attachment to nematodes. *Mycological Research, 96*, 401-412.

TURNER, S., ROWE, J., PERRY, R. & MOENS, M. 2006. Cyst nematodes. *Plant nematology*, 91-122.

ULLBERG, J. & ÓLAFSSON, E. 2003. Free-living marine nematodes actively choose habitat when descending from the water column. *Marine Ecology Progress Series, 260*, 141-149.

UNITED NATIONS, D. E. S. A., POPULATION DIVISION 2015. World population prospects: The 2015 revision, key findings and advance tables. *In:* ESA/P/WP.241., W. P. N. (ed.). New York.

VAID, A., BISHOP, A. & DAVIES, K. 2002. The polypeptide components of the parasporal fibres of *Pasteuria penetrans*. *World Journal of Microbiology and Biotechnology, 18*, 151-157.

VAN VALEN, L. 1974. Molecular evolution as predicted by natural selection. *Journal of Molecular Evolution, 3*, 89-101.

VENKATESH, R., HARRISON, S. K. & RIEDEL, R. M. 2009. Weed Hosts of Soybean Cyst Nematode (*Heterodera glycines*) in Ohio1.

VIALE, A. M., ARAKAKI, A. K., SONCINI, F. C. & FERREYRA, R. G. 1994. Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. *International Journal of Systematic and Evolutionary Microbiology, 44*, 527-533.

VINTER, V. 1959. Sporulation of bacilli. *Folia Microbiologica, 4*, 1-6.

VÖLKER, U. & HECKER, M. 2005. From genomics via proteomics to cellular physiology of the Gram-positive model organism *Bacillus subtilis*. *Cellular Microbiology, 7*, 1077-1085.

WALKER, J. R., GNANAM, A. J., BLINKOVA, A. L., HERMANDSON, M. J., KARYMOV, M. A., LYUBCHENKO, Y. L., GRAVES, P. R., HAYSTEAD, T. A. & LINSE, K. D. 2007. *Clostridium taeniosporum* spore ribbon-like appendage structure, composition and genes. *Molecular Microbiology, 63*, 629-43.

WALLER, L. N., FOX, N., FOX, K. F., FOX, A. & PRICE, R. L. 2004. Ruthenium red staining for ultrastructural visualization of a glycoprotein layer surrounding the spore of *Bacillus anthracis* and *Bacillus subtilis*. *Journal of Microbiological Methods,* 58**,** 23-30.

WAR, A. R., TAGGAR, G. K., WAR, M. Y. & HUSSAIN, B. 2016. Impact of climate change on insect pests, plant chemical ecology, tritrophic interactions and food production. *International Journal of Clinical and Biological Sciences,* 1**,** 16-29.

WEI, L., LIU, Y., DUBCHAK, I., SHON, J. & PARK, J. 2002. Comparative genomics approaches to study organism similarities and differences. *Journal of Biomedical Informatics,* 35**,** 142-150.

WELLER, D. M., RAAIJMAKERS, J. M., GARDENER, B. B. M. & THOMASHOW, L. S. 2002. Microbial populations responsible for specific soil suppressiveness to plant pathogens 1. *Annual Review of Phytopathology,* 40**,** 309-348.

WESTPHAL, A. 2005. Detection and description of soils with specific nematode suppressiveness. *Journal of Nematology,* 37**,** 121.

WIEDENBECK, J. & COHAN, F. M. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews,* 35**,** 957-976.

WILLIAMSON, V. M. & GLEASON, C. A. 2003. Plant–nematode interactions. *Current Opinion in Plant Biology,* 6**,** 327-333.

WILLIAMSON, V. M. & HUSSEY, R. S. 1996. Nematode pathogenesis and resistance in plants. *The Plant Cell,* 8**,** 1735.

WILSON, M. J. & JACKSON, T. A. 2013. Progress in the commercialisation of bionematicides. *BioControl,* 58**,** 715-722.

WINSLOW, C.-E., BROADHURST, J., BUCHANAN, R., KRUMWIEDE JR, C., ROGERS, L. & SMITH, G. 1920. The families and genera of the bacteria: final report of the committee of the Society of American Bacteriologists on characterization and classification of bacterial types. *Journal of Bacteriology,* 5**,** 191.

WISHART, J., BLOK, V. C., PHILLIPS, M. & DAVIES, K. S. 2004. *Pasteuria penetrans* and *P. nishizawae* attachment to *Meloidogyne chitwoodi, M. fallax* and *M. hapla*. *Nematology,* 6**,** 507-510.

WOESE, C. R. 1987. Bacterial evolution. *Microbiological Reviews,* 51**,** 221.

WOUTS, W. 1972. A revision of the family Heteroderidae (Nematoda: Tylenchoidea). 1. The family Heteroderidae and its subfamilies. *Nematologica,* 18**,** 439-446.

WRAY, W., BOULIKAS, T., WRAY, V. P. & HANCOCK, R. 1981. Silver staining of proteins in polyacrylamide gels. *Analytical Biochemistry,* 118**,** 197-203.

WRIGHT, D., BIRTLE, A., CORPS, A. & DYBAS, R. 1983. Efficacy of avermectins against a plant parasitic nematode. *Annals of Applied Biology,* 103**,** 465-470.

WUNSCHEL, D., FOX, K. F., BLACK, G. E. & FOX, A. 1995. Discrimination among the B. cereus group, in comparison to *B. subtilis*, by structural carbohydrate profiles and ribosomal RNA spacer region PCR. *Systematic and Applied Microbiology,* 17**,** 625-635.

WYSS, U., GRUNDLER, F. M. & MUNCH, A. 1992. The parasitic behaviour of second-stage juveniles of *Meloidogyne incognita* in roots of *Arabidopsis thaliana*. *Nematologica,* 38**,** 98-111.

XU, Y., KEENE, D. R., BUJNICKI, J. M., HOOK, M. & LUKOMSKI, S. 2002. Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *Journal of Biological Chemistry,* 277**,** 27312-8.

XU, Y., LIANG, X., CHEN, Y., KOEHLER, T. M. & HÖÖK, M. 2004. Identification and biochemical characterization of two novel collagen binding MSCRAMMs of *Bacillus anthracis*. *Journal of Biological Chemistry,* 279**,** 51760-51768.

XUE, Q., GU, C., RIVERA, J., HÖÖK, M., CHEN, X., POZZI, A. & XU, Y. 2011. Entry of *Bacillus anthracis* spores into epithelial cells is mediated by the spore surface protein BclA, integrin α2β1 and complement component C1q. *Cellular Microbiology,* 13, 620-634.

YAMAMOTO, S., KASAI, H., ARNOLD, D. L., JACKSON, R. W., VIVIAN, A. & HARAYAMA, S. 2000. Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology,* 146**,** 2385-2394.

YANEZ, M., CATALAN, V., APRAIZ, D., FIGUERAS, M. & MARTINEZ-MURCIA, A. 2003. Phylogenetic analysis of members of the genus *Aeromonas* based on *gyrB* gene sequences. *International Journal of Systematic and Evolutionary Microbiology,* 53**,** 875-883.

YAP, W. H., ZHANG, Z. & WANG, Y. 1999. Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *Journal of Bacteriology,* 181**,** 5201-5209.

YASOTHORNSRIKUL, S., DAVIS, W. J., CRAMER, G., KIMBRELL, D. A. & DEAROLF, C. R. 1997. viking: identification and characterization of a second type IV collagen in Drosophila. *Gene,* 198**,** 17-25.

YEATES, G., BONGERS, T., DE GOEDE, R., FRECKMAN, D. & GEORGIEVA, S. 1993. Feeding habits in soil nematode families and genera—an outline for soil ecologists. *Journal of Nematology,* 25**,** 315.

YELTON, A. P., THOMAS, B. C., SIMMONS, S. L., WILMES, P., ZEMLA, A., THELEN, M. P., JUSTICE, N. & BANFIELD, J. F. 2011. A semi-quantitative, synteny-based method to improve functional predictions for hypothetical and poorly annotated bacterial and archaeal genes. *PLoS Computational Biology,* 7**,** e1002230.

YORKE, W. & MAPLESTONE, P. A. 1926. The nematode parasites of vertebrates. *The Nematode Parasites of Vertebrates.*

YU, X.-J., ZHANG, X.-F., MCBRIDE, J. W., ZHANG, Y. & WALKER, D. H. 2001. Phylogenetic relationships of *Anaplasma marginale* and 'Ehrlichia platys' to other Ehrlichia species determined by GroEL amino acid sequences. *International Journal of Systematic and Evolutionary Microbiology,* 51**,** 1143-1146.

ZHENG, L., DONOVAN, W. P., FITZ-JAMES, P. C. & LOSICK, R. 1988. Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. *Genes and Development,* 2**,** 1047-1054.

ZILVERSMIT, M. M., VOLKMAN, S. K., DEPRISTO, M. A., WIRTH, D. F., AWADALLA, P. & HARTL, D. L. 2010. Low-complexity regions in *Plasmodium falciparum*: missing links in the evolution of an extreme genome. *Molecular Biology and Evolution,* 27**,** 2198-2209.

ZODY, M. C. 2007. Comparative genomics: Identifying functional elements through conservation. *Computational Genomics: Current Methods.*

Appendix I a: R-script for hierarchical clustering of selected CLPs based on GXY repeats

################ Installing and loading required packages ###############

```
if (!require("gplots")) {
  install.packages("gplots", dependencies = TRUE)
  library(gplots)      }
if (!require("RColorBrewer")) {
  install.packages("RColorBrewer", dependencies = TRUE)
  library(RColorBrewer)       }
library("ape")
```

############# Reading in data and transform it into matrix format #############

```
data <- read.csv("GXYpercentcompositionmatrix.csv", comment.char="#")
rnames <- data[,1]                 # assign labels in column 1 to "rnames"
mat_data <- data.matrix(data[,2:ncol(data)])  # transform column 2-5 into a matrix
rownames(mat_data) <- rnames           # assign row names
rownames(mat_data)
```

############# Customizing the data and parameters for heatmap #############

```
# creates a own color palette from red to green
my_palette <- colorRampPalette(c("red", "yellow"))(n = 299)
# (optional) defines the color breaks manually for a "skewed" color transition
col_breaks = c(seq(-1,0,length=100),  seq(0,0.8,length=100), seq(0.8,1,length=100))          # for red, yellow and green
respectively
distance = dist(mat_data, method = "manhattan")
cluster = hclust(distance, method = "ward.D")
cc<-sapply(1:108, function(i) ifelse(substr(row.names(mat_data)[i], 1, 1)=="B",  "skyblue", ifelse(substr(row.names(mat_data)[i],
1, 4)=="Clos", "blue",ifelse(substr(row.names(mat_data)[i], 1, 5)=="Paeni",  "orange", ifelse(substr(row.names(mat_data)[i], 1,
6)=="Viridi",  "orange",ifelse(substr(row.names(mat_data)[i], 1, 4)=="Sedi", "orange",ifelse(substr(row.names(mat_data)[i], 1,
5)=="Pitho", "green4", ifelse(substr(row.names(mat_data)[i], 1, 4)=="Mega", "green",ifelse(substr(row.names(mat_data)[i], 1,
4)=="Ppcl", "red", ifelse(substr(row.names(mat_data)[i], 1, 3)=="Pcl", "pink", "black")))))))))))
```

######################### Heatmap generation #########################

```
heatmap <- heatmap.2(mat_data,       density.info="none",  # turns off density plot inside color legend

                                     trace="none",        # turns off trace lines inside the heat map

                                     margins =c(12,9),    # widens margins around plot

                                     col=my_palette,      # use on color palette defined earlier

                                     dendrogram="row",    # only draw a row dendrogram

                                     Colv="NA",           # turn off column clustering

                                     RowSideColors = cc)
```

Extracting the tree from the heatmap and exporting in Newick format

```
tree <- as.phylo(cluster)
write.tree(phy=tree, file = 'GXYcomposition.tree')
```

Appendix I b: Heatmap for the cluster analysis of G-X-Y region in Ppcl sequences with those in selected CLPs based on the percentage amino acid composition of their low complex G-X-Y repeat regions
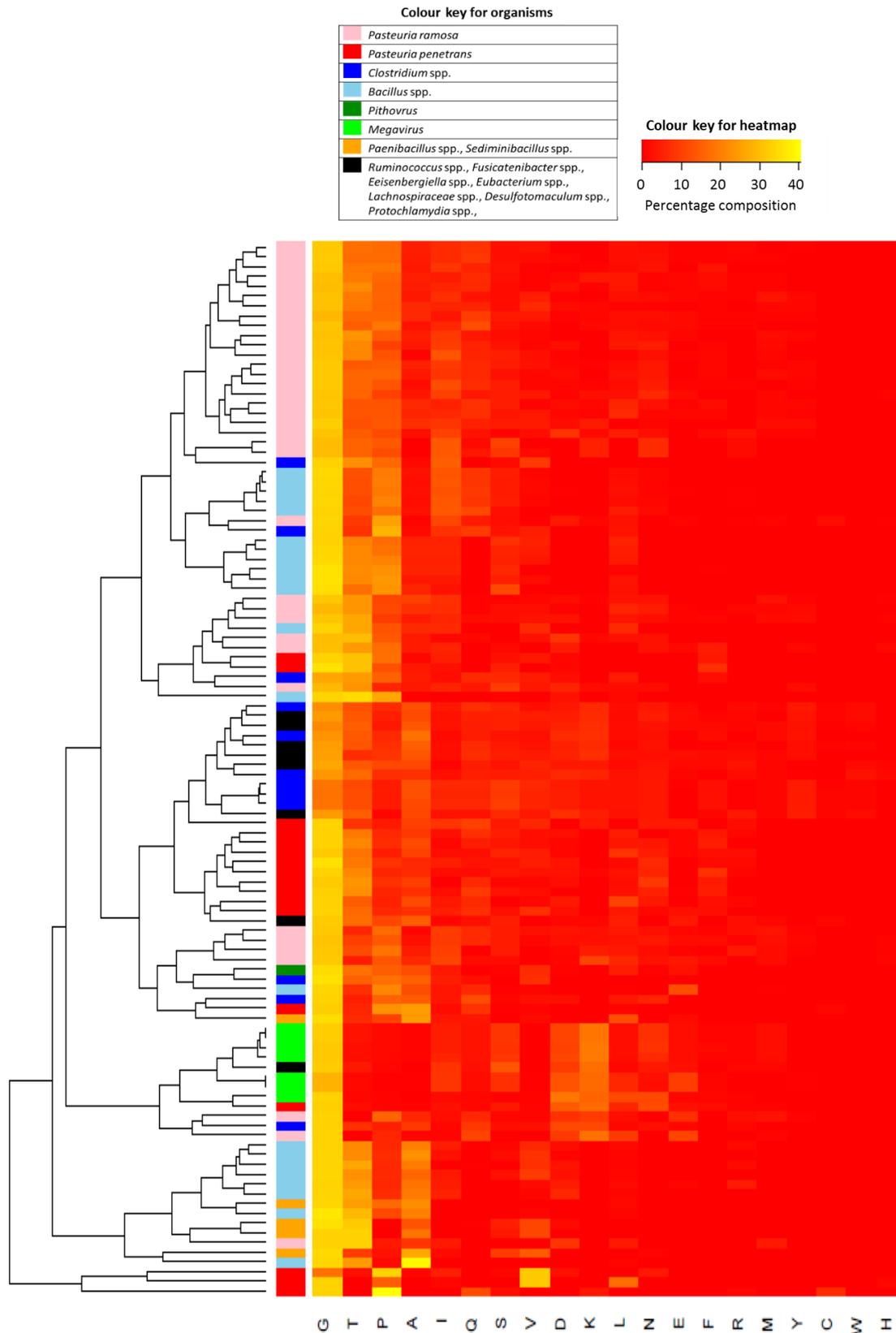
. See Appendix III for list of CLP sequences used for this analysis.

