Pathogenicity & A Bedside Real-Time Detection Assay For *Clostridium difficile* In The Faeces of Hospitalized Patients

A thesis submitted for the degree of

Philosophiae Doctor

(Ph.D)

by

Lovleen Tina Joshi

June 2012

Welsh School of Pharmacy

Cardiff University

ACKNOWLEDGMENTS

Firstly I would like to thank my supervisor Prof. Les Baillie for supervising me throughout the last three years. Most certainly you have been a constant support and are a great teacher.

"The teacher who is indeed wise does not bid you to enter the house of his wisdom but rather leads you to the threshold of your mind"

Kahil Gibran

I would like to thank Dr. Jean Yves Maillard for his support and guidance through the duration of the PhD. I also extend my gratitude to Cardiff University for funding this research project. I would like to also thank Dr. Chris Geddes at the Institute of Fluorescence for agreeing to let me spend three months at the labs in Baltimore. Furthermore I would like to express my sincere thanks to Dr. N. Waterfield for agreeing to collaborate with the invertebrate studies and his generous assistance. I also want to thank Dr. Anthony Hann and Mr. Guy Pitt for their expert guidance and advice in electron microscopy studies. To my colleagues in Labs 1.49 & Lab 1.11, past and present, thank you for your support and for making these last three years enjoyable.

I sincerely want to thank my close friends for being there for me. Finally to my family; Aunty Abi, Uncle Narish, Aunty Rita and Grandmother thank you for your love and unwavering support. You all spurred me onwards to success and positivity, and without you none of this would have been possible.

Veni Vidi Vici

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

aa	Amino Acid
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BHI	Brain Heart Infusion
BLAST	Basic Local Alignment Search Tool
*cdt*A	Binary toxin gene A
*cdt*B	Binary toxin gene B
bp	Base pairs
BSA	Bovine Serum Albumin
CaDPA	Calcium dipicolinate
CCFA	Cycloserine cefoxitin fructose agar
CCTA	Cell cytotoxicity assay
CDAD	*C. difficile* associated diarrhoea
CDI	*C. difficile* infection
CDT	*C. difficile* Binary toxin
cfu	Colony forming units
CRISPR	clustered regularly interspaced short palindromic repeat
CROP	Clostridial Repetitive Oligo-peptides
Cwp	Cell Wall Protein
cfu	Colony forming units
DEFRA	Department for environment, food and rural affairs
DIG	Digoxigenin
DNA	Deoxyribonucleic Acid
dH_2O	Deionised Water
DMEM	Dulbecco's Modified Eagle's Medium
dNTPS	deoxyribonucleotides
EIA	Enzyme immuno assays
EM	Electron Microscopy
g	G- Force (Relative centrifugal force)
GDH	Glutamate dehydrohenase

GT	Glucosyltransferase
HGT	Horizontal gene transfer
kb	kilo base pairs
kDa	kiloDaltons
LCT	Large Clostridial toxins
MAMEF	Microwave accelerated metal enhanced fluorescence
MATH	Microbial adhesion to hydrocarbon
Mcf	Makes caterpillars floppy
MIC	Minimum Inhibitory Concentration
Min	Minutes
mM	Millimole
mm	Millimetres
ml	Millilitre
MLST	Multilocus sequence typing
MSA	Multiple Sequence Alignment
n	Number of replicates
Nm	Nanometre
NPV	Negative predictive value
^{o}C	Degrees Celsius
ORF	Open reading frame
PAI	Pathogenicity Island
PaLoc	Pathogenicity Locus
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFGE	Pulsed field gel electrophoresis
PMC	Pseudomembranous colitis
PMN	Polymorphonuclear leucocytes
Polished H_2O	Double distilled water
PPV	Positive predictive value
REA	Restriction endonuclease analysis
RFLP	Restriction Fragment Length Polymorphisms
RH	Relative Hydrophobicity

ROS	Reactive Oxygen Species
RT	Room Temperature
SASP	Small acid soluble proteins
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
sdw	Sterile deionised water
SSC	Saline sodium citrate
Sec	Seconds
SEM	Scanning Electron Microscopy
SiF	Silver island Films
SLP	Surface Layer Proteins
spp	species
ST	Sodium taurocholate
tcdA	Toxin A encoding gene
tcdB	Toxin B encoding gene
TC	Toxin complexes- insecticidal complexes associated with *Photorhabdus* species
TEM	Transmission Electron Microscopy
UV	Ultra Violet
VRE	Vancomycin resistant Enterococci
w/v	Weight per volume
x	Amount of times
µl	Microlitre

PUBLICATIONS

Joshi, L. T., Phillips, D. S., Williams, C. F., Alyousef, A. & Baillie, L. The contribution of the spore to the ability of *Clostridium difficile* to adhere to surfaces. *Applied & Environmental Microbiology*. AEM 01862-12.

PATENT (Pending)

Regarding the development of a real time detection assay for *Clostridium difficile* as part of postgraduate study at Cardiff University. (2012).

COMMUNICATIONS:

The characterisation of the spore surface of *Clostridium difficile* and its role in spore adherence. 3rd International *Clostridium difficile* Symposium, Bled, Slovenia. **September 2010. Invited Speaker. Oral Presentation**.

Vegetative cells of *Clostridium difficile.* Printed canvas picture in Welsh School of Pharmacy, Cardiff University. False-coloured SEM image of vegetative cells of *C. difficile* strain DS1748. **2010.**

How does *Clostridium difficile* stick around? Plant Microbiology Wales Conference, Aberystwyth. **2010**. **Oral Presentation.**

A bedside real time detection assay for *Clostridium difficile* in the faces of hospitalised patients. Welsh School of Pharmacy Postgraduate Research Day, Cardiff. **2010. Poster presentation.**

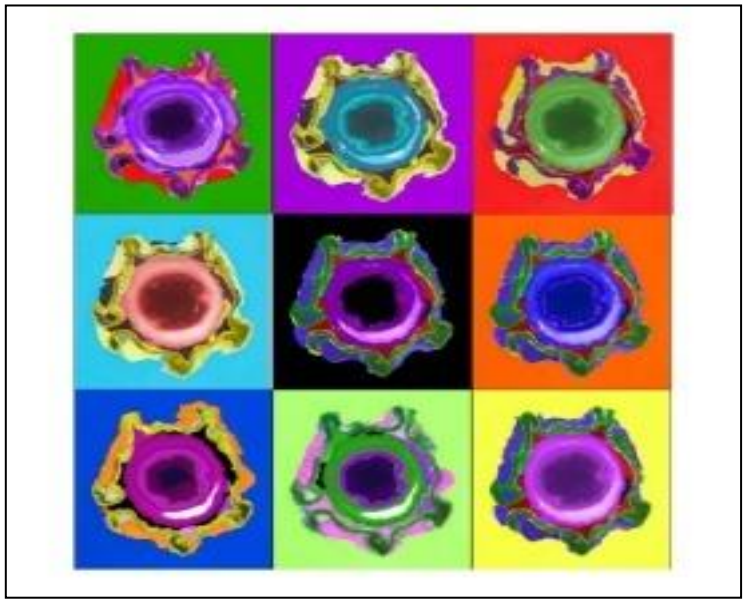
Beauty of the Beast! A single spore of the hypervirulent 027 strain of *Clostridium difficile* false-coloured and arranged in the style of an Andy Warhol pop art image. 2010. **Society for Applied Microbiology 2010 Calendar. December 2010.**.

Update on *C. difficile* research at Cardiff University: A bedside real time detection assay for *Clostridium difficile* in the faces of hospitalised patients. Welsh Microbiology Association Meeting. Brecon Castle Hotel, Brecon. **December 2009. Oral Presentation.**

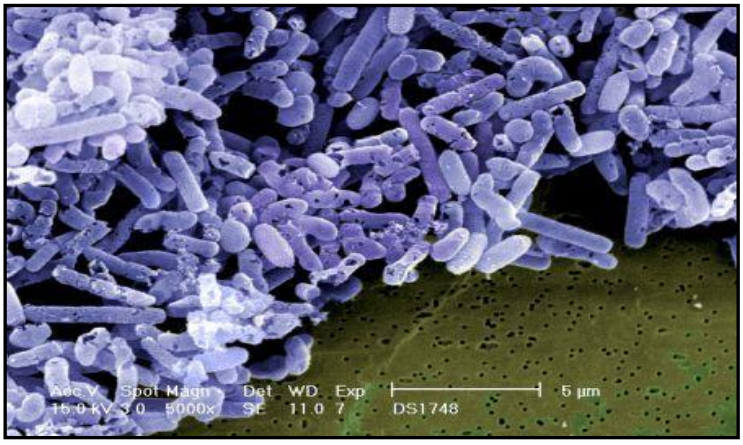
Characterisation of *Clostridium difficile* spores. 6th ClostPath International Conference: Clostridia: The Impact of Genomics on Disease Control, Rome, Italy. **October 2009**. **Poster Presentation.**

The characterisation of *Clostridium difficile* spores. 9th Tripartite Meeting of the Celtic Microbiology Associations, Angel Hotel, Cardiff. **June 2009. 1st Prize Poster Presentation**.

Germination, sporulation and hydrophobicities of *Clostridium difficile* spores. Speaking of Science Conference, Cardiff University. **April 2009**. **Oral Presentation.**



Society for Applied Microbiology 2010 Calendar. December 2010.



Printed canvas picture in School of Pharmacy, Cardiff University

Vegetative cells of *Clostridium difficile*

Make. It. So

SUMMARY

Clostridium difficile, a Gram positive, anaerobic, spore-forming bacterium is the commonest cause of hospital acquired infection in the UK. The organism initiates infection through spore formation and attachment, germination in the gut and then the production of two potent cytotoxins; toxins A and B. While the contribution of toxins A and B to infection is beyond dispute the relative importance of each toxin is a subject of debate. Thus diagnostic assays capable of rapidly detecting the presence of both toxins are needed.

To develop such an assay we first characterised the structure of *C. difficile* spores to better understand their role in pathogenicity and adherence to organic and inorganic surfaces. Following attachment the spore germinates and the resulting vegetative bacteria express toxins. To facilitate the development of an assay capable of detecting both toxins, we employed a bioinformatics based approach which identified highly conserved nucleotide sequences within regions of each toxin which we hypothesised were under strict selective pressure.

The specificity of the probes identified was confirmed using a panel of 58 clinical *C. difficile* isolates, related *Clostridium* isolates, non-related species and human gut metagenomic DNA samples. Selected probes were incorporated into a metal enhanced fluorescent assay platform and their ability to detect the organism in various organic backgrounds was determined. We were able to detect as few as 10 bacteria in 500 µl of human faecal material within 40 seconds, suggesting that this approach has the potential to be developed into a commercial assay.

To support the development of this assay we sought to develop an insect infection model using the worm *Manduca sexta*. Our inability to initiate infection, inspite of the fact that bioinformatic analysis revealed the presence of genes with homology to known insect virulence factors, suggests that *C. difficile* may have potential evolutionary association to invertebrates.

CHAPTER 1

GENERAL

INTRODUCTION

1.1. HISTORY OF *CLOSTRIDIUM DIFFICILE*

First isolated from neonates in 1935 by Hall and O'Toole, *Clostridium difficile* was originally known as *Bacillus difficilis* due to its difficulty to isolate from human faeces. Hall and O'Toole established the toxigenic nature of *C. difficile* when experimenting with rabbits and guinea pigs, finding that when broth culture of the organism was injected in the animals, lesions, respiratory arrest and occasionally death occurred. At the time the toxin was thought to be a type of neurotoxin (Lyerly *et al.*, 1988). The only knowledge available on *B. difficilis* was a few papers between 1934 and 1962, which were documented by Willis in his book "Clostridia of wound infection" in 1969 (Duerden *et al*., 1992).

It was not until 1978 that *Bacillus difficilis*- renamed *Clostridium difficile* - was identified as the causative agent of pseudomembranous colitis (PMC). Prior to this, clinicians were aware that antibiotic therapies induced PMC and associated diarrhoea. Antibiotics such as clindamycin and lincomycin were implicated in causing disease (Small *et al.*, 1968). Meanwhile in the USA, Green (1974) had described a cytotoxin present in the stool of guinea pigs after penicillin treatment, and Tedesco *et al.* (1974) found a significant association between PMC development in patients receiving clindamycin treatment.

Even so the correlation between *C. difficile* and the toxic effects observed by Green and Tedesco *et al.* (1974) were not established until a few years later. Bartlett *et al.* (1978), with their letter pertinently titled "Will the real *Clostridium* species responsible for antibiotic-associated colitis please step forward?", were the first to contest the conclusions from two previous papers naming *Clostridium sordellii* as the cause of PMC (Rifkin *et al.*, 1977; George *et al.*, 1978). Rifkin *et al.* (1977) described a heat labile cytotoxin capable of killing hamsters and causing vascular permeability in the skin of rabbits. The cytotoxin was neutralised by the *C. sordellii* antitoxin but not by other clostridial antitoxins, or antitoxins to *Escherichia coli* or *Vibrio cholerae* toxins, thus leading to the conclusion that *C. sordellii* caused PMC (George *et al.*, 1978).

Interestingly Bartlett *et al.* (1978) were reluctant to ascribe *C. sordellii* with a pathogenic role in the disease as they had not isolated the bacterium from stool cultures. PMC sufferers were known to have intestinal lesions and although injections of *C. sordellii* into the caecum of hamsters resulted in death, the autopsies did not produce evidence of

lesions or cell cytotoxicity. Bartlett *et al.* (1978) had in fact isolated *C. difficile* from infected patients. Broth cultures of *C. difficile* were injected into the caecum of healthy hamsters and gave rise to colitis symptoms seen in humans. Additionally the toxin was found to be cytopathic and was neutralised when mixed with the gas gangrene producing C. *perfringens* antitoxin and *C. sordellii* antitoxin. Thus as more reports showed correlations between *C. difficile* infection and PMC, *C. difficile* was fully implicated with a role in PMC (Burdon & George, 1978; Keighley *et al.*, 1978; Larson *et al.*, 1978).

Early literature on *C. difficile* can be confusing as the cytotoxin was considered to be the only toxin produced from the organism. Taylor *et al.* (1981) actually found that another enterotoxin existed which elicited a positive response in biological assays and degraded on polyacrylamide electrophoresis into two molecules- one of which appeared to migrate with the purified cytotoxin band. This toxin was designated toxin A and was thought to play an important role in *C. difficile* associated disease (CDAD). This was indeed confirmed by further research (Banno *et al*., 1981; Sullivan *et al.*, 1981). It was then firmly established that *C. difficile* produced two distinct toxins.

1.2. *CLOSTRIDIUM DIFFICILE* ASSOCIATED DIARRHOEA

1.2.1. Pathogenesis & Infection

C. difficile manifests in varying severity from mild diarrhoea to fatal colitis in antibiotic-treated patients (Voth & Ballard, 2005). The gut microbiota of antibiotic-treated patients is disrupted allowing *C. difficile*, which may be exogenously acquired or endogenous to the colon, to proliferate. Carriage of *C. difficile* can be asymptomatic and occurs in 40-60% of neonates and 1-3% of healthy adults (Kuijpers & Surawicz, 2008). In the hospital environment the organism is primarily acquired exogenously through the faecal-oral route (Figure 1.1; Salyers & Whitt, 2002).

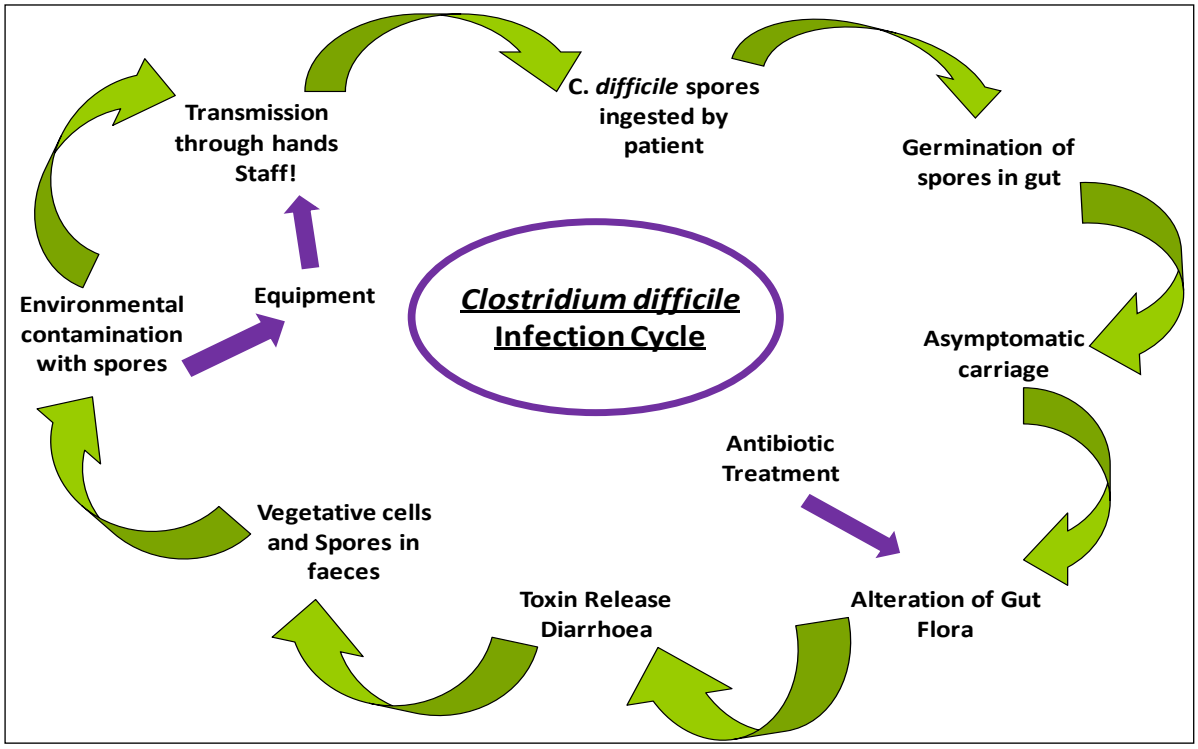


Figure 1.1: *Clostridium difficile* Infection Cycle

The transmission of *Clostridium difficile* from faeces to infection can be seen.

Following uptake spores pass through the stomach, are thought to germinate in response to bile salts and other environmental factors in the gut, and attach to the gut wall to colonise this environment (Panessa-Warren *et al.*, 2007; Sorg & Sonenshein, 2008). Disease arises when there are enough clostridia present to produce copious amounts of toxin.

Toxin A causes necrosis of cells, inducing tumour necrosis factor alpha (TNF-α), interleukins IL-1β, IL-6 and IL-8 and toll-like receptors (TLR-2) to be produced which

signals for transcription factor NFκB (Poutanen & Simor, 2004; Giannasca *et al.,* 2004; Janeway *et al.,* 2005). The human immune system also produces high levels of immunoglobin G (IgG) in response to toxin A post colonisation, suggesting that a robust antibody response contributes to protecting the host from CDAD (Voth & Ballard, 2005). The toxins activate intestinal nerves to release neuropeptides and a calcitonin gene-related peptide (CGRP), thought to have proinflammatory properties.

Neutrophil response and polymorphonuclear leucocytes are stimulated by the release of inflammatory mediators such as macrophage inflammatory protein 2 (MIP-2) identified in rats, from the macrophage and resident mast cells. Mast cells are degranulated post exposure to toxin A. These factors all contribute to the associated inflammatory response and pseudomembrane formation, increased intestinal permeability, causing electrolyte imbalance and inhibition of protein synthesis (Figure 1.2). This may result in damage to the intestinal mucosa, villi and mast cells, thus causing a bloody, viscous fluid to be released- i.e. diarrhoea and pseudomembranous colitis (PMC). This inflammatory response has been demonstrated *in vivo* using rabbit or rat ileal loop models (Voth & Ballard, 2005).

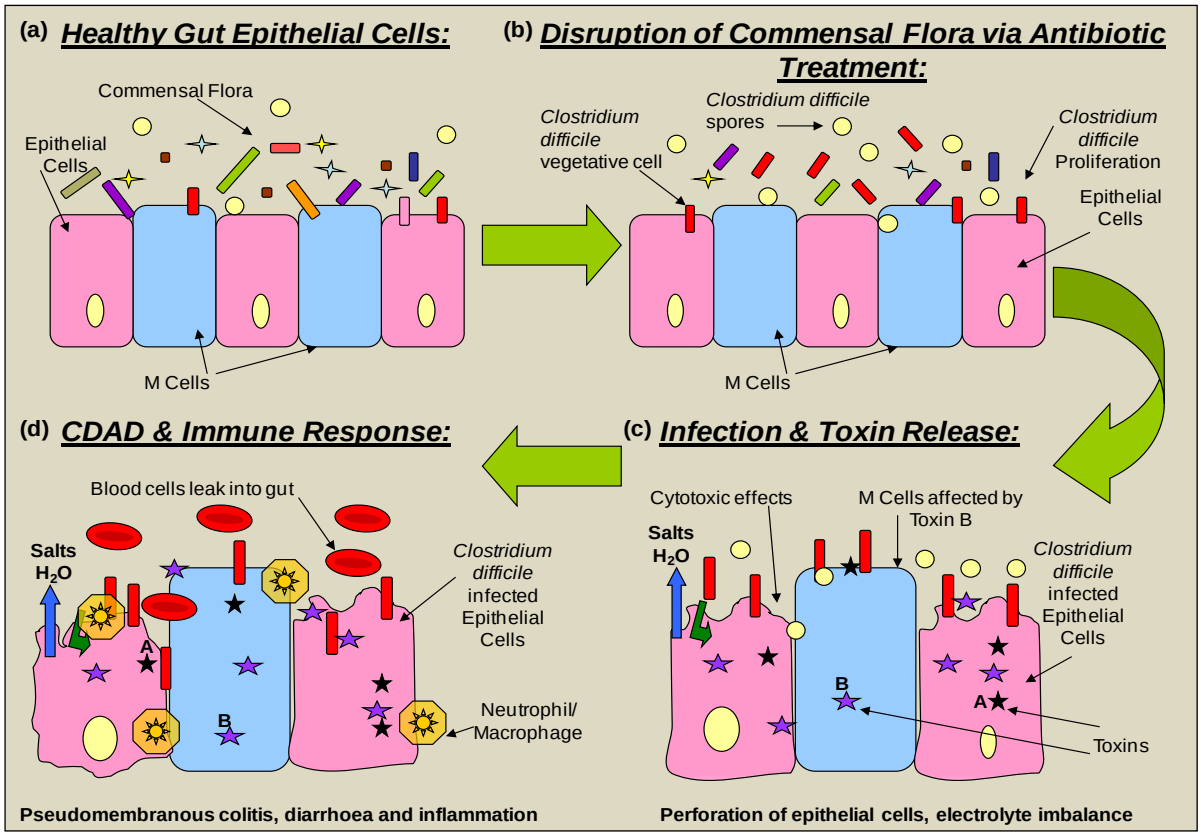


Figure 1.2: Pathophysiology of *C. difficile* infected gut

(a) Healthy gut epithelial cells with normal microbiota. (b) Microbiota disrupted by antibiotic treatment. *C difficile* spores and vegetative cells start to proliferate. (c) Toxins A and B are released from vegetative cells. Immune response is initiated and M cells are damaged. Electrolyte imbalance occurs. (d) Inflammation results in CDAD and pseudomembranous colitis.

Toxin B is known as a cytotoxin and does not elicit fluid release however *in vitro* it has been shown to be lethal to cells (Genth *et al.*, 2008). Polymorphonuclear leucocytes migrate between mucosal cells promoting damage and perforation, thus allowing toxin B to diffuse across the mucosal membrane and cause further damage to the intestinal wall. This pore formation has been found to be cell-specific and dependent on cholesterol (Giesemann *et al.*, 2006). If the damage to the intestinal wall becomes too extensive bacteria can enter the blood stream to cause septicaemia. In an endoscopy, yellow-white raised plaques are seen on the pseudomembranes, with inflamed mucosa (Kuijpers & Surawicz, 2008). These yellow plaques (Figure 1.3) are accumulations of fibrin, mucin and dead host cells, initially forming a few lesions, and progressing to form spreading pseudomembranes, which have been found to form a sheath over the entire colonic mucosa in some cases (Lyerly *et al.*, 1988). This is what distinguishes the severity of the

disease from colitis (Salyers & Whitt, 2002; Poutanen & Simor, 2004). Ultimately PMC can lead to toxic mega colon and even death- as the disease rapidly increases in severity within the space of two weeks (Wren *et al.*, 2009). Thus rapid detection methods are needed to elucidate the presence of *C. difficile* and diagnose the infected patient.

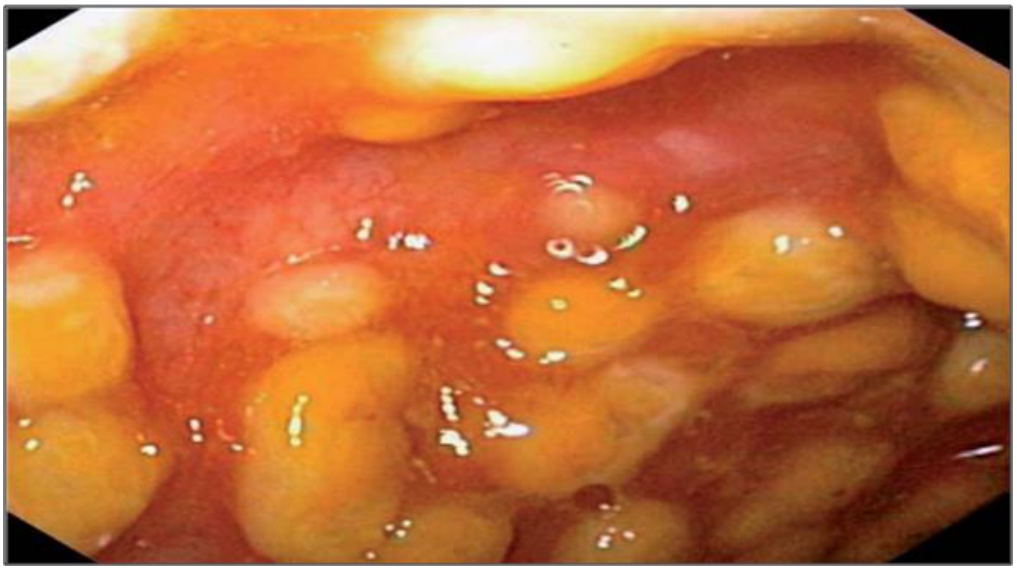


Figure 1.3: Pseudomembrane formation in a patient infected with *C. difficile*
Pseudomembrane formation can be seen in the form of raised yellow plaques on the lining of the gut wall. These are accumulations of fibrin, mucin and dead host cells. Taken from Kuijpers & Surawicz (2008).

Indeed the role of toxins A and B in CDAD have been debated. A study by Lyras *et al*. (2009) suggested that *C. difficile* strains producing only toxin A did not elicit CDAD, and that only toxin B production was essential for disease. The impact of this study was the move within the healthcare industry to produce diagnostics which focused only on toxin B detection. A more recent study by Kuehne *et al*. (2010) however re-established that a mutated strain of *C. difficile* which produced only toxin A could in fact elicit disease, meaning that both toxins A and B have a role in CDI. Thus it is important to develop diagnostics able to accurately detect both toxins A and B of *C. difficile*.

1.2.2. Prevention and treatment of CDAD

In order to prevent *C. difficile* infection control measures need to be employed within the hospital environment (Gerding *et al.*, 1995). These measures include strict practices such as hand washing rather than using alcohol hand rubs as *C. difficile* spores are resistant to alcohol (Bartlett, 2006). Furthermore use of gloves and disinfection with sodium

hypochlorite has resulted in decreased incidence of CDAD (Department of Health & Health Protection Agency, 2008). An increased level of surveillance to detect new cases and isolate patients with CDAD can also help limit spread of the bacteria (McFarland, 2005).

The most important risk factor for *C. difficile* infection is antibiotic use. Broad spectrum antibiotics including clindamycin, aminopenicillins, cephalosporins (cefoxitins) and fluoroquinolones are used to treat pathogens which infect patients (Pepin *et al.,* 2005). Simultaneously, however, they disrupt human colonic microbiota and the host can become susceptible to opportunistic pathogens such as *C. difficile* (Mulligan *et al.,* 1984; Deneve *et al.,* 2008). Fluoroquinolones such as moxifloxacin were found to increase CDAD when North American epidemic strain (NAP1) PCR ribotype 027 emerged (Biller *et al.,* 2006; Deneve *et al.,* 2008).

To successfully treat CDAD broad spectrum antibiotic usage must be discontinued and replaced with a narrow spectrum antibiotic (McFarland, 2005).The primary antibiotics used to manage CDAD are vancomycin and metronidazole which are bactericidal to *C. difficile;* although this treatment is associated with relapse (Hinkson *et al.,* 2008). Vancomycin is given orally at 125 mg four times daily for 10-14 days while metronidazole is given orally at a dosage of 250 mg respectively four times daily (or 500 mg three times daily) for 10-14 days depending on severity of disease (Kuipers & Surawicz, 2008). Original studies have recommended metronidazole as first-line treatment of CDAD, as it is significantly less expensive than vancomycin, and reduces selective pressure for the emergence of vancomycin- resistant *Enterococci* (VRE). Guidelines set out by the UK Department of Health recommend that metronidazole be used in mild cases of disease, and vancomycin be used in severe incidences (Department of Health & Health Protection Agency, 2008).

Ramoplanin is a glycolipodepsipeptide used to treat VRE infections and is currently being evaluated as another measure to combat CDAD (Freeman *et al.,* 2005). Recently tolevamer, a novel toxin-binding anionic polymer, has been found to improve CDAD without affecting normal guy microbiota. Tolevamer works by binding to toxins A and B via a styrene sulphonate polymer (Miller, 2007). Phase 2 clinical trials with tolevamer demonstrated a trend for reduced re-occurrence of *C. difficile* infection (Hinkson *et al.,*

2008). Low risk of *C. difficile* infection has also been found with use of tigecycline (Wilcox, 2007).

Faecal transplants and bowel irrigations are currently being used to replace the original microbiota of the gut (Macconnachie *et al.,* 2009). Studies have shown that patients who have had this treatment do not have any reoccurrence of infection (McFarland, 2005). In cases of toxic mega colon where the patient develops severe PMC, a total colectomy is a last resort, as hemi-colectomies have a 100% mortality rate (Miller, 2007).

Probiotics are increasingly being used as a means of restoring normal gut microbiota following antibiotic treatment to prevent *C. difficile* proliferation (Miller & Fraser, 2009). Probiotics are not as yet widely accepted as an effective means for prevention and is the subject of much debate. Even so, some scientists recommend probiotics use in CDAD prevention to attempt to restore normal colonic microbiota. In 1989 Surawicz *et al.* found that *Saccharomyces boulardii,* a non-pathogenic yeast, prevented infection and relapses in clindamycin treated animals. It is normally given twice a day to infected patients, but is not given to immuno-suppressed patients as fungaemia can occur (Kuipers & Surawicz, 2008). Other probiotics include *Lactobacillus rhamnosus* GG, *Lactobacillus acidophilus, Lactobacillus bulgaricus* and *Lactobacillus plantarum* 299v (Miller, 2007).

Immunomodulation via active and passive immunity is being considered to prevent CDI. Active immunity involves vaccination with toxin gene-derived vaccines; however there is currently no direct evidence to demonstrate the efficacy of this approach (Miller, 2007). Even so, the company Acambis (based in Cambridge) is developing the world's first vaccine against CDAD which commenced clinical trials in 2008. Passive immunity involves the administration of preformed antibodies capable of neutralising the activity of toxin A and B (Miller, 2007).

New therapies for CDAD and recurrence of CDAD include the production of human monoclonal antibodies against the two toxins of *C. difficile*. Lowy *et al.* (2010) have recently performed a randomised, double blind placebo controlled study of two neutralising, fully human monoclonal antibodies against toxins A and B. The antibodies were administered in an established hamster model of *C. difficile* infection, as well as in phase 1 with human healthy volunteers, and significant efficacy of these antibodies was

found. A phase 2 randomised trial involving 200 patients with CDAD who were treated intravenously with the antibodies, along with standard antibiotics, reduced the reoccurrence rate of infection by 70% (Lowy *et al.,* 2010). However the cost of IV antibody therapy is likely to be high. This antibody combination is currently in the process of being licensed to Merck.

1.2.3. Epidemiology

The number of death certificates (Figure 1.4) in England and Wales which mentioned the bacterium *C. difficile* in 2007 was 8,324 - an increase of 28% from 2006 (Office of National Statistics, 2010). Most deaths occurred in the elderly with mortality rates in patients over the age of 85 between 3,429 and 3,396 per million populations (Office of National Statistics, 2010). Importantly approximately 80% of reported cases of *C. difficile* were in patients 65 years of age or older. Before the emergence of the BA/NAP1 027 strain in 2003, the cases of CDAD were not deemed significant across North America or Europe (Figure 1.5; Kuijpers *et al.,* 2008). However between October 2003 and June 2004 an outbreak caused by the ribotype 027 strains at Stoke Mandeville Hospital involved 174 cases and 19 deaths due to *C. difficile* (Freeman *et al.,* 2010). A second outbreak occurred in October 2004 again at Stoke Mandeville involved 160 and another 19 deaths occurred. An investigation afterwards concluded that the outbreaks were the result of poor patient environments and infection control practice (Freeman *et al.,* 2010).