Appendix II: Amino acid sequences of 23 putative collagen-like proteins in *P.penetrans* Res148

>Ppcl1
MSNLELLHRLCCSCCPTAPTGPTGPTCVCPPGPPGPVGPQGPQGPAGPQGPPGPAGSTCVCPPGPTGPPGPQGP
AGTPGAQGPQGPQGPQGPQGAQGLQGPAGTPGAAGPAGTPGAAGPAGTPGAAGPAGTPGAAGPAGPAGTPGAPG
TPGTPGAAGPAGAAGAAGPAGAAGTPGAAGPAGPAGAAGPAGAQGLPGPAGAAGPAGAAGPAGAAGPAGAAGPA
GPAGAAGPAGPAGAAGPAGPAGATGPQGPQGPQGIQGIQGIQGIQGIQGTTGATGPCCAVSSAYYQIISFDSTL
PITNIIYENGPNIRSFIGSTVISLEPNRIYEVIVGFTNGSNGVLGVIHPTLNGVFLPFNVGTPINQELERSFLV
PTPPGGNSLLSLVITRNQGFLAGRQVVVIELPSGN

>Ppcl8
MPNHSGLRGSPLDPEISPKRPIQEFTPIVGFCRSSSCGSTGLVGFCVSTDSCGFIGPSVFCIPTGPSGLCVSGV
PTDFGGPMDPAAGPCALGIFADPSGPTDFAGLCVPNVFTVLVSPVGPVSPVGLVGPVGLVGLVGPVGLVGLVGP
VGLVGLVGPVGLVGPVGPVGPVGLVGLVGLVGPVGLVGLVGPVGLVGPVGPVGPVGLVGPVGHIIFIGFVGFTG
PVDPLELVVFFGPVGFGGFVGLVENRGGL

>Ppcl9
MISVVVTMTSPLPNIVPLGVTRVIGVVVFGASVWMSPAVEGITVTSVVMTGPVGPVTPVTPVIPVIPVIPVIPV
GPVDPVAPGSPVGPVGPVGPVNPVNPVGPVGPVGPVGPVDPVGPVGPVNPVGPVMMLILHMEGKPQGGNKFKIM
YIYPFLNKFLLFYLFFLKVNSDKITFFYDDDCRESFKKRISWGFGSLGRNLDPRDYPMFPPRTSRSPHAEMDYL
P

>Ppcl16
MYHNDYQGKMSDYYKEYYDFYYDYYKNIENNDCKHRGDKNYHNYHQQQPHHEQQCCPPPQECCPIVGPPGPPGP
PGSRGCPGPQGSPGPQGCPGKQGPPGPQGCPGPQGCPGKQGPPGPQGCPGKQGPPGPQGCPGPQGPPGPPGPPG
PPAPPCPPPCPPPCPPPPPPCPPPCPPPPPPCPPPCPPPCPPPCPPPPYPHREY

>Ppcl17
MKRSTKYPFLAMLALVSTVRAQDCPSFAVVTPQCATALNTLQALISPAQAGAPITGDTYMKVEDAHALIKEAFN
VLNIIESDQSNYQKIAQSLHEYQNDLDNGLHVVDAGYVSTRNCKVVDSVCVKNSLGVGGNAQVRGQVSAGNVVA
SNVVTSSATADLVSAFLLDGQIINASSIFVSEINGIPLSASGLNIFGTRGATGPQGTTGNTGPDGNTGPGSTGA
AGAVGLTGNTGIQGITGNTGPVGIVGSTGPDGGFSGAPAGYLNRFNNVSAPVGSPGGFAPITFASSALAENGWT
TVNNITFTCQIAGLYLISFSGNFLLPLVDQFARIGMRIVVSVGDNQAYYGRVNETSPIAGGISTFLTNSALVNC
IVGTTVRVEMTASVADVILFNPALQSTVNTQGQAANLIIRRVF

>Ppcl18
MKIKTLLLLFILGSISFHLSAQEEHLQKICVDKELIDALKTIKSTCRGIISPAIFATCDAYDYGYTQLPSCMVSS
AIDDALFYVEQLKNDQAQLLKTFFQSYQLLLQDRSSKKNKVINNLIVQESAKVKNLVVGGSIIGSFAGTNSVVN
GVTGLTGFTGNTGSQLLARGATGSSGETGNTGATGSTGVTGLIGNIGAAGALGATGPTGVTGLTGFTGPRGDAG
LQGPSGNTGAPGVRLPFAYRFNTTNPLTFTRGQVLSMPNVIFFGMTASTPDTVIFNTSGVFEVFYIINGFRATA
GNVPLSPEVQIYAINAAGNIIDGSTYGYRILNAFPVDFGQQSICGQFIMQANAGDSVRLVNNTGAPTNLLNVNA
QAGNAATTTSISMYVRQIA

>Ppcl19
MIMKAILNIYLIFGYVFCFQLIAMHATEQILVPEYVKHVLLDIKDNNCIRVYDNAFDELCTALKSSDNVDMSVI
VAGINHLLAFENSLPEYSQELHSLKHYKECISTDEQSCVVSKSKCKEYCRLCAETLKVAGNLCVGGLICAPEVS
DPSTVGLDGARGPQGNTGALGQTGDTGITGLTGLTGLIGALGAQGAQGQTGDTGFTGLTGFTGPVGNAGNPGIA
GPLGAANLQTGITGPTGASVTPQAYAMFLVTGVSGLVVPINTGINFAGSIPTVPVGLSLNGDTITIAEPGTYEI
TYIVTEGGGSGLQLGLLVNGVVDPNFNRVSNNPYSQMYGQGLLSVTQPNTQIVLFNFFLLTLNGNLGGDNLGTA
ASLLIKRIAS

>Ppcl20
VRGNARIGGNLIVCGTICPDPRTGQAAAGDPGATGATGSTGFTGFTGPQGAVGALGAAGQTGNTGPTGSTGFTG
PIGTQGAAGAQGETGNTGSTGLTGFTGLTGNVGPAGTPANESLYASYYFTGATSVTGGAATDAVGAASVPFTTQ
GPVNGFNLIGGTDIQVTTTGIYELTFQVLTFQEANLFAVTVNGVIVTRYTSGQKTSSPYGLGRILVNANAGDII
NIKNIGVALSPSVVVSVPLSYGSGDRATASVMIRQIF

>Ppcl21
MLEFHLPESSYIKDTVAISDEANLVIGSFATPEHESCALCFDLHGNFLSYIVAQDKVTKKACKIFNNVCVRNNV
KICGDLLVCGRIINPDCFIRRSITECSNGITGATGPQGNTGPAGLSITGAQGNTGPAGNTGNTGPTGPQGQIGL
QGNTGSAGATGIQGEPGNTGTTGPTGLQGLQGTTGNTGLQGIQGNTGATGATGNTGSIGAQGSTGPVGATGFTG
ATGFTGPQGQTGNTGFTGLTGATGAVGNTGFTGPQGQTGFTGPQGATGFTGPTGFTGVQGPQGNTGFTGAQGNQ
GSTGSTGFTGPTGFTGATGIQGNTGLTGVTGAQGIQGNTGLQGSTGFTGATGIQGSTGLQGSTGFTGATGFTGP
TGFTGATGNTGFTGPQGAAGAVGAQGPQGTTGNTGLQGTTGFTGSTGYTGFTGPQGPTGVTGVTGPQGAVGSTG
FTGATGDAQGQTGNTGPQGATGNTGPQGVTGSTGNTGERGNTGPTGTALEPANFVSSYSLISATAPQNAYTGIQF
DQNAVPPAGWTKTADTFICNQAGIYEMSYVITAAAVTGFNTQLVYSRILQNGVTVVPGSVISVSWNSTNTGNLT
NILAQTVLVSASVNDVFQLQFGATNANAIFVRPITTGGLGTSSGASFTIRRVA

>Ppc123
MLAVLLSAPLCAQEFDNCDELSLAEQRYIASVQEEGVFIPEFISYSIQTVAKFFPRAVYNDSFLELRCHAKDNV
RIIALSKAIRAVDQALQLAAQREELADVVAYLTAYKSQILSGQAIISCDRKRLKTYCQACARILKVCNLTVCGA
ISIPGLSSSSGSLGGNGAASLGDPGATGFTGATGLTGLTGLTGFTGAQGAVGAQGEPGQTGITGQTGPTGSVGS
RGLQGPTGAQGDSGTTGFTGPIGSALADDAYGYFYYNSSSSVGNGSPVPLNVQGTVVGTSLQANGVINIVPAGT
YKISFAVATNNPNKFGLSINGAAPEARTIFAQGASNAMNFGEVILTVGANTNVQVVNATGSSILVWGLLGGAGA
PPNSISASVLVRRIA

>Ppc124
MNEVTQLSQADYDLIAEAADKSFVYIPDGVKAALSEANTEKSMSSDTLQLMKWLEQGNHLYSMSDTKNIFKEYK
DIPATNQHELCLVHFAQLVDQEDSLVVIDDSIDRKKPKVYTNLLAHILQIGGNLIVNGLINVPSAGTISGGATG
ADGFTGAKGPTGPIGQTGNTGPDGFTGFTGFTGNTGFLGNTGVTGATGPNGSTGFTGFTGFTGNIGATGSTGST
GAVGDTGPVGLRGAQGLTGNTGPSSSITGSTGQTGALLGAAAAYGFFYTTGAVTGVAETVTLLSAFTFRGSIPA
VTPGIANGGSIITIADAGTYKFNYIVQGVTSSVIGLQVNFVTDPVTIFAQSSANSQNVGEAILTVPAGASIRLM
NLNTANLNVLNSLGGDVQGTAFSLMIRRLN

>Ppc125
MKKIIIYLLLISSVVQSTITASFLLPHDLVSALQEIVRLSGQEVLSPALQLLCSHIEEHDYVVSEDIVGHALPQ
ALQYVQDHKDQLQSDRYQAALDAINQTMSDIEHNNVVTVDDATTRKHVHNKAFDNLIVRGNVKVRDLIVCGALN
GTVNCGTGIGINGQTGISITGVQGLQGVAGATGDTGATGVTGDTGLTGVIGPVGIAGAQGEQGQTGPTGITGLS
GFTGNQGPTGIPGSTGFTGEPGVALSYAYIYNTAPQTAVSFVAFNSNGVLQNITHPLNDIDITVNNTGVYEITY
EVMAINTNGGIGVNNNAQFYLTINGADQNDTRNGVGFGDDPNTPLQIRGQAVFTLNAGDIVNLAARTGTFSVDL
VPPQTGFAGSSINASILIRQIS

>Ppc126
MASLNKVRVQLLDTITGAVLQEVDVLTSADAVTFSDGQTFQQKLDSGLLKGATGAQGVQGIQGEAFSISKVYSS
VSAMNSGYATDGVAQGKFVLIDTGNINDTDNAKLYVKGTSSYTYLTDLSGANGIQGPQGPQGIQGVQGIKGENG
ATGSTGAVGATGAKGEIGTRGSVWYSGTLITGTSTTATTFSGSGITSALVNDQYFNTSTANVYVCTIDGNANTA
KWVYSICLKGATGATGAQGIQGAKGDVGATGATGETVRVGTTYSTATQANLFFKLV

>Ppc128
VILNLFPPCGFPSVCNIPIITGPTGFTGPTGFTGPTGSTGPTGFTGSTGSTGPTGFTGFTGFTGPTGPTGPTGD
PGATGSTGGTGITGITGATGATGPGLNTDVTIVAGGGADNQIITPTPNPGPGATGTAVILQTGNGQIYGSGDIL
LIGTTDILLPSTGTYLMSFHIDANYSTTTGTAPAAGAQGSYVAYFRQFTADFFFNQIVAFWVGPALFGDAFDSS
ISNTVLGCVRDIPPHGLNNHFRLEIQFSSFNPAVPVNLNVQLPSTMVTITKYSDSICS

>Ppc129
MILNLFPPCGFPSICNISIITGPTGFTGPTGPTGSTGPTGPTGPTGPTGFTGFTGPTGPTGPTGDPGATGSTGP
TGITGITGITGITGVTGVTGPTGPVITTDVTVIPSTAGDIQTLAPNTTTPITLVTPSGTIFGSGDVIVTTTDII
LPRTGTYVVSFSLQGRLASNVPGTAGDIRAFILQSNTGASFTQALGTNGDSLGVATGAVVQPNVPSSGISSTGL
ACVSDTGGGNLSNILQLIVAWASGPPLAGIPTSLDVLLFETVVTVFQYSTNICISQ

>Ppc130
LLIGGNLFVNGTIIQPAQASVSSGGATGPNGATGPQGATGPNGPTGPTGDTGQTGFTGFTGNQGATGATGNTGN
TGPQGQTGNTGFTGNAGAQGATGNTGPDTFTGLTGASGVLLGAAASYGYFYTTGLATGVSGTILAGSPATFAGS
LPVVTPGLVLGGGGTTITVAAAGTYKLSYIVQGITSSVFGLRINGVDDPTTVFAQSSLNAQDIGEVIITLPAGA
VVQLINRDTANLSVLNSMGGDQQGTAFSLTIIRLN

>Ppc133
MSRSQNNIINYVDNRINRLYDRINSERFSAIIPGPRGIKGENGKIGLKGNNGDKGNNGNTGQKGNFGNDGLKGD
KGDSGNDGLKGDKGDFGNDGLKGNKGDLGNDGLKGDKGEVGNDGLKGDKGDLGNDGLKGDIGSKGDIGSKGDTG
SKGESGSTAEKGDKGNDGNKGNDGNKGDKGMQGKGFKYRCNYDPCKIYYYNDVVTVNTCCCNSIYVYTCKTKTG
PLSRCTNICNVPGWKLMLKICNKNPCDINSDSSDSDSCIDTCVKQLMCETNYPNIKINNDNYRNICLPICNNNY
LAKNNYLKYTGIWRRNYYYDIGDLTMLNNIIYIATTNNNDEIPHNGSLYWDVFINNNVIYQGIWKMQKTYSVNN
IVIHNGYTYISIDYPPKGTDIQNQNYWALLNFDNNTNNNNNTNKDNNIIETYNGSDIISSIDINSNESCSIIKS
PSDALYYAYINEDVLIPALNSLHVFDLKFEKIENNINIIPEEQYIIFFNPGTYKITLNINFSGINNLTSNGYIY
AKNNLQEVKEIIPANNEIIFNKTLNNTLQHIFPVTITYPNSKIWLRINWSNQSDIITIKSQKTWILIEQIY

Appendix III: List of CLP sequences used for comparative studies with *Pasteuria* PpcI sequences

Protein	Organism	Accession Number
Pcl18	*P.ramosa*	ADU04102.1
Pcl38	*P.ramosa*	ADU04122.1
Pcl37	*P.ramosa*	ADU04121.1
Pcl36	*P.ramosa*	ADU04120.1
Pcl35	*P.ramosa*	ADU04119.1
Pcl34	*P.ramosa*	ADU04118.1
Pcl33	*P.ramosa*	ADU04117.1
Pcl32	*P.ramosa*	ADU04116.1
Pcl31	*P.ramosa*	ADU04115.1
Pcl30	*P.ramosa*	ADU04114.1
Pcl29	*P.ramosa*	ADU04113.1
Pcl28	*P.ramosa*	ADU04112.1
Pcl27	*P.ramosa*	ADU04111.1
Pcl26	*P.ramosa*	ADU04110.1
Pcl25	*P.ramosa*	ADU04109.1
Pcl24	*P.ramosa*	ADU04108.1
Pcl23	*P.ramosa*	ADU04107.1
Pcl22	*P.ramosa*	ADU04106.1
Pcl21	*P.ramosa*	ADU04105.1
Pcl20	*P.ramosa*	ADU04104.1
Pcl19	*P.ramosa*	ADU04103.1
Pcl17	*P.ramosa*	ADU04101.1
Pcl16	*P.ramosa*	ADU04100.1
Pcl15	*P.ramosa*	ADU04099.1
Pcl14	*P.ramosa*	ADU04098.1
Pcl13	*P.ramosa*	ADU04097.1
Pcl12	*P.ramosa*	ADU04096.1
Pcl11	*P.ramosa*	ADU04095.1
Pcl10	*P.ramosa*	ADU04094.1
Pcl8	*P.ramosa*	ADU04092.1
Pcl7	*P.ramosa*	ADU04091.1
Pcl6	*P.ramosa*	ADU04090.1
Pcl5	*P.ramosa*	ADU04089.1
Pcl4	*P.ramosa*	ADU04088.1
Pcl3	*P.ramosa*	ADU04087.1
Pcl2	*P.ramosa*	ADU04086.1
Pcl1	*P.ramosa*	ADU04085.1
Proto	*Protochlamydia naegleriophila*	WP_059059638.1
Eisen	*Eisenbergiella tayi*	WP_069429780.1
Rumino1	*Ruminococcus torques*	WP_070103488.1
Eubac	*Eubacterium dolichum*	WP_004799529.1
Fusica	*Fusicatenibacter* sp.	CUQ46188.1
Desulfo	*Desulfotomaculum guttoideum*	SEU24404.1
Lachno	*Lachnospiraceae* bacterium.mt14	WP_053983733.1
Rumino2	*Ruminococcus* sp. JC304	WP_019163773.1
Ba.BclA	*B.anthracis*	WP_000069710.1
Bt1	*B.thuringiensis* IBL200	EEM98224.1
Bt2	*B.thuringiensis* serovar israelensis	EAO54405.1
Bt3	*B.thuringiensis* serovar huazhongensis	WP_001288630.1
Bt4	*B.thuringiensis* serovar Berliner	WP_003271363.1
Bc1	*B.cereus*	WP_050567719.1
Bc2	*B.cereus*	WP_059303846.1
Bc3	*B.cereus* AH1273	WP_002079900.1
Bc4	*B.cereus* AH603	EEL67218.1
Bc5	*B.cereus* AH1272	EEL88100.1
Bc6	*B.cereus* AH1273	WP_002079900.1
Bc.ExsJ	*B.cereus* ATCC10876	AAN85822.1
B.acidicola	*B.acidicola*	WP_066270930.1
B.LL01	*Bacillus* sp. LL01	WP_047972494.1
B.JH7	*Bacillus* sp. JH7	WP_061139390.1
B.pumilus1	*B.pumilus*	WP_060597094.1

B.pumilus2	*B.pumilus*	WP_045209866.1
B.safensis1	*B.safensis*	WP_075612411.1
B.wied	*B.wiedmannii*	WP_064459749.1
B.velez	*B.velezensis*	WP_015240391.1
B.weihen1	*B.weihenstephanensis.KBAB4*	WP_012261565.1
B.weihen2	*B.weihenstephanensis*	WP_038626174.1
B.amylo1	*B.amyloliquefaciens*	KJD59259.1
B.amylo2	*B.amyloliquefaciens.EBL11*	WP_032866501.1
Clos1	*Clostridium arbusti*	WP_010235513.1
Clos2	*Clostridium formicaceticum*	WP_070965690.1
Clos3	*Clostridium botulinum*	WP_052705854.1
Clos4	*Clostridium beijerinckii* ATCC 35702	WP_012059564.1
Clos5	*Clostridium tyrobutyricum*	WP_023625409.1
Clos6	*Clostridium pasteurianum* NRRL	ALB45444.1
Clos7	*Clostridium diolis*	WP_039769245.1
Clos8	*Clostridium formicaceticum*	WP_070968717.1
Clos9	*Clostridium paraputrificum*	WP_027098858.1
Clos10	Clostridiales bacterium.VE202-06	WP_049923917.1
Clos11	*Clostridium botulinum*	WP_061312809.1
Clos12	*Clostridium* sp. CAG:265	CDB74255.1
Sedimini	*Sediminibacillus albus*	SDK36124.1
Paeni1	*Paenibacillus swuensis*	WP_068611280.1
Paeni2	*Paenibacillus pabuli*	WP_068964107.1
Paeni3	*Paenibacillus mucilaginosus*	WP_063634302.1
Paeni4	*Paenibacillus* sp. ov031	WP_072735197.1
Pitho	*Pithovirus sibericum*	YP_009001250.1
Mega1	*Megavirus IBa*	AGD92195.1
Mega2	*Megavirus chiliensis*	YP_004894340.1
Mega3	*Megavirus courdo7*	AEX61385.1
Mega4	*Megavirus courdo11*	AFX92324.1
Mega5	*Megavirus IBa*	AGD92520.1
Mega6	*Megavirus chiliensis*	YP_004894639.1
Mega7	*Megavirus courdo11*	AFX92662.1
Mega8	*Megavirus courdo7*	AEX61765.1

Appendix IV: Heatmaps for sequence similarity matrices of selected motifs identified in Ppcl sequences with respect to other CLPs

23_CTD_motif1

	Ppcl23	B.LL01	B.acidicola	Clos1	Bt3	Bc5	Bt4	Bt2	Bc6	B.weihen1	Bc2	B.wied	Bc1	Paeni1	Sedimini	B.safensis1	B.pumilus1	B.pumilus2	B.velez	B.amylo1	B.amylo2
Ppcl23	✕	64.58%	60.42%	61.22%	62.00%	64.00%	64.00%	64.00%	63.83%	63.83%	66.00%	66.00%	68.00%	54.17%	56.25%	62.50%	57.14%	59.57%	59.57%	59.57%	59.57%
B.LL01	64.58%	✕	75.00%	79.17%	75.00%	75.00%	75.00%	75.00%	74.47%	74.47%	75.00%	75.00%	77.08%	75.00%	77.08%	79.17%	75.00%	74.47%	74.47%	74.47%	74.47%
B.acidicola	60.42%	75.00%	✕	77.08%	83.33%	77.08%	77.08%	77.08%	76.60%	76.60%	75.00%	75.00%	77.08%	85.42%	87.50%	83.33%	83.33%	82.98%	80.85%	80.85%	80.85%
Clos1	61.22%	79.17%	77.08%	✕	83.67%	83.67%	83.67%	83.67%	82.98%	85.11%	85.71%	85.71%	85.71%	91.67%	85.42%	85.42%	81.63%	80.85%	82.98%	82.98%	82.98%
Bt3	62.00%	75.00%	83.33%	83.67%	✕	94.00%	94.00%	94.00%	93.62%	91.49%	92.00%	92.00%	92.00%	89.58%	91.67%	91.67%	91.84%	91.49%	89.36%	89.36%	89.36%
Bc5	64.00%	75.00%	77.08%	83.67%	94.00%	✕	100%	100%	100%	97.87%	98.00%	98.00%	96.00%	83.33%	85.42%	87.50%	83.67%	82.98%	85.11%	85.11%	85.11%
Bt4	64.00%	75.00%	77.08%	83.67%	94.00%	100%	✕	100%	100%	97.87%	98.00%	98.00%	96.00%	83.33%	85.42%	87.50%	83.67%	82.98%	85.11%	85.11%	85.11%
Bt2	64.00%	75.00%	77.08%	83.67%	94.00%	100%	100%	✕	100%	97.87%	98.00%	98.00%	96.00%	83.33%	85.42%	87.50%	83.67%	82.98%	85.11%	85.11%	85.11%
Bc6	63.83%	74.47%	76.60%	82.98%	93.62%	100%	100%	100%	✕	97.87%	97.87%	97.87%	95.74%	82.98%	85.11%	87.23%	82.98%	82.98%	85.11%	85.11%	85.11%
B.weihen1	63.83%	74.47%	76.60%	85.11%	91.49%	97.87%	97.87%	97.87%	97.87%	✕	100%	100%	97.87%	82.98%	80.85%	82.98%	78.72%	78.72%	80.85%	80.85%	80.85%
Bc2	66.00%	75.00%	75.00%	85.71%	92.00%	98.00%	98.00%	98.00%	97.87%	100%	✕	100%	98.00%	85.42%	83.33%	85.42%	81.63%	80.85%	82.98%	82.98%	82.98%
B.wied	66.00%	75.00%	75.00%	85.71%	92.00%	98.00%	98.00%	98.00%	97.87%	100%	100%	✕	98.00%	85.42%	83.33%	85.42%	81.63%	80.85%	82.98%	82.98%	82.98%
Bc1	68.00%	77.08%	77.08%	85.71%	92.00%	96.00%	96.00%	96.00%	95.74%	97.87%	98.00%	98.00%	✕	85.42%	83.33%	85.42%	81.63%	80.85%	82.98%	82.98%	82.98%
Paeni1	54.17%	75.00%	85.42%	91.67%	89.58%	83.33%	83.33%	83.33%	82.98%	82.98%	85.42%	85.42%	85.42%	✕	93.75%	89.58%	89.58%	89.36%	87.23%	87.23%	87.23%
Sedimini	56.25%	77.08%	87.50%	91.67%	85.42%	85.42%	85.42%	85.42%	85.11%	80.85%	83.33%	83.33%	83.33%	93.75%	✕	95.83%	95.83%	95.74%	93.62%	93.62%	93.62%
B.safensis1	62.50%	79.17%	83.33%	85.42%	91.67%	87.50%	87.50%	87.50%	87.23%	82.98%	85.42%	85.42%	85.42%	89.58%	95.83%	✕	95.83%	95.74%	97.87%	97.87%	97.87%
B.pumilus1	57.14%	75.00%	83.33%	81.63%	91.84%	83.67%	83.67%	83.67%	82.98%	78.72%	81.63%	81.63%	81.63%	89.58%	95.83%	95.83%	✕	100%	97.87%	97.87%	97.87%
B.pumilus2	59.57%	74.47%	82.98%	80.85%	91.49%	82.98%	82.98%	82.98%	82.98%	78.72%	80.85%	80.85%	80.85%	89.36%	95.74%	95.74%	100%	✕	97.87%	97.87%	97.87%
B.velez	59.57%	74.47%	80.85%	82.98%	89.36%	85.11%	85.11%	85.11%	85.11%	80.85%	82.98%	82.98%	82.98%	87.23%	93.62%	97.87%	97.87%	97.87%	✕	100%	100%
B.amylo1	59.57%	74.47%	80.85%	82.98%	89.36%	85.11%	85.11%	85.11%	85.11%	80.85%	82.98%	82.98%	82.98%	87.23%	93.62%	97.87%	97.87%	97.87%	100%	✕	100%
B.amylo2	59.57%	74.47%	80.85%	82.98%	89.36%	85.11%	85.11%	85.11%	85.11%	80.85%	82.98%	82.98%	82.98%	87.23%	93.62%	97.87%	97.87%	97.87%	100%	100%	✕

VII

23_CTD_motif2

	Ppcl23	Clos1	Sedimini	B.velez	B.amylo1	B.amylo2	B.safensis1	B.pumilus1	B.pumilus2	Bt3	Bt4	Bc5	Bc6	Bt2	B.weihen1	Bc1	Bc2	B.wied	Paeni1	B.acidicola	B.LL01
Ppcl23	✕	64.71%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	80.39%	78.43%	78.43%	78.43%	84.31%	84.31%	84.31%	84.31%	82.35%	80.39%	84.31%
Clos1	64.71%	✕	88.00%	88.00%	88.00%	88.00%	88.00%	88.00%	88.00%	88.00%	88.00%	90.00%	90.00%	90.00%	88.00%	86.00%	86.00%	86.00%	84.00%	90.00%	84.00%
Sedimini	82.35%	88.00%	✕	98.00%	98.00%	98.00%	98.00%	100%	100%	100%	100%	98.00%	98.00%	98.00%	98.00%	98.00%	100%	100%	98.00%	98.00%	98.00%
B.velez	82.35%	88.00%	98.00%	✕	100%	100%	100%	100%	100%	100%	100%	98.00%	98.00%	98.00%	100%	98.00%	100%	100%	98.00%	98.00%	98.00%
B.amylo1	82.35%	88.00%	98.00%	100%	✕	100%	100%	100%	100%	100%	100%	98.00%	98.00%	98.00%	100%	98.00%	100%	100%	98.00%	98.00%	98.00%
B.amylo2	82.35%	88.00%	98.00%	100%	100%	✕	100%	100%	100%	100%	100%	98.00%	98.00%	98.00%	100%	98.00%	100%	100%	98.00%	98.00%	98.00%
B.safensis1	82.35%	88.00%	98.00%	100%	100%	100%	✕	100%	100%	100%	100%	98.00%	98.00%	98.00%	98.00%	98.00%	100%	100%	98.00%	98.00%	98.00%
B.pumilus1	82.35%	88.00%	100%	100%	100%	100%	100%	✕	100%	100%	100%	98.00%	98.00%	98.00%	98.00%	98.00%	100%	100%	98.00%	98.00%	98.00%
B.pumilus2	82.35%	88.00%	100%	100%	100%	100%	100%	100%	✕	100%	100%	98.00%	98.00%	98.00%	98.00%	98.00%	100%	100%	98.00%	98.00%	98.00%
Bt3	82.35%	88.00%	100%	100%	100%	100%	100%	100%	100%	✕	100%	98.00%	98.00%	98.00%	100%	98.00%	100%	100%	98.00%	98.00%	98.00%
Bt4	80.39%	88.00%	100%	100%	100%	100%	100%	100%	100%	100%	✕	98.00%	98.00%	98.00%	98.00%	98.00%	100%	100%	98.00%	98.00%	98.00%
Bc5	78.43%	90.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	✕	100%	100%	98.00%	96.00%	98.00%	98.00%	96.00%	96.00%	96.00%
Bc6	78.43%	90.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	100%	✕	100%	98.00%	96.00%	98.00%	98.00%	96.00%	96.00%	96.00%
Bt2	78.43%	90.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	100%	100%	✕	96.00%	96.00%	98.00%	98.00%	96.00%	96.00%	96.00%
B.weihen1	84.31%	88.00%	98.00%	100%	100%	100%	98.00%	98.00%	98.00%	100%	98.00%	98.00%	98.00%	96.00%	✕	98.00%	100%	100%	98.00%	96.00%	98.00%
Bc1	84.31%	86.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	96.00%	96.00%	96.00%	98.00%	✕	98.00%	98.00%	98.00%	96.00%	96.00%
Bc2	84.31%	86.00%	100%	100%	100%	100%	100%	100%	100%	100%	100%	98.00%	98.00%	98.00%	100%	98.00%	✕	100%	100%	98.00%	98.00%
B.wied	84.31%	86.00%	100%	100%	100%	100%	100%	100%	100%	100%	100%	98.00%	98.00%	98.00%	100%	98.00%	100%	✕	100%	98.00%	98.00%
Paeni1	82.35%	84.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	96.00%	96.00%	96.00%	98.00%	98.00%	100%	100%	✕	98.00%	96.00%
B.acidicola	80.39%	90.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	96.00%	96.00%	96.00%	98.00%	96.00%	98.00%	98.00%	98.00%	✕	94.00%
B.LL01	84.31%	84.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	98.00%	96.00%	96.00%	96.00%	98.00%	96.00%	98.00%	98.00%	96.00%	94.00%	✕

VIII

25_CTD_motif1

	Clos5	Clos2	B.amylo1	B.velez	B.pumilus2	B.safensis	Paeni3	Paeni2	Paeni4	Bt3	B.weihen2	Bc3	B.JH7	Bt2	Bc1	Bc2	B.wied	Pitho	Clos3	Clos1	B.LL01	Ppcl25
Clos5	—	59.26%	44.44%	44.44%	44.44%	48.15%	59.26%	48.15%	48.15%	55.56%	55.56%	55.56%	55.56%	55.56%	55.56%	55.56%	55.56%	55.56%	59.26%	62.96%	55.56%	55.56%
Clos2	59.26%	—	75.00%	75.00%	75.00%	78.57%	82.14%	67.86%	60.71%	78.57%	71.43%	71.43%	71.43%	71.43%	75.00%	75.00%	75.00%	75.00%	75.00%	82.14%	67.86%	60.71%
B.amylo1	44.44%	75.00%	—	100%	100%	96.43%	71.43%	71.43%	57.14%	85.71%	78.57%	78.57%	78.57%	78.57%	82.14%	82.14%	82.14%	67.86%	71.43%	75.00%	75.00%	57.14%
B.velez	44.44%	75.00%	100%	—	100%	96.43%	71.43%	71.43%	57.14%	85.71%	78.57%	78.57%	78.57%	78.57%	82.14%	82.14%	82.14%	67.86%	71.43%	75.00%	75.00%	57.14%
B.pumilus2	44.44%	75.00%	100%	100%	—	96.43%	71.43%	71.43%	57.14%	85.71%	78.57%	78.57%	78.57%	78.57%	82.14%	82.14%	82.14%	67.86%	71.43%	75.00%	75.00%	57.14%
B.safensis	48.15%	78.57%	96.43%	96.43%	96.43%	—	75.00%	75.00%	60.71%	89.29%	82.14%	82.14%	82.14%	82.14%	85.71%	85.71%	85.71%	71.43%	75.00%	78.57%	78.57%	60.71%
Paeni3	59.26%	82.14%	71.43%	71.43%	71.43%	75.00%	—	71.43%	64.29%	78.57%	75.00%	75.00%	75.00%	75.00%	75.00%	75.00%	75.00%	89.29%	75.00%	82.14%	71.43%	53.57%
Paeni2	48.15%	67.86%	71.43%	71.43%	71.43%	75.00%	71.43%	—	85.71%	78.57%	78.57%	78.57%	78.57%	78.57%	78.57%	78.57%	78.57%	75.00%	75.00%	75.00%	60.71%	64.29%
Paeni4	48.15%	60.71%	57.14%	57.14%	57.14%	60.71%	64.29%	85.71%	—	64.29%	64.29%	64.29%	64.29%	64.29%	64.29%	64.29%	64.29%	60.71%	60.71%	60.71%	60.71%	60.71%
Bt3	55.56%	78.57%	85.71%	85.71%	85.71%	89.29%	78.57%	78.57%	64.29%	—	96.43%	96.43%	96.43%	96.43%	96.43%	96.43%	96.43%	75.00%	82.14%	85.71%	71.43%	64.29%
B.weihen2	55.56%	71.43%	78.57%	78.57%	78.57%	82.14%	75.00%	78.57%	64.29%	96.43%	—	100%	100%	100%	96.43%	100%	100%	71.43%	82.14%	82.14%	67.86%	60.71%
Bc3	55.56%	71.43%	78.57%	78.57%	78.57%	82.14%	75.00%	78.57%	64.29%	96.43%	100%	—	100%	100%	96.43%	100%	100%	71.43%	82.14%	82.14%	67.86%	60.71%
B.JH7	55.56%	71.43%	78.57%	78.57%	78.57%	82.14%	75.00%	78.57%	64.29%	96.43%	100%	100%	—	100%	96.43%	100%	100%	71.43%	82.14%	82.14%	67.86%	60.71%
Bt2	55.56%	71.43%	78.57%	78.57%	78.57%	82.14%	75.00%	78.57%	64.29%	96.43%	100%	100%	100%	—	96.43%	100%	100%	71.43%	82.14%	82.14%	67.86%	60.71%
Bc1	55.56%	75.00%	82.14%	82.14%	82.14%	85.71%	75.00%	78.57%	64.29%	96.43%	96.43%	96.43%	96.43%	96.43%	—	96.43%	96.43%	71.43%	78.57%	82.14%	67.86%	60.71%
Bc2	55.56%	75.00%	82.14%	82.14%	82.14%	85.71%	75.00%	78.57%	64.29%	96.43%	100%	100%	100%	100%	96.43%	—	100%	71.43%	78.57%	82.14%	67.86%	60.71%
B.wied	55.56%	75.00%	82.14%	82.14%	82.14%	85.71%	75.00%	78.57%	64.29%	96.43%	100%	100%	100%	100%	96.43%	100%	—	71.43%	78.57%	82.14%	67.86%	60.71%
Pitho	55.56%	75.00%	67.86%	67.86%	67.86%	71.43%	89.29%	75.00%	60.71%	75.00%	71.43%	71.43%	71.43%	71.43%	71.43%	71.43%	71.43%	—	82.14%	85.71%	60.71%	60.71%
Clos3	59.26%	75.00%	71.43%	71.43%	71.43%	75.00%	75.00%	75.00%	60.71%	82.14%	82.14%	82.14%	82.14%	82.14%	78.57%	78.57%	78.57%	82.14%	—	82.14%	57.14%	64.29%
Clos1	62.96%	82.14%	75.00%	75.00%	75.00%	78.57%	82.14%	75.00%	60.71%	85.71%	82.14%	82.14%	82.14%	82.14%	82.14%	82.14%	82.14%	85.71%	82.14%	—	64.29%	64.29%
B.LL01	55.56%	67.86%	75.00%	75.00%	75.00%	78.57%	71.43%	60.71%	60.71%	71.43%	67.86%	67.86%	67.86%	67.86%	67.86%	67.86%	67.86%	60.71%	57.14%	64.29%	—	57.14%
Ppcl25	55.56%	60.71%	57.14%	57.14%	57.14%	60.71%	53.57%	64.29%	60.71%	64.29%	60.71%	60.71%	60.71%	60.71%	60.71%	60.71%	60.71%	60.71%	64.29%	64.29%	57.14%	—