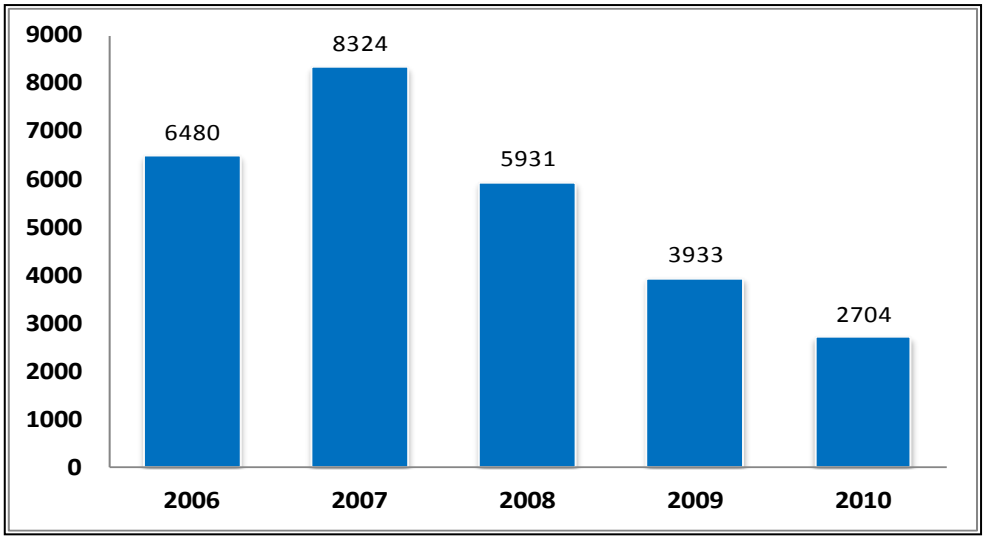


Figure 1.4: National Statistics of *C. difficile*

Graph from National Statistics data depicting the number of death certificates mentioning *Clostridium difficile* as the cause of deaths in England and Wales 2006-2010. (Data taken from Office of National Statistics online). After 2007 there is a steady decrease in mentions of *C. difficile*. There are fewer mentions in 2010 compared to 2007, attributed to better implementation of cleaning practice.

Currently in England there is a *C. difficile* PCR Ribotyping Network (CDRN) which allows for rapid ribotyping of samples (HPA, 2010). Indeed in 2008- 2009 the CDRN processed 4682 samples in England, the most prevalent being 027 (36%), 106 (13%) and 001 (7%) (Freeman *et al.*, 2010). The costs of *C. difficile* infection are estimated by the European Centre for Disease Prevention and Control (ECDPC) to be between €5.000-15.000 per case in England and $1.1 billion per year in the USA. Across Canada and the USA the incidences of *C. difficile* infection (CDI) have changed radically, increasing from 36.5 cases per 100 000 population in 1991 to 156.3 cases per 100 000 population in 2003 (Office of National Statistics, 2010). A similar epidemiological increase occurred from 2.7 cases per 1000 in 1997 to 6.8 per 1000 discharges in 2001.

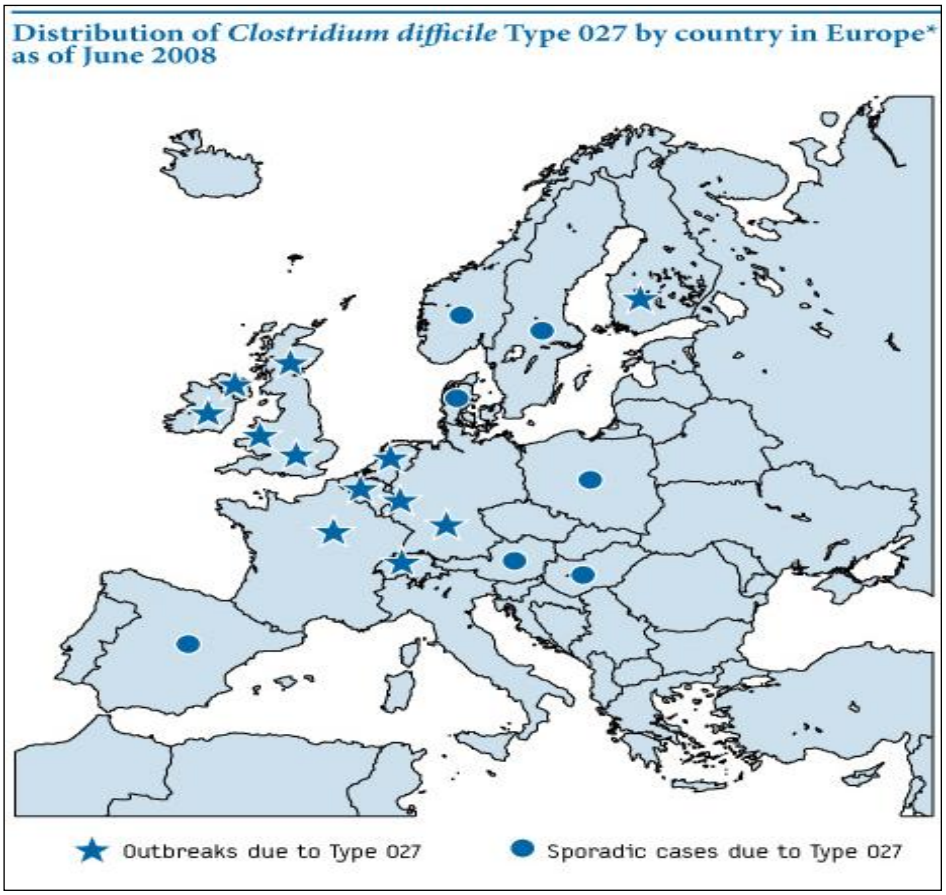


Figure 1.5: Distribution of *Clostridium difficile* type 027 strain in Europe by country

This image shows the areas across Europe where outbreaks of 027 ribotype occurred and the sporadic cases of *C. difficile*. (Taken from Eurosurveillance, June 2008: Kuijpers *et al.*, 2008.)

1.2.4. Incidence in Animals

There have been many documented cases of CDI in a range of livestock and domestic animals. *C. difficile* has been suggested to be an enteric pathogen of horses, associated with acute colitis post antibiotic treatment- usually with erythromycin (Baverud, 2002; Taha *et al.*, 2007). The organism has also been identified as a cause of enteritis in calves, with toxins A and B found to be in 25.3% and 22.9%, respectively, of diarrhoeic calves (Hammit *et al.*, 2008). The dominant isolate in the calf was found to be of ribotype 078. Ribotype 078 has been implicated as a cause of enteritis in piglets. Further genetic evaluation via multiple locus variable number tandem repeat analysis (MLVA) revealed porcine 078 strains and human 078 strains are genetically related (Debast *et al.*, 2009). Therefore there is a high degree of similarity between human toxinotype V strains and animal toxinotype V strains (Jhung *et al.*, 2008; Alvafrez-Perez *et al.*, 2009).

C. difficile has been found to cause enteritis in captive ostriches. In other species enteritis occurs in stressed neonates or antibiotic treated animals, however in ostriches *C. difficile* is presumed to be an important commensal (Frazier *et al.*, 1993). In this particular case described by Frazier, the presence of an antibiotic compound called sulfamerazine in feed could have induced clinical symptoms. Replacement of sulfamerazine with oral penicillin was shown to reduce mortality by 50% in the ostriches, but it seems that CDAD is a common problem in captive ostriches. Another interesting incidence of *C. difficile* is in desert locust *Schistocerca gregaria Forskål*, where it was found to be a commensal of the gut flora. *C. difficile* has also been found in retail meat and salads, indicating food could be involved in transmission from animals to humans (Bakri *et al.*, 2009). *C. difficile* has also been found in the environment. Studies of the distribution of *C. difficile* within the environment of South Wales have found its presence in river water at a percentage of 87.5%, sea water (44%) and in 46.7% of lake samples taken throughout South Wales (Al Saif & Brazier, 1996). *C. difficile* was also recovered from soil samples and 50% of swimming pools examined in the same study (Al Saif & Brazier, 1996).

1.3. GENERAL MICROBIOLOGY

1.3.1 *Clostridium difficile*

C. difficile is a member of the *Firmicutes* phylum, and the family *Clostridiaceae*. The bacteria grow optimally at 37^{o}C and are heterotrophic. *C. difficile* is an anaerobic spore forming bacillus, possessing subterminal spores and peritrichous flagella conferring its motility (Hafiz & Oakely, 1976). Bacilli have dimensions of 0.5-1.9 μm by 3.0-16.9 μm (Hatheway, 1990). Bacteria also stain Gram positive; however older colonies sometime do exhibit Gram stain variability (Aktories & Wilkins, 2000). *C. difficile* is known to have a pleomorphic nature, and varying morphologies have been observed with strains (Reynolds *et al.*, 2010; Siani *et al.*, 2010). Indeed dwarf colonies have been documented in culture, representing colonies from inoculated spores which have subsequently germinated (Wilkins, 2000; J. S. Brazier*: personal observation*).

Colonies have a ground glass appearance on agar with an irregular lobate edge. Typically on blood agar media the diameter of colonies ranges from 3-5 mm (Figure 1.6; Aktories & Wilkins, 2000). Colonies change form with age and sporulation is noticeable on cultures >72 hours post incubation. This is when the colonies have reached stationary or decline phase of growth. Interestingly these colonies have been described to have a "fried egg" appearance at 48-72 hours of incubation due to the raised whitish- grey centre spot visible in the colony.

On selective agar media colonies appear to have a bluish/ green tinge due to an alpha- type haemolysis (Aktories & Wilkins, 2000). The most common selective agar used for *C. difficile* isolation from stool specimens is cycloserine cefoxitin fructose agar (CCFA). Under ultraviolet light colonies can characteristically exhibit a yellow-green fluorescence (chartreuse) and exude a pungent horse manure-like odour (Delmee, 2001). The distinct odour is attributed to *C. difficile's* ability to produce isovaleric acid, isocaproic acid and *p*-cresol (Levett, 1984). Indeed *C. difficile* can withstand a concentration of 0.5% *p*-cresol, unlike other anaerobic bacteria which are inhibited at this concentration (Dawson *et al.*, 2008).

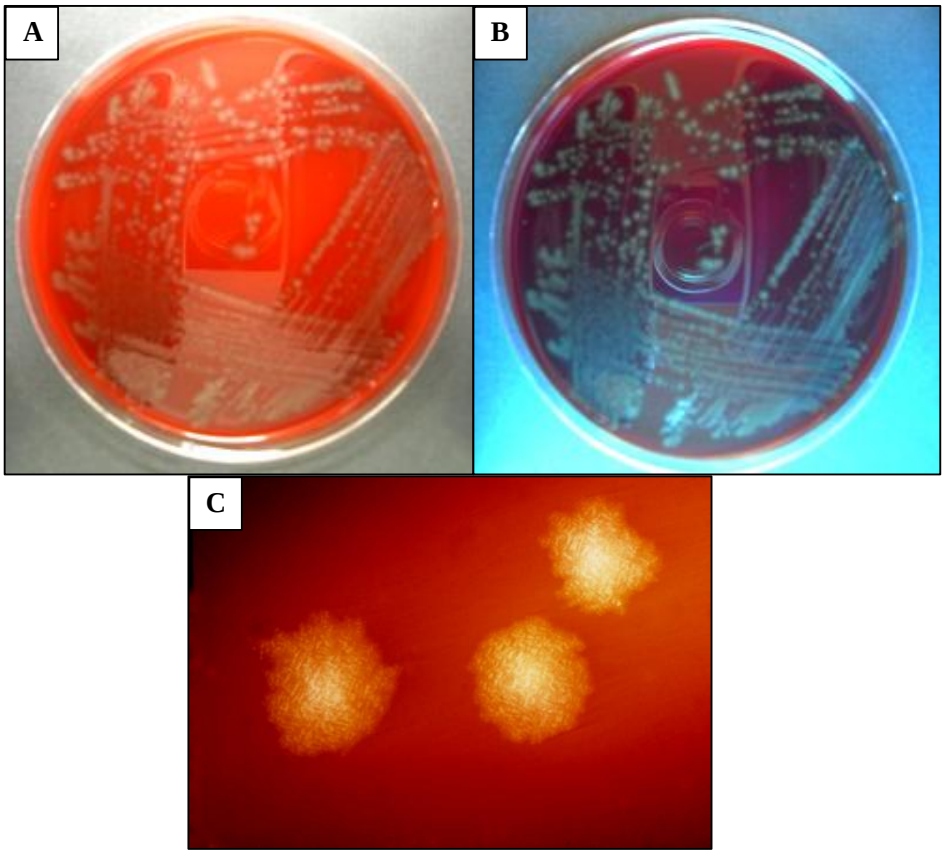


Figure 1.6: *Clostridium difficile* active cultures.
A: *C. difficile* grown on CCFA under white light. B: *C. difficile* grown on CCFA under UV light. Chartreuse colour visible. (Images courtesy of Daniel Phillips) C. Individual colonies of *C. difficile* showing characteristic lobate edges. (Image courtesy of Sciencephoto.com).

1.3.2. Vegetative cells

Vegetative *C. difficile* cells have complex structures, some of which serve as potential virulence factors (Figure 1.7). Certain strains *C. difficile* possess flagella which confer motility. Flagella are thought to be involved in gut colonisation and have been shown to play a role in cell adherence (Tasteyre *et al.*, 2000; 2001). Tasteyre *et al.* (2001) found that crude flagella were able to adhere 10x more strongly to mouse caecum tissue compared to a non-flagellated strain of the same serogroup.

C. difficile vegetative cells also possess fimbriae. Fimbriae are 4 - 9 nm in diameter and 6 nm in length and are thought to be involved in adhesion; however as yet no association between fimbriae and adherence ability of strains has been determined (Borriello, 1998). Some strains of *C. difficile* produce a polysaccharide capsule (glycocalyx), which is

thought to protect the bacterium from phagocytosis by polymorphonuclear leucocytes (PMN) i.e. it is an antiphagocytic factor. It has been suggested that toxigenic strains which produce more fimbriae and capsular material may cause more severe cases of disease; however this is still the subject of some debate (Bongaerts & Lyerly, 1997).

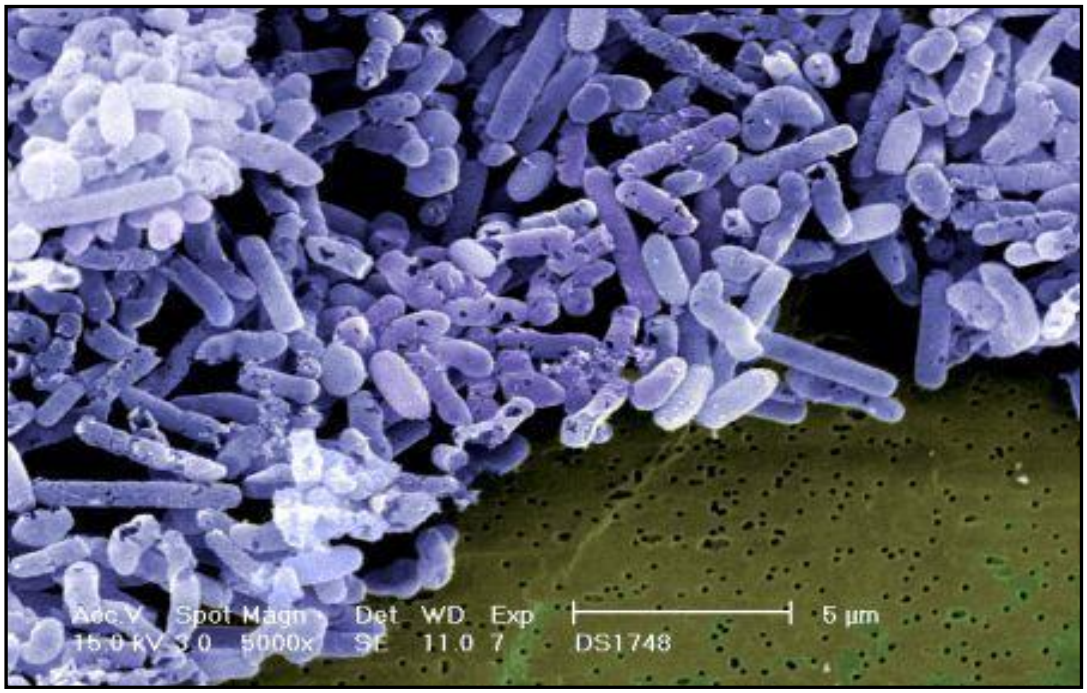


Figure 1.7: Vegetative cells of *Clostridium difficile.*

False-coloured SEM image of vegetative cells of *C. difficile* strain DS1748. Characteristic bacilli rods (purple) are visible. Scale bar represents 5 µm.

1.3.3. Vegetative cell proteins

The surface layer (S-layer) of *C. difficile* is a proteinaceous para-crystalline structure surrounding the vegetative cell (Fagan *et al.*, 2009). The S-layer has been implicated in adhesion and colonisation of cells within the gut and as a protector of the cell against parasitic attack by bacteriophages. The S-layer consists of S-layer proteins (SLP) (which act as a molecular sieve) and cell wall proteins (Cwp) (Waligora *et al.*, 2001). Eleven *cwp* genes are found to be clustered around the *slp* locus, and those downstream of *slpA* (*cwp*2-*cwp*7) are conserved between strains (Calabi *et al.*, 2001; Sebaihia *et al.*, 2006). Of all the Cwp, CwpV is the largest and is known to promote *C. difficile* aggregation (Reynolds *et al.*, 2011).

The S-layer comprises of two peptide subunits encoded by the single *slpA* gene and is the major surface antigen to which the immune system is exposed (Ni Eidhin *et al.,* 2008). The pre-protein of *slpA* contains a signal peptide which guides translocation across the cell membrane and production of mature SLP post cleavage (Fagan *et al.,* 2009). Mature SLP are high molecular weight (HMW) and low molecular weight proteins (LMW), ~40 kDa and ~35 kDa respectively, which facilitate adhesion to cell culture lines (Calabi *et al.,* 2001). The LMW protein is an immunodominant antigen (i.e. this protein is most easily recognised by the immune system, and thus influences the specificity of the induced antibody) and so antibodies against this protein are most common in sera from patients with CDAD.

Interestingly there are differences in sequence homology between strains of *C. difficile* particularly in the LMW protein which has been found to induce very limited immunological cross reactivity. Indeed the LMW antigen is the main basis of serotyping of *C. difficile,* which comprises of 21 main serogroups. The HMW sequences do not show any sequence identity to the LMW sequences, are highly conserved between strains and are immunologically cross reactive (Fagan *et al.,* 2009). The HMW sequences are 70-71% identical, whereas the LMW proteins possess 34-40% identity. The reason for sequence conservation in the HMW proteins has yet to be elucidated, although it has been noted that the variable regions of the LMW proteins appear to resemble those of flagellin sequences, which possess highly variable central sequence segments (Calebi *et al.,* 2002). Alternatively, this lack of sequence identity may prevent the host mounting a protective immune response and enable reinfection, thus conferring an evolutionary advantage (Calebi *et al.,* 2002). Hence as the S-layer surrounds the cell almost completely there is potential for it to be used as a vaccine, and an antibody response may also be used to eliminate carriage of *C. difficile* (Ni Eidhin *et al.,* 2008).

1.3.4. Spores

C. difficile produces spores in times of stress, such as nutrient depletion and in aerobic environments. Spores are highly resistant to desiccation, extreme temperatures and chemicals and contaminate the environment through faecal soiling from infected patients, often surviving in the environment for months (Gerding *et al.,* 2008). The spores themselves appear to be composed of a number of concentric layers as can be see in Figure 1.8. These layers consist of an inner core surrounded by an outer coat layer and a

cortex peptidoglycan layer. Indeed not all spores exhibit the same structure, with some spores possessing an outer exosporium layer hypothesised to play a role in spore adherence (Figure 1.8; Joshi *et al.*, Unpub.).

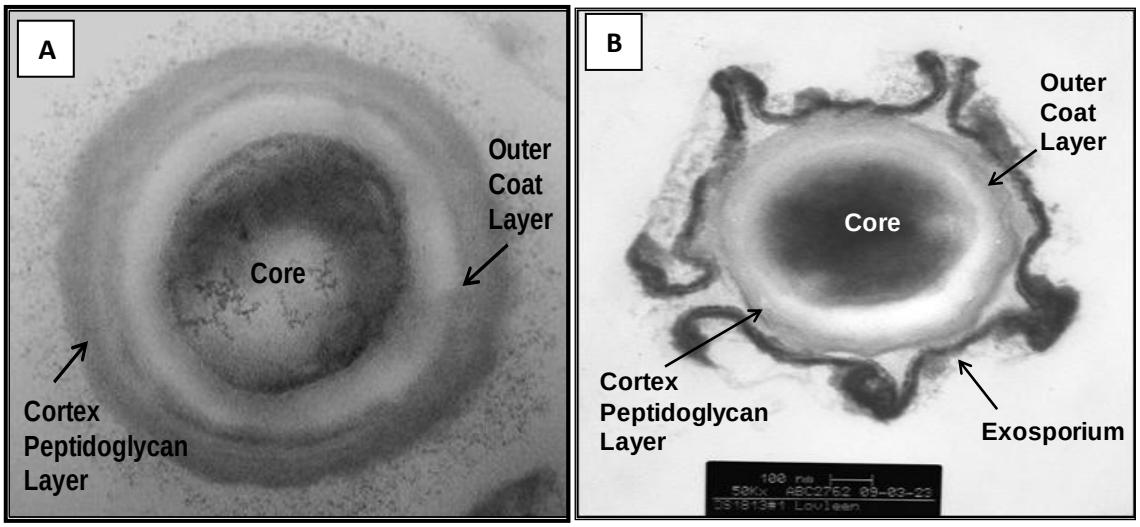


Figure 1.8: TEM images of *Clostridium difficile* spores.

A. TEM image of a spore of strain DS1748. Concentric layers of the spores can bee seen. B. TEM image of a spore of DS1813 with exosporium visible. The layers of the spore coat are visible. Scale bar represents 100 nm. Both images are of clinical isolates examined during this study.

1.3.4.1. Sporulation

Sporulation is induced by limitations in certain key nutrients in a population of cells (Durre, 2005). In *Bacillus* and *Clostridium* species the mechanism of sporulation is quite similar; however *C. difficile* sporulation has not been extensively studied. In times of nutrient depletion, vegetative cells generate spores. Initially the vegetative cells split asymmetrically producing two septa which both contain a copy of the chromosome (Sorg & Sonenshein, 2008). The mother cell subsequently engulfs the smaller forespore, assisting it with maturation, and addition of layers such as the peptidoglycan cortex and coat proteins (Figure 1.8) allows formation of the endospore. The mother cell lyses releasing the spore into the environment, where it lays dormant until conditions are adequate for germination to ensue (Durre, 2005).

Recently proteins expressed in the outer spore coat layer of strain CD630 have been identified as carrying catalase, peroxiredoxin and chitinase activity. These coat proteins

are CotCB, CotD and CotE respectively (Permpoonpattana *et al.*, 2011). CotCB and CotD proteins express putative manganese catalases and CotE act as a bi-functional protein with peroxiredoxin and chitinase activity (studied in Chapter 4), which is hypothesised to play a role in assembly of the spore coat via protein monomer polymerisation. However the exact functional activity of these proteins has yet to be fully elucidated.

1.3.4.2. Sporulation genetics

C. difficile lacks the genes present in most bacilli which govern the phosphor-relay system triggering sporulation. As *C. difficile* sporulation mechanisms are not yet clear, *B. subtilis* is used as a model organism for sporulation. In *B. subtilis* there are 5 sigma factors associated with sporulation working in a series to direct RNA polymerase to transcribe genes in varying stages (Figure 1.9; Piggot & Losick, 2002; Piggot & Hilbert, 2004). An early event in sporulation is the synthesis and activation of sigma factor H (σ^H), which joins the core polymerases and directs them to promoter sites: the Spo0A- genes.

Spo0A products negatively or positively regulate the genes and are transcription factors (Durre, 2005). These products allow for formation of the septa and forespore as described earlier (Section 1.3.4.1). In *C. perfringens* Spo0A assists in both toxin production and sporulation; however in *C. difficile* the role of Spo0A has been debated. Indeed Underwood *et al.* (2009) generated Spo0A mutants by inactivating genes encoding Spo0A and associated sensor kinases to understand their contribution to sporulation, revealing links between toxin production and the sporulation pathway. Sigma factors E and F are active in forespore development and are later replaced by sigma factors G and K. In *B. subtilis* the *sigG* gene region downstream shares synteny with *C. difficile* but has different genes inserted into this area (Durre, 2005). Interestingly, the genome of *C. difficile* possesses a prophage-like element known as *skin* inserted in the sigma-factor K (σ^K) gene. This 14.6 kb *skin* insertion encodes a RNA polymerase sigma factor required for efficient sporulation (Haraldsen & Sonenshein, 2003).

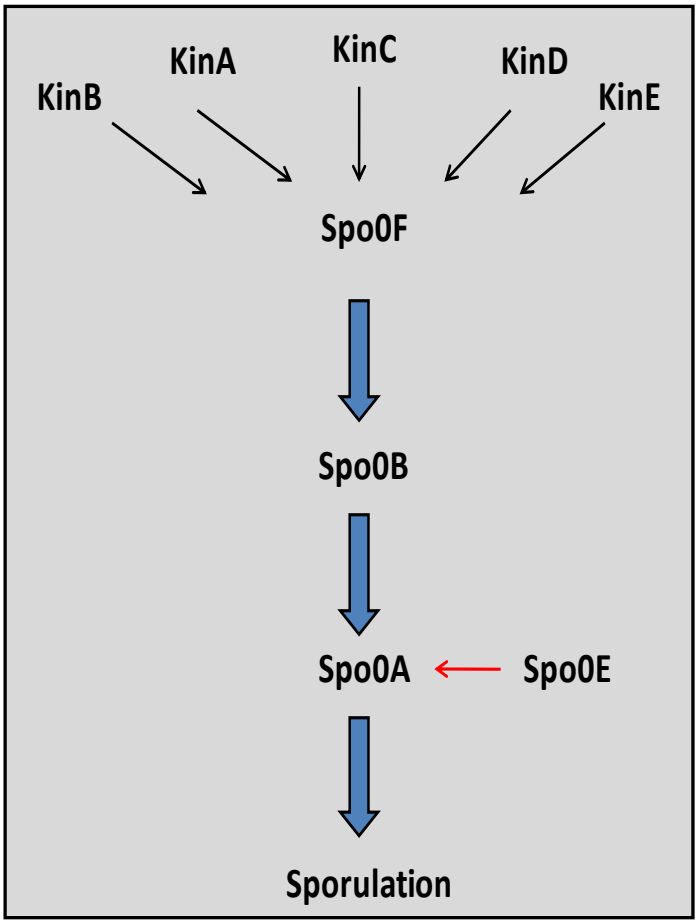


Figure 1.9: Sporulation Initiation Spo0A Pathway (Adapted from Piggot & Hilbert, 2004)
Kinases (KinA - KinE) phosphorylate the Spo0 genes: Spo0F- the sporulation initiation phosphotransferase. Spo0B the response regulator and the phosphoryl group is then passed onto the master regulator Spo0A (Paredes *et al.,* 2005). Spo0A is subsequently phosphorylated on an aspartate active site, resulting in binding to a target sequence known as the "0A box" near the promoters of genes controlled by Spo0A, leading to gene activation or repression (Underwood *et al.,* 2009). Some spore-forming bacteria have evolved to utilise Spo0A to adapt their virulence and ability to survive.

1.3.4.3. Spore germination

Germination occurs when there is change in optical density caused by spore rehydration and calcium dipicolinate (CaDPA) release (Sorg & Sonenshein, 2008). The stimulus of spore germination has yet to be established in *C. difficile*; however the primary bile salts cholate, taurocholate and glycolate have been shown *in vivo* to cause spore germination (Giel *et al.,* 2010). Indeed early experiments involved supplementing selective agar with sodium taurocholate and other bile salts such as deoxycholate (Wilson *et al.,* 1982; Wilson, 1983; Buggy *et al.,* 1985). The normal physiological concentration of sodium taurocholate in the duodenum is approximately 6.9 mmol l^{-1} and 1.2 mmol l^{-1} in the

jejunum (Leverrier *et al.*, 2003) which, interestingly, corresponds with the optimal range for germination of *C. difficile* previously found by Sorg & Sonenshein (2008) and Wheeldon *et al*. (2008). Glycine has also been found to be a potent co-germinant of *C. difficile* spores (Sorg & Sonenshein, 2008; Wheeldon *et al.*, 2008).

Conversely, chenodeoxycholate has been found to inhibit germination in the presence of taurocholate and cholate (Sorg & Sonenshein, 2009). Another inhibitor of spore germination *in vivo* is cholestyramine- a bile salt sequestrant, responsible for the removal of excess bile salts from the lower digestive tract (Giel *et al.*, 2010). Cholestyramine is used as a treatment for CDAD, but the efficacy of cholestyramine has been debated as bile acid synthesis increases 42-46 fold once cholestyramine is ingested. In addition cholestyramine is able to bind to the antibiotic vancomycin potentially competing with its activity (Giel *et al.*, 2010).

Spore germination is initiated in the presence of germinants, such as nutrients, surfactants, lysozyme, peptidoglycan and dipicolinic acid. There are no known germination receptors in *C. difficile*, and as such *Bacillus* spp are used as study examples. In *B. subtilis*, germination occurs in response to asparagine, glucose, fructose and potassium (Setlow, 2003). Three sensory receptors are located within the *B. subtilis* spore inner membrane which activates the germination process; *gerA*, *gerB* and *gerK* (Durre, 2005). Once germination is initiated there is no way of stopping the spore from converting into a vegetative cell, even if the germinants are removed. After this, there is cation (H^+, Na^+, K^+) and CaDPA release leading to hydrolysis of the spore peptidoglycan layer (Figure 1.8). At this point the water content of the spore increases, the core of the spore expands and enzymatic activity, metabolism and spore outgrowth is able to occur (Setlow, 2003).

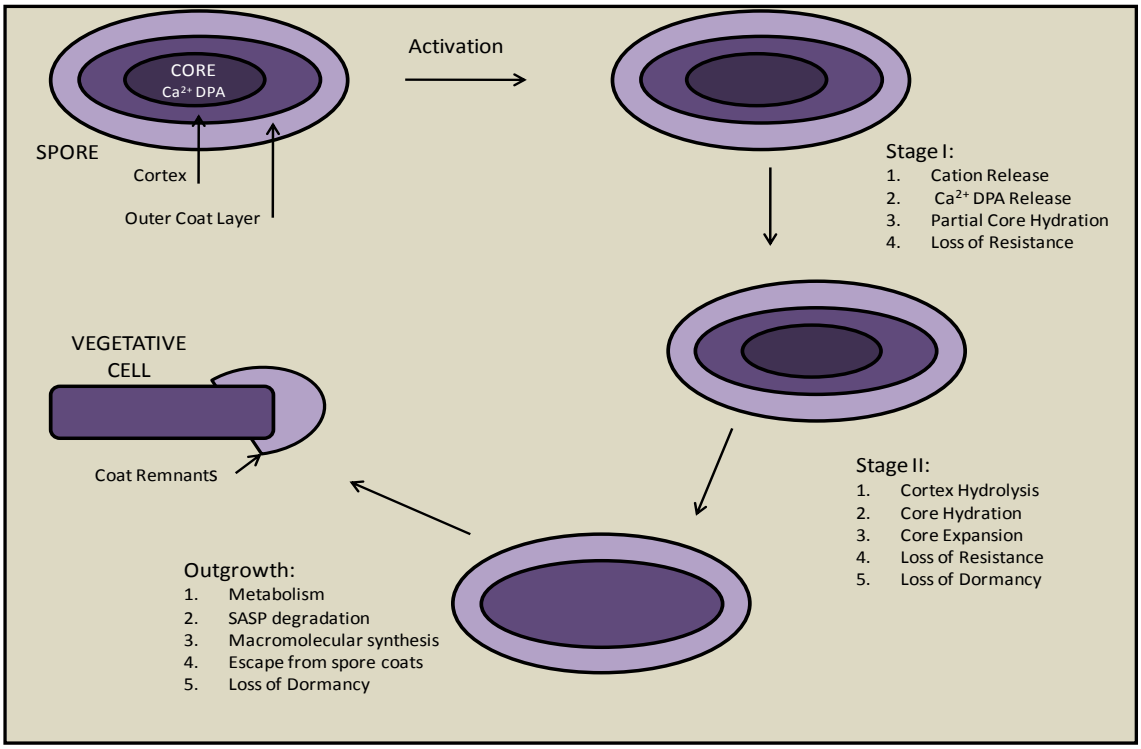


Figure 1.10: Main events in Spore germination (Adapted from Setlow, 2003)

The cycle of spore germination has distinct stages. The first is activation; the mechanisms of which have yet to be elucidated. Secondly Stage 1 involves core hydration and ion release, leading to stage 2 where there is total core hydration, expansion and loss of characteristic spore resistance. At the outgrowth stage there is degradation of small acid soluble proteins (SASP), degradation of spore coats and finally emergence of a vegetative cell capable of metabolism.

1.3.4.4. Enzymes necessary for spore germination

Spore germination enzymes have been well studied in *C. perfringens*. The enzymes associated with spore hydrolysis are SleB and CwlJ, known specifically as spore cortex-lytic enzymes (SCLEs). CwlJ breaks down the peptidoglycan layers leading to outgrowth, however the mechanisms of activation for SleB remains unknown. CwlJ is thought to be activated by CaDPA from the spore coat. Another enzyme –SleC- is essential for complete hydrolysis to occur, and enzyme SleM is able to degrade peptidoglycan in the spore cortex, but is not essential for spore germination (Burns *et al.*, 2010).