25_CTD_motif2

	Ppcl25	Clos5	B.LL01	Clos2	Paeni3	B.weihen2	Bc3	Bc1	Bc2	Bt3	B.JH7	Bt2	B.amylo1	B.velez	B.wied	B.pumilus2	B.safensis	Pitho	Paeni4	Paeni2	Clos1	Clos3
Ppcl25	—	46.67%	61.29%	48.39%	61.29%	51.61%	54.84%	54.84%	54.84%	54.84%	54.84%	54.84%	58.06%	58.06%	51.61%	54.84%	54.84%	56.67%	58.06%	51.61%	51.61%	54.84%
Clos5	46.67%	—	58.62%	51.72%	48.28%	44.83%	48.28%	48.28%	48.28%	48.28%	48.28%	48.28%	48.28%	48.28%	48.28%	48.28%	48.28%	46.43%	51.72%	44.83%	41.38%	41.38%
B.LL01	61.29%	58.62%	—	76.67%	86.67%	83.33%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	82.76%	80.00%	83.33%	86.67%	83.33%
Clos2	48.39%	51.72%	76.67%	—	80.00%	80.00%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	75.86%	80.00%	80.00%	76.67%	80.00%
Paeni3	61.29%	48.28%	86.67%	80.00%	—	96.67%	100%	100%	100%	100%	100%	100%	100%	100%	96.67%	96.67%	96.67%	89.66%	86.67%	90.00%	90.00%	93.33%
B.weihen2	51.61%	44.83%	83.33%	80.00%	96.67%	—	96.67%	96.67%	96.67%	96.67%	96.67%	96.67%	96.67%	96.67%	96.67%	96.67%	96.67%	89.66%	83.33%	86.67%	86.67%	90.00%
Bc3	54.84%	48.28%	86.67%	83.33%	100%	96.67%	—	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	89.66%	86.67%	90.00%	90.00%	93.33%
Bc1	54.84%	48.28%	86.67%	83.33%	100%	96.67%	100%	—	100%	100%	100%	100%	100%	100%	100%	100%	100%	89.66%	86.67%	90.00%	90.00%	93.33%
Bc2	54.84%	48.28%	86.67%	83.33%	100%	96.67%	100%	100%	—	100%	100%	100%	100%	100%	100%	100%	100%	89.66%	86.67%	90.00%	90.00%	93.33%
Bt3	54.84%	48.28%	86.67%	83.33%	100%	96.67%	100%	100%	100%	—	100%	100%	100%	100%	100%	100%	100%	89.66%	86.67%	90.00%	90.00%	93.33%
B.JH7	54.84%	48.28%	86.67%	83.33%	100%	96.67%	100%	100%	100%	100%	—	100%	100%	100%	100%	100%	100%	89.66%	86.67%	90.00%	90.00%	93.33%
Bt2	54.84%	48.28%	86.67%	83.33%	100%	96.67%	100%	100%	100%	100%	100%	—	100%	100%	100%	100%	100%	89.66%	86.67%	90.00%	90.00%	93.33%
B.amylo1	58.06%	48.28%	86.67%	83.33%	100%	96.67%	100%	100%	100%	100%	100%	100%	—	100%	100%	100%	100%	89.66%	86.67%	90.00%	90.00%	93.33%
B.velez	58.06%	48.28%	86.67%	83.33%	100%	96.67%	100%	100%	100%	100%	100%	100%	100%	—	100%	100%	100%	89.66%	86.67%	90.00%	90.00%	93.33%
B.wied	51.61%	48.28%	86.67%	83.33%	96.67%	96.67%	100%	100%	100%	100%	100%	100%	100%	100%	—	100%	100%	89.66%	83.33%	86.67%	86.67%	90.00%
B.pumilus2	54.84%	48.28%	86.67%	83.33%	96.67%	96.67%	100%	100%	100%	100%	100%	100%	100%	100%	100%	—	100%	89.66%	83.33%	86.67%	86.67%	90.00%
B.safensis	54.84%	48.28%	86.67%	83.33%	96.67%	96.67%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	—	89.66%	83.33%	86.67%	86.67%	90.00%
Pitho	56.67%	46.43%	82.76%	75.86%	89.66%	89.66%	89.66%	89.66%	89.66%	89.66%	89.66%	89.66%	89.66%	89.66%	89.66%	89.66%	89.66%	—	86.21%	89.66%	86.21%	93.10%
Paeni4	58.06%	51.72%	80.00%	80.00%	86.67%	83.33%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	86.67%	83.33%	83.33%	83.33%	86.21%	—	93.33%	86.67%	93.33%
Paeni2	51.61%	44.83%	83.33%	80.00%	90.00%	86.67%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	86.67%	86.67%	86.67%	89.66%	93.33%	—	90.00%	96.67%
Clos1	51.61%	41.38%	86.67%	76.67%	90.00%	86.67%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	90.00%	86.67%	86.67%	86.67%	86.21%	86.67%	90.00%	—	93.33%
Clos3	54.84%	41.38%	83.33%	80.00%	93.33%	90.00%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	90.00%	90.00%	90.00%	93.10%	93.33%	96.67%	93.33%	—

26_NTD_motif1

	Eisen	Fusica	Desulfo	Eubac	Clos8	Lachno	Clos11	Clos9	Ppcl26	Clos7	Clos4	Clos6
Eisen		97.56%	78.05%	75.61%	85.37%	90.24%	85.37%	82.93%	90.24%	87.80%	87.80%	87.80%
Fusica	97.56%		80.49%	78.05%	87.80%	87.80%	87.80%	80.49%	87.80%	85.37%	85.37%	85.37%
Desulfo	78.05%	80.49%		85.37%	87.80%	85.37%	85.37%	82.93%	80.49%	85.37%	85.37%	85.37%
Eubac	75.61%	78.05%	85.37%		87.80%	87.80%	87.80%	87.80%	82.93%	87.80%	87.80%	87.80%
Clos8	85.37%	87.80%	87.80%	87.80%		90.24%	95.12%	87.80%	90.24%	95.12%	95.12%	95.12%
Lachno	90.24%	87.80%	85.37%	87.80%	90.24%		97.56%	95.12%	95.12%	100%	100%	100%
Clos11	85.37%	87.80%	85.37%	87.80%	95.12%	97.56%		90.24%	92.68%	95.12%	95.12%	95.12%
Clos9	82.93%	80.49%	82.93%	87.80%	87.80%	95.12%	90.24%		90.24%	92.68%	92.68%	92.68%
Ppcl26	90.24%	87.80%	80.49%	82.93%	90.24%	95.12%	92.68%	90.24%		95.12%	95.12%	95.12%
Clos7	87.80%	85.37%	85.37%	87.80%	95.12%	100%	95.12%	92.68%	95.12%		100%	100%
Clos4	87.80%	85.37%	85.37%	87.80%	95.12%	100%	95.12%	92.68%	95.12%	100%		100%
Clos6	87.80%	85.37%	85.37%	87.80%	95.12%	100%	95.12%	92.68%	95.12%	100%	100%	

26_NTD_motif2

	Ppcl26	Clos7	Clos4	Clos6	Rumino1	Clos10	Eubac
Ppcl26		93.33%	93.33%	93.33%	96.67%	96.67%	96.67%
Clos7	93.33%		100%	100%	95.00%	98.33%	95.00%
Clos4	93.33%	100%		100%	95.00%	98.33%	95.00%
Clos6	93.33%	100%	100%		95.00%	98.33%	95.00%
Rumino1	96.67%	95.00%	95.00%	95.00%		96.67%	96.67%
Clos10	96.67%	98.33%	98.33%	98.33%	96.67%		96.67%
Eubac	96.67%	95.00%	95.00%	95.00%	96.67%	96.67%	

IX

26_CTD_motif1

	Ppcl26	Clos6	Clos7	Clos4	Rumino1	Eisen	Fusica
Ppcl26		96.67%	93.33%	96.67%	83.33%	88.33%	88.33%
Clos6	96.67%		96.67%	100%	90.00%	91.67%	88.33%
Clos7	93.33%	96.67%		96.67%	90.00%	91.67%	90.00%
Clos4	96.67%	100%	96.67%		90.00%	91.67%	90.00%
Rumino1	83.33%	90.00%	90.00%	90.00%		91.67%	91.67%
Eisen	88.33%	91.67%	91.67%	91.67%	91.67%		95.00%
Fusica	88.33%	88.33%	90.00%	90.00%	91.67%	95.00%	

26_CTD_motif2

	Ppcl26	Clos11	Clos7	Clos4	Clos6	Clos9	Desulfo	Eubac	Lachno	Clos8	Eisen	Fusica
Ppcl26		100%	90.00%	90.00%	90.00%	95.00%	90.00%	90.00%	85.00%	95.00%	85.00%	90.00%
Clos11	100%		85.00%	90.00%	90.00%	95.00%	85.00%	85.00%	85.00%	95.00%	85.00%	95.00%
Clos7	90.00%	85.00%		100%	100%	90.00%	80.00%	80.00%	80.00%	90.00%	85.00%	95.00%
Clos4	90.00%	90.00%	100%		100%	90.00%	80.00%	80.00%	80.00%	90.00%	85.00%	95.00%
Clos6	90.00%	90.00%	100%	100%		90.00%	80.00%	80.00%	80.00%	90.00%	85.00%	95.00%
Clos9	95.00%	95.00%	90.00%	90.00%	90.00%		95.00%	95.00%	90.00%	90.00%	85.00%	90.00%
Desulfo	90.00%	85.00%	80.00%	80.00%	80.00%	95.00%		90.00%	85.00%	80.00%	85.00%	85.00%
Eubac	90.00%	85.00%	80.00%	80.00%	80.00%	95.00%	90.00%		90.00%	80.00%	80.00%	80.00%
Lachno	85.00%	85.00%	80.00%	80.00%	80.00%	90.00%	85.00%	90.00%		80.00%	75.00%	75.00%
Clos8	95.00%	95.00%	90.00%	90.00%	90.00%	90.00%	80.00%	80.00%	80.00%		95.00%	100%
Eisen	85.00%	85.00%	85.00%	85.00%	85.00%	85.00%	85.00%	80.00%	75.00%	95.00%		95.00%
Fusica	90.00%	95.00%	95.00%	95.00%	95.00%	90.00%	85.00%	80.00%	75.00%	100%	95.00%	

Ppcl33 whole sequence

	Ppcl33	Rumino2	Mega7	Mega7	Mega5	Mega6	Mega1	Mega2	Mega4	Mega3
Ppcl33		93.31%	85.56%	66.37%	84.46%	83.78%	31.86%	31.95%	71.28%	63.36%
Rumino2	93.31%		97.01%	74.29%	96.22%	96.84%	27.95%	28.05%	53.02%	50.14%
Mega7	85.56%	97.01%		83.33%	98.47%	98.02%	28.67%	28.75%	49.71%	48.76%
Mega7	66.37%	74.29%	83.33%		81.77%	81.38%	23.10%	23.19%	49.71%	48.34%
Mega5	84.46%	96.22%	98.47%	81.77%		97.42%	28.33%	28.41%	48.84%	48.13%
Mega6	83.78%	96.84%	98.02%	81.38%	97.42%		28.50%	28.58%	48.55%	48.34%
Mega1	31.86%	27.95%	28.67%	23.10%	28.33%	28.50%		99.91%	32.93%	37.60%
Mega2	31.95%	28.05%	28.75%	23.19%	28.41%	28.58%	99.91%		33.08%	37.70%
Mega4	71.28%	53.02%	49.71%	49.71%	48.84%	48.55%	32.93%	33.08%		96.59%
Mega3	63.36%	50.14%	48.76%	48.34%	48.13%	48.34%	37.60%	37.70%	96.59%	

IX

Appendix V: Media, Chemicals and Reagents

(A) Culture Media

Nutrient Broth
Dehydrated culture media from Oxoid (Thermo Scientific)
Steps to prepare:
  i.   13 g of dehydrated media was dissolved in 1 litre de-ionized water.
  ii.   The media was sterilized in an autoclave at 121°C for 15 minutes.

Luria Broth
Tryptone   10 g
Yeast Extract   5 g
NaCl   10 g
De-ionized water   1 litre
The above ingredients were mixed into a solution and autoclaved at 121°C for 15 minutes.

Sporulation broth
Nutrient broth (Oxoid)   8 g
KCl (10% w/v)   10 ml
$MgSO_4.7H_2O$ (1.2%)   10 ml
NaOH (1M)   1.5 ml
$Ca(NO_3)_2.4H_2O$ (1M)   1 ml
$MnCl_2$ (0.01M)   1 ml
$FeSO_4.7H_2O$ (1 mM)   1 ml
        Steps to prepare:
  i.   Nutrient broth, KCl, $MgSO_4.7H_2O$, NaOH, $Ca(NO_3)_2.4H_2O$, $MnCl_2$, $FeSO_4.7H_2O$ were dissolved in 1 litre d.w. and pH was adjusted to 7.6.
  ii.   The above solution was sterilized in an autoclave at 121°C for 15 minutes.
  iii. $Ca(NO_3)_2.4H_2O$, $MnCl_2$ (0.01M) and $FeSO_4.7H_2O$ (1 mM) solutions were made separately in de-ionized water and filter sterilized.
  iv.   The autoclaved and filter sterilized portions of the media were mixed together.

(B) Reagents for SDS-PAGE

Tris/HCl, 1.5 M pH 8.8
Tris base   18.15 g
De-ionized water   100 ml
pH adjusted to 8.8 with HCl
Stored at 4°C

Tris/HCl, 0.5 M pH 6.8
Tris base   6 g
De-ionized water   100 ml
pH adjusted to 6.8 with HCl
Stored at 4°C

Acrylamide Stock, 30%
Readymade
Stored at 4°C

TEMED
Readymade
Stored at 4°C

10% APS (Ammonium persulphate) in de-ionized water
Made fresh

SDS (Sodium dodecyl suphate) 10%
Stored at RT

Electrode Buffer (2X)

Tris	6.06 g
Glycine	28.8 g
SDS	2 g
De-ionized water	1 litre

Steps to prepare:
  i. Tris and Glycine were dissolved in ~800 ml d.w.
  ii. SDS was added to the above.
  iii. Volume was made up to 1 litre.

Sample Buffer

Tris/HCl, 0.5 M pH 6.8	2.5 ml
Glycerol	2 ml
SDS (10%)	4 ml
Bromophenol blue (1%)	0.4 ml
DTT	0.1 ml
De-ionized water	1 ml
Total	10 ml

The sample buffer was made without DTT and stored at RT.
DTT from stock was added freshly before use.

Resolving Gel (For 2 gels)

	5%	10%	12%
Acrylamide stock 30%	1.67 ml	3.33 ml	4 ml
1.5 M Tris/HCl, pH 8.8	2.5 ml	2.5 ml	2.5 ml
De-ionized water	5.67 ml	4.015	3.345 ml
SDS (10%)	100 µl	100 µl	100 µl
APS (10%)	50 µl	50 µl	50 µl
TEMED	5 µl	5 µl	5 µl

Stacking Gel (For 2 gels)

	4%
Acrylamide stock 30%	1.33 ml
1.5 M Tris/HCl, pH 8.8	1.25 ml
De-ionized water	2.34 ml
SDS (10%)	50 µl
APS (10%)	25 µl
TEMED	5 µl

(C)   Reagents for Coomassie Staining

Coomassie Stain, 0.1%

Coomassie Brilliant Blue R-250	5 g
Methanol	2.5 litres
Glacial acetic acid	500 ml
De-ionized water	2 litres

Steps to prepare:
  i. Coomassie brilliant blue was dissolved in methanol while stirring continuously till dissolved completely.
  ii. Glacial acetic acid was added and stirred again.
  iii. Water was added and stirred.
  iv. The solution was filtered through two layers of No.1 Whatman filter paper.

  Stored at RT

De-staining solution

Methanol	500 ml
Glacial acetic acid	350 ml
De-ionized water	150 ml

Stored at RT

(D)   Reagents for Silver Staining

Fixing Solution

Methanol	50%
Acetic acid	10%
De-ionized water	40%

Stored at RT

Washing Solution

Methanol	5%
Acetic acid	7%
De-ionized water	88%

Stored at RT

Glutaraldehyde 10%
1:5 dilution from 50% readymade stock
Made fresh

Dithiothreitol (DTT) 5 µg/ml
10 mg in 2 litres of in de-ionized water
Stored at 4°C

Silver Nitrate (0.1%)
2 g in 2 litres d.w.
Stored at RT

Developer

Sodium bicarbonate (3%)	2 litres
Formaldehyde (37%)	1 ml

Stored at RT

Citric acid (2.3 M)

Citric acid monohydrate	24.2 g
De-ionized water	50 ml

Made fresh

(E)   Reagents for Western Blot

Blotting Buffer

Tris	6.06 g
Glycine	28.83 g
Water	1600 ml
Methanol	~400 ml

Steps to prepare:
   i.   Tris and Glycine were dissolved in water.
   ii. Methanol was added.
   iii. Volume was made up to 2 litres.
      Stored at 4°C

Methanol
To soak PVDF membranes before blotting
Stored at 4°C

Washing Solution (PBS-Tween 1X)

Sodium Chloride	8.0 g

Potassium Chloride   0.2 g
Disodium hydrogen orthophosphate (anhydrous) 1.44 g
Potassium dihydrogen orthophosphate   0.24 g
Tween 80*   0.5 ml

Steps to prepare:
- i   The above ingredients (except tween 80) were dissolved in ~800 ml de-ionized water.
- ii   pH of the solution was adjusted to 7.4.
- iii   Volume was made up to 1 litre.
- iv   Tween 80 added gradually to rapidly-stirred solution

Stored at 4°C

Blocking Solution (2% Tween 80 in PBS)

Washing Solution (1X)   100 ml
Tween 80   1.95 ml
Skimmed Milk Powder*   1%
      Stored at 4°C
Skimmed milk powder to be added just before use.

Substrate for colourimetric detection

BCIP*   2 mg/ml in DEA buffer, 10 mg needed per blot
**Made fresh*
   10% DEA buffer
      Diethanolamine   97 ml
      Magnesium chloride hexahydrate   100 mg

Steps to prepare DEA buffer:
- i.   The above two ingredients were mixed in 800 ml d.w.
- ii.   pH was adjusted to 9.8 with HCl.
- iii.   Volume was made up to 1 litre.

DEA buffer stored at 4°C.

(F)   **Antibodies for Immunodetection**

Primary Antibodies
- Anti-PpWS
  A polyclonal antibody to whole Pasteuria endospore, raised in rabbit
  (Davies *et al.,* 1992)
- Col1981
  A polyclonal antibody to custom synthesized peptide GTPGTPGPAGPAGPA
  (Sigma-Genosys Ltd., UK)
- Col1982
  A polyclonal antibody to custom synthesized peptide GPQGPQGTQGIQGIQ
  (Sigma-Genosys Ltd., UK)

Secondary Antibody
- Anti-rabbit IgG (A3687) Aproduced in goat (conjugated with Alkaline phosphatase) for Western blot
- Anti-rabbit IgG (F0382) produced in goat (conjugated with FITC) for immunofluorescence microscopy
  Both of the above were obtained from Sigma-Aldrich, UK

(G)   **Lectin**
- Wheat germ agglutininin (biotinylated) for western blot
- Wheat germ agglutininin (fluorescein-labelled) for immunofluorescence microscopy
- Wheat germ agglutininin (unconjugated) for blocking NAG residues in indirect immunofluorescence microscopy and attachment tests
  All the above were obtained from Vector Labs, Peterborough, UK

(H)   **Enzymes for Western blot and attachment tests**

Collagenase (C0773) from *Clostridium histolyticum* (Sigma-Aldrich, UK): 1500 units/ml stock

β-N-acetylglucosaminidase (A6805) from *Strptococcus pneumoniae* (Sigma-Aldrich, UK): 50 units/ml stock

(I)   VECTASTAIN® ABC-AP Reagent (Vector Laboratories, UK)

Two drops of Reagent A (Avidin DH solution) was mixed with two drops of Reagent B (biotinylated alkaline phosphatase) in the provided reagent mixing bottle and allowed to stand for 30 minutes at room temperature before use.