Sebaihia *et al.* (2006) found genes homologous to the CwlJ and SleB genes of *B. subtilis* (CD3563), and the SleC genes of *C. perfringens* (CD0551), within the genome of *C. difficile* CD630 (Sebaihia *et al.*, 2006). In fact, in a recent paper Burns *et al.* (2010) found

SleC mutants in *C. difficile* R20191 were unable to germinate and form colonies, even in the presence of germinants such as taurocholate, whereas the parent strains were capable of producing colonies. This suggests that SleC is in fact necessary for germination of *C. difficile* spores (Burns *et al.*, 2010).

1.3.4.5. Exosporium- a role in germination?

An exosporium layer has been found to surround spores of the highly plastic genome sequenced *C. difficile* strain 630 (Sebaihia *et al.*, 2006; Lawley *et al.*, 2009). Other investigations into the exosporium of *C. difficile* ATCC 9689 and 43594 have revealed that during early germination endospores produce strain-specific exosporial projections which appear to facilitate attachment of the spore to agar or to other spores (Panessa-Warren *et al.*, 1997). In a later paper (2007) Panessa-Warren also reported bumps and knobs covering the spore at the initial stages of attachment, which later developed into more pronounced protrusions. Interestingly the visualization of spore structure during germination and the structural role of the exosporium have not yet been defined. However it is clear that bile salts, spore structure and the function of certain spore layers must all play a part in germination of the spore in the gut.

1.4. MOLECULAR BIOLOGY OF TOXINS

1.4.1. Large clostridial toxins

Pathogenic isolates of *C. difficile* usually produce two toxins: toxin A (TcdA- enterotoxin) and toxin B (TcdB- cytotoxin). These toxins are part of the large clostridial toxin (LCT) family which target the rho-GTPases and the Ras-GTPases (guanosine triphosphases) responsible for regulation of actin cytoskeleton dependent cell functions. Furthermore these GTPases are also involved in cell apoptosis, transcription and transformation (Table 1.1; Aktories & Just, 2005; Alouf & Popoff, 2006). These single chained toxins are composed of three distinctive functional domains and are characterised by their cytopathic effects on cell culture (Von Eichel-Streiber *et al.*, 1996). Other classification properties include: (a) high molecular weight (ranging from 250-308 kDa); (b) an amino- terminal enzymatic domain with characteristic repeats; (c) a central hydrophobic region; and (d) a carboxy-terminal domain with carbohydrate recognition sequence repeats (CROPs) (Rupnik *et al.*, 2009).

Table 1.1: Properties of large clostridial toxins

The main properties of toxins from species within the family of LCTs are shown and compared.

Clostridium species	Toxin	Size (kDa)	Homology (%) (No. oligonucleotide repeats)	Cytopathic effect typical for:	Cellular target proteins	Position of glucosylation	Enzymatic Activity
C. difficile	TcdA	308	48/15	TcdB	Rho, Rac, Cdc42	Thr37 (Rho)	UDP-glucosylation
	TcdB	270	100	TcdB	Rho, Rac, Cdc42	Thr37 (Rho)	UDP-N-acetyl-glucosylation
	TcdB-1470	270	93/2	TcsL		-	
C. sordellii	TcsH	300	-	-	Rac, Ras, Rap, Ral		
	TcsL	270	76/14	TcsL	Rac, Ras, Rap1, Rap2	Thr35 (Rho)	UDP-glucosylation
C. novyi	Tcnα	250	32/16	TcdB	Rho, Rac, Cdc42	-	UDP-N-acetyl-glucosylation
C. perfringens	TcpL	-	-	-	Ha- Ras, Rap1B, Ras1	-	UDP-glucosylation
	TpeL	-	-	TcdB, TcsL		Thr35 (Rho)	UDP-N-acetyl-glucosylation

Toxins from other species classed within this family include *Clostridium novyi*'s alpha toxin (TcnA), *Clostridium sordellii*'s haemorrhagic (TcsH) and lethal toxins (TcsL), and *Clostridium perfringens* (types B and C) toxin TcpL, due to their structural and functional relatedness (Figure 1.11; Rupnik *et al.*, 2009). Recently *C. perfringens* has been shown to possess another LCT, TpeL (Nagahama *et al.*, 2011). This has implications in terms of *C. difficile* toxin evolution, as it was previously thought that *C. difficile* toxins were only related to two other species of clostridia; *C. novyi* and *C. sordellii* (Amimoto *et al.*, 2007). It is well documented that *C. difficile*'s two toxins share high similarity with *C. sordellii*'s toxins. Indeed, *C. difficile* TcdA is a functional homolog of TcsH, and TcdB is a homolog of TcsL (Figure 1.11).

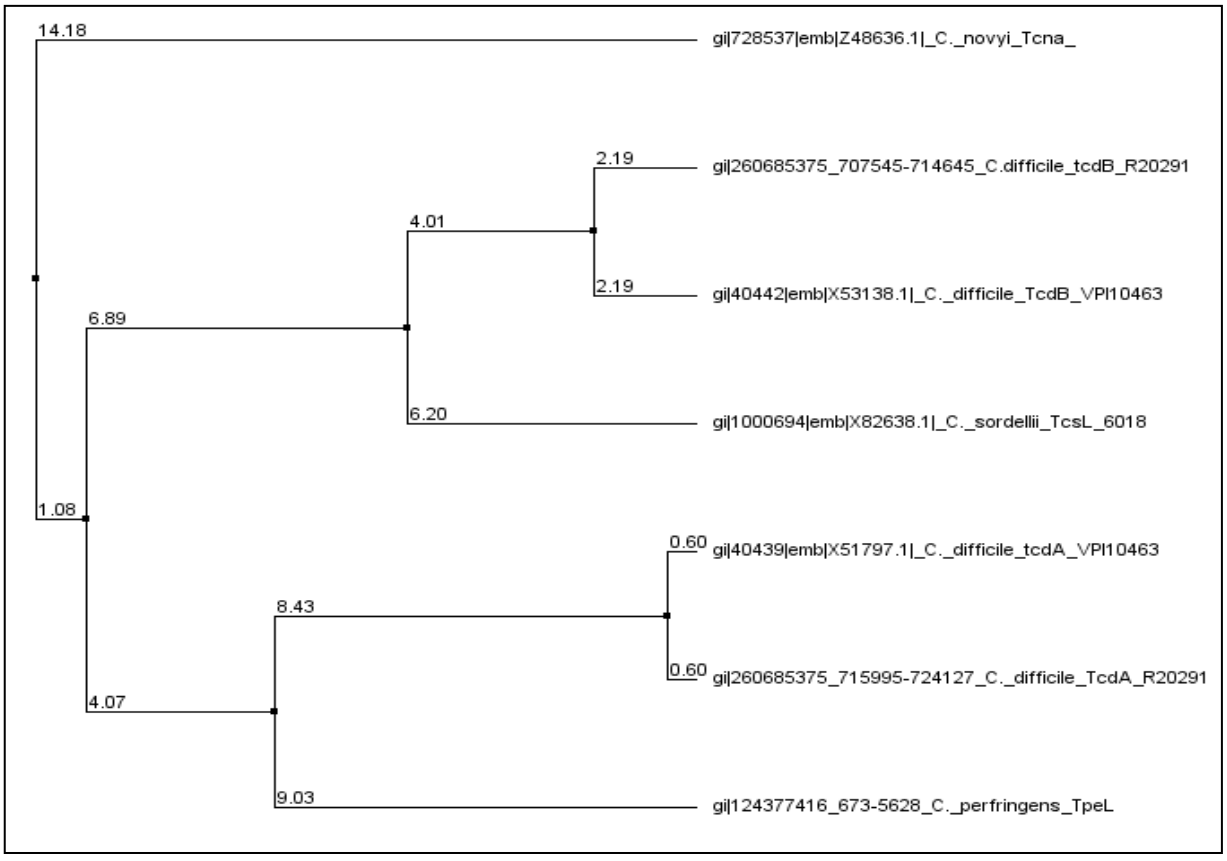


Figure 1.11: Phylogenetic tree demonstrating relatedness of large clostridial toxins.
Genetic relatedness is determined from nucleotide sequences of the toxins using average percentage distance (%). Tree constructed using the Jalview™ programme for multiple sequence alignments. Interestingly TcdA of *C. difficile* appears to be more closely related to the TpeL toxin of *C. perfringens*. TcdB is more closely related to the TcsL toxin of *C. sordellii*, and *C. novyi* appears to have diverged from the main toxin family, and thus has its own separate lineage.

1.4.2. Structure of the large clostridial toxins

The N- terminus of *tcdA* and *tcdB* is 543 aa and harbours the glucosyltransferase (GT) domain which is responsible for the glucosylation of Rho proteins. The central translocation domain is characterised by a hydrophobic region (956-1128 aa) which is thought to have a role in membrane penetration during the infective process by translocation of proteins across cellular membranes (Von Eichel-Streiber *et al.,* 1992; Aktories & Barbieri, 2005). The cysteine protease domain is also in this region. The combined translocation and receptor binding domains are classed as the delivery domain (Figure 1.12; Genth *et al.,* 2008).

The C- terminus harbours the receptor binding domain of *tcdA* and *tcdB* and a putative transmembrane domain which controls receptor- binding and pore formation. This region consists of short homologous aa residues known as CROPS- combined repetitive oligopeptides and exhibits homology to the carbohydrate-binding regions of GTs from *Streptococcus mutans* (Just & Gerhard, 2004).

CROPs within TcdA are 21 to 50 residues long and can be repeated throughout the protein. These repeats are isolate-specific, giving rise to the theory that this region of the toxin may have been generated from multiple duplication and recombination events in the gene (Rupnik *et al.,* 2005). The increased repetitiveness of sequence could also explain the agglutination activity of TcdA in the infective process (Voth & Ballard, 2005). Interestingly these repeats are only conserved at nucleotide level within *tcdA*, resulting in frequent homologous recombinations such as deletions and insertions (Alouf & Popoff, 2006). The *tcdB* encodes CROPs that are homologous to four of the CROPs of *tcd*A (Voth & Ballard, 2005).

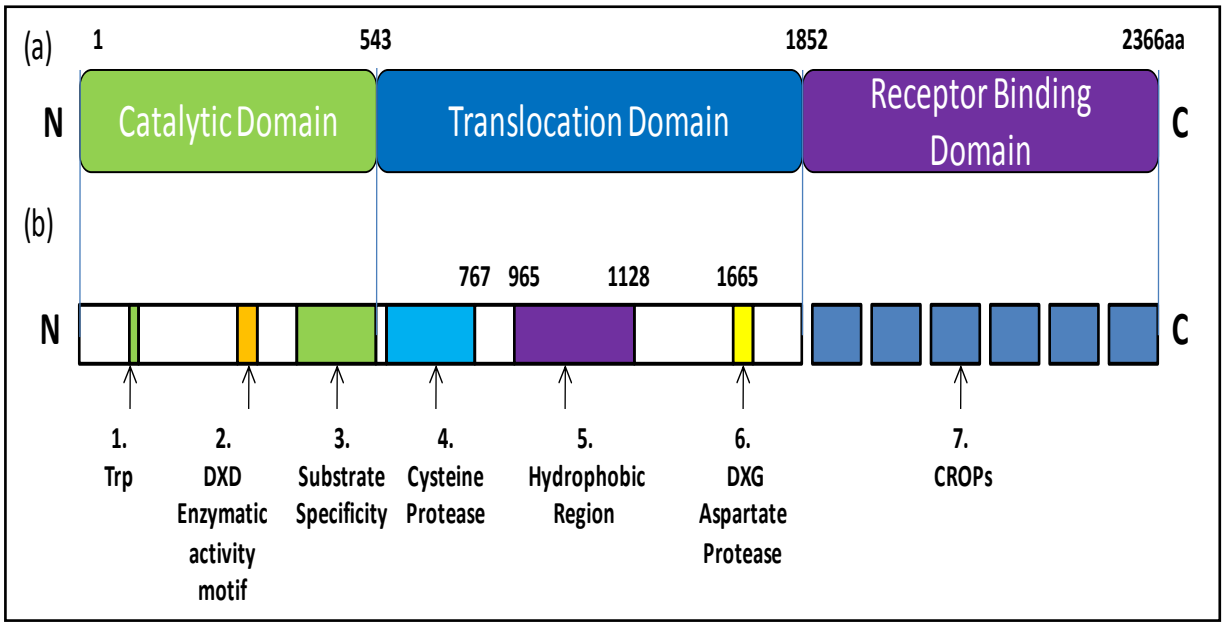


Figure 1.12: Detailed domain structure of large clostridial toxin of Toxin B.

Tripartite Structure of TcdB: (a) Three functional domains are visible. (b) Putative Functions of regions within domains given:

1. Trp: Binding site for phosphatidylserine

2. DXD Motif Glucosyltransferase Domain: Short conserved motif is found in many families of glycosyltransferases and is expected to play a carbohydrate binding role in sugar-nucleoside diphosphate and manganese dependent glycosyltransferases.

3. GTPase Substrate specificity: Catalytic domain harbours enzymatic activity

4. Cysteine Protease: Homology to RTX toxins of *Vibrio Cholerae*

5. Hydrophobic Region: Possible involvement in pore formation and delivery of the catalytic domain into the cytosol of the cell

6. DXG Aspartate Protease: DXG (1665aa) motif is suggested to be part of an aspartate protease domain, possibly involved in processing of the toxins.

7. CROPs: clostridial Repetitive Oligopeptides- repeating units.

1.4.3. Mechanism of action of the large clostridial toxins

LCT are glucosyltransferases that modify the host cell membrane G proteins which control cell processes- mainly targeting actin polymerisation, resulting in host cell death (apoptosis) (Poxton *et al.*, 2001; Salyers & Whitt, 2002; Voth & Ballard, 2005). The toxins glucosylate the Rho, Rac and Cdc42 GTPases at the threonine residue (Thr35/37) which is essential for the switch function of the GTPases (Giesemann *et al.*, 2008). This renders them inactive (Salyers & Whitt, 2002). The G proteins cycle between two states; the GDP (guanosine diphosphate) binding state and the GTP (guanosine triphosphate)

binding state (Salyers & Whitt, 2002). Toxins A and B act upon the GDP-bound state, as this configuration allows the threonine to be uncovered, thus allowing easier glucosylation by the toxins. This reduces the activity of the GTPase in the G protein, resulting in an imbalance of cellular function (Figure 1.13; Salyers & Whitt, 2002).

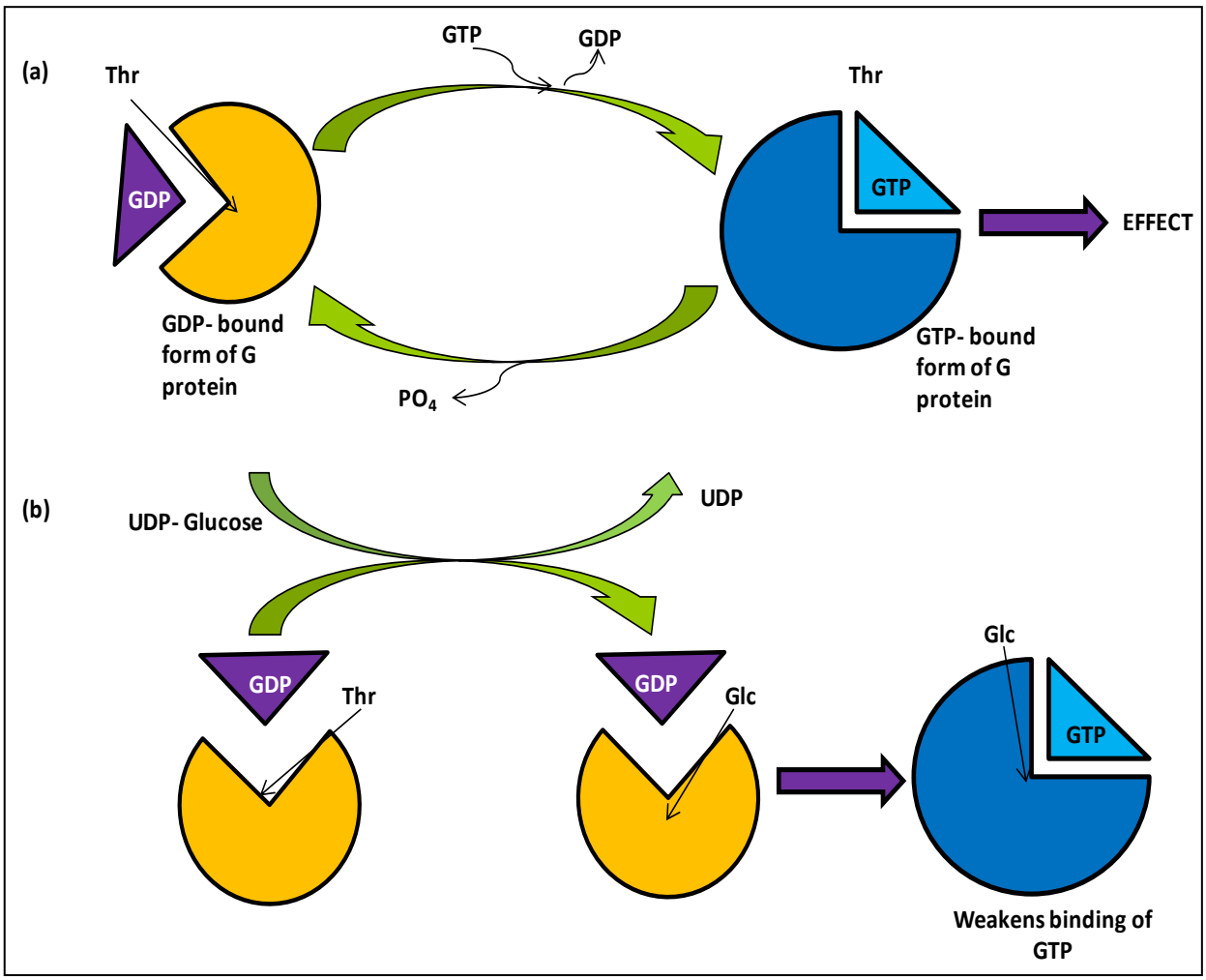


Figure 1.13: Toxin Mode of Action:

(a) The normal GTP cycle with the threonine residues being glucosylated. (b) The effect of glucosylation on the G proteins by toxins A and B. The glucosylated GTP bound form residue has a lower affinity for GTP, disrupting the normal functions of the G protein causing changes to host cells.

1.4.4. Mode of action of toxins A and B

The mode of action of the toxins is specific. Toxins A and B are both cytotoxic; however toxin B is more potent than toxin A. Both cause severe vascular permeability by disrupting junctions between cells and causing haemorrhaging (Poutanen & Simor, 2004). The toxins

bind to specific receptors. The receptor for toxin A in animals is the trisaccharide Gal-β1-4GlcNAc and Gal-α1-3GlcNAc which is not present in humans. In humans the co-receptor for toxin A is glycoprotein 96 (gp96), binding to the apical membrane of the host cell (Rupnik *et al.,* 2009). Toxin B binds to an unknown receptor on the baso-lateral membrane of the host cell (Rupnik *et al.*, 2009).

Upon binding both toxins are endocytosed by low pH endosomes. This in turn induces structural changes leading to pore-formation. Qa'Dan *et al.* (2000) established this by using bafilomycin- a macrolide antibiotic derived from *Streptomyces griseus*, finding that endosomes with a low pH may cause a conformational change in the protein regions in the toxins. The toxin N termini are translocated into the cytosol across the membrane by the translocation domain and then transferred into the cytosol through pores (Poxton *et al.,* 2001). The glucosyltransferase domain in the cytosol is released by autoproteolysis and targets GTPases (Rho, Rac). The delivery domain is then cleaved by autoproteolysis triggered by inositol hexakisphosphate ($InsP_6$) (Davies *et al*., 2011; Reinike *et al.,* 2007).

The cysteine protease domain is also thought to play a part in the auto-proteolytic cleavage, as it is located adjacent to this site (Ergerer *et al.,* 2007). This cysteine protease is linked to the RTX toxin of *Vibrio cholerae*, which possesses a catalytic triad comprising an aspartate-1665 residue (also present in *tcdB*), a cysteine 698 residue and a histidine 653 residue. Hence it has been suggested that the clostridial glucosylating toxins belong to a new family of cysteine proteases. The catalytic domain of the toxin is translocated into the cytosol of target cells, which supports the theory of toxin autocatalytic cleavage (Egerer *et al.,* 2010). Dithiothreitol (DTT) induces cysteine protease activity by increasing the synergistic effect of $InsP_6$, which in turn activates the cysteine protease domain (Reineke *et al.,* 2007).

1.4.5. Pathogenicity Locus

Toxin production is encoded by the *tcdA* and *tcdB* genes located on a 19.6 kb pathogenicity locus (PaLoc) that encompasses three additional small open reading frames (ORFs): *tcdR*, *tcdE* and *tcdC* (Figure 1.14; Cohen *et al.,* 2000; Spigaglia & Mastrantonio, 2002). Toxin A is 8.1 kb in length and is 308 kDa consisting of 2710 aa residues and toxin B is 7 kb in length and is 270 kDa consisting of 2366 aa residues (Bongaerts *et al.,* 1994; Faust *et al.,* 1998).

The PaLoc is located between two insertion sequences- *cdu 2/2* and *cdd 2-3* located upstream and downstream respectively of the PaLoc (Cohen *et al.*, 2000). The locus also contains ORFs for insertion sequence *cdu-4* (Spigaglia & Mastrantonio, 2002). In every toxigenic *C. difficile* strain the PaLoc has been found to be present at the same chromosomal integration site, and in non-toxigenic strains this site is replaced by a 115 bp non-coding sequence (Rupnik *et al.*, 2009; Stare & Rupnik, 2010). This gene organisation is not conserved in all strains, and usually *C. difficile* strains are compared to reference strain VPI10463 (Figure 1.14) via a system known as toxinotyping, as described in Section 1.5.2.

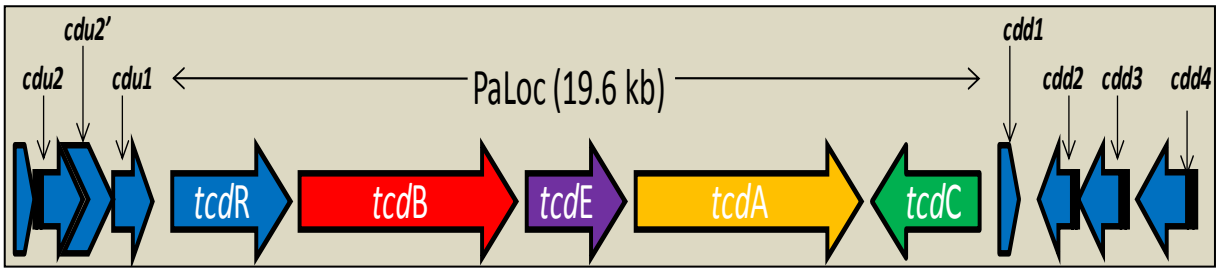


Figure 1.14: Pathogenicity Locus of *C. difficile*

PaLoc of *C. difficile* strain VPI 10643 (toxinotype 0). The PaLoc includes two large toxin genes (*tcd*A, *tcd*B) and three accessory genes involved in various regulatory and transportation functions. The PaLoc is located between the Open Reading Frames (ORFs) of insertion sequences *cdu 2/2* and *cdd 2-3* and *cdu-4* in non-toxigenic strains (Adapted from Bruggeman & Gottschalk, 2009).

The toxins genes *tcdA* and *tcdB* have ~66% sequence homology, functional homology and a low GC content. They are also similar in their biological functional activity, using their highly conserved N-terminal region to glucosylate substrates (Voth & Ballard, 2005). Due to this it has been proposed by Von Eichel-Streiber *et al.* (1992) and others that the genes arose from a duplication event.

Other genes located in the PaLoc include *tcdR* (formerly known as *tcdD* and as TxeR), *tcdE* and *tcdC* genes (Braun *et al.*, 1996). Gene *tcdR* encodes an alternative sigma factor involved in positive transcriptional regulation of toxin expression. Indeed *tcdR* was first found to be an unconventional positive regulator in expression experiments using *E. coli*, (Moncreif *et al.*, 1997). These results were confirmed by experiments using *C. perfringens* and then *C. difficile* (Mani & Dupuy, 2001). These experiments assessed *tcdR*'s ability to

act as a sigma factor, and established that a combination mixture of core RNA polymerase and *tcdR* is able to bind to the *tcdA*, *tcdB* and *tcdR* promoter regions (Mani, 2002).

Hence *tcdR* functions as an RNA polymerase sigma factor for toxins A and B and activates its own expression in a similar manner to its effect on the toxin promoters (Mani & Dupuy, 2001; Dupuy *et al.,* 2008). The *tcdR* gene is also homologous to the *cntR* gene (previously known as the TetR and BotR genes) regulator of the toxin *cnt*ABCDE genes of *C. botulinum* and *cnt*A gene of *C. tetani* respectively (Sebaihia *et al.,* 2007). These proteins act as alternative sigma factors and are so different in their structure and function that they have been classed into their own subgroup of sigma-^{70} family (Dupuy *et al.,* 2008). This family includes the *Uvi*A gene of *C. perfringens*, which regulates an UV-induced bacteriocin (Bruggeman, 2005).

The negative regulator is encoded by the *tcdC* gene and is responsible for down regulation of toxin A and B production (Hinkson *et al.,* 2008). Gene *tcdC* is a 231 aa and 693 bp (Govind *et al.,* 2006). The *tcdC* gene is not similar to any regulatory protein known as it is highly variable. Some strains possess deletions in *tcdC* of 18 bp and/or 39 bp which are associated with toxin overproduction. This is evident in toxinotype III, of which the hypervirulent 027 strain B1/NAP1 of *C. difficile* is part (Spigaglia & Mastrantonio, 2002; Loo *et al.,* 2005; McDonald *et al.,* 2005).

Transcription of *tcdC* occurs early in the growth phase of *C. difficile* (Cohen *et al.,* 2000). *tcdC* has been studied by Matamouros *et al.* (2007), who found that *tcdC,* when co-expressed with *tcdR*, caused less activity of the toxins, resulting in reduced expression of *tcdA*. This occurred as *tcdR* activated expression from the *tcdA* promoter, giving the first direct evidence that *tcdC* negatively regulates *tcdA* transcription by interfering with the *tcdR*- containing RNA polymerase (Figure 1.15; Dupuy *et al.,* 2008). Indeed the putative transmembrane domain located with the *tcdC* region has been proven by Govind *et al*. (2006) to be a membrane associated protein.

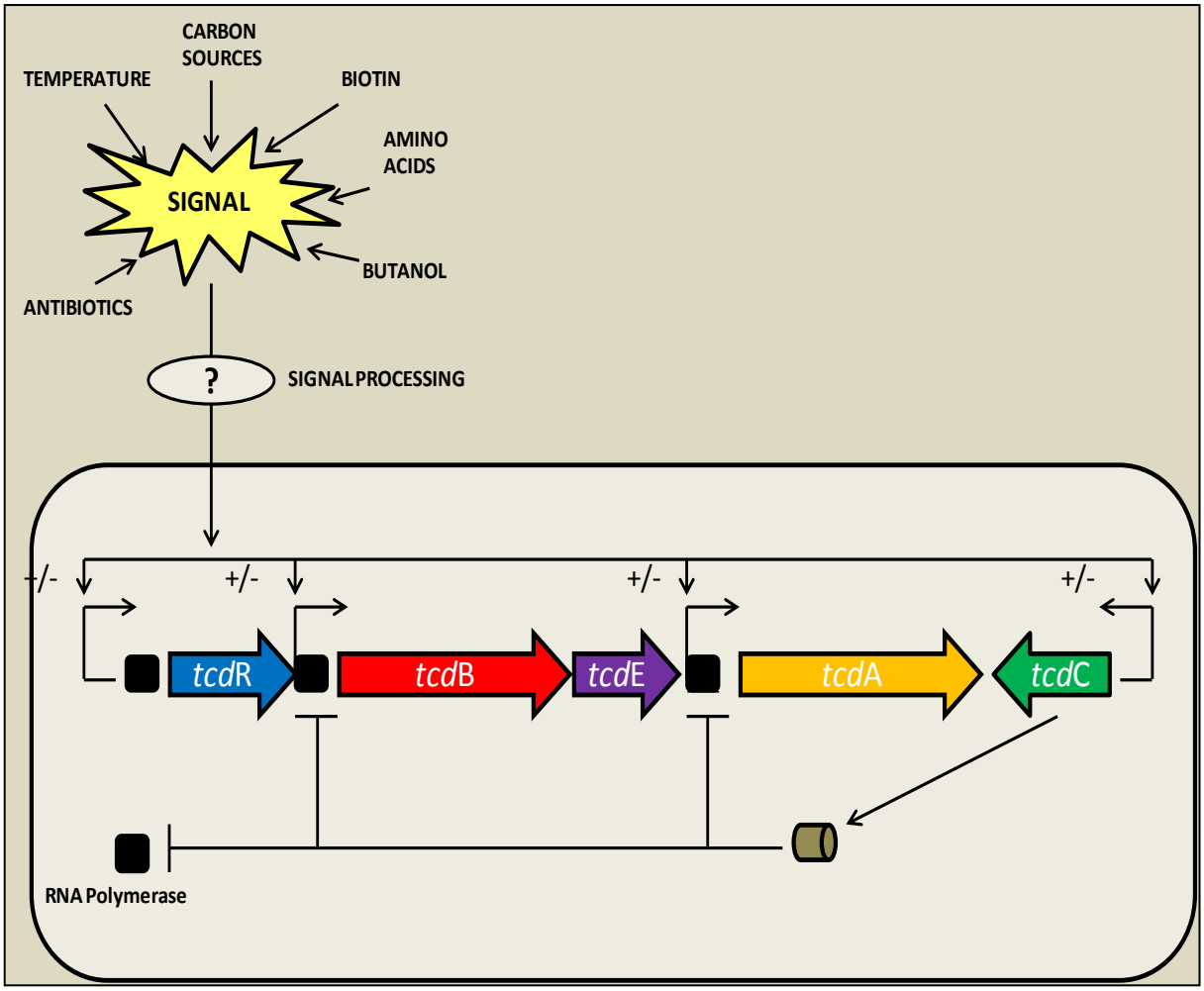


Figure 1.15: Schematic representation of *C. difficile* toxin gene regulation by *tcdC*

The *tcdC* gene represses transcription of *tcdA* and *tcdB* via inhibition of *tcdR* RNA polymerase interaction or the ability of *tcd*R to recognise the toxin promoters. When *tcdC* is absent, and there is presence of positive stimuli, *tcdR* transcribes toxin production (Matamouros *et al.*, 2007). Adapted from Dupuy *et al.* 2008.

The *tcdE* gene is a 501 bp ORF between *tcdA* and *tcdB*, which encodes for a 170 aa hydrophobic protein. Gene *tcdE* appears to have high homology with a phage holin-like pore forming protein thought to contribute to release of the toxin from bacterial cells (Bartlett, 2006; Jank *et al.*, 2007; Rupnik *et al.*, 2009). Holins themselves are cytolytic proteins which cause lysis of the bacterial host in order to release progeny phage. Indeed expression of *tcdE* in *E. coli* caused bacterial cell death, which suggests that *tcdE* in some way functions as a lytic protein, potentially facilitating release of toxins A and B into the extracellular environment (Tan *et al.*, 2001).

There is high homology between PaLoc genes and phage genes. The most notable is *tcdE*'s phage – like holin. Interestingly *tcdA* also shares homology with phage gene ϕ CT2 in *Clostridium tetani* strain E88, 55% sequence identity with *Lactobacillus casei* phage ϕ A2 (ORF 22) and homology with a 103 aa section of *tcdC* in strain VPI 10463 (Canchaya *et al.*, 2003; Goh *et al.*, 2005). This suggests that the origins of the PaLoc may have been an ancient prophage (Fortier & Moineau, 2007).

1.4.5.1. Evolutionary origins of the PaLoc

The PaLoc is a genetic element hypothesised to have been incorporated into the *C. difficile* genome from an unrelated strain via a process called horizontal gene transfer (HGT). This hypothesis has yet to be proven and thus the genetic element is still designated a "pathogenicity locus" (Rupnik *et al*., 2005). The locus was defined by Braun *et al.* (1996) as a distinct genetic element incapable of mobility. Braun *et al.* also proposed that the PaLoc may be a self-governing part of a larger genetic element linked with virulence.

Thus there is no surprise that many scientists are reluctant to speculate upon the origins of the PaLoc. Indeed there is a distinct possibility that at some point in the organism's evolution that the PaLoc may have originally been a pathogenicity island (PAI) acquired via HGT into the *C. difficile* genome, but has lost its mobility over time. Interestingly the PaLoc is replaced by a 115 bp fragment in non-toxigenic strains which supports this argument. The flanking regions of PAIs are characterised by direct repeats which are involved in recombination events. This can lead to deletion of the PAI (Alouf & Popoff, 2006). Characteristics representative of a PAI include presence in pathogenic strains and absence in less pathogenic strains of the same species, encoding of toxins, mobility between genomes (in some cases mediated by bacteriophage) and genetic instability (prone to deletions) (Woodford & Johnson, 2004).

Sebaihia *et al.* (2006) reported that *C. difficile* strain CD630 possesses a highly plastic genome consisting of 11% mobile genetic elements (transposons, prophages, insertion sequences, introns). It is understood that pathogens with high genetic flexibility, such as *C. difficile*, may be more competent in colonisation of new ecological niches and may have a selective advantage over other organisms with less plastic genomes (Hacker *et al.*, 2003). Thus this foreign DNA may have been originally integrated to give *C. difficile* an

advantage over other organisms due to selective pressure. We may already be seeing the evidence of this in the form of the hypervirulent ribotypes recently emerging to cause increased severity of disease.

1.4.5.2. Evolutionary origins of the PaLoc and horizontal gene transfer

Horizontal gene transfer (HGT) also occurs between microbes within the host environment. There are a number of studies demonstrating the transfer of antibiotic resistant determinants between Gram positive and Gram negative bacteria within the intestinal tracts of humans and animals (Shoemaker *et al.,* 2001; Igimi *et al.,* 2008). Indeed the human gut is an environment where HGT should occur frequently, as there is a constant supply of nutrients, numerous bacteria (10^{12}/g) and surfaces for bacterial adherence (Shoemaker *et al*., 2001).

Transfer of genes can occur in many ways. The first process is usually via transformation, involving the stable uptake of extracellular DNA, under natural bacterial growth conditions called natural transformation. The ability of DNA acquisition via transformation amongst a broad range of bacteria also indicates that this ability is conserved amongst many species and thus is important in evolution (Thomas & Nielson, 2005). Conjugative transfer is a process by which bacteria acquire plasmids. Bacteria are known to form a pilus to allow the transfer of plasmid material to other cells, which results in a junction between the donor bacteria and a pore through which the plasmid DNA and genes can be transferred to the recipient cell. Plasmids are small genetic elements which can be transferred quickly, and are facilitated by transposons, insertion sequences and other mobile elements which allow integration of the DNA into the host genome (Madigan & Martinko, 2005). DNA can also be incorporated into a bacterium by bacteriophage, in a process known as transduction (Elliott *et al.,* 2007). Lytic bacteriophage incorporate host DNA into their genome, and this gets transferred as the bacteriophage replicates, resulting in a transducing particle which is released along with normal phage.

Bacterial genomes are highly plastic, and thus some genes have the ability to be horizontally transferred. For example, it has been hypothesised that species related to *C. difficile* may have evolved from a common ancestor as many genes shared by these species have high sequence identity (Von Eichel-Streiber *et al.,* 1992; He *et al.,* 2010). *C. difficile*'s close relatives include *C. sordellii* known to cause toxic shock, and *C. novyi*

which is present within soil, manure and also produces toxins. The toxin gene *tcdB* has 85% homology and 74% identity with TcsL of *C. sordellii* and the G+C content of *C. difficile* toxin A and B encoding regions are ~28% G+C, which is similar to that of the entire *C. difficile* genome (G+C content of ~29%) (Voth & Ballard, 2005). A feature of a typical PAI is a different G+C content from the rest of the genome and the *C. difficile* PaLoc does not show this variation. Therefore it may be plausible that the acquisition of virulence genes by *C. difficile* was not a recent event, and that the toxins which are so prevalent within the clostridial family may have in the past been part of a plastic, mobile gene pool where HGT may have occurred frequently within ancestors of the organism, resulting in the spectra of toxin- possessing clostridia which we see today.

1.4.5.3. Evolutionary origins of the PaLoc and codon usage

Codon usage bias refers to the frequency at which synonymous codons occur within DNA, which is important when attempting to understand how an organism has evolved (Bulmer, 1991). Codons are recognised by several tRNAs, and those recognised by the most abundant tRNA are used more often than those recognised by the rarer tRNAs. In *E. coli* the quantification of cytoplasmic tRNA revealed that the relative abundance values of varying tRNA molecules were different (Ikimura, 1981). This difference is thought to be associated with amino acid abundance in proteomes, where the tRNA abundance correlated with the abundance of codons in genes of unicellular organisms, such as *E. coli* and *Saccharomyces cerevisiae*. This lead to the theory that codon usage and tRNA genes co-evolved, allowing for optimal translation in organisms. Altered codon usage in organisms is often indicative of some sort of HGT having occurred.

There currently is little literature regarding the codon usage within the *C. difficile* PaLoc and no evidence of tRNA sequences either within or outside of the PaLoc (Braun *et al.*, 1996). Even so, some species of clostridia do show signs of HGT within their genomes, such as *C. botulinu*m type A strain that possesses a chromosomally encoded BoNT locus, whose downstream boundary is flanked by two genes with high sequence similarity to IS element IS1069 of *Lactococcus lacti*s (Bruggeman, 2005).

A recent paper by He *et al.* (2010) examined the phylogeny, HGT, recombination and evolutionary history of *C. difficile* at the whole genome level. It has been found that

strains associated with disease evolved from multiple lineages, suggesting that *C. difficile* is more dynamic than previously thought. This has refuted the idea that these virulent isolates evolved from a single lineage. Indeed the paper suggests that the *C. difficile* and *C. tetani* lineages diverged around 2.34 billion years ago, and *C. difficile* itself 85 million years ago (He *et al.,* 2010). Interestingly although this common ancestor of *C. difficile* is millions of years old, the lineages all contained isolates of 017, 078 and 027 ribotype, which implies that there may be genetic elements conferring virulence characteristics which are common to all strains. Indeed this poses questions about how host-pathogen interactions have changed in the past thirty years, as *C. difficile* was only implicated as a cause of CDAD since then. Have changes in human activity, such as antibiotic usage affected the recent evolution of *C. difficile*?

1.4.6. Variant *C. difficile* toxins

While the 19.6 kb PaLoc is present in all toxigenic isolates of *C. difficile* with a defective PaLoc, truncated *tcdA* genes are still capable of causing clinical disease (Stabler *et al.,* 2008). Indeed the DNA sequence of the PaLoc varies between *C. difficile* strains; a factor which has been exploited by Rupnik *et al.* (2001) in the construction of the toxinotyping classification system (as described in Section 1.5.2). Currently 31 toxinotypes have been identified (Rupnik *et al.,* 2009).

Within the PaLoc variations such as deletions are usually located in the repetitive regions of *tcdA* 3' and 5' ends (Bruggeman & Gottschalk, 2009). Interestingly in *tcdB* variations only occur at the nucleotide level, and no truncated forms of *tcdB* have been found yet. This suggests that *tcdB* is essential for disease manifestation when compared to *tcdA* where deletions can range from 100 bp to 5.9 kb in toxinotype X (Soehn *et al.,* 1998). Toxinogenic *C. difficile* strains usually produce both toxins (A^+B^+); however some variant toxin strains have been reported. There are seven toxin production types of *C. difficile* identified to date as listed in Table 1.2.

Table 1.2: Seven types of toxin production in *C. difficile*

The ability of strains to produce toxin A, toxin B and the binary toxin CDT are displayed. The corresponding toxinotype and molecular background of the strains is listed. Adapted from Rupnik *et al.* (2008).

Toxin Production Type	Toxinotype	Molecular background
A+ B+ CDT-	0 Minor types: I, II, XII, XIII, XVIII, XIX, XX Major types: XXI	CDT- : absence of entire/ large part of CDT locus
A+ B+ CDT+	Minor types: XXIV Major types: III, IV, V, VI, VII, IX, XIV, XV, XXII, XXIII	CDT+ : presence of full length CDT locus
A- B+ CDT-	VIII, some 0 strains	A-: nonsense mutation at aa position 47 in tcdA gene A-: mechanism unknown
A- B+ CDT+	X XVI, XVII, some V-like strains	A-: rearrangement in PaLoc and large deletion causing probably changes in regulation and low or no transcription of truncated tcdA A-: mechanism unknown
A+ B- CDT+	IX-like	B-: mechanism unknown
A- B- CDT+	XIa, XIb some strains without PaLoc	A-, B-: only small non-functional part of PaLoc present A-, B-: no PaLoc
A- B- CDT-	PaLoc and CDT locus negative strains	Complete absence of tcd and cdt genes

The best characterised strains are the atypical 8864 strain and serogroup F strain 1470, which both have truncated 3' ends on their *tcdA* genes, resulting in a 6.0 kb deletion in 8864 and a 1.7 kb deletion in the immunodominant repetitive ligand domain of 1470 (Letournier *et al.*, 2003). Other studies have indicated that some variant strains in serogroups F and X have a mutation in 5' end of *tcdA*, which results in the untimely termination of *tcdA* translation. Although strain 1470 is not known to cause CDAD, other strains belonging to the serogroups X and F have been involved in CDAD (Letournier *et*

al., 2003). The cytotoxins produced by these strains display unusual cytopathic effects in cell culture reminiscent of that exhibited by the lethal toxin of *C. sordellii* (Sambol *et al.,* 2000).

Deletions in the *tcdC* gene have recently become important as this gene encodes the negative regulator of the toxin (Dupuy, 2008). These types of deletion can be differentiated into four groups from 18 bp to 56 bp (Curry *et al.,* 2007). The BA/NAP1 027 ribotype strain which demonstrates high toxin A and B production and binary toxin production is characterised by a deletion in the *tcdC* locus (Jank *et al.,* 2007). These hyper-virulent toxin variant strains demonstrate increased morbidity and mortality compared to other isolates (Drudy *et al.,* 2007). Only in two toxinotypes is the conformation of the PaLoc totally changed- in toxinotypes X and XI, the latter of which only has *tcdC* and two thirds of *tcdA* left.

Insertions within the PaLoc, although rare, do occur in several toxinotypes (Bruggeman & Gottschalk, 2009). Toxinotype X has an insertion of 1.1 kb between *tcdE* and *tcdA* in strain 8864, and a 2 kb insertion in *tcdA* in strains of toxinotype XIV (Alouf & Popoff, 2004). This 2 kb insertion at the 5' end of *tcdA* was found to be an IStron: CdISt1, excised from mRNA and specific only to *C. difficile,* and has now been found in other toxinotypes (XVII, XXII, and XXIII) (Mehlig *et al.,* 2001; Bruggeman & Gottschalk, 2005). It has been suggested that truncations and deletions within a genome, as in the *C. difficile tcdA* and *tcdC* genes, may represent a specific form of adaptation for the pathogen to certain niches, hosts, tissues or during transition of the pathogen from a chronic state of infection to an acute state or *vice versa* (Alouf & Popoff, 2004).

1.4.7. Binary toxin genes

Several *C. difficile* strains express a binary toxin (CDT) in addition to toxins A and B which is part of a family of adenosine diphosphate (ADP) ribosylating toxins. The CDT is unrelated to *tcdA* and *tcdB,* but is homologous to the toxins expressed by the VIP1 genes in *Bacillus cereus,* CST of *C. spiroforme* and the C2 toxin in *C. botulinum* (Table 1.3; Durre, 2005). It is also closely related to the *C. perfringens* Iota Ia and Ib toxins, sharing approximately 80% and 82% amino acid sequence identity respectively. Differences within the ADP ribosylating toxins family include varying substrate specificity for actin isoforms and in the steps of cellular uptake (Barth *et al.,* 2008). The protein components

of CDT and toxins from other clostridial species, such as *C. sporogenes* and *C. perfringens*, are interchangeable, generating biologically active chimeras. As these toxins commonly cause gastrointestinal infections in animals and humans this indicates potential evolutionary links between these pathogens (Barth *et al.*, 2004; Geric *et al.*, 2006). This relatedness demonstrates that binary toxin synthesis in these species may have resulted from a shared evolutionary past, supporting the hypothesis that there may have been a common ancestor from which *C. difficil*e evolved.

Table 1.3: Binary Toxins that modulate the actin cytoskeleton

The binary toxins depolymerise actin filaments and alter cell barriers using ADP- ribosylation.

Pathogenic Species	Toxin	Components	Cellular Targets	Substrate
C. difficile	CDT	CDTa (Enzyme, 48 kDA) CDTb (Binding, 94 kDa)	Cellular & Muscular G-actin [Modification sites all at Arg 177]	α- and β/γ Actin
C. perfringens E	Iota	Ia (Enzyme, 48 kDa) Ib (Binding, 98 kDa)		α- and β/γ Actin
C. spiroforme	CST	Sa (Enzyme, 44 kDa) Sb (Binding, 92 kDa)		α- and β/γ Actin
C. botulinum C, D	C2 Toxin	C2I (Enzyme, 49 kDa) C2II (Binding, 81 kDa)	Cellular G-actin -	β/γ Actin
B. cereus	VIP	VIP1 (Enzyme, 52 kDa) VIP2 (Binding, 80 kDa)	-	α- (?) and β/γ Actin

CDT is located on a 4.3 kb chromosomal locus, and is composed of two unlinked molecules needed for toxin activity; CDTa and CDTb, of which CDTa is the enzymatic component and CDTb is the binding component (Perelle *et al.*, 1996; Geric *et al.*, 2006). The binary toxin is an ADP (adenosine diphosphate) ribotransferase which affects the actin cytoskeleton of cells and is known to cause fluid accumulation in rabbit ileal loops (Geric *et al.*, 2006). In *C. difficile*, binary toxins are usually produced by variant strains (Table 1.3), and it has been suggested that binary toxin production contributing to the virulence of emerging epidemic strains although the significance of this contribution is widely debated (Barth *et al.*, 2004).

The contribution of CDT to strain virulence is widely debated. Recently CDT has been shown to cause rearrangements in human colonic carcinoma cells, forming long (150 mm) microtubules that wrap and embed clostridia in the gut, thus leading to the theory that CDT assists colonisation by increasing adherence of the pathogen (Schwan *et al.*, 2009). Indeed this phenomenon is not confined to CDT and is shared with the C2 toxin of *C. botulinum* and the Ia toxin of *C. perfringens* suggesting that binary ADP- ribosylating toxins play an important role in pathogenesis (Schwan *et al.*, 2009).

CdtB binds to host cells via specific receptors as "B" monomers which are generated post proteolysis of the "B" precursor molecule, which then form heptamers as the toxin binds to the cell surface/ solution. The heptamer complex then acts as a docking platform that translocates CdtA into the cytosol by acidified endosomes, where it ADP-ribosylates G actin molecules (Barth *et al.*, 2004; Rupnik *et al.*, 2009). ADP ribosylation of G actin occurs at Arginine 177, thus blocking actin polymerisation (Schwan *et al.*, 2009). This results in cytoskeletal disarray, cell death, proteolysis of mitogen activated protein kinases (MAPK) which subsequently inhibit cell signalling; or result in an increase of intracellular levels of cyclic AMP (adenosine monophosphate) leading to oedema and immunosuppression (Barth *et al.*, 2004). The prevalence of binary toxin producing strains of *C. difficile* is 1-6% (Stare *et al.*, 2007). *C. difficile* strain CD630 is CDT negative, but has regions which possess similarity to the binary toxin regions, thus representing a truncated CDT locus.

1.4.8. Genetic plasticity of the *C. difficile* 630 genome

The sequencing of the *C. difficile* 630 chromosome has revealed a highly plastic and mobile genome with relatively high number of genetic elements such as introns, transposons, prophage and insertion sequence (IS) elements (Bruggeman & Gerhard, 2009). Lateral gene transfer of mobile elements such as these brings about sequence reassortments in bacterial genomes. In human bacterial pathogens the majority of genes acquired in this way confer pathogenicity and resistance to antibacterial agents (Woodford & Johnson, 2004).

The existence of ten CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) elements within the genome supports the idea that *C. difficile* has an acquired immunity (Sebaihia *et al.*, 2006; Karginov & Hannon, 2010). The spacer regions in the CRISPR

elements are homologous to bacteriophage and plasmid sequences and are thought to confer resistance to prophage infection (Sorek *et al.*, 2008). Sebaihia and colleagues have described CRISPR sequences as sampling agents, preserving a record of all foreign DNA encountered by the cell. The CRISPR element in *C. difficile* is encoded on mobile elements: two prophage (CD0904-979 and CD2889-2952), and a *skin* element discussed in Section 1.3.4.2. The genes located near the CRISPR regions are known as *cas* genes. These genes are thought to be linked to the CRISPR–mediated acquired immunity through RNA interference systems, recognised as the CRISPR-*cas* system (CASS) (Bruggeman & Gottschalk, 2009). The overall function of CRISPRs is yet to be elucidated. Interestingly, their frequency in the 630 genome is the highest identified in any bacterium to date (Bruggeman & Gottschalk, 2009).

Seven conjugative transposons have been also been identified in strain 630 which carry the majority of antibiotic resistance genes in the genome and exhibit 58-77% similarity to the vancomycin resistance locus in *Enterococcus faecalis* (Sebaihia *et al.*, 2006). One particular transposon *Tn5397* harbours a group II intron capable of entering into the genome at specific sites via self splicing from mRNA (Mullany *et al.*, 1996; Bruggeman & Gerhard, 2009). Sebaihia *et al.* (2006) have also identified several copies of chimeric genetic elements within the *C. difficile* 630 genome called IStrons which comprise of an insertion sequence and a group 1 intron. IStrons merge the high mobility of insertion sequences with the exact splicing of introns for excision by mRNA and are not usually found in bacterial genomes (Bruggeman & Gerhard, 2009). These were first identified in strain C34 by Braun *et al.* (2000) where a copy inserted itself into gene *tcdA*. The IStron, however, did not affect toxin A production as it was excised accurately from the primary transcript, thus allowing toxin expression. This demonstrates the high genomic plasticity within the 630 genome and as further sequencing projects are completed (R20291 and several other strains - http://www.sanger.ac.uk/Projects/C_difficile/) more light will be shed about the *C. difficile* genome.

1.5. MOLECULAR GENOTYPING OF *C. DIFFICILE*

1.5.1. General typing methods

Genotyping of pathogenic bacteria, such as *C. difficile*, enables the source of outbreaks to be identified and aids in understanding the epidemiology of disease. Currently there are a range of techniques used to type *C. difficile.* These include toxinotyping, pulsed field gel electrophoresis (PFGE), PCR ribotyping, multilocus sequence typing (MLST), multilocus variable-number tandem-repeat analysis (MVLA) and restriction endonuclease analysis (REA). Each is described in this section in detail (Table 1.4; Rupnik *et al.,* 1998; Marsh *et al.,* 2010).

Pulsed field gel electrophoresis (PFGE) is based on the digestion of chromosomal DNA by restriction enzymes that cleave and produce HMW fragments (~10Mb) that are separated in agarose gels by a complex form of electrophoresis. The electrical field regularly changes orientation causing the DNA to also alter its direction (Primrose & Twyman, 2007). PFGE has been widely used to type outbreaks of *C. difficile*; however many strains (Serogroup G) are not typable by PFGE due to rapid degradation of DNA during the procedure resulting in undistinguishable smears (Bidet *et al.,* 2000; Alonso *et al.,* 2005).

PCR ribotyping is the most common method for typing of *C. difficile*, and is based on identification of mutations within the 16S-23S rRNA intergenic spacer region (Stubbs *et al.,* 1999). Primers complementary to the 3' end of the 16S rRNA gene and to the 5' end of the 23S rRNA gene are used to amplify the variable length intergenic spacer region (Bidet *et al.,* 2000). PCR ribotyping is faster than PFGE and is able to type strains from a range of serogroups. Indeed, Stubbs *et al.* (1999) have created a library of 116 distinct types of *C. difficile* based on PCR ribotyping.

Multi locus sequence typing (MLST) has been used to type *C. difficile*, but its discriminatory power is the subject of debate. Indeed Killgore *et al.* (2008) completed a review of seven typing methods for *C. difficile* and concluded that MLST was better suited to investigations of population structure (Marsh *et al.,* 2010). MLST genotyping has been developed for investigation of clonal relationships between bacterial populations and for global epidemiological analyses using nucleotide sequences of housekeeping genes (Woodford & Johnson, 2004; Griffiths *et al.,* 2010). Each allele is assigned a unique

sequence type number and laboratories can perform their own MLST analyses and keep control of their data using MLST databases (http://pubmlst.org/). Surprisingly there is only a single report of *C. difficile* MLST which was conducted by Lemee *et al.* (2004), despite its high throughput and portability between research laboratories (Dawson *et al.*, 2009).

Multilocus variable-number tandem repeat analysis (MVLA) is based on PCR amplification and fragment analysis of repeat loci. Amplification is cheap and yields results quickly (Marsh *et al.*, 2006). Indeed MVLA is a highly discriminatory method for *C. difficile* genotyping and useful to distinguish between strains from different outbreaks (Marsh *et al.*, 2010). Restriction endonuclease analysis (REA) is able to distinguish sub-types of *C. difficile* (Dawson *et al.*, 2009). REA is a standard in *C. difficile* genotyping and has been instrumental in the typing of the current B1/NAP1 clone strains (Marsh *et al.*, 2006). The technique however is not readily transportable between laboratories as interpretation of REA banding patterns is subjective. For accurate comparison bacterial strains must be run on the same gel which requires exchange of reference strains between research groups (Zaisz *et al.*, 2009).

Amplified fragment length polymorphism (AFLP) analysis has also been used for typing of *C. difficile*. AFLP involves amplification of multiple genomic restriction fragments by PCR, resulting in high-resolution sub-genomic fingerprints (Klaassen *et al.*, 2002). It has been found that AFLP is more reliable at genotyping *C. difficile* than PFGE (Klaassen *et al.*, 2002). Arbitrarily primed PCR has also been used as a genotyping technique; however the technique does not provide adequate discriminatory power to distinguish genotypes and suffers from lack of reproducibility (Marsh *et al.*, 2006). Surface layer protein (Slp) gene A sequence typing has been used as a reproducible method for genotyping (Kato *et al.*, 2005).

An analysis carried out by Killgore and colleagues (2008) compared seven genotyping techniques; REA, PFGE, PCR-ribotyping, MLST, MVLA and *SlpA* sequence typing, and found there is no single method that is easily transportable between research laboratories which has a high level of discrimination (Killgore *et al.*, 2008).

Table 1.4: Comparison of Molecular Genotyping Techniques

A general comparison of molecular genotyping techniques used for *C. difficile* is shown. The most widely used method of genotyping is PCR ribotyping which is able to recognise strains.

Typing Technique	DNA Target	Advantage of Typing Method	Disadvantage of Typing Method
PFGE	*smaI* Restriction sites	Molecular fingerprinting technique for typing clinical isolates and distinguishes between subtypes	Laborious and time consuming. Inter-laboratory comparisons difficult. DNA degradation results in untypeable results
PCR Ribotyping	16S-23S rRNA Interspacer Region	Gold-Standard, recognises many strains, easily reproducible and portable between laboratories	Limited discriminatory power, doesn't distinguish between sub-types
MLST	Housekeeping genes (7 Loci)	Easily Reproducible	Limited discriminatory power, time consuming, doesn't distinguish between sub-types, limited research conducted in *C. difficile* typing
MVLA	DNA Repeat Units	Can sub-type strains	Can be too discriminatory
REA	*Hind*III Restriction sites	Distinguishes sub-types	Laborious and time consuming. Inter-laboratory comparisons difficult.
AFLP	*PstI* and *MseI* Restriction sites	Based on PCR so DNA degradation is not an issue	Uses Fluorescent labels, requires appropriate analysis equipment
***SlpA* sequence type**	Surface layer protein A gene	PCR amplification is transferable between laboratories. Distinguishes between sub-types	Only uses the sequence of one locus

1.5.2. Toxinotyping of *C. difficile*

The toxinotyping methodology was established after 1980 when variant strains of *C. difficile* emerged (Figure 1.16). Subsequent restriction fragment length polymorphism analysis of the toxin A and B genes in the PaLoc showed variations in the regions encoding toxin A (Bruggeman & Gottschalk, 2009). These variations are the basis for distinguishing strains into toxinotype. Toxinotyping is now considered a highly important method for understanding the toxin production in *C. difficile* isolates, and as such has changed laboratory diagnosis of disease and scientific understanding of the role of toxins in pathogenesis and evolution (Rupnik, 2001; 2008).

In 1985 Delmee *et al* grouped *C. difficile* strains into 30 serotypes, of which 9 were of variant toxinotype. Currently 15 serogroups are known to be variant (Rupnik, 2008).

Further studies conducted using PCR ribotyping methods and REA typing show high correlation to toxinotyping. Strains of a given ribotype tend to belong to a certain toxinotype, however ribotyping remains the more discriminative method as a single toxinotype can often include several ribotypes (Stubbs *et al.*, 1999). There are now 31 known toxinotypes of *C. difficile*, grouped from toxinotype 0 which is the reference strain VPI10463, to XXXI (http://www.mf.uni-mb.si/mikro/tox/) [accessed on 12/2011].

Toxinotyping is conducted via a restriction fragment length polymorphism- polymerase chain reaction – RFLP PCR- method. The 19.6 kb PaLoc region is amplified in order to produce ten PCR fragments that are then compared to reference strain VPI 10463 for length and restriction site polymorphisms (Rupnik *et al.*, 2008). All toxin genes are individually amplified in three fragments (defined by Rupnik *et al.*, 1999) corresponding to the catalytic domain (A1 & B1), putative translocation domain (A2 & B2) and the repetitive domain of each protein (A3 & B3) (Rupnik *et al.*, 1997). However this method has been modified to only amplify the two most variable regions of the PaLoc- regions B1 and A3 with the restriction enzymes *Hinc*II/*Acc*I and *EcoR*I, and the restriction patterns determined. The strains with toxin genes similar to the reference strain VPI 10463 belong to toxinotype 0. Strains with similar toxin changes were classified in to variant toxinotypes from I - XXVII (Rupnik *et al.*, 2008; Rupnik *et al.*, 2009).

Deletions in the PaLoc are usually found in *tcdA* and to date no form of significant deletion in *tcdB* is known (Figure 1.17). This may be due to the larger region of repeat regions within *tcdA* compared to *tcdB*. The presence of such short homologous regions of DNA in *tcdA* may assist in insertion and deletion events in these repeat regions (Rupnik, 2008). Minor toxinotypes are defined as having variations in only one part of the toxin genes, with the sole difference usually being an insertion or deletion in the A3 fragment of *C. difficile*. Major toxinotypes show changes in all three fragments or at least in two fragments of both toxin genes (Rupnik, 2008).

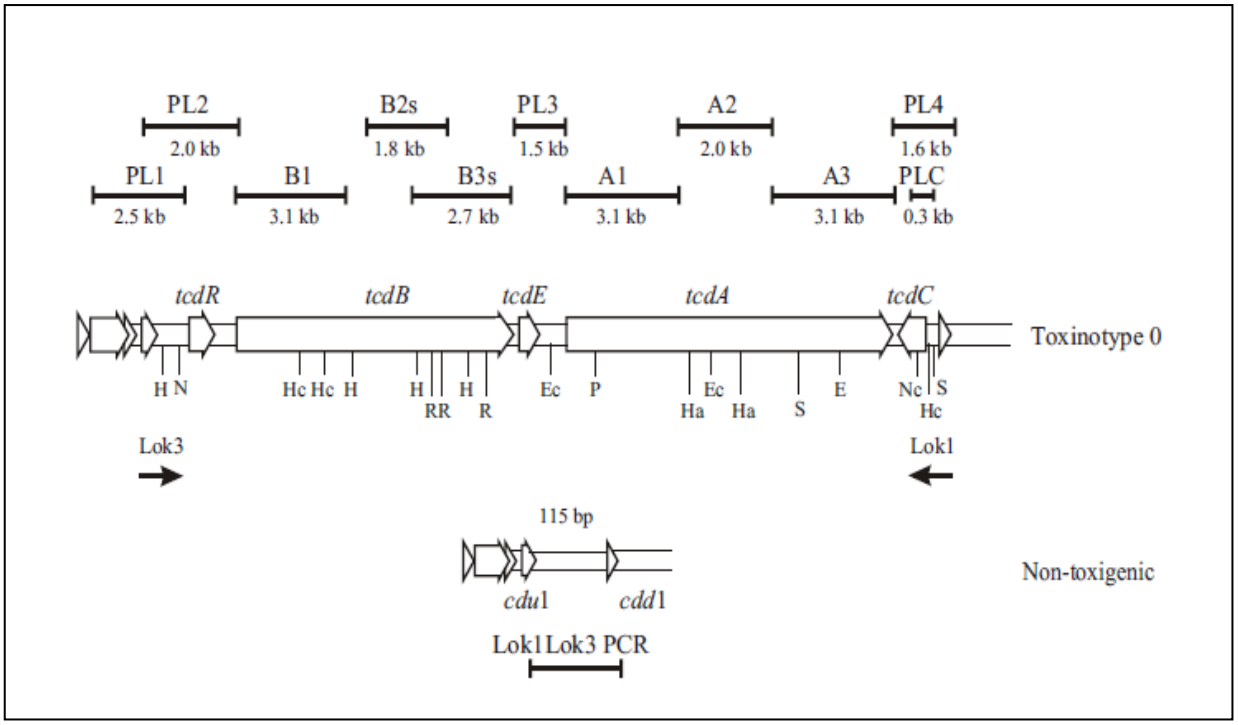


Figure 1.16: Toxinotyping ORFs of *C. difficile*

The restriction fragment areas are shown here in the schematic representation of the 5 open reading frames of *C. difficile.* The toxin A and B genes and accessory genes *tcdR*, *tcdC* and *tcdE* and varying areas where restriction enzymes act upon these genes are depicted. Restriction enzymes used: A – AccI, E – EcoRI, Ec – EcoRV, H –HindIII, Hc – HincII, N – NsiI, Nc – NcoI, P – PstI, R –RsaI, S – SpeI, X – XbaI

Taken from http://www.mf.unimb.si/mikro/tox/images/PaLoc%20PCR%20in%20nontox.pdf

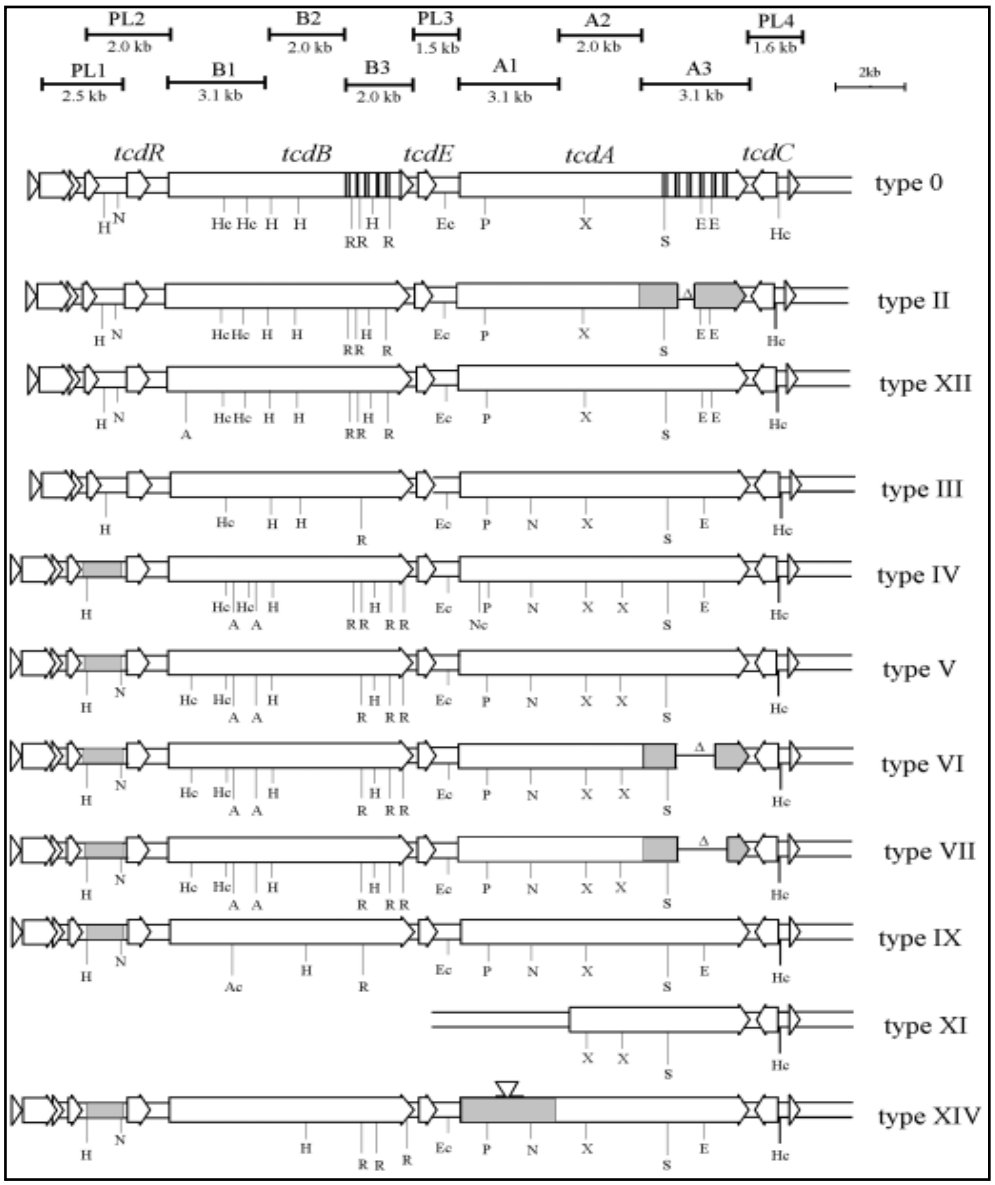


Figure 1.17: Detailed schematic representation of the PaLoc region

In toxinotype 0 and representative variant toxinotypes the presence or the absence and the exact positions of restriction site are shown: A – AccI, E – EcoRI, Ec – EcoRV, H –HindIII, Hc – HincII, N – NsiI, Nc – NcoI, P – PstI, R –RsaI, S – SpeI, X – XbaI. Taken from Rupnik, (2008).