(J)   Other Buffers

- **PBS Buffer**
  NaCl   8 g
  KCl   0.2 g
  Na_2HPO_4 1.44 g
  KH_2PO_4 0.24 g
  pH adjusted to 7.4 with HCl

- **PBS-Ca^{2+} buffer**

  PBS with 0.1mM Calcium chloride

Appendix VI: Images from Schaeffer-Fulton Differential staining

XVII
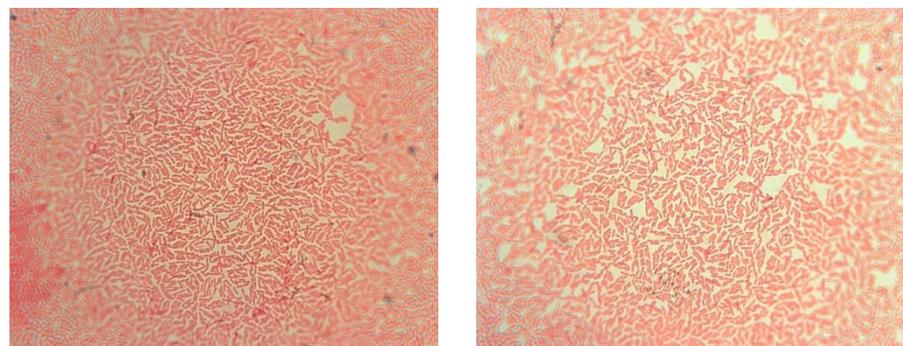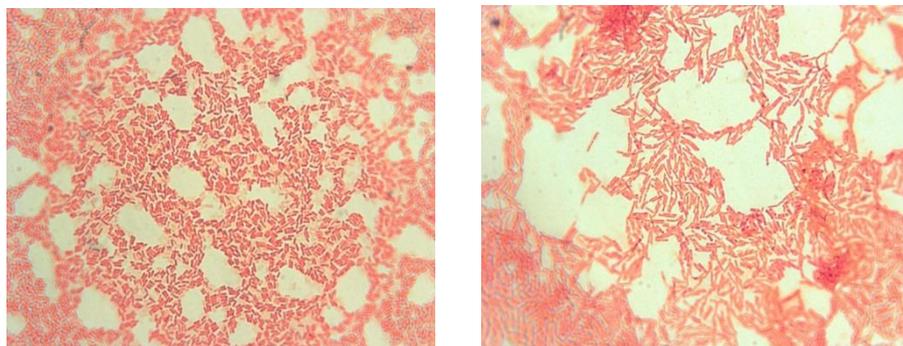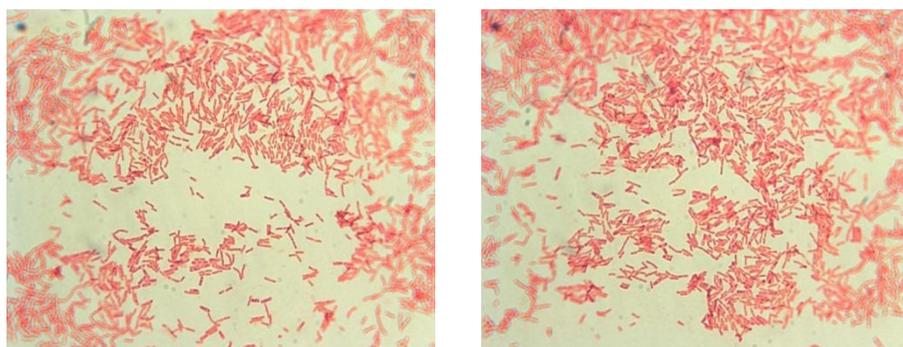
Vegetative cells of *B.thuringiensis* (2 days after inoculation)
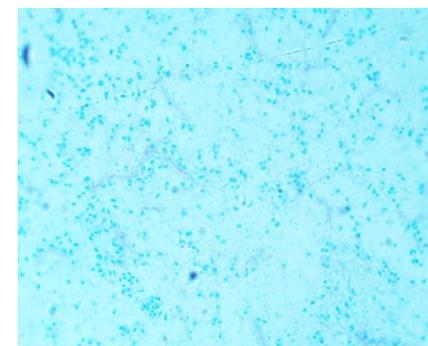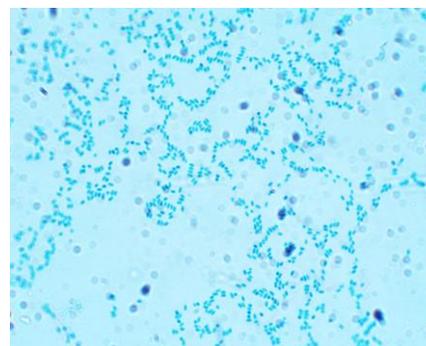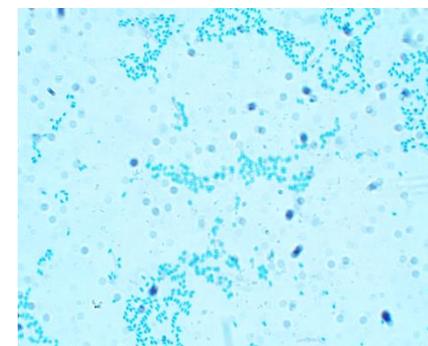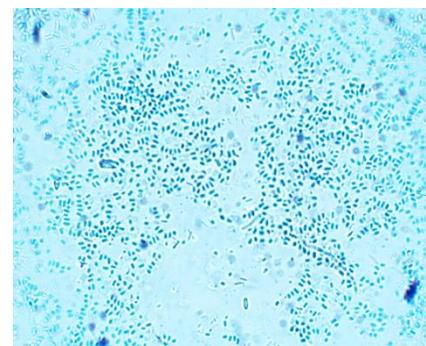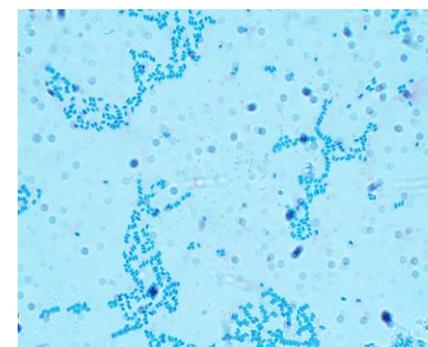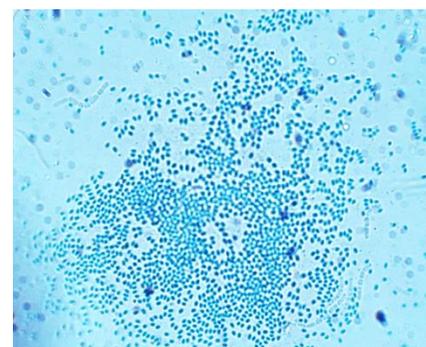
Endospores of *B.thuringiensis* (8 days after inoculation)

Appendix VIIa: Calibration graphs for the determination of Molecular weights of proteins bands in SDS-PAGE gels and Western blots

The graphs were plotted in MS-Excel using the Rf values of molecular markers on x-axis and logarithmic values of their molecular weight (log MW) on y-axis. See Appendix VIIb for the Rf and log MW values and the final calculated molecular weights of the unknown protein bands.

XVIII
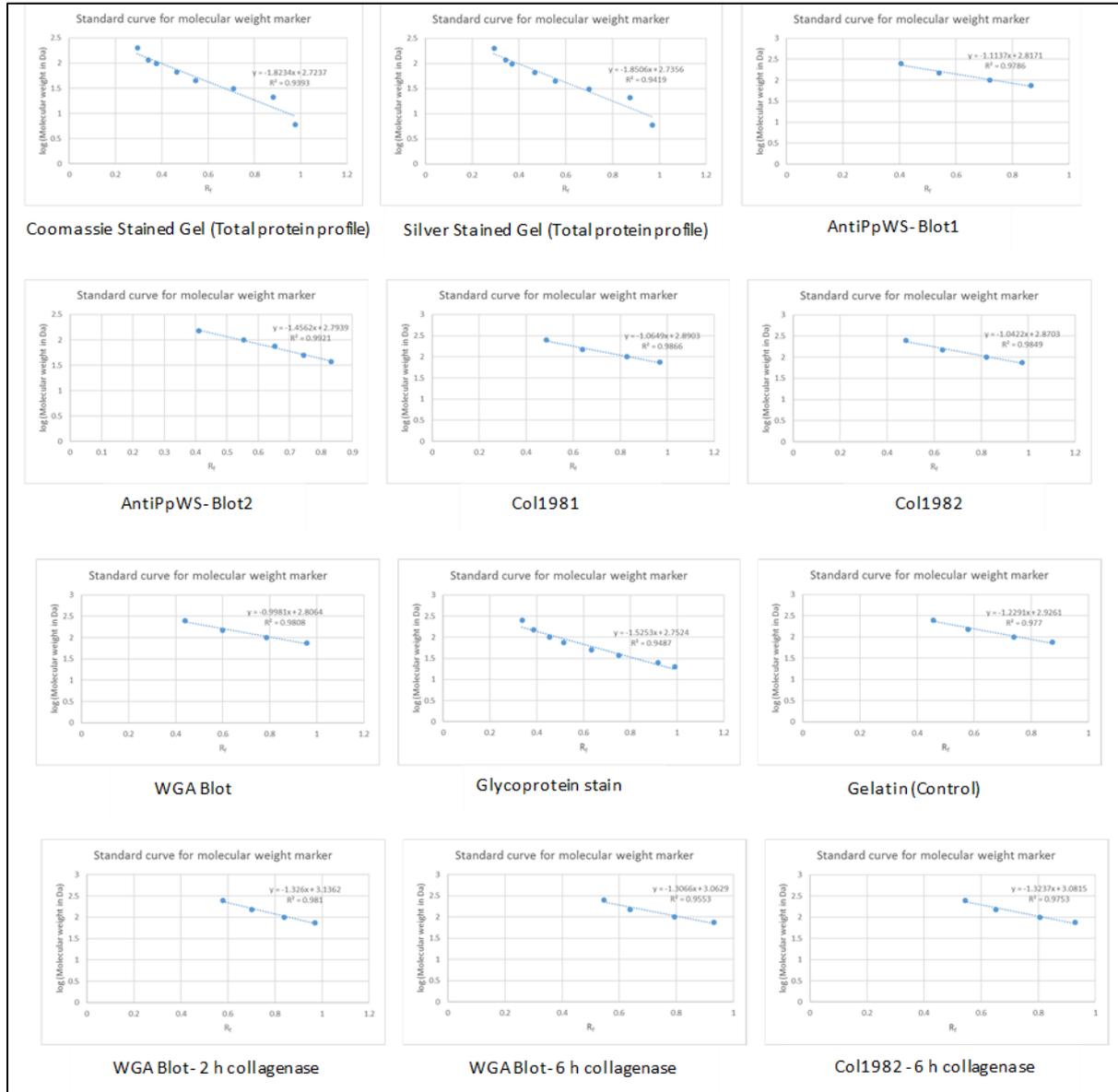
Appendix VIIb: Calculations for the determination of Molecular weights of proteins bands in SDS-PAGE gels and Western blots

The following tables show the distance of migration 'Y' of each band (as measured from the images of the gels and plots), the relative distance of migration (R_f), the molecular weight of each molecular marker, and all the calculations done to determine the molecular weight of the unknown proteins in the endospores of *B. thuringienisis* Al Hakam, *B thuringienisis* kurstaki cry-, *B.thuringienisis* Berliner and *P.penetrans*. The molecular weights that were to be determined are shown as '?'. Calibration graphs (shown in Appendix VIIa) were plotted in MS-Excel using the R_f values of molecular markers on x-axis and logarithmic values of their molecular weight (log MW) on y-axis (values used to plot the graphs are shaded in grey in the tables below). For each graph, a line of best fit was drawn and the straight-line equation derived from it was used to calculate the log MW values for unknown protein bands. Finally, the molecular weights of unknown (shown in bold in the table) were determined from their log MW values.

Table 1: Coomassie Stained Gel (Total protein profile)

S.No.	Lane	Band	Y	R_f	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	227.9	0.294978	200	2.30103	--
2	Molecular marker	Band 2	264	0.341703	116	2.064458	--
3	Molecular marker	Band 3	290.4	0.375874	97	1.986772	--
4	Molecular marker	Band 4	358.9	0.464535	66	1.819544	--
5	Molecular marker	Band 5	422.5	0.546855	45	1.653213	--
6	Molecular marker	Band 6	548.6	0.71007	31	1.491362	--
7	Molecular marker	Band 7	680.6	0.880922	21	1.322219	--
8	Molecular marker	Band 8	754.9	0.97709	6	0.778151	--
9	**BT-Al Hakam**	**Band 2**	359.9	0.46583	?	1.874306	**74.8697162**
10	**BT-Al Hakam**	**Band 4**	392.1	0.507507	?	1.798312	**62.85090268**
11	**BT-Al Hakam**	**Band 5**	419.5	0.542972	?	1.733645	**54.15583418**
12	**BT-Al Hakam**	**Band 8**	452.8	0.586073	?	1.655054	**45.19126427**
13	**BT-Al Hakam**	**Band 10**	535.9	0.693632	?	1.458932	**28.76945319**
14	**BT-Al Hakam**	**Band 11**	555.5	0.719001	?	1.412674	**25.86270725**
15	**BT-Al Hakam**	**Band 13**	687.5	0.889852	?	1.101143	**12.62243228**
16	**BT-Al Hakam**	**Band 14**	701.2	0.907585	?	1.06881	**11.71682422**
17	**BT-Al Hakam**	**Band 17**	740.3	0.958193	?	0.976531	**9.473940997**
18	**BT-kurstaki cry-**	**Band 5**	416.6	0.539218	?	1.740489	**55.01606064**
19	**BT-kurstaki cry-**	**Band 7**	424.4	0.549314	?	1.722081	**52.73280144**
20	**BT-kurstaki cry-**	**Band 16**	722.7	0.935413	?	1.018068	**10.42480964**
21	**BT-berliner**	**Band 1**	342.3	0.443049	?	1.915844	**82.38414613**
22	**BT-berliner**	**Band 2**	350.1	0.453145	?	1.897435	**78.96506527**
23	**BT-berliner**	**Band 3**	377.5	0.48861	?	1.832769	**68.04069311**
24	**BT-berliner**	**Band 4**	384.3	0.497411	?	1.81672	**65.57226203**
25	**BT-berliner**	**Band 5**	411.7	0.532876	?	1.752054	**56.50070878**
26	**BT-berliner**	**Band 6**	419.5	0.542972	?	1.733645	**54.15583418**
27	**BT-berliner**	**Band 10**	540.8	0.699974	?	1.447367	**28.01348896**
28	**BT-berliner**	**Band 12**	670.8	0.868237	?	1.140556	**13.82153993**
29	**BT-berliner**	**Band 15**	707	0.915092	?	1.055121	**11.35328223**
30	**BT-berliner**	**Band 18**	757.9	0.980973	?	0.934993	**8.609803063**
31	**Pasteuria**	**Band 9**	494.8	0.640435	?	1.555931	**35.96921934**

Reference front = 772.6

*the length of resolving gel was used as reference front for the calculation of R_f using the formula shown in Chapter 4, section 4.2.16)

Table 2: Silver Stained Gel (Total protein profile)