1.6. CLINICAL DIAGNOSIS OF *C. DIFFICILE*

1.6.1. Diagnostic sensitivity

Current detection of *C. difficile* is conducted either invasively in the patient or non-invasively in the laboratory. Usually the effectiveness of such tests is defined by their sensitivity, specificity, and predictive values. The sensitivity in terms of a diagnostic test is the proportion of persons with a disease identified as positive for a disease by the assay (Altman & Bland, 1994). The specificity in terms of a diagnostic test is the proportion of persons without a given condition/disease who are identified by the assay to be negative for the condition; thus the assay is specific (Saah & Hoover, 2007).

The diagnostic specificity of an assay is usually lessened by sample contamination; however the diagnostic sensitivity and specificity of an assay is very different to the analytical sensitivity and specificity. To explain, the analytical sensitivity of an assay represents the smallest amount of a substance (bacteria in this case) that can be accurately measured by an assay. Thus the analytical specificity of an assay is the ability of an assay to measure one particular organism in a sample rather than others (Saah & Hoover, 2007). Thus the diagnostic sensitivity and specificity of an assay defines its operational ability- but cannot be extrapolated to a clinical setting.

Predictive values are used to measure if an assay will give the correct clinical diagnosis (Altman & Bland, 1994). The positive predictive value (PPV) of a test is the probability that a person who has a positive test result truly has the disease. To be precise, this means the PPV is the number of patients with the disease out of every 100 patients with a positive test result who are correctly diagnosed (Altman & Bland, 1994; Saah & Hoover, 2007). The negative predictive value (NPV) is the probability that a person with a negative test result truly does not have the disease; i.e. the NPV is the number of patients without the disease out of every 100 patients with a negative test result who are correctly diagnosed (Altman & Bland, 1994; Saah & Hoover, 2007).

All these statistical definitions are linked: the higher the diagnostic sensitivity of an assay, the better the NPV, and the higher the specificity, the better the PPV (Saah & Hoover, 2007). Accordingly, if a diagnostic assay has perfect sensitivity and specificity (100%) then all persons with positive test results would have the disease, and all persons with negative test results would not have the disease.

1.6.2. False positives and false negatives

When performing diagnostic tests it is possible that false positive and false negative results can be obtained (Table 1.5). False negative results, where the disease has not been detected by the assay can lead to infected patients not being treated for potential CDI and increased transmission risk (Planche & Wilcox, 2010). False positive results can also have consequences. Importantly they can diminish the diagnostic sensitivity of an assay, cause unnecessary introduction of CDAD antibiotic treatment to patients, and in certain cases false positive patients may be cohorted with patients who truly have CDI which in turn can lead to acquisition of genuine infection (Planche & Wilcox, 2010). Thus an effective *C. difficile* diagnostic assay must be able to define true positives and true negatives. This would require a standard reference assay to be developed in CDI diagnosis.

Table 1.5: False Positive and False Negative results

This table lists the possible explanations of false positives and false negative results. It is important to have a diagnostic assay capable of distinguishing true positives and negative cases of CDI (Adapted from Woodford & Johnson, 2004).

Reasons for False – positive results	Reasons for False - negative results
In PCR- carry over contamination from previous amplified products	Inhibition of PCR
Presence of exogenous target DNA in kits	Inadvertent loss of template nucleic acid targets due to poor extraction methods
Poor primer design (Non specificity of primers)	Poor primer design- (non conserved regions at primer sites in variant strains)
Unfavourable amplication conditions	Poor sensitivity of nucleic acid amplification/ detection in assay
Contamination through laboratory staff	Digestion of nucleic acid template with DNAses and RNAses from exogenous sources
	Poor specificity/ sensitivity
	Unfavourable amplification conditions

1.6.3. Diagnostic methods for *C. difficile* detection

The sensitivities and specificities for the tests described below are detailed in Table 1.6.

1.6.3.1. Culture

C. difficile can be detected via its colony morphology and its definitive odour- which has previously been likened to horse manure (Aktories *et al.*, 2000). There is a high NPV of

82-92% of odour assessment for the presence of *C. difficile* infection (Rupnik *et al.*, 2009). On media colonies are typically cream in colour, and as they age they can fade to white and display a fried- egg appearance on agar due to spore formation. Also, when put under ultraviolet light, at ~365 nm in wavelength, the colonies can fluoresce *chartreuse* in colour, however this is not universal across strains. It must be noted that colonies grown on selective agar called cycloserine-cefoxitin fructose agar (CCFA) first described in 1979, will auto-fluoresce due to the presence of a neutral red pH indicator (Aktories *et al.*, 2000).

Culture on CCFA is not very specific as non-toxinogenic isolates can grow, but is very sensitive for *C. difficile* growth, and requires between 40-48 hours incubation (Delmee *et al.*, 2005). Culture then requires additional testing to determine if the strain is toxigenic. Also ethanol shock has been shown to enhance the isolation of *C. difficile* pre-inoculation. Another alternative of CCFA is CDMN agar: *Clostridium difficile* moxalactam norfloxacin agar, found to be 20% more effective at isolating *C. difficile* (Oxoid) than CCFA (Aspinall *et al.*, 2002). CDMN contains 32mg/l moxalactam, 16 mg/l norfloxacin and 0.5 g/l cysteine hydrochloride and has been found to inhibit growth of faecal streptococci, other *Clostridium* species and *Bacteroides*, unlike CCFA (Aspinall *et al.*, 2002; Jobstl *et al.*, 2009).

Stool culture, however, is labour intensive with results not available until 76-92 hours, thus few laboratories routinely perform stool cultures. Also stool culture does not differentiate between toxigenic and non-toxigenic strains; therefore stool culture diagnostic tests are usually conducted in conjunction with toxigenic tests in a two step algorithm (Goldenberg *et al.*, 2009). This improves sensitivity to 90% and specificity to 98% (Poutanen & Simor, 2004). Stool culture does however provide an advantage when strain typing during an outbreak (Poutanen & Simor, 2004). Toxigenic culture is the selective culture of *C. difficile* followed by confirmation of toxin production (Goldenberg & French, 2011). This assay measures toxin production *in vitro*, which is non-comparable to *in vivo* toxin production, leading to lower specificity. This assay takes 4-5 days to confirm the presence of *C. difficile* (Planche & Wilcox, 2011).

1.6.3.2. Cytotoxicity assay

There is currently no universal standard test for *C. difficile*, but the cell cytotoxicity assay (CCTA) is considered to be the reference "gold standard" (Rupnik *et al.*, 2009). A cytotoxic assay is conducted by inoculating a faecal filtrate sample onto cultured cell monolayers, which should display a cytopathic effect that is neutralised by a specific cross reacting *C. sordellii* antitoxin if toxin B is present (Delmee *et al.,* 2005). Cultures are examined microscopically at between 24-48 hours and either Vero or Hep2 cell lines are used (Planche & Wilcox, 2011). CCTA is specific and sensitive, taking between 48-72 hours for results to be obtained, specifically 24 hours first for cytotoxicity demonstration, and 24 hours secondly for neutralisation of toxicity (Poutanen *et al.,* 2004; Goldenberg *et al.,* 2009). The diagnostic sensitivity is relatively high; 94- 100% and specificity is 99% (Snell *et al.,* 2004). The CCTA however is labour- intensive and expensive as tissue culture facilities are required and results are known to vary according to the type of cell line supplied and reagents used and only detects toxin B (Poutanen *et al.,* 2004). Thus it has been suggested that a two step algorithm of stool culture with CCTA can increase sensitivity of CDI detection (Arnold *et al.,* 2011).

1.6.3.3. Enzyme Immuno-Assays

Enzyme immunoassay (EIA) kits detect either toxins A and B, or only toxin A. EIAs are widely available from companies, are easy to perform and are rapid, yielding results within 20 minutes to an hour (Tenover *et al.*, 2011). This method is less sensitive than the CCTA; even so studies evaluating commercial kits have found most tests to be reliable when compared with CCTA (Belanger *et al.,* 2003; Planche *et al.*, 2008; Tenover *et al.*, 2010). Previous studies have found combined use of EIA and CCTA gives a PPV of 99.5 % leading to the proposal of a two-step algorithm (She *et al.,* (2009). More recent studies have found that EIA kits can deliver false positive results and have sensitivities of approximately 60% (Tenover *et al.,* 2010). Indeed some EIAs tested by Planche (2008) were found to have overestimated cases of *C. difficile* by 20%. The findings and concerns from Planche *et al* (2008) were echoed by Eltringham (2009) finding only a 10% PPV. Thus EIAs should not be used as standalone tests for CDI.

1.6.3.4. Antigen Enzyme Immunoassays: Glutamate dehydrogenase

Glutamate dehydrogenase (GDH) is a common antigen produced by all *C. difficile* isolates (Lyerly et al., 1991; Carmen *et al.,* 2012). Early commercial GDH tests cross reacted with other species, especially *C. botulinum*, *C. sporogenes*, and *Peptostreptococcus anaerobius* (Williams & Spencer, 2009). Current commercial tests detect GDH within 15 minutes, and cost approximately £6 per test (Inverness Medical, 2009). Although the tests are rapid, easy to perform and sensitive, they are not specific for diagnosis of toxinogenic strains, thus detection algorithms are still used (Snell *et al.,* 2004; Gilligan, 2008; Eastwood *et al.,* 2009).

It has been widely suggested that GDH assays should be used in conjunction with CCTAs which yields better PPV and NPV (Ticehurst *et al.,* 2006; Fenner *et al.,* 2008). Commercial assays, such as the CDIFF Quik Chek Complete (TechLab) have GDH detection alongside toxin detection in a lateral flow type of assay (Quinn *et al.,* 2010). Swindells *et al.* (2010) found that this particular commercial assay was highly sensitive (100%) and could be relied upon for exclusion of patient carriage of *C. difficile* as the NPV was 100% when compared to both CCTA and stool culture.

Conversely, more recent studies which tested the GDH, EIAs and/ or CCTA combinations and/or CCTA, found they missed 23-24% of positive specimens, and some even had reduced sensitivity (83.1%) (Sloan *et al.,* 2008; Novak-Weekley *et al.,* 2010). Real time PCR and PCR have also been used in these algorithms, invariably having high sensitivity and high specificity compared to all other diagnostic methods (Doing *et al.,* 2010). An issue with these two/three step algorithms is the turnaround time and cost as the combined use of CCTA and GDH assays, or RT-PCR and GDH assays could be costly in terms of diagnosis. An ideal assay would also have a rapid turnaround time. Even so, these disadvantages can be offset by reducing unnecessary antibiotic usage and infection control measures in the hospital and streamlining of isolation facilities within a hospital (Williams & Spencer, 2009). Interestingly it has recently been observed that the sensitivity of the GDH assays may vary according to PCR ribotype, and that GDH is less sensitive than PCR in detection of ribotypes 002, 007 and 106 (Tenover *et al.,* 2010). Hence there is some debate regarding the effectiveness of GDH assays; however despite these concerns GDH detection does offer a high NPV and may in future partly replace culture methods if used in effective algorithms (Fenner *et al.,* 2008).

1.6.3.5. Faecal Lactoferrin

Lactoferrin is a glycoprotein secreted by mucosal membranes and is a component of polymorphonuclear leucocytes (Kane *et al.,* 2003). Lactoferrin is highly expressed in patients with severe bowel inflammation and also those with *C. difficile* infection (Wren *et al.,* 2009). A test for lactoferrin could be used in conjunction with other tests as an algorithm, as with the GDH assays. Indeed Wren *et al.* (2009) examined the potential of faecal lactoferrin tests for use in *C. difficile* diagnosis, as the correlation between lactoferrin production and CDAD is highly significant (Williams & Spencer, 2009). It was established that faecal lactoferrin is a useful marker for indication of the presence of inflammatory diarrhoea (Wren *et al.,* 2009).

1.6.3.6. Polymerase Chain Reaction

PCR methods for the detection of *C. difficile* toxin A or B or both have been developed with high specificity and sensitivity, but there is no standardisation of the PCR methods and no commercial assays are available (Poutanen *et al.,* 2004). PCR for toxin B has a minimum detection limit of 1 x 10^5 *C. difficile* per gram of stool, with a sensitivity of 95% when compared to CCTA (Goldenberg *et al.,* 2009). The turnaround time though, is rapid within 4 hours, but when compared to stool culture sensitivity is reduced to 83.6% with a specificity of 98.2%. PCR methods offer greater specificity and sensitivity, although their cost is high, but in the future it is possible that detection of toxin B may be reduced due to mutations within the PaLoc (Rupnik *et al.,* 2009).

1.6.3.7. Real Time Polymerase Chain Reaction

Real Time PCR is a very sensitive and specific, targeting specific genes that are associated with toxin production (Knetsch *et al.,* 2011). RT PCR requires labour intensive purification of the sample before PCR can be conducted resulting in the total time for sample preparation and detection to be in excess of 8 hours (Rinttila *et al.,* 2004). The method itself is more powerful than conventional PCR and the *C. difficile* genes detected include *tcdA, tcdB, tcdC* regulator and *gluD* and 16S genes (Rinttila *et al.,* 2004; Van den Berg *et al.,* 2005). RT PCR has high concordance with toxin culture, and thus is widely used to detect toxigenic *C. difficile* strains directly from faeces (Belanger *et al.,* 2003; Van den Berg *et al.,* 2007).

The commercial assays all target *tcdB* and include the ProGastro assays for use with the Cepheid Smart Cycler, the similar GeneXpert ™ (Cepheid), and the BD Gene Ohm assay (Doing *et al.,* 2010). The Xpert and BD Gene Ohm assays have been directly compared to stool culture and CCTA, both giving 100% sensitivity but individually for stool culture the Xpert and BD Gene Ohm is 97% specific (Table 1.7; Swindells *et al.,* 2010). These new RT PCR assays have a relatively rapid turnaround time and when used in conjunction with other assays, such as GDH, have high sensitivity (Larson *et al.,* 2004). The commercial assays developed to address the issue of turnaround time have led to improved, faster and more sensitive assays despite the cost. Even so, a 2 hour assay turnaround is still relatively long for a diagnosis to be made.

1.6.3.8. Diagnostic Imaging Methods

Flexible sigmoidoscopy is used to detect PMC in patients, and is superior to the cytotoxin tests (Johal *et al.,* 2004). In rare cases where stool cytotoxin tests may have produced false negative results, invasive flexible sigmoidoscopies are used to examine the sigmoid region of the colon (Johal *et al.,* 2004). The process takes between 20-30 minutes and causes the patient slight pain in the bowel due to the presence of air.

Areas which escape detection via sigmoidoscopies are detected via colonoscopies (Bartlett & Gerding, 2008). PMC is often not present using this visualisation technique and as such colonoscopies are deemed insensitive. There are also risks of perforation occurring in severe cases of colitis. Computed tomography scans are also used for severe cases of CDAD, and characteristic features which can be seen in the scans include inflammation of the mucosa which can be easily identified by clinicians (Bartlett & Gerding, 2008).

Detection Target	Diagnostic Specificity	Diagnostic Sensitivity	Time to perform assay	Advantages	Disadvantages	References
Toxin B	99-100%	85-100%	48-72 hours	Gold Standard Cheap; Highly sensitive and specific	Labour intensive; Requires tissue culture facilities; Requires skilled technicians; long turnaround time	Goldenberg *et al.*, 2009 Poutanen *et al.*, 2004 Snell *et al.*, 2004 Turgeon *et al.*, 2003;
Strain	>98%	>90%	48-96 hours	Cheap; detects non-toxinogenic strains allowing for molecular typing and storage of strain; next standard method after CTN	Labour intensive; detects non-toxinogenic strains	Goldenberg *et al.*, 2009 Zheng *et al.*, 2004; Poutanen *et al.*, 2004
Toxin A only, or toxin A and Toxin B	90-96%	75-99%	15 minutes- 1hour	Rapid turnaround time; ; generally cheap when compared to PCR; easy to use	Expensive when compared to CTN; cross reactivity with other species; sensitivities and specificities widely debated amongst groups, and vary according to manufacturer	She *et al.*, 2009 Williams & Spencer, 2009 Planche *et al.*, 2008 Snell *et al.*, 2004 Poutanen *et al.*, 2004
Glutamate dehydrogenase	84%	92%	20 minutes	Rapid turnaround time; cheap to acquire; easy to use	sensitivities and specificities widely debated amongst groups, and vary according to manufacturer	Novak-Weekley *et al.*, 2010 Doing *et al.*, 2010 Swindells *et al.*, 2010 Gilligan, 2008 Fenner *et al.*, 2008 Sloan *et al.*, 2008
Genes: *tcdA*, *tcdB*, Triose phosphate isomerise, Glutamate dehydrogenase	100%	97%	> 2 -8 hours	Newer commercial assays are very rapid; highly sensitive and specific; relatively easy to use	Some are Labour intensive; Expensive; Newer commercial assays only detect toxin B	Larson *et al*, 2010 Novak-Weekley *et al.*, 2010, Swindells *et al.*, 2010 Peterson *et al.*, 2007; Belanger *et al.*, 2003
Genes: *tcdA*, *tcdB*, 16S, *gluD*	100%	92-97%	> 4 hours	Highly sensitive and specific	Labour intensive; Expensive	Goldenberg *et al.*, 2009 Sloan *et al.*, 2008 Zheng *et al.*, 2004; Poutanen *et al.*, 2004

Table 1.6: Comparison of current *C. difficile* detection methods

The sensitivities and specificities (as described in Section 1.6.1) for each detection method are compared. The times to perform the assay and the advantages of each assay are listed.

1.6.3.9. Summary of *C. difficile* diagnostics

The gold standard for *C. difficile* diagnosis is the cell cytotoxin neutralisation assay against which all other *C. difficile* diagnostics are measured. There is considerable debate as to the effectiveness of the various methods currently used for *C. difficile* diagnosis, and although varying algorithms have been implemented across some laboratories, there is a pressing need for the diagnostic methodology to be standardised to enable accurate diagnosis and to enable comparisons between laboratories. The widely differing sensitivities and specificities of the various assays is a cause for concern as the literature and relevant statistics regarding the diagnostics is often confusing, conflicting and limited. This is not to say, however, that assays such as RT PCR are not effective- the problem is their long turnaround time. Current assays target toxin B of *C. difficile* (RT PCR and CCTA), and EIAs can target toxin B or both toxins A and B; however these tests have been suggested to be used in two/three step algorithms to enable accurate diagnosis to be made and avoid false negative or positive results. Indeed the limitations of current diagnostic assays for *C. difficile* detection have a considerable impact upon treatment, as early treatment of the disease may reduce the spread of the organism and reduce patient mortality. Thus there is a pressing need to develop an assay which rapidly detects toxigenic *C. difficile* with ease and has high specificity, sensitivity and predictive values.

1.7. TECHNOLOGY FOR PATHOGEN DETECTION

1.7.1. Microwave-accelerated metal - enhanced fluorescence

Detection of human pathogens using DNA based assays is often labour intensive and costly due to the need for the inclusion of a purification step to remove inhibitory biological materials (Section 1.6) (Belanger *et al.,* 2003; Rintilla *et al.,* 2004; Eastwood *et al.,* 2009; Knetsch *et al.,* 2011). A new platform technology called microwave accelerated metal enhanced fluorescence (MAMEF) has recently been developed which eliminates the need for extensive sample purification and has been used successfully to detect pathogenic spore formers such as *Bacillus anthracis* ((Geddes & Lakowicz, 2002; Aslan *et al.,* 2007; 2008). The approach combines two technologies, metal enhanced fluorescence (MEF) and the low power microwave heating of metal, to yield a system with a level of sensitivity comparable to that of RT PCR, but with a much shorter turnaround time of 60 seconds. There are two key processes involved in MAMEF: firstly DNA extraction from the microorganism and secondly the MAMEF itself. To release DNA the microorganism is

lysed using low power microwaves in a cavity. The cavity itself comprises a glass slide with aluminium triangle structures adhered to the surface and silicone isolators use to hold the aqueous solution of microorganisms. The metal used in this activity can be changed and more recently a gold "bow tie" triangle has been used for microwave lysis as the power generated is higher than aluminium (Figure 1.18; Geddes & Lakowicz, 2002; Zhang *et al.,* 2011). Indeed this method is capable of extracting DNA from whole spores of *B. anthracis* and vegetative cells of a range of pathogens (Aslan *et al.,* 2007; 2008; Tennant *et al.*, 2011; Zhang *et al.,* 2011).

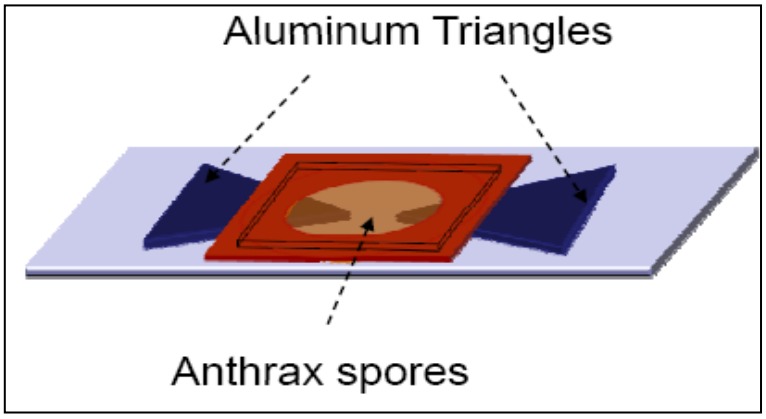


Figure 1.18: Aluminium "Bow tie" Structure used in MAMEF

The aluminium triangles are used to extract DNA from the target microorganism. To release DNA the microorganism is lysed using low power microwaves in a cavity. The cavity itself comprises a glass slide with aluminium triangle structures adhered to the surface and silicone isolators use to hold the aqueous solution of microorganisms. Image taken from Aslan *et al.* (2008).

The MAMEF bioassay comprises a silvered glass slide (Silver island film- SiF) which contains silver nanoparticles, biomolecules and aqueous media. The silver nanoparticles act as a platform for the anchorage of biomolecules, such as DNA probes (specific to the microorganism), which is facilitated through the addition of stable thiol groups to the DNA (Aslan *et al.,* 2008). Once target DNA from the microorganism adheres to the anchored DNA a further reporter biomolecule fluoresces and this fluorescence is enhanced by the silver nanoparticles (Geddes & Lakowicz, 2002; Geddes *et al.,* 2003; Aslan *et al.,* 2007; 2008; 2010; Tennant *et al.*, 2011; Zhang *et al.*, 2011).

Microwave power is used to increase fluorescence (hence metal enhanced fluorescence) as the excited fluorophores transfer their energy to the silver nanoparticles and are subsequently thought to become more photostable. The energy is transferred via surface plasmons which are fluctuations in electron density at the surfaces of metal, which become excited when they react with a photon (Aslan & Geddes, 2008). For MEF silver is currently used as the metal, but in the rationale for MEF can apply to any metal that supports plasmon resonance (Aslan & Geddes, 2008). Silver is employed firstly as platform for bimolecular recognition events to occur, secondly as an amplifier of fluorescent signatures, and thirdly a low temperature surface (during microwave heating) to create a temperature gradient. As the metals are not heated by the microwaves, they remain at a lower temperature than the water, thus causing mass transport of warmer biomolecules to the colder silver nanoparticles on the surface (Aslan *et al.*, 2005).

In the case of the *B. anthracis* DNA assay focussed microwave heating with pulses of 20 seconds was found to be sufficient to extract DNA from vegetative organisms and whole spores (Aslan *et al.*, 2008). Indeed the focussed microwaves were capable of inducing structural changes in the spores which was visualised via transmission electron microscopy (TEM) (Figure 1.19; Aslan *et al.*, 2010). Specifically for the MAMEF component of the assay a single stranded nucleotide anchor probe (Figure 1.20), with a nucleotide sequence specific to *B. anthracis* DNA, was attached to the silvered film and target *B. anthracis* oligonucleotide and a fluorophore labelled oligonucleotide were incubated alongside an anchor DNA probe and subsequently low power microwave heating for 30 seconds was employed for MEF (Aslan *et al*., 2010). Thus the total time for *B. anthracis* detection was < 60 seconds.

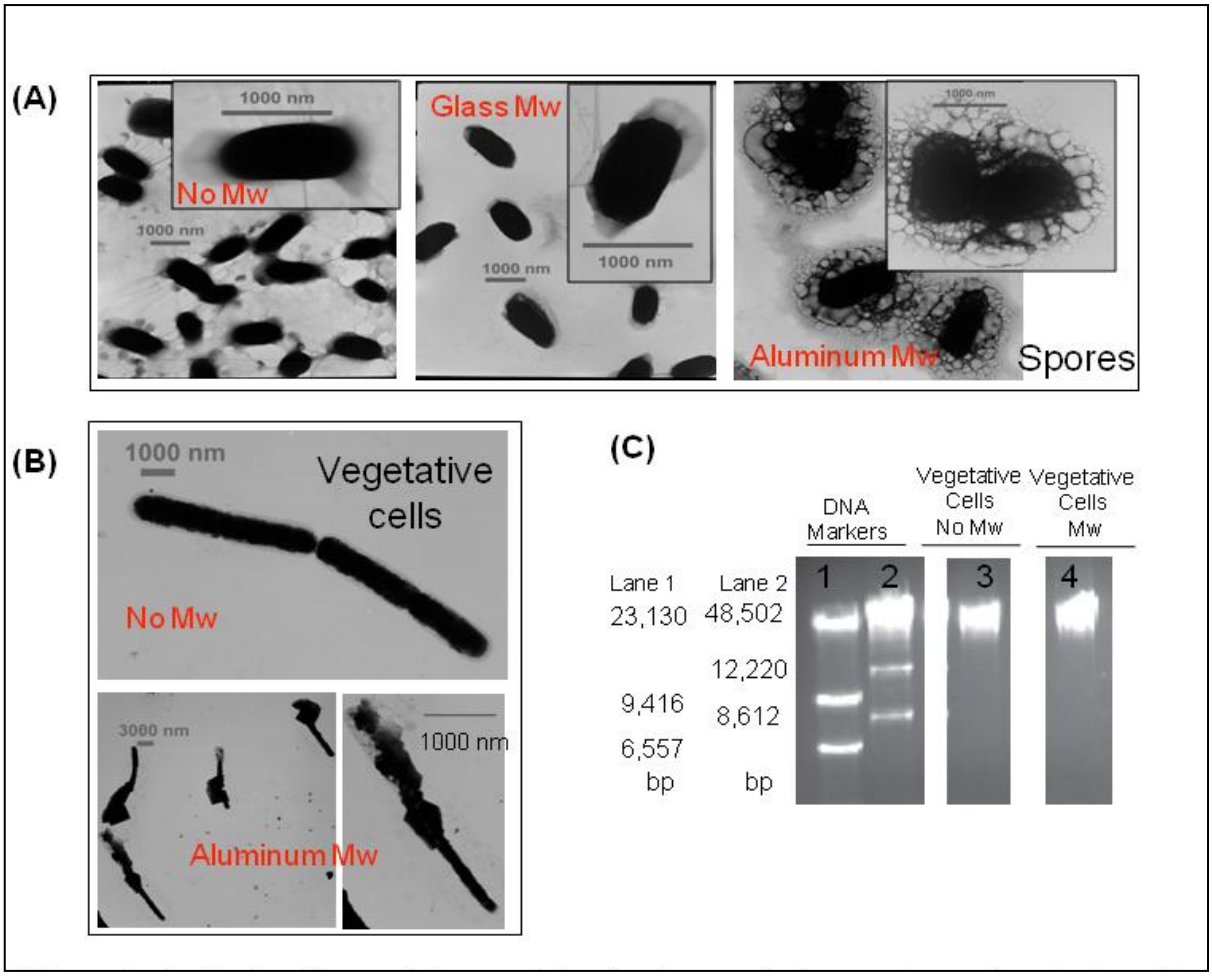


Figure 1.19: Extraction of *B. anthracis* DNA

(a) Transmission electron microscopy (TEM) images of *B. anthracis* spores before and after low power microwave heating. (b) TEM images of vegetative organisms both before and after 20 low power microwave exposure between aluminium triangles. (c) Gel electrophoresis study of DNA extracted from vegetative cells. The markers are shown in lanes 1 and 2, and in lane 3 DNA can be seen from unmicrowaved vegetative cells, and in lane 4 DNA can be seen from microwaved spores. (Taken directly from Aslan *et al.*, 2010).

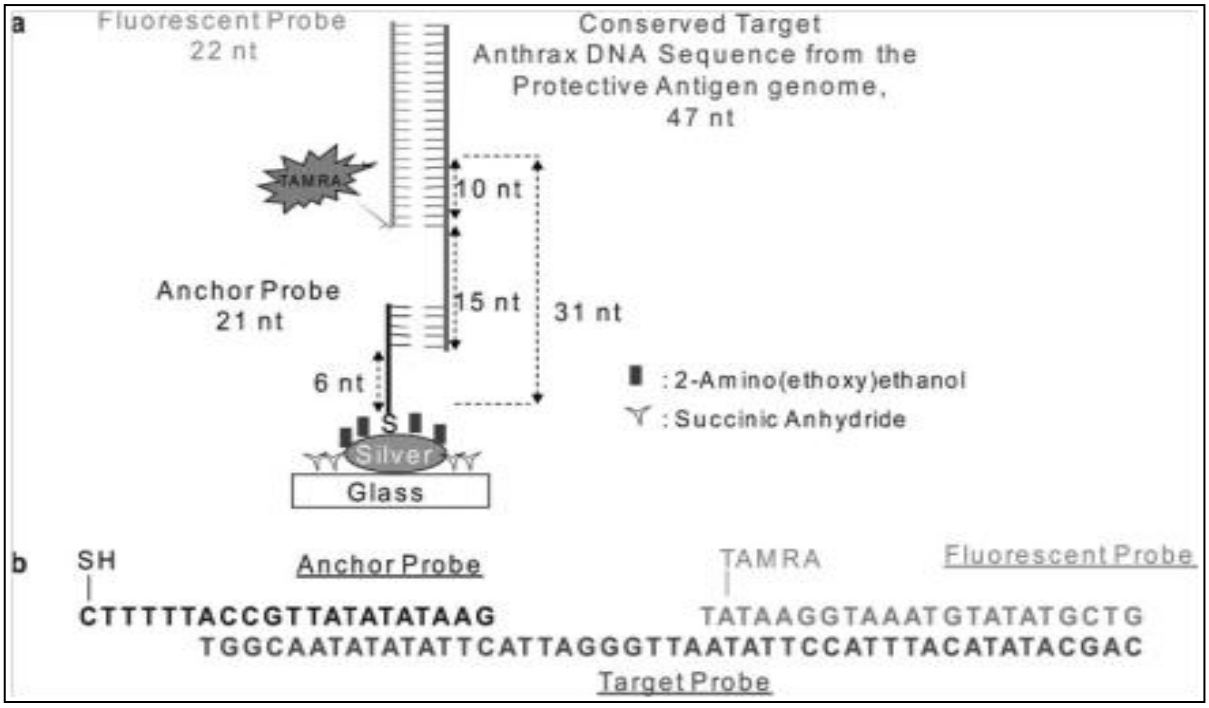


Figure 1.20: Probe design for *B. anthracis* assay

This image shows the configuration for the DNA based hybridisation MAMEF assay designed specifically for *B. anthracis*. The anchor probe is adhered to silvered glass, and when hybridisation occurs the fluorescent probes fluoresces. Image taken directly from Aslan *et al.*, (2007).

1.8. AIMS AND OBJECTIVES OF STUDY

1.8.1. Hypothesis

- We hypothesise that the structure and physical properties of the spore form of *C. difficile* contributes to the virulence of the organism and to the ability to release DNA following microwave treatment.

- Based on bioinformatic analysis of the whole genome of *C. difficile* we hypothesise that the pathogen may have evolved from an ancestor which was an insect pathogen.

- We hypothesise that the toxin genes within *C. difficile* may have been acquired through horizontal gene transfer over its evolution from other bacteria.

- We hypothesise that the toxin genes may also have been a relic of an insect associated toxin ancestor gene, and indeed genes within the whole genome may also have some insect associations.

- Our final hypothesis is that the nucleotide sequences of toxin A and toxin B genes contain conserved regions encoding structures essential to the biological function of each toxin. The conserved nature of these sequences make them ideal target for detection probes.

1.8.2. Aims

The main aim of this thesis was to design an assay based on conserved highly specific regions of toxin A and B capable of detecting the presence of toxigenic *C. difficile* in a faecal samples within 60 seconds. The clinical application of such a detector would be to rapidly diagnose potential *C. difficile* infection within patients, thus allowing clinicians to determine the best course of treatment for the patient, preventing cross infection and further development of PMC.

The specific aims of this PhD are:

1. To characteristic the structure of *C. difficile* spores with a view to understanding its contribution to adherence and pathogenicity

2. To employ a bioinformatic based approach to investigate the evolutionary origins.

3. Determine the ability of *C. difficile* to infect and colonise invertebrates with a view to:

i) Understanding the biological history of the pathogen. Did it originally evolve as an insect pathogen?

ii) Determine the feasibility of establishing a non-mammalian infection and colonisation model.

iii) To provide a model system with which to assess the efficacy of a *C. difficile* detection assay.

4. To identify conserved genetic signatures within toxins A and B of *C. difficile* using a bioinformatics based approach which could form the basis of a specific diagnostic assay.

5. To determine the specificity of conserved genetic signatures against a range of *C. difficile* isolates, related species and unrelated species for specificity.

6. Incorporate specific toxin A and B probes within the MAMEF technological platform and determine its ability to detect the presence of bacterium in a range of organic matrices including faeces. The sensitivity of the assay will also be determined.