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	229.3	0.293974	200	2.30103	--
2	Molecular marker	Band 2	268	0.34359	116	2.064458	--
3	Molecular marker	Band 3	289.3	0.370897	97	1.986772	--
4	Molecular marker	Band 4	365.3	0.468333	66	1.819544	--
5	Molecular marker	Band 5	433.3	0.555513	45	1.653213	--
6	Molecular marker	Band 6	545.3	0.699103	31	1.491362	--
7	Molecular marker	Band 7	682.7	0.875256	21	1.322219	--
8	Molecular marker	Band 8	756	0.969231	6	0.778151	--
9	**BT-Al Hakam**	**Band 2**	234.7	0.300897	?	2.178759	**150.9243**
10	**BT-Al Hakam**	**Band 3**	249.3	0.319615	?	2.14412	**139.3541**
11	**BT-Al Hakam**	**Band 6**	298.7	0.382949	?	2.026915	**106.3935**
12	**BT-Al Hakam**	**Band 7**	333.3	0.427308	?	1.944824	**88.06927**
13	**BT-Al Hakam**	**Band 8**	354.7	0.454744	?	1.894052	**78.35226**
14	**BT-Al Hakam**	**Band 9**	374.7	0.480385	?	1.8466	**70.24254**
15	**BT-Al Hakam**	**Band 10**	406.7	0.52141	?	1.770678	**58.97639**
16	**BT-Al Hakam**	**Band 12**	502.7	0.644487	?	1.542912	**34.90696**
17	**BT-Al Hakam**	**Band 13**	522.7	0.670128	?	1.495461	**31.29398**
18	**BT-Al Hakam**	**Band 14**	549.3	0.704231	?	1.432351	**27.06142**
19	**BT-Al Hakam**	**Band 15**	573.3	0.735	?	1.375409	**23.73608**
20	**BT-Al Hakam**	**Band 17**	629.3	0.806795	?	1.242545	**17.48016**
21	**BT-Al Hakam**	**Band 18**	657.3	0.842692	?	1.176114	**15.00077**
22	**BT-Al Hakam**	**Band 21**	690.7	0.885513	?	1.09687	**12.49885**
23	**BT-Al Hakam**	**Band 22**	712	0.912821	?	1.046334	**11.12588**
24	**BT-Al Hakam**	**Band 24**	757.3	0.970897	?	0.938857	**8.686748**
25	**BT-Al Hakam**	**Band 25**	773.3	0.99141	?	0.900896	**7.95969**
26	**BT-kurstaki cry-**	**Band 1**	221.3	0.283718	?	2.210552	**162.3871**
27	**BT-kurstaki cry-**	**Band 2**	233.3	0.299103	?	2.182081	**152.083**
28	**BT-kurstaki cry-**	**Band 5**	297.3	0.381154	?	2.030237	**107.2103**
29	**BT-kurstaki cry-**	**Band 10**	414.7	0.531667	?	1.751698	**56.45438**
30	**BT-kurstaki cry-**	**Band 12**	505.3	0.647821	?	1.536743	**34.41465**
31	**BT-kurstaki cry-**	**Band 13**	532	0.682051	?	1.473396	**29.74376**
32	**BT-kurstaki cry-**	**Band 15**	566.7	0.726538	?	1.391068	**24.60752**
33	**BT-kurstaki cry-**	**Band 16**	613.3	0.786282	?	1.280506	**19.07684**
34	**BT-kurstaki cry-**	**Band 17**	636	0.815385	?	1.226649	**16.85191**
35	**BT-kurstaki cry-**	**Band 19**	668	0.85641	?	1.150727	**14.14905**
36	**BT-kurstaki cry-**	**Band 21**	693.3	0.888846	?	1.090701	**12.32257**
37	**BT-kurstaki cry-**	**Band 25**	774.7	0.993205	?	0.897575	**7.899045**
38	**BT-berliner**	**Band 1**	220	0.282051	?	2.213636	**163.5445**
39	**BT-berliner**	**Band 2**	234.7	0.300897	?	2.178759	**150.9243**
40	**BT-berliner**	**Band 9**	374.7	0.480385	?	1.8466	**70.24254**
41	**BT-berliner**	**Band 10**	409.3	0.524744	?	1.76451	**58.14462**
42	**BT-berliner**	**Band 12**	505.3	0.647821	?	1.536743	**34.41465**
43	**BT-berliner**	**Band 13**	530.7	0.680385	?	1.47648	**29.95575**
44	**BT-berliner**	**Band 15**	566.7	0.726538	?	1.391068	**24.60752**
45	**BT-berliner**	**Band 17**	630.7	0.80859	?	1.239224	**17.34698**
46	**BT-berliner**	**Band 18**	656	0.841026	?	1.179198	**15.10769**
47	**BT-berliner**	**Band 20**	676	0.866667	?	1.131747	**13.54399**
48	**BT-berliner**	**Band 23 (start of smear)**	686.7	0.880385	?	1.10636	**12.77498**
49	**BT-berliner**	**Band 23 (end of smear)**	758.7	0.972692	?	0.935536	**8.620563**
50	**BT-berliner**	**Band 25**	770.7	0.988077	?	0.907065	**8.073556**
51	**Pasteuria**	**Band 4**	289.3	0.370897	?	2.049217	**111.9998**
52	**Pasteuria**	**Band 10**	405.3	0.519615	?	1.774	**59.42918**
53	**Pasteuria**	**Band 11**	469.3	0.601667	?	1.622156	**41.89437**
54	**Pasteuria**	**Band 16**	609.3	0.781154	?	1.289997	**19.4983**

Reference front = 780

*the length of resolving gel was used as reference front for the calculation of R$_f$ using the formula shown in Chapter 4, section 4.2.16)

Table 3: AntiPpWS

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	232	0.405594	250	2.39794	--
2	Molecular marker	Band 2	309	0.54021	150	2.176091	--
3	Molecular marker	Band 3	412	0.72028	100	2	--
4	Molecular marker	Band 4	495	0.865385	75	1.875061	--
5	**BT-Al Hakam**	**Band 1**	228	0.398601	?	2.373178	**236.1444**
6	**BT-Al Hakam**	**Band 3**	492	0.86014	?	1.859162	**72.30399**
7	**BT-kurstaki cry-**	**Band 1**	227	0.396853	?	2.375125	**237.2054**
8	**BT-kurstaki cry-**	**Band 2**	422	0.737762	?	1.995454	**98.95875**
9	**BT-kurstaki cry-**	**Band 4**	498	0.870629	?	1.84748	**70.38499**
10	**BT-berliner**	**Band 1**	221	0.386364	?	2.386807	**243.6727**
11	**BT-berliner**	**Band 2**	413	0.722028	?	2.012977	**103.0333**
12	**BT-berliner**	**Band 4**	498	0.870629	?	1.84748	**70.38499**
13	**Pasteuria**	**Band 4**	493	0.861888	?	1.857215	**71.98056**

Reference front = 572

*the length of resolving gel was used as reference front for the calculation of R_f using the formula shown in Chapter 4, section 4.2.16)

Table 4: Col1981

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	366.7	0.485888	250	2.39794	--
2	Molecular marker	Band 2	482.7	0.639592	150	2.176091	--
3	Molecular marker	Band 3	624.8	0.827879	100	2	--
4	Molecular marker	Band 4	731.3	0.968994	75	1.875061	--
5	**BT-Al Hakam**	**Band 2**	340.7	0.451438	?	2.409564	**256.7817**
6	**BT-Al Hakam**	**Band 4**	530.1	0.702398	?	2.142316	**138.7765**
7	**BT-kurstaki cry-**	**Band 1**	329	0.435935	?	2.426073	**266.7307**
8	**BT-kurstaki cry-**	**Band 3**	496	0.657215	?	2.190432	**155.0358**
9	**BT-berliner**	**Band 1**	329	0.435935	?	2.426073	**266.7307**
10	**BT-berliner**	**Band 5**	620.9	0.822711	?	2.014195	**103.3225**
11	**BT-berliner**	**Band 6**	651.2	0.862859	?	1.971441	**93.6356**
12	**Pasteuria**	**Band 1**	330	0.43726	?	2.424662	**265.8655**

Reference front = 754.7

*the length of resolving gel was used as reference front for the calculation of R_f using the formula shown in Chapter 4, section 4.2.16)

Table 5: Col1982

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	361.2	0.478601	250	2.39794	--
2	Molecular marker	Band 2	479.4	0.635219	150	2.176091	--
3	Molecular marker	Band 3	620	0.821518	100	2	--
4	Molecular marker	Band 4	735.2	0.974162	75	1.875061	--
5	**BT-Al Hakam**	**Band 2**	490	0.649265	?	2.193636	**156.184**
6	**BT-berliner**	**Band 4**	597.5	0.791705	?	2.045185	**110.9647**
7	**BT-berliner**	**Band 5**	620.9	0.822711	?	2.012871	**103.0079**
8	**Pasteuria**	**Band 1**	353.4	0.468266	?	2.382274	**241.1424**
9	**Pasteuria**	**Band 3**	519.4	0.68822	?	2.153037	**142.2449**
10	**Pasteuria**	**Band 6**	661	0.875845	?	1.957495	**90.67648**
11	**Pasteuria**	**Band 7**	734.2	0.972837	?	1.856409	**71.84713**

Reference front = 745.7

*the length of resolving gel was used as reference front for the calculation of R_f using the formula shown in Chapter 4, section 4.2.16)

Table 6: Glycoprotein stain

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	223	0.337368	250	2.39794	--
2	Molecular marker	Band 2	255	0.385779	150	2.176091	--
3	Molecular marker	Band 3	301	0.455371	100	2	--
4	Molecular marker	Band 4	340	0.514372	75	1.875061	--
5	Molecular marker	Band 5	419	0.633888	50	1.69897	--
6	Molecular marker	Band 6	496	0.750378	37	1.568202	--
7	Molecular marker	Band 7	607	0.918306	25	1.39794	--
8	Molecular marker	Band 8	654	0.98941	20	1.30103	--
9	**BT-Al Hakam**	**Band 2**	211	0.319213	?	2.265504	**184.2909**
10	**BT-kurstaki cry-**	**Band 2**	213	0.322239	?	2.260889	**182.3429**
11	**BT-kurstaki cry-**	**Band 3 (start of smear)**	405	0.612708	?	1.817836	**65.74102**
12	**BT-kurstaki cry-**	**Band 3 (end of smear)**	486	0.73525	?	1.630924	**42.74878**
13	**Pasteuria**	**Band 1**	203	0.30711	?	2.283964	**192.2934**

Reference front = 661

*the length of resolving gel was used as reference front for the calculation of R_f using the formula shown in Chapter 4, section 4.2.16)

Table 7: WGA Blot

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	322	0.438692	250	2.39794	--
2	Molecular marker	Band 2	440	0.599455	150	2.176091	--
3	Molecular marker	Band 3	577	0.786104	100	2	--
4	Molecular marker	Band 4	703	0.957766	75	1.875061	--
5	**BT-Al Hakam**	**Band 4**	488	0.66485	?	2.142813	**138.9355**
6	**Pasteuria**	**Band 1**	247	0.336512	?	2.470527	**295.4793**
7	**Pasteuria**	**Band 2**	335	0.456403	?	2.350864	**224.3179**
8	**Pasteuria**	**Band 3**	412	0.561308	?	2.246159	**176.262**
9	**Pasteuria**	**Band 5**	520	0.708447	?	2.099299	**125.6896**
10	**Pasteuria**	**Band 6**	654	0.891008	?	1.917085	**82.61991**
11	**Pasteuria**	**Band 7**	713	0.97139	?	1.836856	**68.68407**

Reference front = 734

*the length of resolving gel was used as reference front for the calculation of R_f using the formula shown in Chapter 4, section 4.2.16)

Table 8: Gelatin (Control)

Reference front = 757
 *the length of resolving gel was used as reference front for the calculation of Rf using the formula shown in Chapter 4, section 4.2.16)

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	346	0.457067	250	2.39794	--
2	Molecular marker	Band 2	438	0.5786	150	2.176091	--
3	Molecular marker	Band 3	559	0.738441	100	2	--
4	Molecular marker	Band 4	662	0.874505	75	1.875061	--
5	**Gelatin untreated**	**Band 1**	446	0.589168	?	2.201954	**159.204**
6	**Gelatin untreated**	**Band 2**	476	0.628798	?	2.153245	**142.313**
7	**Gelatin treated**	**Band 3**	559	0.738441	?	2.018482	**104.3475**
8	**Gelatin treated**	**Band 4**	672	0.887715	?	1.83501	**68.39272**

Table 9: WGA Blot- 2 h collagenase treatment

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	379	0.578626	250	2.39794	--
2	Molecular marker	Band 2	458.8	0.700458	150	2.176091	--
3	Molecular marker	Band 3	550.1	0.839847	100	2	--
4	Molecular marker	Band 4	635.2	0.969771	75	1.875061	--
5	**BT-Al Hakam (untreated)**	**Band 2**	486.1	0.742137	?	2.152126	**141.9469**
6	**BT-Al Hakam (treated)**	**Band 4**	488.2	0.745344	?	2.147875	**140.5641**
7	**Pasteuria (untreated)**	**Band 1**	364.3	0.556183	?	2.398701	**250.4385**
8	**Pasteuria (untreated)**	**Band 2**	418.9	0.639542	?	2.288167	**194.1634**
9	**Pasteuria (untreated)**	**Band 4**	505	0.770992	?	2.113864	**129.9763**
10	**Pasteuria (treated)**	**Band 1**	363.2	0.554504	?	2.400928	**251.7259**
11	**Pasteuria (treated)**	**Band 2**	418.9	0.639542	?	2.288167	**194.1634**
12	**Pasteuria (treated)**	**Band 4**	505	0.770992	?	2.113864	**129.9763**
13	**Pasteuria (treated)**	**Band 5**	548	0.836641	?	2.026814	**106.3687**

Reference front = 655
 *the length of resolving gel was used as reference front for the calculation of Rf using the formula shown in Chapter 4, section 4.2.16)

Table 10: WGA Blot- 6 h collagenase treatment

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	360.1	0.547264	250	2.39794	--
2	Molecular marker	Band 2	419.9	0.638146	150	2.176091	--
3	Molecular marker	Band 3	521.8	0.793009	100	2	--
4	Molecular marker	Band 4	613.1	0.931763	75	1.875061	--
5	**Pasteuria (untreated)**	**Band 1**	325.5	0.494681	?	2.41655	**260.9456**
6	**Pasteuria (untreated)**	**Band 2**	384.2	0.583891	?	2.299989	**199.521**
7	**Pasteuria (untreated)**	**Band 3**	451.4	0.686018	?	2.166549	**146.74**
8	**Pasteuria (untreated)**	**Band 6**	559.6	0.850456	?	1.951694	**89.47347**
9	**Pasteuria (untreated)**	**Band 8**	616.3	0.936626	?	1.839104	**69.04056**
10	**Pasteuria (treated)**	**Band 4**	475.6	0.722796	?	2.118494	**131.3694**
11	**Pasteuria (treated)**	**Band 5**	505	0.767477	?	2.060114	**114.8456**
12	**Pasteuria (treated)**	**Band 7**	604.7	0.918997	?	1.8621386	**72.80120**

Reference front = 658
 *the length of resolving gel was used as reference front for the calculation of Rf using the formula shown in Chapter 4, section 4.2.16)

Table 11: Col1982 - 6 h collagenase treatment

S.No.	Lane	Band	Y	Rf	M.W (kDa)	log MW	MW of unknown
1	Molecular marker	Band 1	356.9	0.543227	250	2.39794	--
2	Molecular marker	Band 2	428.3	0.651903	150	2.176091	--
3	Molecular marker	Band 3	529.1	0.805327	100	2	--
4	Molecular marker	Band 4	610	0.928463	75	1.875061	--
5	**Pasteuria (untreated)**	**Band 1**	448.3	0.682344	?	2.178281	**150.7583**
6	**Pasteuria (treated)**	**Band 2**	494.5	0.752664	?	2.085199	**121.6744**
7	**Pasteuria (treated)**	**Band 3**	587.9	0.894825	?	1.89702	**78.88968**

Reference front = 657
 *the length of resolving gel was used as reference front for the calculation of Rf using the formula shown in Chapter 4, section 4.2.1

Appendix VIII: Raw data for the calculation of Mean CTCF values in Immunofluorescence Microscopy experiments

Table 1

Microscopic fields	Ensdospore	BTAlHakam Anti-PpWS				BTberliner Anti-PpWS				BTkurstaki cry- Anti-PpWS				Pas Anti-PpWS			
		Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF
Field 1	Spore1	378	45.638	17251	13664.35	192	32.427	6226	4336.72	197	45.244	8913	3370.799	832	29.011	24137	10981.21
	Spore2	256	38.031	9736	7306.944	96	35.208	3380	2435.36	165	45.424	7495	2853.055	832	33.787	28111	14955.21
	Spore3	369	52.881	19513	16011.74	154	35.247	5428	3912.64	156	41.135	6417	2028.252	832	29.805	24798	11642.21
	Spore4	322	59.649	19207	16151.7	143	35.266	5043	3635.88	155	44.619	6916	2555.385	1052	40.19	42280	25645.51
Field 2	Spore1	323	53.263	17204	13739.66	98	27.663	2711	1392.729	163	54.491	8882	3987.599	1023	46.738	47813	29132.76
	Spore2	200	50.33	10066	9855.275	83	27.217	2259	1142.505	156	48.064	7498	2813.788	1023	39.528	40437	21756.76
	Spore3	246	62.423	15356	12717.53	101	29.95	3025	1666.373	212	41.892	8881	2515.276	832	34.238	28486	13293.47
	Spore4	224	47.304	10596	8193.488	159	31.164	4955	2816.172	140	44.864	6281	2077.22	832	33.12	27556	12363.47
Field 3	Spore1	196	56.801	11133	9030.802	132	34.659	4575	3222.561	205	48.99	10043	3884.903	690	28.636	19759	8500.96
	Spore2	183	60.814	11129	9166.234	153	30.68	4694	3126.4	204	49.407	10079	3950.942	780	27.401	21373	8646.52
	Spore3	196	40.765	7990	5887.802	85	32.741	2783	1912.111	153	53.902	8247	3650.957	780	27.641	21560	8833.52
	Spore4	178	51.32	9135	7225.861	164	35.037	5746	4065.697	122	47.533	5799	2134.181	882	32.706	28847	14456.29
Field 4	Spore1	277	66.264	18355	16670.08	139	45.604	6339	4734.627	216	39.491	8530	2334.742	912	31.396	28633	14825.32
	Spore2	304	53.556	16281	14431.84	88	46.045	4052	3036.282	147	44.721	6574	2357.783	832	33.523	27891	15294.52
	Spore3	255	64.388	16419	14867.9	104	45.471	4729	3528.606	184	40.538	7459	2181.558	928	27.623	25634	11584.08
	Spore4	230	58.248	13397	11997.97	52	37.154	1932	1331.803	136	43.794	5956	2055.282	882	36.294	32011	18657.52
Field 5	Spore1	237	63.806	15122	10980.37	123	49.195	6051	4330.876	164	55.341	9076	3244.816	807	41.603	33574	16352.22
	Spore2	212	71.292	15114	11409.25	119	43.261	5148	3483.815	165	54.333	8965	3098.26	912	43.816	39960	20497.46
	Spore3	252	58.841	14828	10424.24	96	45.51	4369	3026.464	116	54.009	6265	2140.504	988	44.784	44247	23162.59
	Spore4	239	60.28	14407	14393.32	110	39.618	4358	2819.678	146	54.534	7962	2770.824	911	45.662	41598	22156.8
Average CTCF		11706.31754				2997.8649				2800.306238				16136.92043			

Mean= Mean gray value; IntDen = Integrated density
CTCF=Integrated Density-(Area of selected endospore ×Mean background fluorescence)