CHAPTER 2

CONTRIBUTION OF THE SPORE TO THE VIRULENCE OF *CLOSTRIDIUM DIFFICILE*

2.1. ABSTRACT

C. difficile, a Gram positive spore-forming bacterium, is the commonest cause of hospital acquired infection in the United Kingdom. Its prevalence is partly due to its ability to produce resistant spores, enabling the organism to colonise healthcare environments and infect susceptible individuals. In preliminary studies the composition of culture media had a marked effect on the ability of individual clinical *C. difficile* isolates to form spores and germinate.

To determine if spore structure contributed to surface binding we characterised the surface properties of spores produced by 21 clinically relevant isolates with regards to their hydrophobicity and ability to adhere to hospital grade stainless steel. Overall, isolates of hypervirulent ribotype 027 and UK epidemic ribotypes 001, 078 and 106 appeared to be more hydrophobic than isolates occurring less frequently in the UK. We also found that spores of highly hydrophobic isolate DS1813 (ribotype 027) adhered more strongly to human gut epithelial cell lines Caco-2 and HT-29 than spores of lowest hydrophobic isolate DS1748 (ribotype 002).

Examination of the structure of spores from a range of clinical isolates by electron microscopy revealed differences: most notably the presence of a pronounced exosporium layer surrounding hydrophobic DS1813 spores. This outer layer is thought to contribute to hydrophobicity and spore attachment, and appeared to be absent from spores of other strains examined. Based on these results we hypothesise that differences in spore surface properties could contribute to the spread and prevalence of virulent strains of *C. difficile*.

2.2. INTRODUCTION

C. difficile is an anaerobic, spore-forming Gram positive bacterium implicated as the primary cause of antibiotic associated diarrhoea in the UK. Between 2004 - 2008 *C. difficile* was involved in 10.5 per 1000 deaths in England and Wales, compared to 3 per 1000 deaths for Methicillin Resistant *Staphylococcus aureus* (Office of National Statistics, 2010). Alterations in colonic microbiota, usually due to broad spectrum antibiotic treatment, is thought to increase sensitivity to *C. difficile* infection enabling the vegetative organism to produce potent cytotoxins which destroy intestinal epithelial cells, increase vascular permeability and ultimately leads to the formation of pseudomembranes composed of neutrophils, fibrin, mucosa and cellular debris as described in Section 1.2 (Poutanen & Simor, 2004; Rupnik *et al.,* 2009). The subsequent bloody diarrhoea results in the release of large numbers of spores into the environment.

The formation of spores occurs when the vegetative organism is subjected to some type of stress, usually nutritional limitation, and allows the organism to survive in a dormant state until conditions improve- a situation that can be anything from a couple of months to several years (Setlow, 2003). Indeed it is this ability to form spores which enables the bacterium to survive outside of the anaerobic environment of the gut and spread from person to person within healthcare facilities. In addition to protecting the organism from the toxicity of oxygen and a shortage of nutrients the spore also confers increased resistance to general-purpose disinfectants; a property which is thought to contribute to the transmission of the organism as a consequence of ineffectual disinfection (Cardiff & Vale NHS, 2007; Rupnik *et al.*, 2009).

It has been suggested that sub-inhibitory levels of non-chlorine based hospital cleaning agents may exacerbate the problem of transmission by stimulating epidemic strains of *C. difficile* to convert to chemically resistant spores and thereby avoid the attentions of the biocide (Ronner *et al.,* 1990; Fawley *et al.,* 2007). As well as conferring resistance to chemical insult, the spore also contributes to transmission by mediating attachment to surfaces commonly found in the healthcare environment, such as bed linen and stainless steel (Hellickson & Owens, 2007). This ability to attach to surfaces further increases biocide resistance, particularly in the presence of organic matter such as faeces, and as such presents a considerable challenge for those seeking to break the infection cycle.

The attachment of a spore to a specific surface is primarily influenced by the hydrophobic properties of the surface, the spore, and to a lesser degree by spore surface structures such as appendages and an exosporium (Doyle & Rosenburg, 1990; Husmark & Ronner, 1992; Klavenes *et al.*, 2002; Setlow, 2003). The exosporium is a loose outer layer which surrounds spores produced by some, but not all, members of the *Bacillus* and *Clostridium* spp. including *C. difficile* (Henriques & Moran, 2007). The spore architecture of *Bacillus* and *Clostridium* spp. is highly similar as shown in Figure 2.1 and Section 1.3.4. Whilst the morphology of spores has been well understood, its protein composition and structures have yet to be fully elucidated (Husmark & Ronner, 2002).

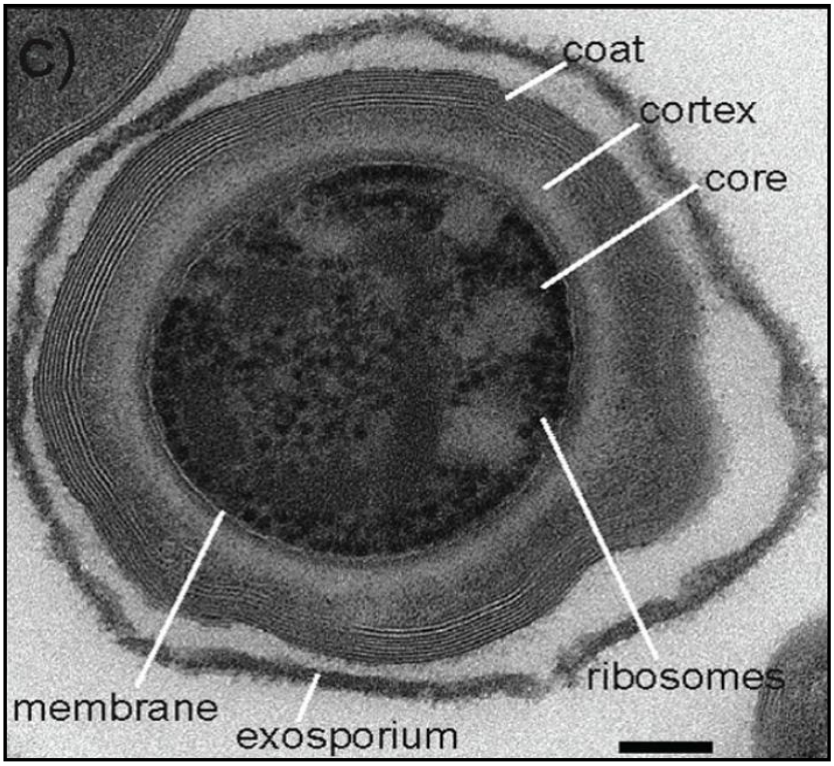


Figure 2.1: *Clostridium difficile* spore of strain 630

This transmission electron microscopy images clearly shows the layers of the spore in detail. The spore comprises of a core which contains ribosomes. Surrounding the electron dense core is a membrane further enveloped by a cortex layer which in itself is formed of concentric layers (Driks, 2004). These areas function to protect the spore and keep it dry (Driks, 2009). The spore is also surrounded by an exosporium layer. This image was taken from Lawley *et al.* (2009).

It has been suggested that the hydrophobicity of the exosporium plays a key role in attachment to environmental and cellular surfaces (Koshikawa *et al.,* 1989). The hydrophobicity of the exosporium of enterotoxigenic isolates of *Bacillus cereus* enables the spore form of this organism to adhere to environmental surfaces and resist cleaning and rinsing procedures (Faille *et al.,* 2002). In addition a correlation has been demonstrated between spore hydrophobicity and the ability of pathogenic isolates of *B. cereus* to adhere to human colon carcinoma cells, germinate and express their exotoxins (Andersson *et al.,* 1998).

Following attachment to the intestine the spores of *C. difficile* must also germinate and express toxins if they are to initiate an infection. The mechanisms by which spores of *C. difficile* determine that conditions have improved sufficiently to support growth have yet to be fully elucidated. In the case of other spore forming organisms such as *Bacillus* spp, specific receptors within the spore have been identified which respond to LMW germinants such as amino acids, nucleosides, sugars, or salts and initiate germination (Moir, 2006). Somewhat surprisingly, given that the proteins involved in germination are remarkably conserved across the *Bacilli* and *Clostridia*, bioinformatic analysis of the genome sequence of *C. difficile* strain 630 has failed to reveal the presence of homologs to known germinant receptor genes (Sebaihia *et al.,* 2006; Lawley *et al.,* 2009). While specific receptors have not been identified studies have shown that *in vitro C. difficile* spore germination is triggered by the amino acids glycine, alanine, cysteine and phenylalanine as well as the bile salt taurocholate. The bile salt chenodeoxycholate inhibits taurocholate-induced *C. difficile* spore germination (Giel *et al.,* 2010).

It has recently been hypothesized that *C. difficile* spores require the presence of taurocholate and glycine to trigger germination in the gut. Thus spores remain dormant in the lower intestine while the available taurocholate is deconjugated to cholate and taurine by members of the commensal microbiota. Treatment with broad spectrum antibiotics subsequently alters the composition of the gut microbiota such that organisms which deconjugate taurocholate are inhibited allowing the local concentration of taurocholate to increase and the concentration of inhibitory chenodeoxycholate to decrease. As a consequence *C. difficile* spores bind both taurocholate and glycine and thus trigger germination (Rameirez *et al.,* 2010).

2.3. RESEARCH AIMS

The aims of this chapter are to:

- Determine if media composition affects the ability of *C. difficile* to form spores and germinate
- Understand if spore surface structure plays a role in virulence
- Establish if spores are able to adhere to a range of surfaces, including hospital grade stainless steel and human adenocarcinoma colonic cell lines
- Examine any potential differences in the surface structure of spores of varying strains using electron microscopy

2.4. MATERIALS & METHODS

2.4.1. Strains, growth conditions and media composition

The clinical isolates of *C. difficile* used in this study are shown in Table 2.1 and were obtained from the National Anaerobic Reference Unit, Cardiff Wales, except strain CD630 which was obtained from the National Collection of Type Cultures (NCTC), Health Protection Agency, UK. *Bacillus atrophaeus* ATCC 9372 was purchased from American Type Culture Collection, (ATCC, Rockville, MD, USA).

Table 2.1: Strains of *Clostridium difficile* used in this study

The bacterial species and strains listed above were used in this study. These strains are clinically relevant isolates obtained from the National Anaerobic Reference Unit, University Hospital Wales, Cardiff, courtesy of Dr. J. S. Brazier and Dr. Val Hall. The strain designation, isolate source and the PCR ribotype are listed.

Clostridium difficile Strains	PCR Ribotype	Source	Toxin Production
DS1759	001	Maidstone	$tcdA^+ tcdB^+$
DS1747	001	St. James, Leeds	$tcdA^+ tcdB^+$
R8652	001	NCTC	$tcdA^+ tcdB^+$
DS1750	001	St. James, Leeds	$tcdA^+ tcdB^+$
DS1813	027	Hinchingbrooke	$tcdA^+ tcdB^+$
DS1801	027	Leicester	$tcdA^+ tcdB^+$
R20291	027	Stoke-Mandeville	$tcdA^+ tcdB^+$
DS1807	027	Salford	$tcdA^+ tcdB^+$
R10459	106	Dudley	$tcdA^+ tcdB^+$
DS1798	106	Poole	$tcdA^+ tcdB^+$
DS1787	106	Leicester	$tcdA^+ tcdB^+$
DS1771	106	Bristol Southmead	$tcdA^+ tcdB^+$
DS1742	014	Bristol Frenchay	$tcdA^+ tcdB^+$
DS1748	002	Leeds	$tcdA^+ tcdB^+$
DS1721	005	Leicester	$tcdA^+ tcdB^+$
DS1752	012	Bradford	$tcdA^+ tcdB^+$
CD630	012	NCTC	$tcdA^+ tcdB^+$
DS1723	078	Leicester	$tcdA^+ tcdB^+$
DS1724	020	Leicester	$tcdA^+ tcdB^+$
DS1684	010	Brighton	$tcdA^- tcdB^-$
DS1665	023	Bath	$tcdA^+ tcdB^+$
Other Species			
Bacillus atrophaeus ATCC 9372		ATCC	

Cultures of *C. difficile* were grown in brain–heart infusion (BHI) broth and agar (Oxoid, Basingstoke, Hampshire, UK) unless otherwise indicated. Others were grown on Columbia agar media. Media was degassed in anaerobic conditions (a mixture of 80% N_2, 10% CO_2 and 10% H_2) for 24 h prior to use. Cultures were incubated anaerobically at 37^{o}C for 48 h in a Bug Box Plus anaerobic cabinet (Ruskinn Technology Ltd, Bridgend,

Wales). Cultures were tested for purity by streaking pure colonies onto BHI agar and re-incubating. To enable germination of *C. difficile* spores, 1% (w/v) of the bile salt sodium taurocholate (ST) (Sigma Aldrich, Dorset, UK) was added to specified media before autoclaving. Colonies were streaked to purity. Columbia agar and broth (Difco, Franklin Lakes, NJ, USA) was also used in experiments in this Chapter.

Media composition: Columbia agar media (Difco, USA) contained approximately per litre: Pancreatic Digest of Casein (12.0 g) , Peptic Digest of Animal Tissue (5.0 g), Yeast Extract (3.0 g), Beef Extract (3.0 g), Corn Starch (1.0 g), Sodium Chloride (5.0 g), Agar (13.5 g), Colistin (10.0 mg), Nalidixic Acid (10.0 mg). Brain Heart Infusion agar (Oxoid Ltd, Basingstoke, UK) contained approximately per litre: Calf brain infusion solids (12.5 mg), Beef heart infusion solids (5.0mg), Protease peptone (10.0mg), Sodium chloride (5.0mg), Glucose (2.0 mg), Disodium phosphate (2.5 mg), Agar (10.0 mg).

2.4.2. Spore production and enumeration

To produce spores, a single colony harvested from the appropriate agar plate was used to inoculate 10 ml of the corresponding degassed broth which was incubated for 10 d at 37^{o}C in the anaerobic workstation. At the end of this period the cultures were centrifuged (Heraeus Biofuge primo R, Fisher, UK) at 5 000 x *g* for 15 min at room temperature and the supernatant discarded. The retained pellet was resuspended in 10 ml sterile deionised water (sdw) and centrifuged at 5 000 x *g* again. This washing cycle was repeated a further two times at the end of which the spore pellet was resuspended in 1 ml of sdw and heated in a heating block (Techne, driblock, UK) to 80^{o}C for 10 min to inactivate any remaining vegetative cells and stored at 4^{o}C.

The number of spores produced following broth culture was determined using a method based on that of Miles & Misra (1938). An aliquot of 20 µl of extracted spores was serially diluted in degassed sterile broth media (180 µl) using a microtitre plate (Sterilin Ltd., Teddington, Middlesex, England) until a dilution down to 10^{-10} was achieved. A total of three 10 µl aliquots from each dilution were spotted onto the surface of the appropriate degassed agar plates (neat to 10^{-10}). Once inoculated, spots were allowed to dry at 25^{o}C and the plates were incubated anaerobically for 48 h at 37^{o}C. Colony forming units (cfu) were counted from each spot on the microtitre plate and the mean number of colony

forming units per millilitre (cfu/ml) determined by multiplying the mean cfu by the dilution factor. Unless otherwise stated all organisms were stored as spores at 4^{o}C.

2.4.3. Transmission Electron Microscopy studies

Spore samples produced as described in Section 2.4.2 (1 ml) were centrifuged at 2 000 x *g* for 4 min. The supernatant was removed and pelleted spores were fixed with 1 ml of a 2% (v/v) glutaraldehyde and 2% (v/v) Osmium tetroxide mix. This was left for 1 h at 22^{o}C, after which the Osmium tetroxide- spore mix was centrifuged (Heraeus Biofuge primo R, Fisher, UK) at 2 000 x *g* for 2 min. BHI agar was prepared (3 g to 50 ml water) and molten agar (50^{o}C) added to the stained spore pellet and left to set. The extra agar around the spores was removed and the agar-embedded spore pellets were dehydrated through an ethanol series from 50% ethanol to 100% alcohol for 10 min each. The 100% dehydration was repeated three times, and the ethanol was replaced with propylene oxide (PO) for 15 minutes.

Epoxy resin was composed of a mixture of araldite CY212, dodecyl succinic anhydride (DDSA) and benzyldimethylamine (BDMA) (Sigma-Aldrich, Dorset, UK). An equal quantity of PO was added to the resin and used to cover each spore pellet. The pellet was left for 36 h to allow resin infiltration. PO mixture was subsequently removed and pure Epoxy resin (Sigma-Aldrich, Dorset, England) was added to the samples which were embedded into a flat moulded tray and left in an oven at 60^{o}C to harden for 48 h. Transverse sections were cut on a Microtome (Reichert-Jung, Depew, NY, USA) and stained on nickel grids with uranyl acetate and lead citrate. Images were taken on the Transmission Electron Microscope (Philips EM 208). Of each strain sample examined under TEM, 10 spores were viewed.

2.4.4. Negative-stain electron microscopy of whole spore

Spore samples were diluted in sdw, placed on Formvar-coated grids and examined after negative staining in 2% (w/v) methylamine tungstate. After 15–30 s the excess stain was withdrawn using filter paper. The grids were washed twice with water and dried by filter paper. Samples were examined immediately using a Phillips EM 208 transmission electron microscope with an accelerating voltage of 80 kilovolts (kV), and magnifications between 20,000 x and 55,000 x were used.

2.4.5. Microbial adhesion to hydrocarbon test

To determine the hydrophobic characteristics of individual *C. difficile* spores from a range of isolates we used a hexadecane hydrocarbon based microbial adhesion to hydrocarbon test (MATH). The MATH assay has been commonly used to ascertain hydrophobic characteristics of bacteria and it is hypothesized that hydrophobic bacteria are more likely to adhere to the hydrocarbon and the hydrophilic bacteria tend to remain in water (Rosenberg *et al.*, 1980). Thus the *C. difficile* spores of each isolate were suspended in 9 ml of sdw to achieve an optical density of between 0.500-0.600 at OD_{600} (Ultrospec 1100 *pro* UV/Visible spectrophotometer, Biochrom, Cambridgeshire, UK). A 4 ml aliquot of spore suspension in a Universal container was vortex mixed (VortexGenie, Fisher Scientific, UK) with 0.4 ml of the hydrocarbon *n*-hexadecane (Sigma-Aldrich, Dorset, England) at full speed for 1 min. The mixture was incubated at room temperature ($22^{o}C$) for 15 min to allow the different phases to partition settle. Subsequently the optical density of the lower aqueous phase was determined. Each assay was repeated twice. Changes in hydrophobicity were calculated as a percentage from original OD_{600} to the final OD_{600} post hexadecane exposure.

2.4.6. Contact angle measurements

To ascertain the hydrophobic characteristics of the surfaces tested in this study we employed a contact angle test upon the hospital grade stainless steel discs and the agar. The contact angle between the test surface and a droplet of water was measured using a horizontal projection technique. Briefly, a droplet of water was put onto the stainless steel disc and the agar and a light was used to project the image. The projection allows a crude measurement of the contact angle between the droplet and the surface using a protractor, the hypothesis being that the contact angle between the water and a hydrophobic surface (steel) would be less than the contact angle between the water and a hydrophilic surface (agar).

2.4.7. Stainless steel plate transfer assay

In order to assess adherence of spores to a hospital grade stainless steel surface and an agar surface considered to be representative of a hydrophilic surface, we employed a spore transfer assay. A stainless steel disc 2 cm in diameter with a Grade 2B finish (Goodfellows Cambridge Ltd, Huntingdon, UK) was inoculated with 10 µl of spore suspension (500 spores per 10 µl in units of cfu/µl) and left to dry at $25^{o}C$ for 20 min. The

inoculated steel discs were aseptically attached to a steel rod plunger and securely fastened into the supporting metal frame as seen in Figure 2.2 (IKA, Labortechnik, Staufen, Germany). An electronic weighing balance was secured to a raising platform and positioned beneath the metal plunger. Either side of the platform, two Bunsen burners created a sterile vortex of air encompassing the platform area. A degassed BHI agar plate was placed directly under the metal plunger, on the tared electronic balance. The platform and balance were raised until the steel disc compressed the agar surface with a force of 100 g (+/-5g) and held for 10 s. The balance was lowered; the agar plate replaced with another and the compressing process repeated a total of sixteen times for each inoculated disc. The compressed agar plates were anaerobically incubated at 37°C for 48 hours and the number of cfu on each plate counted. Sixteen strains of *C. difficile* were tested (Figure 2.8).

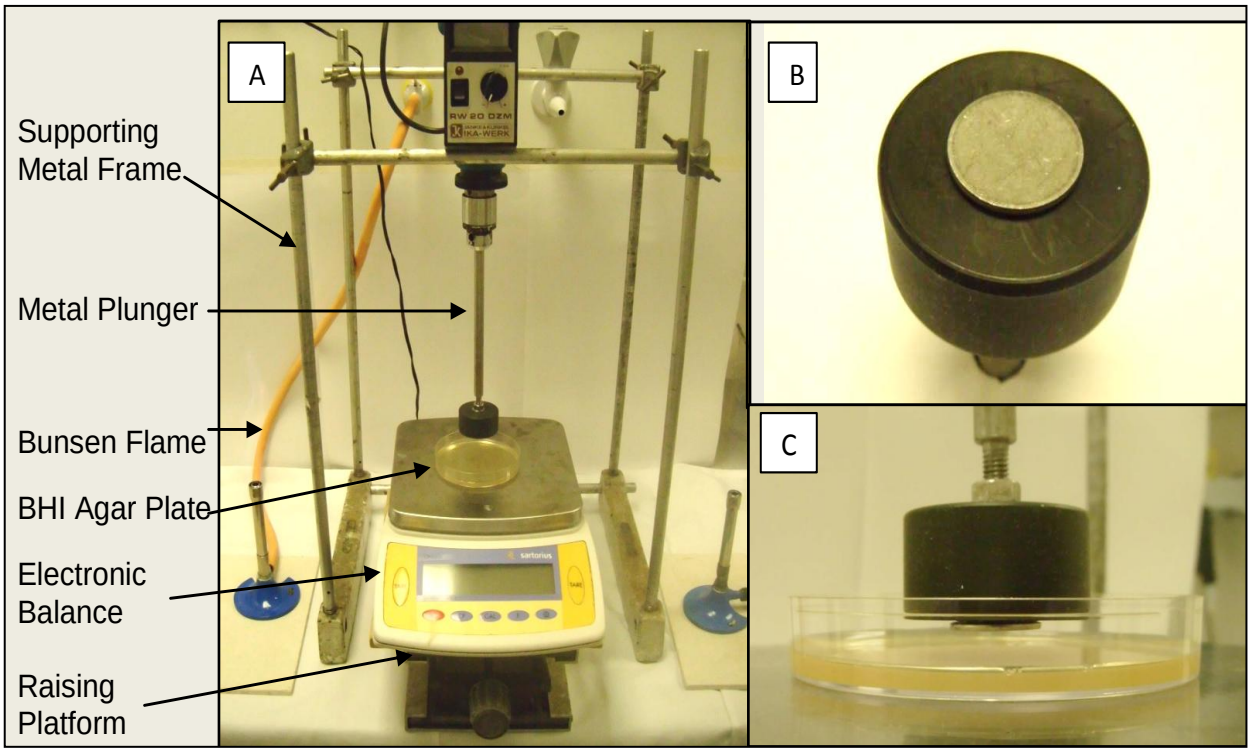


Figure 2.2: Plate transfer assay for spore transfer

This image shows the transfer apparatus used to assess spore transfer from stainless steel. A: Apparatus set up for spore transfer. The plunger is attached to a supporting metal frame. To maintain a sterile environment a Bunsen burner is placed either side of an electronic balance with a BHI agar plate on it. The balance was tared with each agar plate and the spores plunged at a force of 100 g/ 10 s. B: An inoculated stainless steel disc fastened to metal plunger. C: BHI agar plate compressed aseptically with inoculated steel disc. As *per* Williams *et al.* (2009).

2.4.8. Preparation of mammalian cell culture lines

Two human epithelial colorectal adenocarcinoma cell lines, Caco 2 (ATCC HTB-37) and HT 29 (ATCC HTB-38) were grown as monolayers in continuous culture as subsequently described. The two intestinal cell lines were passaged in minimum essential medium with Earle's balanced salt solution (MEM/EBSS, Invitrogen, Paisley, UK) supplemented with 10% foetal bovine serum, 1% non-essential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine and 50 µg/ml penicillin: streptomycin solution (Invitrogen, Paisley, UK) and incubated at 37°C in a 5% CO_2 atmosphere in 25 cm^2 vented flasks. Flasks were sub-cultured after cells reached full confluence (analysed via phase-contrast microscopy according to cell density) every 6-7 days, at which point cells were washed 2 times in PBS (pH 7.2), detached from the plastic flask using a trypsin: EDTA (Ethylene-diamine-tetraacetic acid) solution (Invitrogen) and seeded into new flasks to a cell density of 1-5 for Caco 2 cells and 1-10 for HT 29 cells.

2.4.9. Adherence ability of *C. difficile* spores to Human gut epithelial cell lines

Spores of strain DS1813 at 1.40 x 10^8 cfu/ml and DS1748 at 1.30 x 10^8 cfu/ml were tested for adherence to Caco 2 and HT 29 cells. Caco 2 cells and HT 29 cells were prepared as detailed above. The initial cell density was at 1-2 x 10^5 for each cell line. The cell density can be thus related to the ratio of bacteria per cell line. One hundred µl of 1.30 x 10^8 cfu/ml spores were added to 1 ml 7 day old confluent Caco 2 gut epithelial cells in antibiotic-free Eagles minimum essential medium (EMEM) and left to incubate for 100 min in a Bug Box Plus anaerobic cabinet (Ruskinn Technology Ltd, Bridgend, UK). Post incubation, the cell- spore mixture was washed in phosphate buffered saline (PBS; Sigma-Aldrich, Dorset, UK) following centrifugation at 800 x *g* to remove any unattached spores. This washing step was performed three times in total and the supernatant from each wash step collected for enumeration of any spores not adhering to the cells. The supernatant was enumerated as per Miles & Misra (1938) on degassed BHI + 0.1% ST. Controls included PBS incubated with spores and EMEM incubated with the spores.

This method was repeated for 15 day old confluent Caco 2 cells, which had spontaneously become differentiated and polarized, showing morphological changes, including tight junctions, cell polarity and formation of a microvilli brush boarder (Pinto *et al.*, 1983). The experiment was also repeated using confluent 7 d and 15 d old HT 29 cell monolayers, these cells however remained in an undifferentiated state, but are documented

to contain a small proportion (≤0.2%) of mucin-secreting goblet cells post-confluency (Pinto *et al,*. 1982; Lesuffleur *et al.*, 1990). The human gut epithelial cell lines differentiated into a microvilli brush border at 15 d for Caco 2 and HT 29 cells. All tests were performed in duplicate.

2.4.10. Statistical analysis

Statistical analysis was performed using GraphPad Prism version 5.04 for Windows, (GraphPad Software, La Jolla California USA, www.graphpad.com). Statistical significant differences were tested for using one way analysis of variance (ANOVA) at the 95% confidence interval in conjunction with Dunnett's post test. A P value of <0.05 was considered significant (Bowker & Randerson, 2006).

2.5. RESULTS

2.5.1. Spore germination

We compared the ability of a representative collection of twenty one clinically relevant isolates of *C. difficile* to form spores when cultured in BHI media in the absence of the co-germinant sodium taurocholate. The average spore yield for each isolate following culture and enumeration on BHI based media are shown in Figure 2.3. Each result represents the mean of four individual assays. In the absence of sodium taurocholate, viable counts on BHI agar varied from 10^4 - 10^7 cfu/ml across the range of strains tested. While strain DS1759 (PCR ribotype 001) produced the highest yield of spores, this trait was not shared by other ribotype 001 strains (n=4). Indeed there was variation (Figure 2.3) in spore germination ability between strains of the same PCR ribotype. There is a significant difference in germination ability between each strain as determined via one way ANOVA (P=0.02). Further statistical analysis using Tukeys multiple comparison test confirms DS1759 is better able to germinate than other strains (P<0.05).

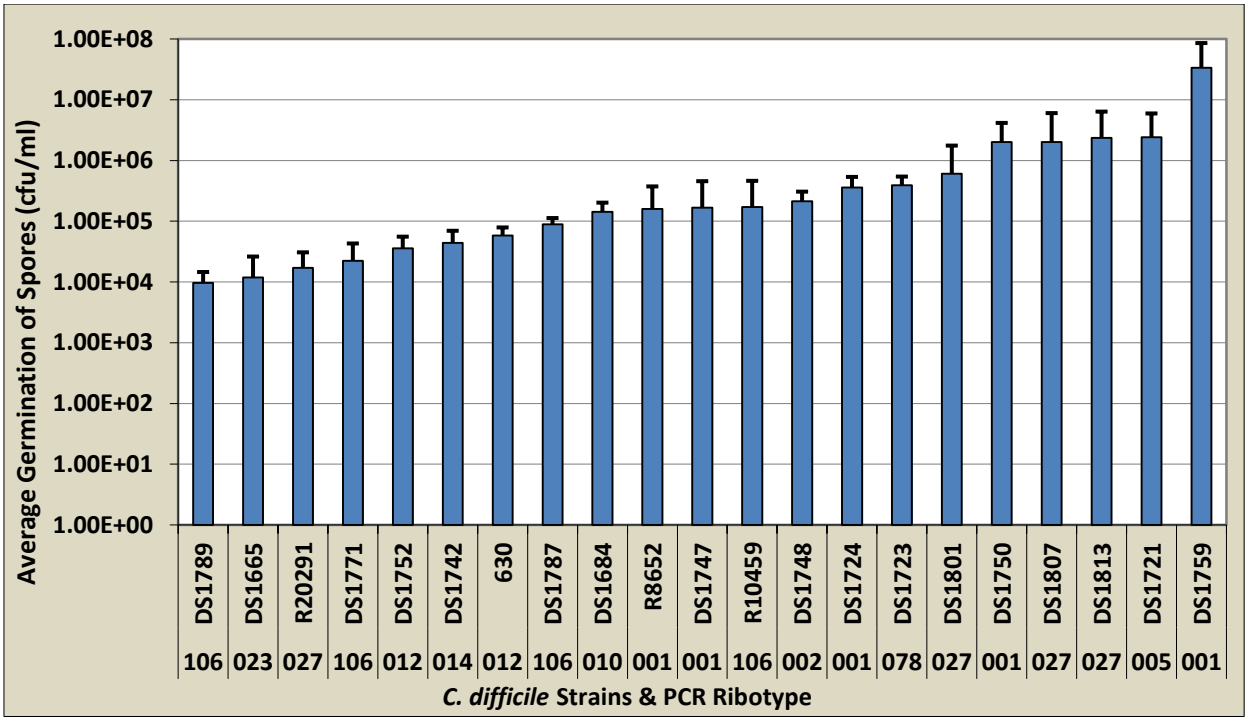


Figure 2.3: Spore formation and germination on BHI media

A range of 21 clinically relevant *C. difficile* isolates were grown in BHI media and cultured without the germinant sodium taurocholate. There were variations in the ability of isolates to germinate on BHI media, and thus the number of viable spores (cfu/ml). Spores were produced in quadruplet and these values represent an average spore yield across strains and the difference is significant between strains as P = 0.02. (n = 4).

2.5.2. The effect of media composition on spore formation

To determine if the ability of *C. difficile* isolates to form spores was influenced by the composition of the media, and the ability of eight clinical isolates of *C. difficile* to form spores in BHI and Columbia broth was determined. Following incubation for 10 d, heat shocked cultures were diluted and plated onto either BHI or Columbia agar containing 1% sodium taurocholate to trigger germination. The ability of certain strains to form spores appears to be affected by the composition of the media. This was most evident for strains R20291 (ribotype 027, Stoke Mandeville isolate) (difference of approximately 3 log as shown in Figure 2.4) and DS1801 (ribotype 027) (difference of approximately 2 log as in Figure 2.4) which show the greatest difference in spore counts. There was no significant difference (one way ANOVA; P = 0.2519) in spore germination ability between BHI and Columbia media when the co-germinant sodium taurocholate.

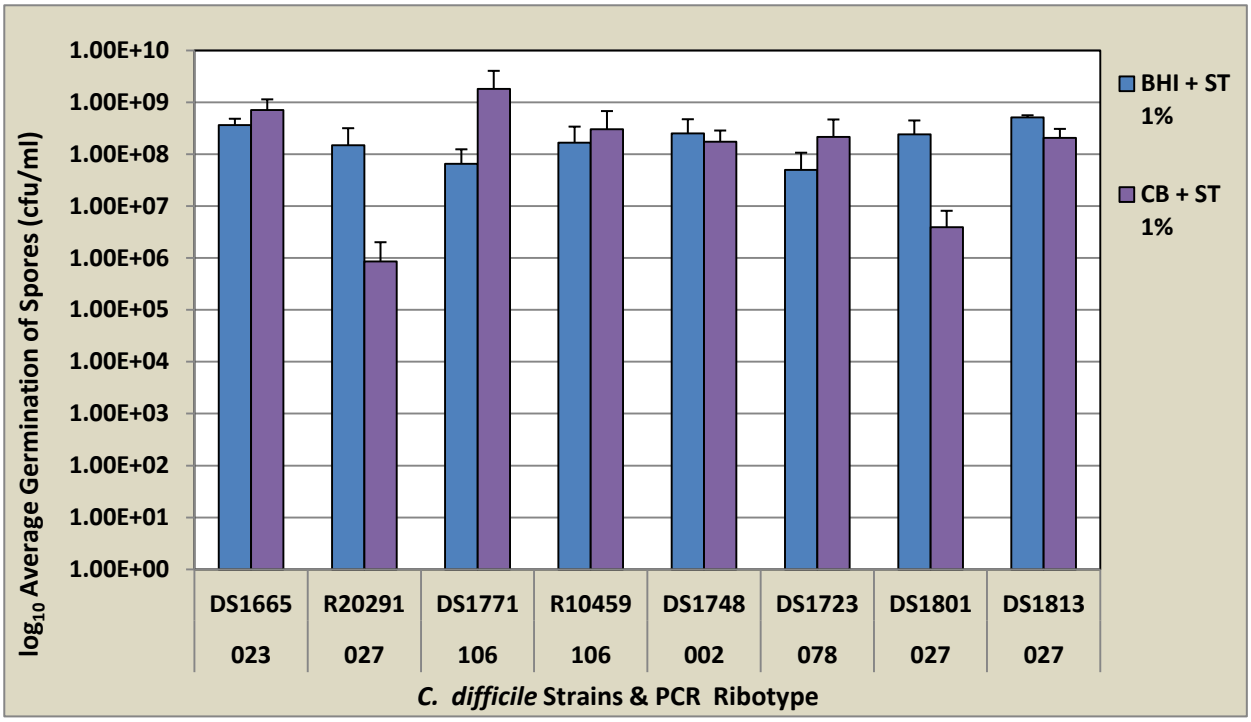


Figure 2.4: Comparison of spore formation and germination of clinical isolates of *C. difficile*

A further 8 clinical isolates were compared for spore germination ability. Spores generated in either Columbia or BHI broth were cultured on the corresponding agar media in the presence of 1% sodium taurocholate. Significant variations in spore germination ability in Columbia media can be seen for two isolates- DS1801 and R20191 in particular (P = 0.2519) (n = 4).

Furthermore the percentage difference in the germination of spores between media with and without the supplement taurocholate was assessed (Figure 2.5). The percentage of spores that germinated in the absence of the co-germinant taurocholate differed markedly depending on the media and the isolate. Indeed the results suggest that Columbia media contains more essential co-germinants than BHI and that the isolates DS1801 and DS1813, both hypervirulent 027 ribotypes, differ in their germinant requirements from the other strains.

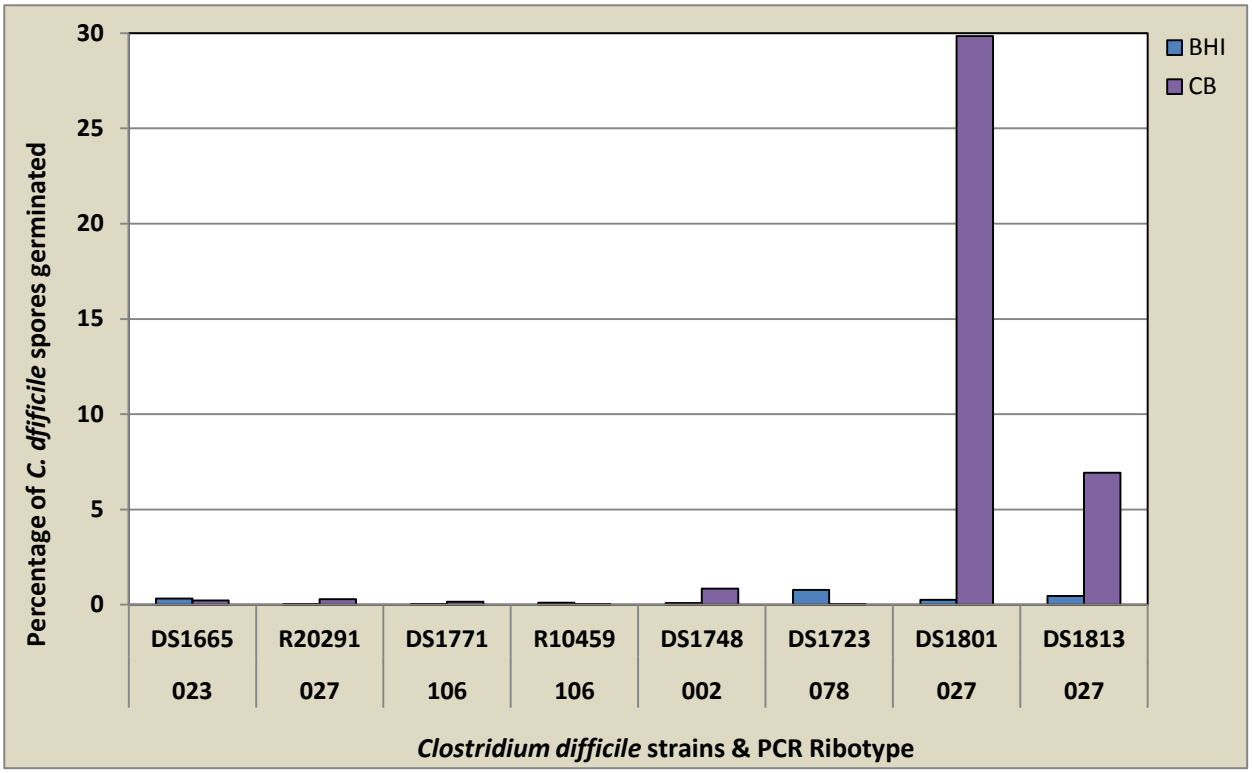


Figure 2.5: Percentage difference in germination of *C. difficile* spores between Columbia or BHI media. The percentage difference in spore germination between media with and without sodium taurocholate supplemented is shown together for clinically relevant *C. difficile* isolates. The isolates with the most marked difference (DS1813 and DS1801) grown in Columbia media germinate better with the germinant sodium taurocholate added.

2.5.3. Hydrophobicity

The ability of a spore to adhere to a surface is influenced by its hydrophobicity. To characterise the relative hydrophobicity (RH) of spores obtained from a range of clinical isolates, we employed the Microbial Adhesion to Hydrocarbon Test (Rosenberg *et al.*, 1990) with hexadecane as the organic solvent. Spores of *B. atrophaeus* ATCC 9372 (Marker) were included as an internal control to enable a comparison to be made to the results from other studies. As expected the RH value for this strain was comparable to that reported by Wiencek *et al.* (1990). As can be seen from Figure 2.6 relative hydrophobicities ranged from 77% (DS1813 ribotype 027) to 14% (DS1748 ribotype 002). Interestingly the most hydrophobic isolates appear to be those of hypervirulent or epidemic ribotypes (027, 001, 078, 106) suggesting a possible association with virulence.

There is a significant difference in the hydrophobicities within each PCR ribotype and between each strain ($P < 0.0001$).

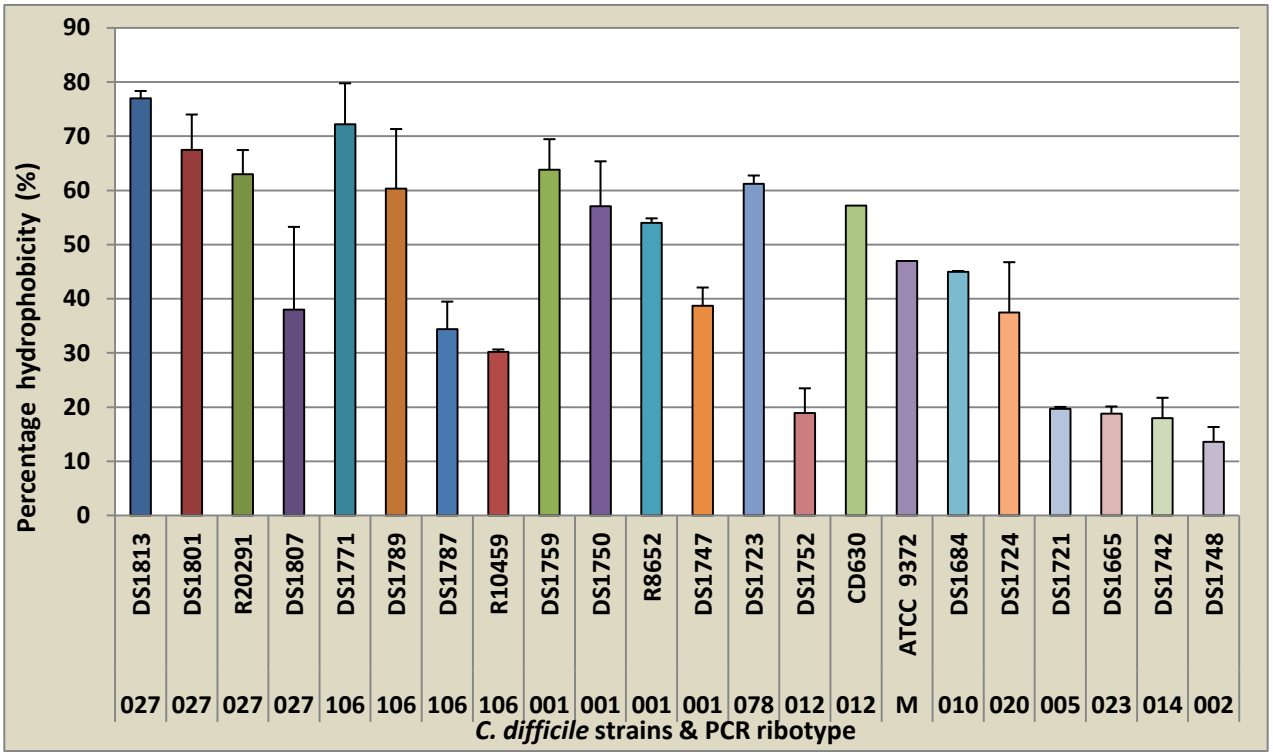


Figure 2.6: Relative hydrophobicity of the spore form of different ribotypes of *Clostridium difficile*

A microbial adhesion to hydrocarbon test was used to determine the hydrophobic characteristics of the surface of *C. difficile* spores for a range of 21 clinical isolates. The highest percentage hydrophobicity demonstrates the most hydrophobic spores and the lowest percentage demonstrates the most hydrophilic spores. The percentage hydrophobicity is grouped according to PCR ribotype and there is significant difference between the strains according to hydrophobicity percentage ($P<0.0001$) ($n = 2$).

2.5.4. Spore surface characterisation by electron microscopy

To determine if variation in relative hydrophobicity could be related to spore surface architecture we employed transmission electron microscopy to characterise the spore forms of a range of *C. difficile* isolates. The most hydrophobic DS1813 (77% RH), DS1771 (72.2 % RH), DS1684 (45% RH) and hydrophilic DS1748 (14%) were examined under TEM. While the spores were structurally similar there were discernable differences between the strains.

Spores produced from strain DS1813 comprised an electron dense cortex surrounded by concentric rings of peptidoglycan, a second thinner outer coat layer and finally an exosporium which surrounded the majority (66%) of the spores observed (Figures 2.7 A, B, C). The exosporial layer is a bag–like structure surrounding the spore which can clearly be seen in Figure 2.7 B. The exosporial layer surrounds the spores of DS1813 (Figure 2.7 A & C). It must be noted that this exosporial structure was not seen in some DS1813 samples, which suggests that it is not strain specific. In contrast we were unable to detect the presence of an exosporial layer surrounding the spores of the other strains. Indeed the common concentric ring structure of the spore coat layers can be seen in each TEM image.

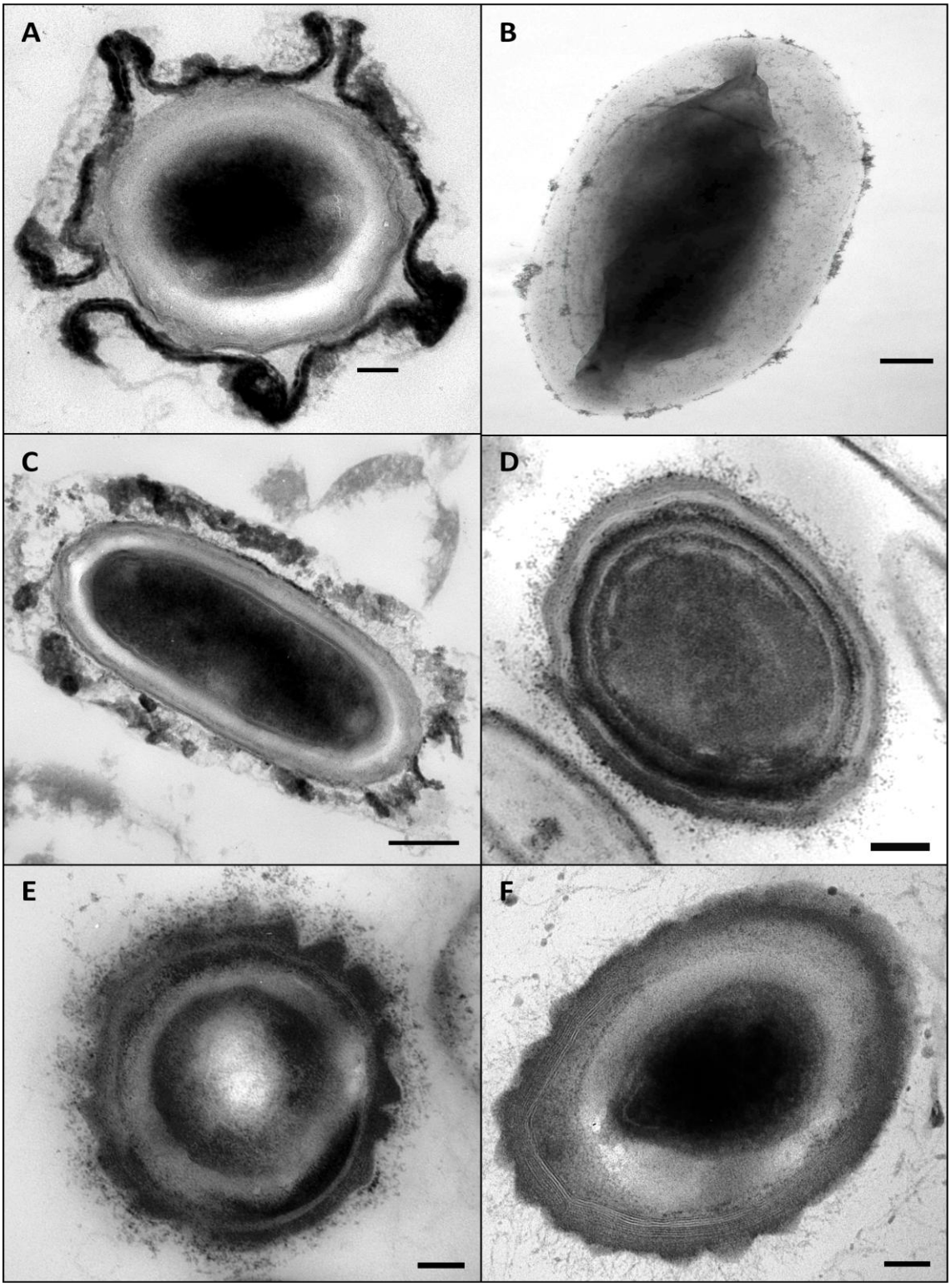


Figure 2.7: Transmission electron microscopy of *C. difficile* spores

A: TEM image of a cross section of a hydrophobic spore produced by *C. difficile* DS1813. The exosporial layer can be seen. B: Negative stain electron microscopy of exosporium sac surrounding a spore of DS1813 C: DS1813 with exosporium visible. Subsequent strains did not exhibit and exosporial layer D: Strain DS1771. E: Strain DS1684 F: Strain DS1748 (Images scale bar = 100 nm).

2.5.5. Ability of *C. difficile* spores to adhere to stainless steel

Stainless steel is a material common in the hospital environment and thus we sought to determine if spores from a range of clinical isolates varied in their ability to adhere to this material. To achieve this, a plate transfer method was employed in which the percentage of spores which remained attached to the surface of a hydrophobic (>45 degree contact angle) steel disc was determined following consecutive impression on the surface of 16 hydrophilic (0 degree contact angle) BHI agar plates. As can be seen in Figure 2.8, the adherence of spores from different clinical isolates varied from 18-100%.

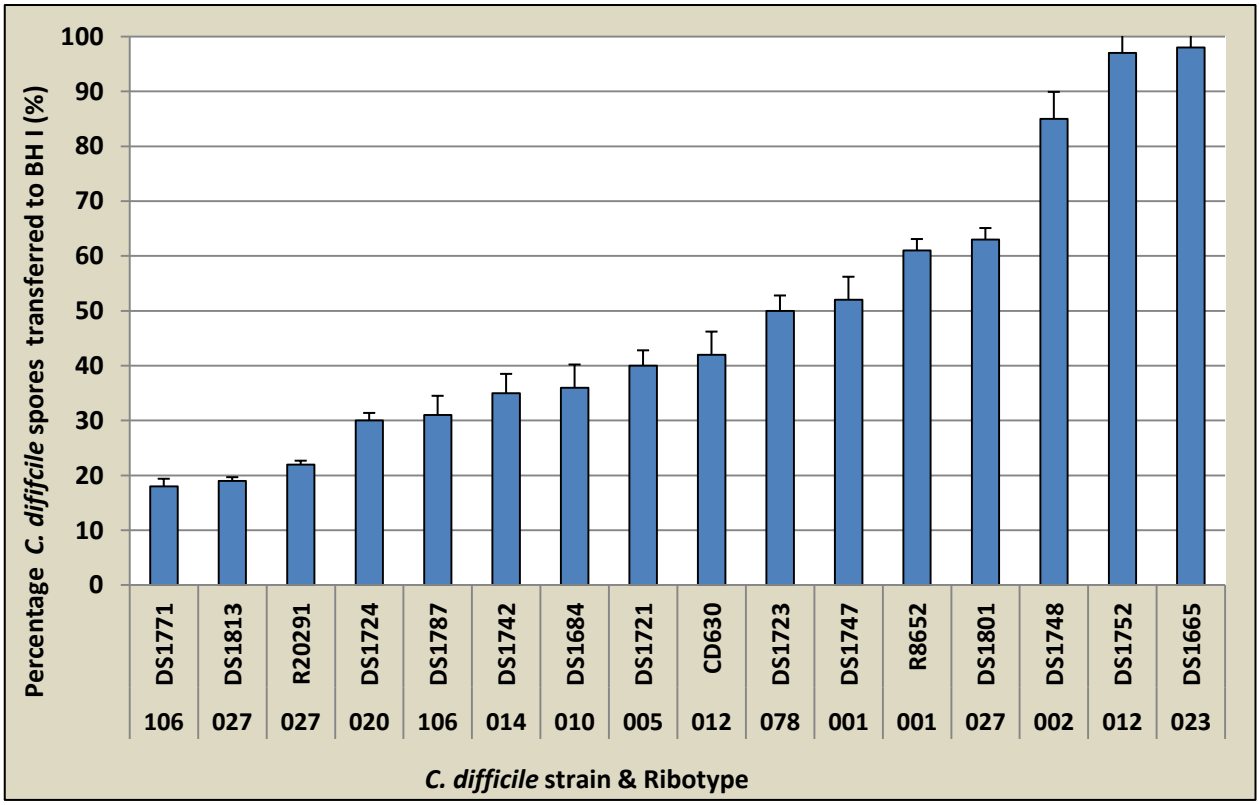


Figure 2.8: Percentage of *C. difficile* spores transferred from stainless steel to BHI plate

The percentage of *C. difficile* spores was determined after transfer from a stainless steel disc onto the surface of BHI agar plates following 16 replica platings. The correlation between surface transfer and strains was not statistically significant (P>0.05). n=2.

Differences in adherence to steel correlated with relative hydrophobicity as determined using the MATH assay, in that hydrophobic isolates adhered better than their hydrophilic counterparts. Thus spores of DS1813 adhered more strongly to steel than those of DS1748.

2.5.6. Adherence of *C. difficile* spores to human gut epithelial cell lines

We employed a human gut epithelial cell model to determine the adhesive ability of spores derived from DS1813 and DS1748. We used two intestinal cell lines Caco 2 and HT 29. Caco 2 cells originate from a colon carcinoma cell line which require 15 d culture to become fully differentiated, at which time they possess a microvilli brush border which is absent at day 7. HT 29 cells originate from an adenocarcinoma cell lines and do not spontaneously differentiate upon culture although some cells can develop goblet cells which produce mucus.

Hence we sought to determine if spores of DS1813 and DS1748 differed in their ability to bind to these cells at different stages of development. There was a significant difference (one way ANOVA; P= 0.042) in the ability of spores of strain DS1813 to adhere to both cell lines in contrast to DS1748 (Figure 2.9). Interestingly there was no significant difference between DS1748 and DS1813 adherence to 7 d Caco-2 cells and 15 d HT 29 cells; however there was a significant difference for both strains at 15 d old Caco 2 cells (P = 0.0393) and 7 d HT 29 cells (P = 0.0079). This difference in adherence ability may be attributed to the differences in surface properties between the spores of both strains.

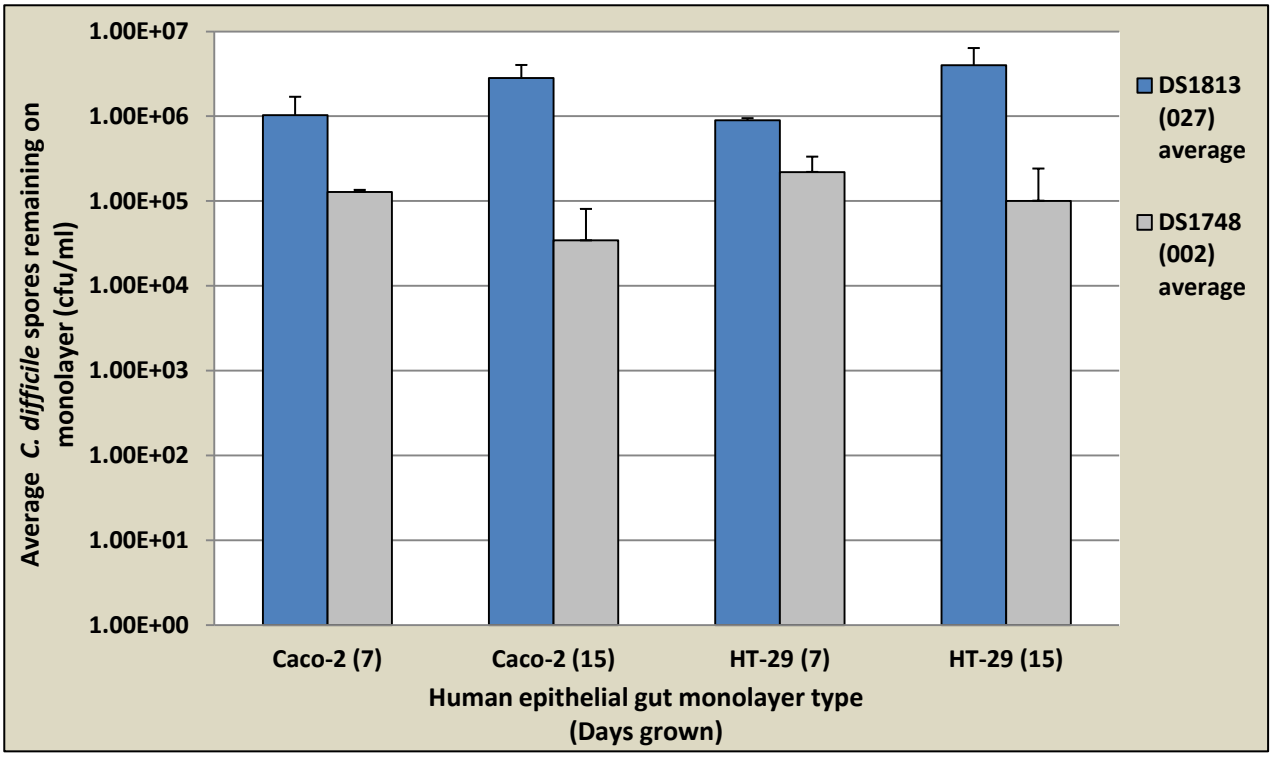


Figure 2.9: Adherence of *C. difficile* spores to two separate human gut epithelial cell monolayers

The spores of isolates DS1813 (hydrophobic) and DS1748 (hydrophilic) were assessed for adherence to monolayers of human colon adenocarcinoma grade II cell line (HT 29) and Caco 2 human colorectal cell lines. Spores were enumerated on 0.1% sodium taurocholate and BHI agar. These values represent an average spore adherence for strains DS1813 and DS1748 (n = 2).

2.6. DISCUSSION

The spore form of *C. difficile* plays a key role in facilitating person to person spread of the pathogen. It allows the organism to survive in the hostile hospital environment protecting it from the toxic effects of oxygen and biocides. Currently there is considerable debate as to whether clinical isolates vary in their ability to form spores: it has been suggested that isolates of the hypervirulent type BI/NAP1/027 strains sporulate better than their non-epidemic counterparts and that this difference results in increased disease occurrence and severity (Akerlund *et al.,* 2008). However a recent conflicting study by Burns *et al.* (2010) demonstrated that sporulation rates of a number of strains of *C. difficile,* including those of non-clinical isolates, were broadly similar. Hence there was no preference of a particular strain/ribotype to possess increased sporulation ability. While we were unable to comment on the rate of spore formation we did observe that certain stains such as R20291 (BI/NAP1/027) failed to achieve the same level of spores at 10 days as other isolates and that these differences were growth media dependent, raising the question as to their relevance *in vivo.*

Once the spore has been released from an infected individual it must adopt strategies which enable it to survive and colonise the environment. The results of our study suggest that the emergence of epidemic and hypervirulent isolates may in part be due to their enhanced ability to adhere to organic (human adenocarcinoma cells) and inorganic (stainless steel) surfaces compared to non-epidemic strains. Using a simple MATH assay developed by Rosenberg *et al.* (1980) we observed what appears to be a relationship between PCR ribotype and spore hydrophobicity. The spores that demonstrated the highest levels of hydrophobicity were predominantly hypervirulent and epidemic ribotypes (027, 001, 106 and 078) while the hydrophilic spores were mainly non-epidemic clinical isolates. A similar association was seen when the binding of spores to stainless steel was examined. This material is widely used in the healthcare environment in the manufacture of surgical equipment and hand washing facilities, and as such represents a potential source of spore contamination. Thus spores from hypervirulent ribotypes may be better adapted to adhere to these surfaces and thus increase the potential for person to person spread.

To determine if the observed differences in hydrophobicity and adherence were due to variations in spore structure we employed electron microscopy to characterise the

morphology of a range of spores with varying relative hydrophobicity as determined by MATH. Spores of hydrophobic DS1813, DS1771, DS1684 and hydrophilic DS1748 strains were examined under TEM. While the spores of each isolate shared a number of common features- an internal core, cortex and outer coat layers- they differed with regards to the presence of an exosporium which was only present in strain DS1813.

This bag like structure was recently described surrounding spores of another strain of *C. difficile* CD630 (Lawley *et al.,* 2009) and is thought to contribute to spore hydrophobicity and play a direct role in adherence (Koshikawa *et al.,* 1989; Charlton *et al.,* 1999). Indeed it has been suggested that the degree to which a spore can adhere to a surface is directly influenced by the presence or absence of the exosporium in a range of *Bacillus* species, such as *B. megatarium* and *B. cereus* (Koshikawa *et al.,* 1989). However this exosporial layer did not surround spores of every isolate examined, suggesting that there is no direct correlation between spore hydrophobicity and the presence of an exosporium.

In addition to mediating binding to steel, the exosporium plays a central role in facilitating the attachment of spores to the cells lining the gut- an important first step in bacterial invasion. The human colonic cell lines Caco 2 and HT 29 are used as *in vitro* models due to their ability to retain key features of normal intestinal epithelial cells and thus were considered to be representative of cells potentially encountered by *C. difficile* during germination *in vivo* (Howell *et al.,* 1992). Caco 2 cells become polarized and fully differentiate into mature cells having microvillus brush-boarder morphology by 15 d of culture, whereas 7 d old cells lack these features (Pinto *et al.,* 1983).

Parental HT 29 cells do not spontaneously differentiate in culture, and thus do not exhibit the same characteristics as Caco 2 cells after 15 d of prolonged cultivation (Pinto *et al.,* 1982; Roussett, 1986). However a small number of goblet cells are present post-confluency capable of producing mucins (Pinto *et al.,* 1982; Lesuffleur *et al.,* 1990). Not surprisingly we found that the spores produced by the hypervirulent DS1813 (027) which possessed an exosporium were more efficient at attaching to both HT 29 and Caco 2 gut monolayers than spores of DS1748. It has previously been observed that spores of *C. difficile* attach to the apical microvilli of human epithelial cells by means of their exosporium which, in addition to facilitating colonisation, acts as an anchor during germination (Panessa-Warren *et al.,* 1997).

As *C. difficile* growth is restricted to oxygen limited environments, such as those encountered in the gut, there is considerable interest in the manner in which *C. difficile* spores regulate their germination. Bile salts such as taurocholate, which has a positive effect on germination, and chenodeoxycholate which has an inhibitory effect on germination, have been posited to have a central role in regulating germination and as such probably represent the means by which *C. difficile* spore "sense" their physical location. In addition to bile salts other amino acid based co-germinants have been identified *in vitro* such as glycine, alanine, cysteine and phenylalanine (Setlow, 2003; Sorg & Sonenshein, 2008; Giel *et al.*, 2010). Indeed combinations of these amino acids are able to stimulate *in vitro* spore germination in the absence of taurocholate (Sorg & Sonenshein, 2008).

In the context of our study differences in the composition of BHI and Columbia media may account for the variation we observed in spore formation and germination. While there appears to be an association between the 027 ribotype and the ability to germinate in Columbia media in the absence of taurocholate, the limited number of isolates examined makes it impossible to draw any solid conclusions at this time. While commercial growth media in no way seeks to mimic the complex conditions encountered in the human gut, they do provide a system in which to compare the ability of different isolates to germinate. Indeed the results from this study suggest that further examination of strain DS1801 could aid our understanding of the factors that regulate germination. However the results obtained using these media must be treated with caution as they may simply represent media specific artefacts and bear no relationship to what actually happens *in vivo*.

In conclusion we found that the surface properties of spores produced by hypervirulent and epidemic isolates of *C. difficile* made them better equipped than their non-epidemic counterparts to adhere to both gut cells and healthcare surfaces and as a consequence may enhance their opportunity to transfer to susceptible hosts and initiate new infections. By fully understanding this process by which spores attach to clinically relevant surfaces we hope to be able to develop strategies which could inhibit attachment of spores and thus block the infection cycle. We also conclude that that there is no direct correlation between spore hydrophobicity and the presence of an exosporium. Finally the study of the factors affecting germination and spore formation are extremely complex, particularly when

attempting to relate laboratory derived observations using a small number of isolates to events that are occurring *in vivo*.

CHAPTER 3

DETECTION ASSAY PROBE DESIGN AND *CLOSTRIDIUM DIFFICILE* TOXIN EVOLUTION

3.1. ABSTRACT

Conserved sequences within the gene encoding regions of *C. difficile* toxins A and B were identified using a restriction enzyme site based typing system developed by Rupnik *et al* (1998). These conserved sequences lacked the restriction sites targeted by the restriction enzymes used in Rupnik's toxinotyping system. Bioinformatic analysis was employed to design probes specific to toxins A and B from these regions. From bioinformatic analysis we found that conserved regions within the toxin A and B gene sequences shared homology to virulence genes associated with insects.

Candidate probes specific to toxins A and toxin B of *C. difficile* through bioinformatic analysis were synthesised and examined for specificity using a rigorous screening process against 58 clinical isolates of *C. difficile*, unrelated species, species from the LCT family and 10 human metagenomic gut DNA extracts. Probes which hit only with toxigenic *C. difficile* isolates were employed in the MAMEF based technological platform as described in Chapter 5.

3.2. INTRODUCTION

As described in Section 1.4.5.1, the genes encoding toxin A and B are thought to have evolved as a result of horizontal gene transfer. Gene transfer events occur in organism evolution, conferring characteristics which enable an organism to survive. Indeed within *C. difficile* areas of sequence conservation have been identified. Thus, on the premise that structurally important regions of each toxin are likely to be highly conserved, the results of the restriction enzyme based toxinotyping method (previously described by Rupnik *et al.*, 1998; Section 1.5.2) were utilised to pinpoint potentially conserved nucleotide regions in *C. difficile* toxin A and toxin B, which were not restricted by the enzymes used by Rupnik. The regions A1, A2 and B3 lacked any of the restriction sites that the range of restriction enzymes Rupnik (1998) specifically used to digest the PaLoc of *C. difficile* VPI 10463. Thus these areas were considered suitable target for probe design (Figure 3.1).