Table 2

Microscopic fields	Ensdospore	BTAlHakam Col1981				BTberliner Col1981				BTkurstaki cry- Col1981				Pas Col1981			
		Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF
Field 1	Spore1	124	36.363	4509	3186.23	165	16.352	2698	1696.986	118	41.076	4847	3446.665	812	41.185	33442	14757.27
	Spore2	129	34.868	4498	3121.893	117	13.564	1587	877.1903	71	45	3195	2352.425	812	38.047	30894	12209.27
	Spore3	92	27.152	2498	1516.59	114	12.956	1477	785.3905	90	37.622	3386	2317.948	812	37.139	30157	11472.27
	Spore4	93	28.645	2664	1671.923	104	16.337	1699	1068.058	135	45.296	6115	4512.921	611	42.1	25723	11663.43
Field 2	Spore1	176	33.102	5826	4042.944	105	12.048	1265	669.0725	106	46.585	4938	3494.307	716	36.024	25793	7351.167
	Spore2	169	34.183	5777	4064.861	109	13.22	1441	822.3705	113	39.796	4497	2957.968	716	37.155	26603	8161.167
	Spore3	152	37.336	5675	4135.088	118	19.941	2353	1683.291	78	41.141	3209	2146.66	716	36.446	26095	7653.167
	Spore4	206	37.762	7779	5692.014	91	17.33	1577	1060.53	110	45.545	5010	3511.828	678	38.124	25848	8384.924
Field 3	Spore1	81	48.506	3929	3112.783	121	16.81	2034	1303.13	76	46.632	3544	2267.295	812	23.006	18681	5614.905
	Spore2	69	40.42	2789	2093.704	158	12.886	2036	1081.641	114	47.991	5471	3555.943	960	27.673	26566	11118.4
	Spore3	71	37.366	2653	1937.551	109	15.826	1725	1066.613	116	45.353	5261	3312.345	711	32.252	22931	11490.12
	Spore4	71	34.408	2443	1727.551	134	11.104	1488	678.6065	109	43.606	4753	2921.936	860	27.755	23869	10030.53
Field 4	Spore1	106	28.321	3002	1933.865	174	9.96	1733	818.195	78	40.205	3136	2383.281	861	35.361	30446	5288.441
	Spore2	135	29.852	4030	2593.195	182	12.462	2268	1311.135	61	34.18	2085	1496.335	861	37.238	32062	6904.441
	Spore3	113	21.708	2453	1250.341	104	13.567	1411	864.22	109	37.284	4064	3012.123	861	34.163	29414	4256.441
	Spore4	102	29.833	3043	1957.414	130	9.485	1233	549.525	122	34.328	4188	3010.67	861	36.375	31319	6161.441
Field 5	Spore1	111	29.865	3315	2175.668	173	12.977	2245	1441.285	105	38.095	4000	2425.709	912	41.479	37829	8370.716
	Spore2	149	27.268	4063	2533.627	99	14.949	1480	1020.071	97	43.402	4210	2755.655	812	46.272	37573	11344.79
	Spore3	140	24.193	3387	1950.005	96	9.354	898	452.008	114	42.526	4848	3138.77	812	44.062	35778	9549.791
	Spore4	130	26.962	3505	2170.648	101	16.178	1634	1164.779	132	41.75	5511	3531.891	812	33.964	27579	1350.791
Average CTCF		2643.39465				1020.704813				2927.633525				8656.673675			

Mean= Mean gray value; IntDen = Integrated density
CTCF=Integrated Density-(Area of selected endospore ×Mean background fluorescence)

Table 3

Microscopic fields	Ensdospore	BTAlHakam Col1982				BTberliner Col1982				BTkurstaki cry- Col1982				Pas Col1982			
		Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF
Field 1	Spore1	145	27.903	4046	2816.291	138	17.261	2382	1609.166	130	29.431	3826	1785.618	668	23.597	15763	4580.179
	Spore2	145	32.117	4657	3427.291	112	23.795	2665	2037.772	128	27.836	3563	1554.008	780	24.1	18798	5740.215
	Spore3	92	27.641	2543	2308.584	98	25.837	2532	1983.176	109	28.505	3107	1396.218	876	23.39	20490	5825.103
	Spore4	156	30.244	4718	3395.003	78	26.372	2057	1620.181	114	25.982	2962	1172.742	812	28.014	22747	9153.511
Field 2	Spore1	105	25.229	2649	1940.303	105	18.933	1988	1528.625	98	26.663	2613	1248.644	832	26.596	22128	7393.904
	Spore2	110	26.855	2954	2211.555	171	17.795	3043	2294.875	109	26.193	2855	1337.502	599	33.152	19858	9250.159
	Spore3	92	24.446	2249	1628.046	120	15.117	1814	1289	126	25.048	3156	1401.828	803	25.234	20263	6042.472
	Spore4	81	26.926	2181	1634.291	83	15.783	1310	946.875	114	25.36	2891	1303.892	689	31.832	21932	9730.327
Field 3	Spore1	98	26.673	2614	1901.516	123	20.423	2512	3104.522	124	24.823	3078	884.223	912	30.221	27562	14570.33
	Spore2	72	26.792	1929	1405.542	141	19.972	2816	2136.768	122	25.254	3081	922.6065	812	31.969	25959	14391.86
	Spore3	70	23.571	1650	1141.083	122	16.484	2011	1423.296	127	23.764	3018	771.1478	812	36.466	29610	18042.86
	Spore4	91	28.198	2566	1904.407	140	18.657	2612	1937.585	79	25.139	1986	588.3518	870	31.479	27387	14993.63
Field 4	Spore1	200	44.645	8929	7361.8	157	23.21	3644	3142.66	130	25.669	3337	1120.435	938	30.284	28406	15740.66
	Spore2	122	33.025	4029	3073.008	89	15.584	1387	1102.801	104	25.442	2646	872.748	938	30.6	28703	16037.66
	Spore3	154	35.383	5449	4242.256	79	18.696	1477	1224.733	107	25.383	2716	891.5965	938	26.767	25107	12441.66
	Spore4	104	35.683	3711	2896.056	66	7.53	497	286.2455	115	23.991	2759	798.1925	994	31.419	31230	17808.52
Field 5	Spore1	112	34.714	3888	3016.388	84	20.369	1711	1351.081	95	25.863	2457	843.4963	780	46.892	36576	20143.94
	Spore2	125	34.384	4298	3325.219	87	19.023	1655	1282.227	108	23.046	2489	654.701	884	35.941	31772	13148.99
	Spore3	88	33.966	2989	2304.162	88	15.864	1396	1018.942	103	25.272	2603	853.6223	884	32.707	28913	10289.99
	Spore4	138	24.42	3370	2296.05	107	15.43	1651	1192.532	98	24.52	2403	738.5435	884	29.759	26307	7683.993
Average CTCF		2711.442429				1625.652963				1057.005738				11650.49714			

Mean= Mean gray value; IntDen = Integrated density
CTCF=Integrated Density-(Area of selected endospore ×Mean background fluorescence)

Table 4

Microscopic fields	Ensdospore	Pas WGA				Pas WGA+Anti-PpWS				Pas WGA+Col1981				Pas WGA+Col1982			
		Area	Mean	IntDen	CTCF	Area	Mean	IntDen	CTCF	882	42.195	37216	13859.98	Area	Mean	IntDen	CTCF
Field 1	Spore1	882	38.474	33934	14517.43	1050	22.947	24094	4537.488	882	35.26	31099	7742.979	860	33.038	28413	8302.115
	Spore2	882	36.347	32058	12641.43	1050	22.743	23880	4323.488	882	38.766	34192	10835.98	860	34.576	29735	9624.115
	Spore3	882	44.41	39170	19753.43	1050	19.442	20414	857.4875	882	32.077	28292	4935.979	860	28.581	24580	4469.115
	Spore4	726	45.842	33281	17298.65	1050	23.329	24495	4938.488	1052	41.997	44181	20463.92	860	30.305	26062	5951.115
Field 2	Spore1	782	34.058	26633	12748.2	1020	22.346	22793	1329.905	1052	44.037	46327	22609.92	749	29.686	22235	4629.006
	Spore2	938	38.203	35834	19179.34	1020	23.899	24377	2913.905	924	37.626	34766	13934.65	749	32.191	24111	6505.006
	Spore3	938	37.305	34992	18337.34	1020	19.072	19453	-2010.1	973	34.406	33477	11540.96	860	32.817	28223	8007.84
	Spore4	749	35.653	26704	13405.13	908	28.728	26085	6978.637	1052	43.651	45921	16150.72	912	29.804	27181	5743.528
Field 3	Spore1	860	38.388	33014	13624.66	938	22.559	21160	4399.347	1052	40.866	42991	13220.72	1106	21.745	24050	5000.533
	Spore2	860	25.291	21750	2360.655	938	23.523	22065	5304.347	1052	47.349	49811	20040.72	1106	24.674	27289	8239.533
	Spore3	860	31.988	27510	27510	800	22.456	17965	3670.2	1052	46.941	49382	19611.72	938	23.587	22125	5969.123
	Spore4	860	24.024	20661	1271.655	800	25.985	20788	6493.2	1023	37.803	38672	10634.13	938	20.495	19224	3068.123
Field 4	Spore1	1020	34.142	34825	16495.6	716	21.27	15229	3501.278	1023	48.273	49383	21345.13	1106	29.741	32894	16322.53
	Spore2	882	48.084	42410	26560.46	716	23.018	16481	4753.278	1162	54.981	63888	32040.49	752	28.876	21715	10447.6
	Spore3	882	32.111	28322	12472.46	716	20.595	14746	3018.278	884	41.743	36901	12672.77	752	26.529	19950	8682.596
	Spore4	882	32.137	28345	12495.46	716	20.856	14933	3205.278	911	42.413	38638	14377.84	752	34.911	26253	14985.6
Field 5	Spore1	861	37.534	32317	14252.14	1052	21.232	22336	3531.237	911	35.175	32044	7783.842	1052	26.413	27787	7774.804
	Spore2	861	41.315	35572	17507.14	1052	26.477	27854	9049.237	911	36.602	33344	9083.842	1052	29.526	31061	11048.8
	Spore3	861	35.829	30849	12784.14	749	28.119	21061	7672.438	691	47.747	32993	14591.5	1052	32.722	34424	14411.8
	Spore4	861	28.468	24511	6446.144	712	20.225	14400	1672.822	882	42.195	37216	13859.98	1232	29.437	36266	12829.66
Average CTCF		14583.07403				4007.012088				14873.88816				8600.626975			

Mean= Mean gray value; IntDen = Integrated density
CTCF=Integrated Density-(Area of selected endospore ×Mean background fluorescence)

Appendix IX: ANOVA and Tukey tests for CTCF values in fluorescence microscopy experiment

R-script for ANOVA and Tukey's *post hoc* test

```
####   start of script   ####
data<-read.csv("filename.csv")
data1<-stack(data)
result<-aov(values~ind,data = data1)
summary(result)
TukeyHSD(result
shapiro.test(result$residual)
####   end of script   ####
## where 'values' are dependent variable (CTCF) and 'ind' are independent variables (bacteria)##
```

Shapiro-Wilk Test for Normality

Fluorescence data: Shapiro-Wilk Test for Normality		
Anti-PpWS	W = 0.92926, p-value = 0.0002678	Not-normal
Col1981	W = 0.87471, p-value = 1.21e-06	Not-normal
Col1982	W = 0.86203, p-value = 4.198e-07	Not-normal
WGA + Anti-PpWS	W = 0.96496, p-value = 0.2465	Normal
WGA + Col1981	W = 0.95226, p-value = 0.09068	Normal
WGA + Col1982	W = 0.96375, p-value = 0.2246	Normal
WGA treated Vs Only Ab	W = 0.97454, p-value = 0.02234	Not-normal

A. Endospores probed with Anti-PpWS followed by FICT-conjugated anti-rabbit IgG

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
ind	3	2.627e+09	875539190	74.09	<2e-16 ***
Residuals	76	8.981e+08	11817662		

> TukeyHSD(95% family-wise confidence level)

	diff	lwr	upr	p adj
BTberliner-AlHakam	-8708.4526	-11564.019	-5852.886	0.0000000
BTkurstaki-AlHakam	-8906.0113	-11761.578	-6050.445	0.0000000
Pasteuria-AlHakam	4430.6029	1575.037	7286.169	0.0006356
BTkurstaki-BTberliner	-197.5587	-3053.125	2658.008	0.9978509
Pasteuria-BTberliner	13139.0555	10283.489	15994.622	0.0000000
Pasteuria-BTkurstaki	13336.6142	10481.048	16192.181	0.0000000

B. Endospores probed with Col1981 followed by FICT-conjugated anti-rabbit IgG

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
ind	3	668198687	222732896	73.12	<2e-16 ***
Residuals	76	231519763	3046313		

> TukeyHSD(95% family-wise confidence level)

	diff	lwr	upr	p adj
BTberliner -BTAlHakam	-1622.6898	-3072.509	-172.8711	0.0221056
BTkurstaki-BTAlHakam	284.2389	-1165.580	1734.0576	0.9552686
Pasteuria-BTAlHakam	6013.2790	4563.460	7463.0977	0.0000000
BTkurstaki- BTberliner	1906.9287	457.110	3356.7474	0.0049069
Pasteuria- BTberliner	7635.9689	6186.150	9085.7876	0.0000000
Pasteuria-BTkurstaki	5729.0402	4279.221	7178.8589	0.0000000

C. Endospores probed with Col1982 followed by FICT-conjugated anti-rabbit IgG

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
ind	3	1.484e+09	494776198	80.93	<2e-16 ***
Residuals	76	4.647e+08	6113939		

```
> TukeyHSD(95% family-wise confidence level)
```

	diff	lwr	upr	p adj
BTberliner-BTAlHakam	-1085.7895	-3139.726	968.1471	0.5103188
BTkurstaki-BTAlHakam	-1654.4367	-3708.373	399.4999	0.1573515
Pasteuria-BTAlHakam	8939.0547	6885.118	10992.9913	0.0000000
BTkurstaki-BTberliner	-568.6472	-2622.584	1485.2893	0.8858934
Pasteuria-BTberliner	10024.8442	7970.908	12078.7807	0.0000000
Pasteuria-BTkurstaki	10593.4914	8539.555	12647.4280	0.0000000

D. Endospores probed with Anti-PpWS before and after WGA pre-treatment

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
ind	1	1.471e+09	1.471e+09	71.15	3.09e-10 ***
Residuals	38	7.858e+08	2.068e+07		

E. Endospores probed with Col1981 before and after WGA pre-treatment

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
ind	1	386537560	386537560	15.24	0.000375 ***
Residuals	38	963595381	25357773		

F. Endospores probed with Col1982 before and after WGA pre-treatment

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
ind	1	93017080	93017080	5.157	0.0289 *
Residuals	38	685353157	18035609		

G. Endospores probed with Anti-PpWS. Col1981 and Col1982 before and after WGA pre-treatment

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
ind	5	2.025e+09	405003921	18.96	1.1e-13 ***
Residuals	114	2.435e+09	21357507		

> TukeyHSD(95% family-wise confidence level)

	diff	lwr	upr	p adj
Pas.Col1981-Pas.Anti-PpWS	-7480.2468	-11716.5793	-3243.9142	0.0000184
Pas.Col1982-Pas.Anti-PpWS	-4486.4233	-8722.7559	-250.0907	0.0312629
Pas.WGA.Anti-PpWS-Pas.Anti-PpWS	-12129.9083	-16366.2409	-7893.5758	0.0000000
Pas.WGA.Col1981-Pas.Anti-PpWS	-1263.0323	-5499.3648	2973.3003	0.9541668
Pas.WGA.Col1982-Pas.Anti-PpWS	-7536.2935	-11772.6260	-3299.9609	0.0000156
Pas.Col1982-Pas.Col1981	2993.8235	-1242.5091	7230.1560	0.3219139
Pas.WGA.Anti-PpWS-Pas.Col1981	-4649.6616	-8885.9942	-413.3290	0.0226619
Pas.WGA.Col1981-Pas.Col1981	6217.2145	1980.8819	10453.5471	0.0006050
Pas.WGA.Col1982-Pas.Col1981	-56.0467	-4292.3793	4180.2859	1.0000000
Pas.WGA.Anti-PpWS-Pas.Col1982	-7643.4850	-11879.8176	-3407.1525	0.0000114
Pas.WGA.Col1981-Pas.Col1982	3223.3910	-1012.9415	7459.7236	0.2433033
Pas.WGA.Col1982-Pas.Col1982	-3049.8702	-7286.2027	1186.4624	0.3015100
Pas.WGA.Col1981-Pas.WGA.Anti-PpWS	10866.8761	6630.5435	15103.2086	0.0000000
Pas.WGA.Col1982-Pas.WGA.Anti-PpWS	4593.6149	357.2823	8829.9475	0.0253424
Pas.WGA.Col1982-Pas.WGA.Col1981	-6273.2612	-10509.5938	-2036.9286	0.0005232

Appendix X: ANOVA and Tukey tests for Endospores attached per J2 in attachment assays

R-script for ANOVA and Tukey's *post hoc* test
```
####  start of script   ####
data<-read.csv("filename.csv")
result<-aov(SporesAttached~Treatment*conc,data = data2)
summary(result)
TukeyHSD(result)
####  start of script   ####
```

Shapiro-Wilk Test for Normality

Attachment test data: Shapiro-Wilk Test for Normality		
Col1981.50 vs PBS	W = 0.94767, p-value = 0.06309	Normal
Col1981.500 vs PBS	W = 0.94393, p-value = 0.04698	Normal
Col1981.1000 vs PBS	W = 0.91659, p-value = 0.006025	Normal
Col1982.50 vs PBS	W = 0.95367, p-value = 0.1014	Normal
Col1981.500 vs PBS	W = 0.96126, p-value = 0.185	Normal
Col1981.1000 vs PBS	W = 0.94062, p-value = 0.03628	Not-normal
PasCollagenase750 vs PBS	W = 0.97459, p-value = 0.4963	Normal
PasCollagenase375 vs PBS	W = 0.96819, p-value = 0.3149	Normal
PasCollagenase250 vs PBS	W = 0.96443, p-value = 0.2367	Normal
J2Collagenase750 vs PBS	W = 0.95757, p-value = 0.1383	Normal
J2Collagenase375 vs PBS	W = 0.97456, p-value = 0.4954	Normal
J2Collagenase250 vs PBS	W = 0.97586, p-value = 0.5394	Normal
PasWGA.50 vs PBS	W = 0.95929, p-value = 0.1584	Normal
PasWGA.500 vs PBS	W = 0.97137, p-value = 0.3973	Normal
PasWGA.1000 vs PBS	W = 0.9661, p-value = 0.269	Normal
J2WGA.50 vs PBS	W = 0.97421, p-value = 0.484	Normal
J2WGA.500 vs PBS	W = 0.97319, p-value = 0.4515	Normal
J2WGA.1000 vs PBS	W = 0.97169, p-value = 0.4064	Normal
PasNAGase.50 vs PBS	W = 0.97048, p-value = 0.3726	Normal
PasNAGase.25 vs PBS	W = 0.97275, p-value = 0.4379	Normal
J2NAGase.50 vs PBS	W = 0.96564, p-value = 0.2597	Normal
J2NAGase.25 vs PBS	W = 0.96203, p-value = 0.1965	Normal
Col1981-Col1982-PasCol vs PBS	W = 0.98457, p-value = 0.04468	Not-normal
PasCol vs J2Col	W = 0.94268, p-value = 6.517e-05	Not-normal
J2WGA vs PasWGA	W = 0.95784, p-value = 0.0008522	Not-normal
J2NAG vs PasNAG	W = 0.92863, p-value = 0.000249	Not-normal

A. PBS Control vs all other treatments

PBS vs Col1981 (1/50)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1221.0	1221.0	76.72	1.19e-10 ***
Residuals	38	604.8	15.9		

PBS vs Col1981(1/500)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	291.6	291.60	16.39	0.000244 ***
Residuals	38	676.0	17.79		

PBS vs Col198(1/1000)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	18.2	18.23	0.917	0.344
Residuals	38	19.88	755.6		

PBS vs Col1982 (1/50)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	476.1	476.1	33.08	1.25e-06 ***
Residuals	38	547.0	14.4		

PBS vs Col1982 (1/500)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	90.0	90.00	6.004	0.019 *
Residuals	38	569.6	14.99		

PBS vs Col1982 (1/1000)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	12.1	12.10	0.777	0.384
Residuals	38	591.8	15.57		

PBS vs PasCollagenase (750 u/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1836.0	1836.0	151.4	7.94e-15 ***
Residuals	38	460.8	12.1		

PBS vs PasCollagenase (375 u/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1946.0	1946	162.8	2.61e-15 ***
Residuals	38	454.4	12		

PBS vs PasCollagenase (250 u/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1368.9	1368.9	90.53	1.34e-11 ***
Residuals	38	574.6	15.1		

PBS vs J2Collagenase (750 u/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1960.0	1960.0	170.2	1.3e-15 ***
Residuals	38	437.6	11.5		

PBS vs J2Collagenase (375 u/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1782.2	1782.2	147.4	1.2e-14 ***
Residuals	38	459.6	12.1		

PBS vs J2Collagenase (250 u/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1500.6	1500.6	118.7	2.99e-13 ***
Residuals	38	480.4	12.6		

PBS vs PasWGA (1/50)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	2045	2044.9	174.6	8.73e-16 ***
Residuals	38	445	11.7		

PBS vs PasWGA (1/500)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1755.6	1755.6	144.9	1.55e-14 ***
Residuals	38	460.4	12.1		

PBS vs PasWGA (1/1000)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	921.6	921.6	61.79	1.74e-09 ***
Residuals	38	566.8	14.9		

PBS vs J2WGA (1/50)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1404.2	1404	108.1	1.14e-12 ***
Residuals	38	493.6	13		

PBS vs J2WGA (1/500)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1755.6	1755.6	141.8	2.14e-14 ***
Residuals	38	470.4	12.4		

PBS vs J2WGA (1/1000)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1040.4	1040.4	78.5	1 8.81e-11 ***
Residuals	38	503.6	13.3		

PBS vs PasNAGase (50 units/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1755.6	1755.6	148.1	1.11e-14 ***
Residuals	38	450.4	11.9		

PBS vs PasNAGase (25 units/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1525.2	1525.2	120.4	2.44e-13 ***
Residuals	38	481.6	12.7		

PBS vs J2NAGase (50 units/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1254.4	1254	83.39	4e-11 ***
Residuals	38	571.6	15		

PBS vs J2NAGase (25 units/ml)

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	1380.6	1381	92.31	1.03e-11 ***
Residuals	38	568.4	15		

B. Endospores attached after endospores were pretreated with Col1981, Col1982 and Collagenase

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	2	2887.6	1443.8	171.67	< 2e-16 ***
conc	2	1004.4	502.2	59.71	< 2e-16 ***
Treatment(concentration)	4	352.4	88.1	10.47	1.3e-07 ***
Residuals	171	1438.2	8.4		

> TukeyHSD(95% family-wise confidence level)
$`Treatment(concentration)`

	diff	lwr	upr	p adj
Col1982(1/50)-Col1981(1/50)	4.15	1.268547	7.03145299	0.0003783
PasCollagenase(750 u/ml)-Col1981(1/50)	-2.50	-5.381453	0.38145299	0.1466000
Col1981(1/500)-Col1981(1/50)	5.65	2.768547	8.53145299	0.0000002
Col1982(1/500)-Col1981(1/50)	8.05	5.168547	10.93145299	0.0000000
PasCollagenase(375 u/ml)-Col1981(1/50)	-2.90	-5.781453	-0.01854701	0.0472003
Col1981(1/1000)-Col1981(1/50)	9.70	6.818547	12.58145299	0.0000000
Col1982(1/1000)-Col1981(1/50)	9.95	7.068547	12.83145299	0.0000000
PasCollagenase(250 u/ml)-Col1981(1/50)	-0.65	-3.531453	2.23145299	0.9986192
PasCollagenase(750 u/ml)-Col1982(1/50)	-6.65	-9.531453	-3.76854701	0.0000000
Col1981(1/500)-Col1982(1/50)	1.50	-1.381453	4.38145299	0.7838314
Col1982(1/500)-Col1982(1/50)	3.90	1.018547	6.78145299	0.0011341
PasCollagenase(375 u/ml)-Col1982(1/50)	-7.05	-9.931453	-4.16854701	0.0000000
Col1981(1/1000)-Col1982(1/50)	5.55	2.668547	8.43145299	0.0000003
Col1982(1/1000)-Col1982(1/50)	5.80	2.918547	8.68145299	0.0000001
PasCollagenase(250 u/ml)-Col1982(1/50)	-4.80	-7.681453	-1.91854701	0.0000169
Col1981(1/500)-PasCollagenase(750 u/ml)	8.15	5.268547	11.03145299	0.0000000
Col1982(1/500)-PasCollagenase(750 u/ml)	10.55	7.668547	13.43145299	0.0000000
PasCollagenase(375 u/ml)-PasCollagenase(750 u/ml)	-0.40	-3.281453	2.48145299	0.9999632
Col1981(1/1000)-PasCollagenase(750 u/ml)	12.20	9.318547	15.08145299	0.0000000
Col1982(1/1000)-PasCollagenase(750 u/ml)	12.45	9.568547	15.33145299	0.0000000
PasCollagenase(250 u/ml)-PasCollagenase(750 u/ml)	1.85	-1.031453	4.73145299	0.5335966
Col1982(1/500)-Col1981(1/500)	2.40	-0.481453	5.28145299	0.1873670
PasCollagenase(375 u/ml)-Col1981(1/500)	-8.55	-11.431453	-5.66854701	0.0000000
Col1981(1/1000)-Col1981(1/500)	4.05	1.168547	6.93145299	0.0005910
Col1982(1/1000)-Col1981(1/500)	4.30	1.418547	7.18145299	0.0001904
PasCollagenase(250 u/ml)-Col1981(1/500)	-6.30	-9.181453	-3.41854701	0.0000000
PasCollagenase(375 u/ml)-Col1982(1/500)	-10.95	-13.831453	-8.06854701	0.0000000
Col1981(1/1000)-Col1982(1/500)	1.65	-1.231453	4.53145299	0.6827053
Col1982(1/1000)-Col1982(1/500)	1.90	-0.981453	4.78145299	0.4960535
PasCollagenase(250 u/ml)-Col1982(1/500)	-8.70	-11.581453	-5.81854701	0.0000000
Col1981(1/1000)-PasCollagenase(375 u/ml)	12.60	9.718547	15.48145299	0.0000000
Col1982(1/1000)-PasCollagenase(375 u/ml)	12.85	9.968547	15.73145299	0.0000000
PasCollagenase(250 u/ml)-PasCollagenase(375 u/ml)	2.25	-0.631453	5.13145299	0.2625631
Col1982(1/1000)-Col1981(1/1000)	0.25	-2.631453	3.13145299	0.9999991
PasCollagenase(250 u/ml)-Col1981(1/1000)	-10.35	-13.231453	-7.46854701	0.0000000
PasCollagenase(250 u/ml)-Col1982(1/1000)	-10.60	-13.481453	-7.71854701	0.0000000

c. **Endospores attached after endospores and J2 were pretreated with Collagenase**

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	0.5	0.53	0.181	0.672
conc	2	80.8	40.41	13.694	4.68e-06 ***
Treatment(concentration)	2	8.1	4.06	1.375	0.257
Residuals	114	336.4	2.95		

> TukeyHSD(95% family-wise confidence level)
$`Treatment(concentration)`

	Diff	lwr	upr	p adj
PasCollagenase(750 u/ml)-J2Collagenase(750 u/ml)	0.45	-1.12467	2.02467	0.9616682
J2Collagenase(375 u/ml)-J2Collagenase(750 u/ml)	0.65	-0.92467	2.22467	0.8376780
PasCollagenase(375 u/ml)-J2Collagenase(750 u/ml)	0.05	-1.52467	1.62467	0.9999990
J2Collagenase(250 u/ml)-J2Collagenase(750 u/ml)	1.75	0.175327	3.32467	0.0201429
PasCollagenase(250 u/ml)-J2Collagenase(750 u/ml)	2.30	0.725327	3.87467	0.0006529
J2Collagenase(375 u/ml)-PasCollagenase(750 u/ml)	0.20	-1.37467	1.77467	0.9990950
PasCollagenase(375 u/ml)-PasCollagenase(750 u/ml)	-0.40	-1.97467	1.17467	0.9769727
J2Collagenase(250 u/ml)-PasCollagenase(750 u/ml)	1.30	-0.27467	2.87467	0.1673996
PasCollagenase(250 u/ml)-PasCollagenase(750 u/ml)	1.85	0.275327	3.42467	0.0114846
PasCollagenase(375 u/ml)-J2Collagenase(375 u/ml)	-0.60	-2.17467	0.97467	0.8785978
J2Collagenase(250 u/ml)-J2Collagenase(375 u/ml)	1.10	-0.47467	2.67467	0.3348415
PasCollagenase(250 u/ml)-J2Collagenase(375 u/ml)	1.65	0.075327	3.22467	0.0342508
J2Collagenase(250 u/ml)-PasCollagenase(375 u/ml)	1.70	0.12532718	3.274672	0.0263710
PasCollagenase(250 u/ml)-PasCollagenase(375 u/ml)	2.25	0.67532718	3.824672	0.0009205
PasCollagenase(250 u/ml)-J2Collagenase(250 u/ml)	0.55	-1.02467282	2.124672	0.9129223

D. **Endospores attached after endospores and J2 were pretreated with WGA**

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	11.4	11.41	3.181	0.07716
conc	2	284.5	142.23	39.657	8.45e-14 ***
Treatment(concentration)	2	52.2	26.11	7.280	0.00106 **
Residuals	114	408.9	3.59		

> TukeyHSD(95% family-wise confidence level)
$`Treatment(concentration)`

	Diff	lwr	upr	p adj
PasWGA(1/50)-J2WGA(1/50)	-2.450000e+00	-4.18597849	-0.7140215	0.0011103
J2WGA(1/500)-J2WGA(1/50)	-1.400000e+00	-3.13597849	0.3359785	0.1877750
PasWGA(1/500)-J2WGA(1/50)	-1.400000e+00	-3.13597849	0.3359785	0.1877750
J2WGA(1/1000)-J2WGA(1/50)	1.650000e+00	-0.08597849	3.3859785	0.0724727
PasWGA(1/1000)-J2WGA(1/50)	2.250000e+00	0.51402151	3.9859785	0.0036272
J2WGA(1/500)-PasWGA(1/50)	1.050000e+00	-0.68597849	2.7859785	0.4999139
PasWGA(1/500)-PasWGA(1/50)	1.050000e+00	-0.68597849	2.7859785	0.4999139
J2WGA(1/1000)-PasWGA(1/50)	4.100000e+00	2.36402151	5.8359785	0.0000000
PasWGA(1/1000)-PasWGA(1/50)	4.700000e+00	2.96402151	6.4359785	0.0000000
PasWGA(1/500)-J2WGA(1/500)	1.554312e-15	-1.73597849	1.7359785	1.0000000
J2WGA(1/1000)-J2WGA(1/500)	3.050000e+00	1.31402151	4.7859785	0.0000206
PasWGA(1/1000)-J2WGA(1/500)	3.650000e+00	1.91402151	5.3859785	0.0000002
J2WGA(1/1000)-PasWGA(1/500)	3.050000e+00	1.31402151	4.7859785	0.0000206
PasWGA(1/1000)-PasWGA(1/500)	3.650000e+00	1.91402151	5.3859785	0.0000002
PasWGA(1/1000)-J2WGA(1/1000)	6.000000e-01	-1.13597849	2.3359785	0.9164379

E. **Endospores attached after endospores and J2 were pretreated with NAGase**

> ANOVA summary

	Degrees of Freedom	Sum of sqaures	Mean	F value	Pr(>F)
Treatment	1	35.1	35.11	6.938	0.0102 *
conc	1	0.6	0.61	0.121	0.7289
Treatment(concentration)	1	10.5	10.51	2.077	0.1536
Residuals	76	384.6	5.06		

> TukeyHSD(95% family-wise confidence level)
$`Treatment(concentration)`

	Diff	lwr	upr	p adj
PasNAGase(50 u/ml)-J2NAGase(50 u/ml)	-2.05	-3.9187557	-0.1812443	0.0258984
J2NAGase(25 u/ml)-J2NAGase(50 u/ml)	-0.55	-2.4187557	1.3187557	0.8663572
PasNAGase(25 u/ml)-J2NAGase(50 u/ml)	-1.15	-3.0187557	0.7187557	0.3756035
J2NAGase(25 u/ml)-PasNAGase(50 u/ml)	1.50	-0.3687557	3.3687557	0.1596990
PasNAGase(25 u/ml)-PasNAGase(50 u/ml)	0.90	-0.9687557	2.7687557	0.5877977
PasNAGase(25 u/ml)-J2NAGase(25 u/ml)	-0.60	-2.4687557	1.2687557	0.8335583

Appendix XI: R-scripts used for plotting graphs in Chapter 5 and 6

Script for *Figure 5.11* in Chapter 5 (Mean CTCF values of endospores probed with Anti-PpWS, Col19 81 and Col1982)

```
####   start of script  ####

data<-read.csv("filename.csv")
data1<-stack(data)
meanCTCF <- tapply(data1$values, list(data1$ind), mean)
sdCTCF <- tapply(data1$values, list(data1$ind), sd)
nCTCF <- tapply(data1$values, list(data1$ind), length)
SE <- sdCTCF/sqrt(nCTCF)
barplot(meanCTCF, xlab = "Bacterial endspores", ylab = "Mean CTCF values", ylim = c(0,20000))
#Assigning values to the midpoints of bars on the x axis
mids <- barplot(meanCTCF, xlab = "Bacterial endspores", ylab = "Mean CTCF values", ylim = c(0,2
0000))
#Use arrows to put error bars on the plot
arrows(mids,meanCTCF-SE,mids, meanCTCF+SE, lwd = 1.5, angle = 90, code = 3, length = 0.05)

####   end of script   ####
```

Script for *Figure 5.14* in Chapter 5 (Mean CTCF values of endospores probed with Anti-PpWS, Col19 81 and Col1982 in untreated and WGA-treated endospores)

```
####   start of script  ####

data<-read.csv("filename.csv")
means <- tapply(data$CTCF, list(data$WGAtreatment, data$Antibody), function(x) c(x = x))
barplot(means, beside = T, legend = T, xlab = "Antibody treatments", ylab = "Mean CTCF values",
ylim = c(0,20000))
barCenters <- barplot(means, beside = T, legend = T, xlab = "Antibody treatments", ylab = "Mean
CTCF values", ylim = c(0,20000))
SE <- tapply(data$SD, list(data$WGA, data$Chemical), function(x) c(x = x))/sqrt(20)
arrows(barCenters,means-SE,barCenters,
       means+SE, lwd = 1.5, angle = 90,
       code = 3, length = 0.05)

####   end of script   ####
```

Script for *Figure 6.2* and *6.3* in Chapter 6 (Effect of different different treatments on endospore attach
ment)

```
####    start of script   ####

data<-read.csv("filename.csv")
Average <- tapply(data$Avg, list(data$Concentration, data$ChemicalTreatment), function(x) c(x =
x))
barCenters <- barplot(Average, beside = T, ylab = "Number of Spores attached per J2",ylim = c(0
,20))
SE <- tapply(data$SD, list(data$Concentration, data$ChemicalTreatment),
         function(x) c(x = x))/sqrt(20)
arrows(barCenters,Average-SE,barCenters, Average+SE, lwd = 1.5, angle = 90,
         code = 3, length = 0.05)
abline(h=15.0,col="black")
abline(h=16.15,col="black", lty=3)
abline(h=14.05,col="black", lty=3)
mtext(c("750 u/ml", "375 u/ml", "250 u/ml", "750 u/ml", "375 u/ml", "250 u/ml"),side=1,line=0.0
1,at=c(1.5,2.5,3.5,5.5,6.5,7.5))
text(c(7.3),c(15.6),labels=c("Untreated Control (PBS)"))

####    end of script   ####
```

Script for *Figure 6.4* in Chapter 6 (Percentage of conventional vs unconventional type of endospore at
tachment)

```
####    start of script   ####

data<-read.csv("filename.csv")
library(ggplot2)
ggplot(data, aes(x=Treatment, y=Percentage, fill=AttachmentType)) +
  geom_bar(stat="identity", colour='black') +         # contour colour
  guides(fill=guide_legend(reverse=TRUE)) +           # reverse legend
  scale_fill_brewer(palette="black")+                 # colour palette
  xlab("Treatment prior to attachment test")+         # x-axis label
  ylab("Percentage of spores attached")               # y-axis label

####    end of script   ###
```