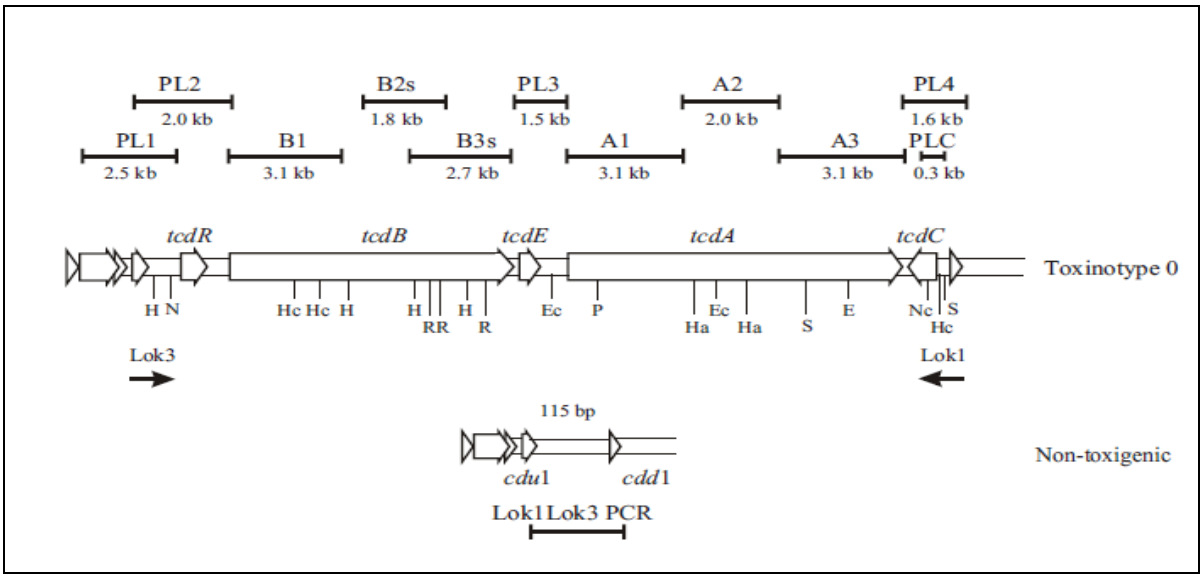


Figure 3.1: PCR amplification region of *C. difficile*

The PCR amplification areas are shown here in a schematic representation. The variable regions include B1, B2 and A3. Thus the areas with highest sequence conservation were B3, A1 and A2. (Image taken from Rupnik *et al.*, 1998). In toxinotype 0 and representative variant toxinotypes the presence or the absence and the exact positions of restriction site are shown: A – AccI, E – EcoRI, Ec – EcoRV, H –HindIII, Hc – HincII, N – NsiI, Nc – NcoI, P – PstI, R –RsaI, S – SpeI, X – XbaI.

To confirm if the nucleotide regions defined above by Rupnik (1998) were indeed highly conserved, individual gene sequences from multiple isolates were directly compared using

a bioinformatics based approach. Bioinformatics is the application of information technology to the field of molecular biology and has enabled information regarding biological structures, such as DNA, proteins and RNA to be analysed (Zvelebil & Baum, 2008). The genomic revolution was considered to have begun in the mid 1990s; however the roots of bioinformatics can be traced back to the 1950's when bovine insulin was sequenced by Frederick Sanger and colleagues (Ryle *et al.,* 1955). Indeed by the 1960's the first database of protein sequences "Atlas of Protein Sequence and Structure" was published by Margaret Dayhoff with >100 protein sequences available (Primrose & Twyman, 2007). During the 1970s and 1980s DNA sequences became more accessible and predicted protein sequences from translation of sequenced genes were available. By 1982 there were so many sequences available that they were collated into a database, known as GenBank, which today contains in excess of 30 million sequences (Benson *et al.,* 1999). Thus as the amount of biological information has increased it has been necessary to employ computer based algorithms to analyse them (Primrose & Twyman, 2007).

The method used to compare the relationship of nucleotide sequences utilises alignment pairing comparisons: two sequences are analysed and the number of shared residues determined (pairwise alignments). This method deduces sequence similarity in DNA and protein sequences which are subject to evolutionary changes including mutations within their genome sequences. Specifically bases, and the amino acids they encode, can change over time as a result of point mutations within the genome. Synthesis of sequences from different organisms and gene duplications are quite common in organism evolution. As such, it is necessary to employ alignment comparisons to reveal any potential similarities, resulting in an alignment score, which represents the alignment quality and the evolutionary closeness of the sequences (Primrose & Twyman, 2007). Those sequences with a high alignment score are deemed as having homology, which suggests that the organism/gene may have evolved from a common ancestor/ ancestral gene. Thus an alignment of two sequences is essentially a hypothesis that the sequence residues of interest have evolved from the same ancestor.

Multiple sequences can be aligned using online programmes such as ClustalW (EBI) (Larkin *et al.,* 2007). ClustalW has a graphic interface which allows viewing of a set of

sequences easily; hence if the sequences show a similar alignment pattern they are likely to have derived from a common ancestral sequence (Woodford & Johnson, 2004). Multiple sequence alignments allow wider analyses of sequences to be made, such as relationships between families to be deduced, predicting sequences, and to construct phylogenetic trees (Zvelebil & Baum, 2008). In the case of this thesis, multiple sequence alignment (MSA) analyses were used to establish areas with conserved residues in the sequences of *C. difficile* toxin A and toxin B.

To establish if a sequence has any homologs, various searching tools are available to detect homologs in databases. One such is BLAST (Basic Local Alignment Search Tool) [NCBI], which finds core similarities (defined by set parameters) between nucleotides and/or proteins and can search for protein sequences using a nucleotide sequence and *vice versa* (Altschul *et al.,* 1997). Firstly suitable matches are located within each database and are indexed according to rank by their similarities. The highest scoring sequences are deemed as having the most homology to the query sequence. These scores have an expectation value (E-value) which is the estimation of the probability of two random sequences aligning with a score higher than the defined parameter of *S* (Primrose & Twyman, 2007). E-values can vary from 0 to *X*, and those with an E-value closer to 0 are also deemed to have significant homology (Altschul *et al.,* 1997). This approach was employed to deduce any probe similarities to database sequences.

3.3. RESEARCH AIMS

The aims of this chapter are to:

- Analyse the toxin encoding genes *tcdA* and *tcdB* within the genome of *C. difficile* to identify conserved regions within each nucleotide sequence of each toxin
- Determine the specificity of these conserved regions using bioinformatic analysis
- Design candidate probes from regions specific to each toxin
- Determine the ability of candidate probes to detect *tcdA* and *tcdB* targets
- Determine the sensitivity of these candidate probes
- Determine the specificity of the candidate probes against a range of *C. difficile* isolates, isolates from related *Clostridium* strains, other bacterial isolates and against human gut metagenomic DNA extracts

3.4. MATERIALS & METHODS

3.4.1. Bioinformatic analysis of *C. difficile* sequences

Submitted nucleotide sequences for the two *C. difficile* toxin genes *tcdA* and *tcdB* were collated from the National Centre for Biotechnology Information (NCBI) genomic database "GenBank" [http://www.ncbi.nlm.nih.gov/Genbank/] [Accessed 01/10/2008]. In total the nucleotide sequences from 20 toxin A entries and 18 toxin B entries, obtained from unrelated viable strains of *C. difficile* were accessed. One further toxin B sequence was obtained from Stabler *et al.* 2008.

3.4.2. Conserved regions in *C. difficile* toxins A and B

Conserved regions were pinpointed as described in Section 1.5.2 (Rupnik *et al.,* 1999).

3.4.3. Multiple Sequence Alignments of *C. difficile* sequences

To confirm that the predicted conserved regions within toxins A and B contained regions of conserved nucleotide sequences, the multiple sequences deposited in the Genebank database were analysed using ClustalW Multiple Sequence Alignment (MSA) program provided by European Bioinformatics Institute (Larkin *et al.,* 2007). The results of this analysis were displayed using Bioedit *v*7.1.3 (Hall, 2001) MSA viewer in which bases are colour coded to highlight conserved regions.

3.4.4. *C. difficile* toxin A and B Probe Design

Probes were designed to recognise nucleotide sequences within conserved regions of toxins A and B. As part of the design process we incorporated features which would enable us to utilise the probes in a future MAMEF assay (as described in Section 1.7). The anchor probe for the *C. difficile* assay was designed to be 17 nucleotides in length and to be separated from the 22 nucleotide fluorescent detector probe by a stretch of 5 nucleotides. The anchor probe binds target DNA while the detector probe subsequently binds at a distance which positions the fluorophore optimally for biomolecular recognition to occur. The use of an anchor and detector probes also allows for two levels of sensitivity within the assay for each toxin.

3.4.5. *In silico* characterisation of probe properties

3.4.5.1. Melting point and GC content analysis

For PCR based annealing reactions to take place optimally, designed probes should ideally have a GC content of approximately 50% so the melting temperature can be adequate for PCR annealing (Primrose & Twyman, 2007). To achieve this aim we employed the following bioinformatic programs:

i) Oligoanalyser 3.1 program (Integrated DNA Technologies, 2009)
[http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx]
[Accessed 01/2009]

ii) Primer 3 [http://frodo.wi.mit.edu/primer3/](Rozen & Skaletsky, 2000)
[Accessed 01/2009]

3.4.5.2. Probe secondary structure analysis

Furthermore, the potential for secondary structures and primer dimers was assessed through manual analysis of the probe sequences. The formation of primer dimers and secondary structures within a PCR reaction are the result of non-specificity, where the probes can bind to themselves or each other, resulting in poor amplification and the generation of PCR products which are not the correct amplicon size. Sequences were also analysed for long repeats of nucleotides, as repeats of >3/4 nucleotides can lead to formations of secondary structures and non-specificity (Primrose & Twyman, 2007). Analysis was performed using the Oligoanalyser 3.1 program (Integrated DNA Technologies, 2009).

3.4.5.3. Probe sequence homology search using BLAST

We employed the Basic Local Alignment Search Tool nucleotide homology search facility: BLASTn: BLAST (NCBI) [http://blast.ncbi.nlm.nih.gov/Blast.cgi] to determine if probes had homology to sequences within the Genbank database.

3.4.6. *C. difficile* toxin A and B- specific Probe Studies

3.4.6.1. Bacterial species, growth conditions, and metagenomic DNA

***C. difficile* isolates**

To enable the designed probes to be tested against a representative collection of *C. difficile* isolates, we supplemented the collection of isolates described in Chapter 2 with additional toxinotypes of *C. difficile* obtained from the National Anaerobic Reference Unit, Cardiff,

Wales, courtesy of Dr. Val Hall. These strains are shown in Table 3.1. These extra isolates also included clinical isolates from blood culture and variant isolates which only produced toxin B (Ribotypes 017, 047, 110). A further three toxinotypes were tested courtesy of Dr. Katie Solomon, University College Dublin, Ireland. Unless otherwise stated all organisms were stored as spores at 4^{o}C as described in Chapter 2. In total 58 *C. difficile* isolates were tested.

The purity of *C. difficile* strains was confirmed by using *Clostridium difficile* Moxalactam Norfloxacin (CDMN) antibiotic selective supplement (Oxoid, Basingstoke, Hampshire, UK). Contents of 1 vial of Oxoid CDMN supplement was reconstituted with 2 ml sdw and added to 500 ml molten (50^{o}C) *C. difficile* agar base (Oxoid, Basingstoke, Hampshire, UK), together with 7% (v/v) defibrinated horse blood (SR0050) (Sigma Aldrich, Dorset, UK). This was mixed well and poured into sterile Petri dishes, left to dry and degassed.

Table 3.1: Table of isolates used in this study

The additional isolates of *C. difficile* are listed. Isolates from blood culture are listed as (B/C). Toxin production for each strain and its PCR ribotype are shown, with additional information including the source. The isolates and information above was provided courtesy of Dr. Jon Brazier and Dr. Val Hall at the Anaerobic Reference Unit, University Hospital Wales, Cardiff, UK, 2008.

C. difficile strain	Year referred	Source	Ribotype	Toxin Production
Representatives of toxinotype VIII				
R9557	1996	Faeces	017	$tcdA^-$ $tcdB^+$
R13695	2000	Faeces	017	$tcdA^-$ $tcdB^+$
R18091	2003	Faeces	017	$tcdA^-$ $tcdB^+$
R10542	1997	Faeces	047	$tcdA^-$ $tcdB^+$
R18045	2003	Faeces	047	$tcdA^-$ $tcdB^+$
R7771	1994	Faeces	110	$tcdA^-$ $tcdB^+$
R17978	2003	Faeces	110	$tcdA^-$ $tcdB^+$
Representatives of toxinotype XII				
R24498	2007	Faeces	056	$tcdA^+$ $tcdB^+$
DS2008/08	2008	Faeces	056	$tcdA^+$ $tcdB^+$
R26796	2008	Faeces	056	$tcdA^+$ $tcdB^+$
Blood culture isolates				
R2139	1985	B/C	017	$tcdA^-$ $tcdB^+$
R9399	1996	B/C	001	$tcdA^+$ $tcdB^+$
R12824	1999	B/C	001	$tcdA^+$ $tcdB^+$
R13400	1999	B/C	014	$tcdA^+$ $tcdB^+$
R15552	2001	B/C	023	$tcdA^+$ $tcdB^+$
R17752	2002	B/C	001	$tcdA^+$ $tcdB^+$
R19058	2003	B/C	078	$tcdA^+$ $tcdB^+$
R19168	2004	B/C	046	$tcdA^+$ $tcdB^+$
R19222	2004	B/C	017	$tcdA^-$ $tcdB^+$
R20408	2004	B/C	045	$tcdA^+$ $tcdB^+$
R22537	2006	B/C	014	$tcdA^+$ $tcdB^+$
R24626	2007	B/C	027	$tcdA^+$ $tcdB^+$
R25028	2007	B/C	005	$tcdA^+$ $tcdB^+$
R25577	2008	B/C	002	$tcdA^+$ $tcdB^+$
R26390	2008	B/C	027	$tcdA^+$ $tcdB^+$
R27038	2008	B/C	005	$tcdA^+$ $tcdB^+$
R27039	2008	B/C	002	$tcdA^+$ $tcdB^+$
R28614	2009	B/C	106	$tcdA^+$ $tcdB^+$
R30061	2010	B/C	014	$tcdA^+$ $tcdB^+$
R30359	2010	B/C	333	$tcdA^+$ $tcdB^+$
Other Strains				
R22680	-	Faeces	017	$tcdA^-$ $tcdB^+$
12727	-	Faeces	001	$tcdA^+$ $tcdB^+$
11204	-	Faeces	001	$tcdA^+$ $tcdB^+$
Isolates with designated toxinotype	Toxinotype			
VPI 10463	0			
Toxinotype XIa	XIa			$tcdA^-$ $tcdB^-$
Toxinotype XIb	XIb			$tcdA^-$ $tcdB^-$
Toxinotype IX	IX			$tcdA^+$ $tcdB^+$

3.4.6.2. Bacterial species unrelated to and related to *C. difficile*

In addition to increasing the range of *C. difficile* isolates tested, the ability of the probes to react with other bacterial species was assessed. Thus to establish the specificity of the designed probes we tested them against a range of bacterial species from closely related *Clostridium* strains, to unrelated strains, as listed in Table 3.2.

Table 3.2: Additional bacterial species used in this study

The additional species and their strain designations used in this study are listed above. The species related to, and those not related to, *C. difficile* are shown. The species were obtained from the NCTC (National Culture Type Collection, HPA, London, UK), unless otherwise stated. Other isolates obtained from the anaerobic reference unit (ARU) at the University Hospital Wales (UHW) are listed as ARU, UHW. Those isolates and information was provided courtesy of Dr. Jon Brazier and Dr. Val Hall at the Anaerobic Reference Unit, University Hospital Wales, Cardiff, UK.

Bacterial species	Strain	Source
Unrelated Bacterial Species		
Methicillin-resistant *Staphylococcus aureus*	NCIMB 9518	NCTC
Staphylococcus aureus		NCTC
Escherichia coli	K12	NCTC
Bacillus subtilis	6051	NCTC
Related *Clostridium* Species		
C. sordellii	R20453	UHW
C. sordellii	13356	UHW
C. novyi	R14479	UHW
C. novyi	277	UHW
C. septicum	R22030	UHW
C. perfringens	13170	NCTC
C. perfringens type C	3180	NCTC
C. coccoides	11035	NCTC
C. leptum	753	DSM
C. innocuum	1280	DSM

3.4.6.3. Human Metagenomic gut DNA

The human gut environment contains approximately 10^{12}/g bacteria. Therefore the probes designed to detect *C. difficile* must be able to specifically detect the toxin genes amongst the numerous bacteria present (Shoemaker *et al*., 2001). Therefore to further confirm the specificity of the designed probes, metagenomic DNA samples from ten human volunteers were obtained from Cardiff School of Biosciences, Wales, Cardiff, UK, courtesy of Dr. Julian Marchesi. The metagenome was extracted from faecal matter (Dr. J. Marchesi-*personal communication*) from humans in Zambia, France and the UK.

3.4.6.4. DNA Extraction from *C. difficile* using Chelex 100 Resin

(Courtesy of Dr. J. Brazier; *Personal communication*)

Chelex 100 Resin (BioRad Laboratories, UK) was dispensed into 2 ml sterile distilled water and vortex mixed well using a VortexGenie (Fisher Scientific, UK). Approximately 100 µl Chelex 100 was added to an Eppendorf tube whilst gently agitated. A 10 µl loopful of *C. difficile* colonies from an overnight (24 h) agar plate culture on CDMN was taken and mixed into the Chelex 100. Subsequently the Eppendorfs were placed on a hot plate at 100^{o}C for 12 min, and then centrifuged for 10 min at 15,000 rpm. The supernatant was aliquoted out and stored as DNA extract at -20^{o}C. Concentration of DNA measured using Biophotometer (Eppendorf, UK) DNA absorbance function calculating absorbance using the relationship that A_{260} of 1.00 = 50 µg/ml pure DNA. DNA purity was calculated at 260 nm/280 nm.

3.4.6.5. Probe Synthesis

Probes were synthesised by Eurofins MWG Operon, Ebersberg, Germany, at HPSF purification, synthesis scale 0.05 µl.

3.4.7. PCR for *C. difficile*

3.4.7.1. Preparation of 1x Tris Boric EDTA Electrophoresis buffer

A 10x TBE buffer stock containing 27.5 g boric acid (Sigma Aldrich, UK), 54 g Tris base and 20 ml 0.5M EDTA was prepared. To make a 1x solution of TBE, 100 ml 10x TBE buffer was added to 900 ml double diH_2O water in a 1L Duran bottle.

3.4.7.2. Preparation of 1% (w/v) Agarose

In a 500 ml Duran bottle 200 ml 1x TBE buffer was added to 2 g molecular biology grade agarose (Sigma Aldrich, UK) and microwaved until all the powder dissolved. The agarose was left to cool and poured into a gel cast mould with comb (Biorad Laboratories, UK).

3.4.7.3. PCR to detect Toxin B within *C. difficile* isolates

As a control a primer known to detect toxin B within *C. difficile* isolates was employed (Wren *et al.,* 1993). Only toxin B was detected via PCR as all *C. difficile* strains contain a stable toxin B gene.

3.4.7.4. PCR thermocycle optimisation conditions for designed probes

To generate probes for use in a dot blot macroarray the hypothesised conserved regions of nucleotide sequences were amplified using PCR. The PCR thermocycle conditions for each of the probes were optimised to obtain definitive amplicons bands of the correct size. When optimising PCR conditions the annealing temperature was assessed by performing a temperature gradient (+/-10^{o}C) thermocycle calculated by the thermocycler (Techne Touch gene gradient thermocycler). PCRs were performed using the Taq PCR core kit (Qiagen, Qiagen House, Crawley, West Sussex, UK). For the optimisation studies DNA extracted from *C. difficile* strain R20291 was used in each 20 µl reaction. Each 20 µl reaction contained: 9.9 µl sdw, 2.5µl PCR buffer, 5 µl Q solution, 0.4 µl of 25 mM $MgCl_2$, 2 µl of 10 mM deoxyribonucleotides (dNTPs), 1 µl primer mix, 1 µl template DNA and 0.2 µl of 5 Units/ µl *Taq* polymerase. The PCR thermocycle programme employed was as follows: 94^{o}C for 5 min and 30 cycles of 94^{o}C (30 s), annealing gradients variable according to probe (Table 3.3) and 72^{o}C (45 s). PCR products were gel electrophoresed in 1% (w/v) agarose (Sigma Aldrich, UK) at 85 V, stained with SafeView^{TM} and photographed in a UV-transilluminator (ChemiDoc, BioRad laboratories, UK).

Table 3.3: PCR Thermocycle annealing temperatures per probe

The probes for each toxin were found to have the above optimal temperatures for PCR to occur. These are the temperature at which all further PCR reactions and dot blot reactions will be conducted.

Clostridium difficile Toxin	Anchor Probe F Primer 5' → 3' (17 nt)	Optimal Temperatures for PCR	Detector Probe F Primer 5' → 3' (22 nt)	Optimal Temperatures for PCR
Toxin A Probe from 50 bp Region	TTTGAATACTTTGCACC	55.6	TGCTAATACGGATGCTAACAAC	66.9
Toxin A Probe 2 from 76 bp	TTTAATACTAACACTGC	37.3	TGTTGCAGTTACTGGATGGCAA	66.9
Toxin B Probe	TCAAGACTCTATTATAG	48.4	TAAGTGCAAATCAATATGAAGT	60.1

3.4.7.5. Staining of PCR products in agarose gel using SafeView™

To visualise amplified DNA bands in a gel we used SafeView™ (NBS Biologicals, Cambridgeshire, UK). SafeView™ is as sensitive as ethidium bromide but does not suffer from concerns over its potential toxicity. Green fluorescence is emitted when it binds to ssDNA, dsDNA and RNA and has a fluorescence excitation maximum at approximately 290-320 nm, emitting at 490 nm). The stain was added to molten agarose before setting, and to 1 x TBE buffer in a gel electrophoresis tank, at the specifications listed below:

Table 3.4: Amount of Safeview™ to use in an agarose gel

The volume of Safeview™ required for visualisation in an agarose gel is shown above. The volume of Safeview™ required to be added to 1 x TBE buffer in a gel electrophoresis tank (Biorad laboratories, UK) is shown. Information is taken from NBS Biologicals, Cambridgeshire, UK).

Size of Gel Cast (cm)	Agarose (ml)	SafeView (µl)
10 x 15	120	2.5
10 x 6	50	1.75
Tank Size	TBE Buffer in tank (ml)	SafeView (µl)
Small	250	10
Medium	800	30

3.4.8. Genomic DNA hybridisation dot blots

3.4.8.1. Preparation of DIG-labelled Hybridisation Probes

After PCR optimisation the product was labelled with digoxigenin (DIG) to enable detection in a larger macroarray via dot blot hybridisation (Figure 3.2). DIG is a non-radioactive molecule with high immunogenicity which is a used as an alternative way of labelling oligonucleotides. The probes were prepared for dot blot hybridisation by performing a PCR reaction by replacing standard dNTPs with DIG-labelled dNTPs in a chemiluminescence-based method (Roche Diagnostics, Charles Avenue, West Sussex, UK). Further advantages of the DIG system is its high sensitivity, short exposure time, safety, and the probes are reusable (Roche Diagnostics: www.roche-applied-science.com/sis/lad/index.jsp?id=lad_040100).

To label the probes a PCR labelling reaction was performed using the DIG-labelled dNTPs. Per 20 µl reaction was: 9.9 µl sdw, 2.5µl PCR buffer, 5 µl Q solution, 0.4 µl of 25

mM $MgCl_2$, 2 µl of 10 mM DIG-labelled dNTPs, 1 µl primer mix, 1 µl template DNA and 0.2 µl of 5 units/ µl *Taq* polymerase. PCR conditions were as in Section 3.4.7.4 above, with annealing temperatures specified in Table 3.3. The template DNA used in the PCR reaction was strain DS1813. The concentration of PCR product was measured using a Biophotometer (Eppendorf, UK) and stored at -20^{o}C, undiluted (50-100 ng/µl).

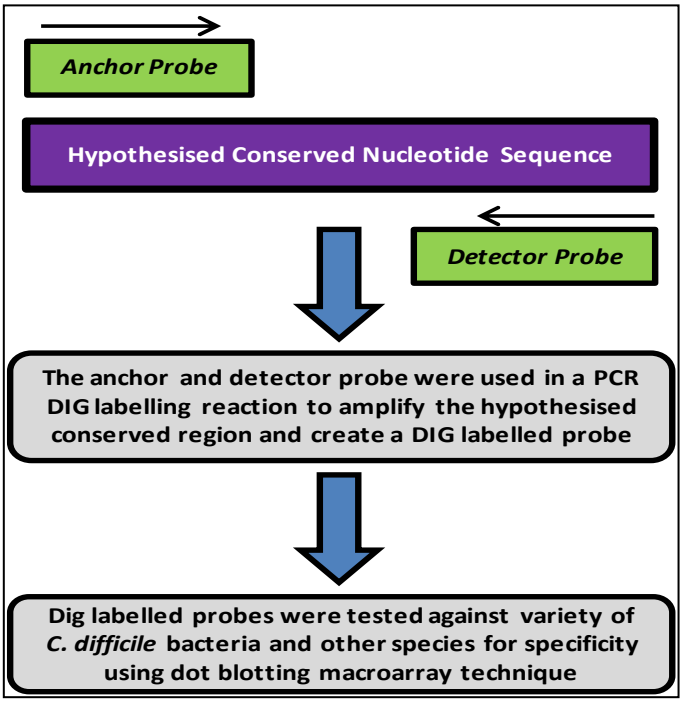


Figure 3.2: Probe generation

Schematic showing the method used to generate DIG labelled probes for use in a macroarray dot blot hybridisation reaction. These DIG labelled probes were screened for specificity against a range of C. difficile isolates, and bacteria from other species. This can be seen later in section 3.5.6.2.

3.4.8.2. Sensitivity of DIG-labelled probes

To establish the sensitivity each DIG-labelled probe, 2 µl of probe was serially diluted in 20 µl sdw, and 2 µl of each dilution sequentially dot blotted onto a positively charged nylon membrane (Roche Diagnostics, Charles Avenue, West Sussex, UK) and fixed with UV light (ChemiDoc, BioRad laboratories, UK) to ascertain the optimum probe concentration for further DNA hybridisation dot blotting experiments.

3.4.8.3. Macro-arraying genomic DNA onto a positively charged nylon membrane

To allow all of the *C. difficile* isolates to be examined easily via dot blot, a Macro-arraying technique was employed to hybridise gDNA samples onto the positively charged nylon membrane. The gDNA extracted from the *C. difficile* isolates (using Chelex 100), was aliquoted at 30 µl per well into a 384 well plate (ABGene, Fisher Scientific, UK) and printed into the positively charged membrane using a Flexys robotic workstation (Genomic Solutions Ltd, UK). The robot was set to blot 4 spots of DNA in a grid format (Figure 3.3). To orientate the membrane a single well was included, containing loading dye (Bromophenol blue, Sigma Aldrich, UK) and 50 ng/ µl lambda phage DNA (Promega, UK). After macroarraying the gDNA was fixed to the nylon membrane using a UV transilluminator (ChemiDoc, BioRad laboratories, UK) for 5 min. The fixed membranes were stored for future use between two sheets of blotting paper at room temperature.

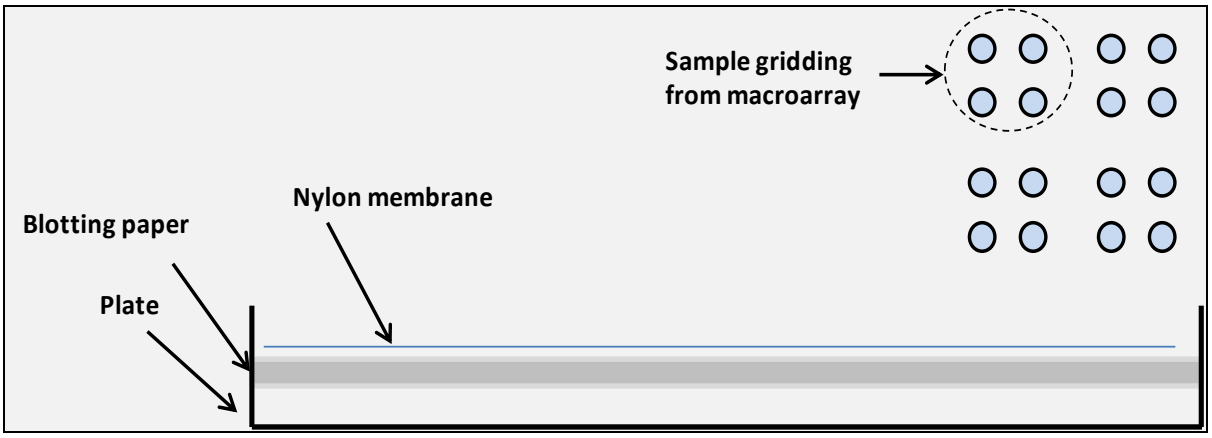


Figure 3.3: Macroarraying experiment

The gDNA from a range *C. difficile* isolates was macroarrayed onto a positively charge nylon membrane using a robotic machine- FLEXYS (Genomic Solutions Ltd, UK). gDNA was macroarrayed in the sample gridding format shown above to give four dots of DNA; thus four replicates for each isolate. After macroarraying the gDNA was fixed onto the membrane using a UV transilluminator (ChemiDoc, BioRad laboratories, UK) and stored on blotting paper in a plate for future use.

3.4.8.4. Solutions for DNA hybridisation dot blots

All solutions used in the dots blots were made according to instructions from Roche. The methods and solutions are specific for use with the DIG detection system (Roche Diagnostics, Charles Avenue, West Sussex, UK). These are listed in detail in the Appendix.

3.4.8.5. DNA hybridisation dot blots

3.4.8.5.1. Pre-hybridisation of membrane

To prepare the membrane for hybridisation the membrane was calibrated in a pre-hybridisation step. The hybridisation tubes were washed thoroughly and the hybridisation oven (Fisher Scientific, UK) warmed to 50^{o}C beforehand. This was checked using a thermometer (Fisher Scientific, UK). For pre-hybridisation, 20 ml of DIG Easy Hyb buffer (Roche Diagnostics, Charles Avenue, West Sussex, UK) was put in each hybridisation tube and left in the hybridisation oven until they reached 50^{o}C. Subsequently the membrane was put into the hybridisation tube with DNA side facing the tube interior and a further 20 ml heated DIG Easy Hyb buffer was added to give 40 ml in each tube.

3.4.8.5.2. Hybridisation of the probes to the membrane

For hybridisation of the DIG labelled probes to any DNA on the membranes, another hybridisation step is undertaken. Each hybridisation experiment was repeated three times. Approximately 5-10 µl of the DIG-labelled PCR product probe (this is dependent on the probe concentration, as described by Roche Diagnostics) to 500 µl of DIG Easy Hyb buffer (Roche Diagnostics, Charles Avenue, West Sussex, UK) and boiled in a beaker on a hot plate (Fisher Scientific, UK) for 10 min (98^{o}C) followed by chilling immediately on ice. The boiled probe was added to the pre-heated hybridisation tube after Prehybridisation stage and subsequently the tubes were rotated in the hybridisation oven (Fisher Scientific, UK) overnight at 50^{o}C.

3.4.8.5.3. Stringency Washes

The stringency washes are employed to remove any unbound probe and is carried out at approximately 10^{o}C above the hybridisation temperature (this is equivalent to the PCR annealing temperatures as described in Table 3.3). In practise the temperature should be 60-68^{o}C depending on the stringency required. The hybridisation oven was set at the appropriate stringency temperature specific for the probe used and the stringency solutions (SSC solutions) were heated manually using a microwave (GE microwave Model No. JE2160BF01, kW 1.65 (M/W) to 60^{o}C. The hybridisation buffer was then discarded from the tube and the stringency washes conducted. The membrane was washed in 40 ml 2x SSC stringency solution twice for 15 min, and then in 40 ml 0.5x SSC stringency solution twice for 25 min. For high stringency the membrane was washed in 40 ml 0.1x SSC stringency solution twice for 15 min.

3.4.8.5.4. Detection of bound probe using DIG chemiluminescence

The bound probes were detected on the membrane using an anti-DIG alkaline phosphatase method. The membrane was equilibrated in washing buffer for 2 min then agitated in 1% blocking solution for 45 min. The membrane was then agitated for 30 min in a 1:10,000 dilution of anti-DIG alkaline phosphatase antibody (Roche Diagnostics, Charles Avenue, West Sussex, UK) in 1% blocking solution. To make the 1:10000 dilution of anti- DIG antibody, 5 µl of antibody was diluted into 50 ml of 1 % blocking solution. The antibody was centrifuged prior to dilution in 1% blocking solution.

The membrane was then washed twice for 15 min in washing buffer to remove any unbound antibody from the membrane, and equilibrated in detection buffer for 2 min. In an eppendorf 10 µl CSPD (Roche Diagnostics, Charles Avenue, West Sussex, UK) was added to 1000 µl detection buffer. The membrane was placed DNA side up onto a clean plastic bag and the 1000 µl CSPD pipetted over the membrane. The plastic bag was folded so there were no bubbles and to ensure the membrane was entirely covered with CSPD. The membrane was incubated at 37^{o}C for 15 min to this makes the light-producing reaction reach equilibrium. The blots were exposed overnight.

3.4.8.5.5. Chemiluminescent Camera Method for detection:

To visualise the blot a chemiluminescent camera was used, courtesy of Dr. Claire Hamilton and Dr. Kelly Berube in Cardiff School of Biosciences. The chemiluminescent camera (Biospectrum® Imaging system, UVP, UK) is sensitive enough to detect any light emission from the membrane. The image was visualized using the VisionWorks® LS analysis software (UVP, UK).

3.5. RESULTS

3.5.1. Bioinformatic analysis of *C. difficile* sequences

Conserved nucleotide regions within the sequences of toxins A and B were identified by employing Rupnik's toxin typing method (1998). Conserved regions from different clinical isolates (20 from *tcdA* and 19 from *tcdB*) of *C. difficile* were collated from the Genbank database and imported into ClustalW (European Bioinformatics Institute) to identify common areas from which to design probes using the Jalview™ MSA viewer (Figure 3.4). Two conserved regions within toxin A and four conserved regions within toxin B were identified (Table 3.5). These results confirmed that the regions defined by Rupnik *et al.* (1998) were indeed conserved within both *C. difficile* toxins, and these conserved regions were suitable for probe design for the MAMEF- based assay.

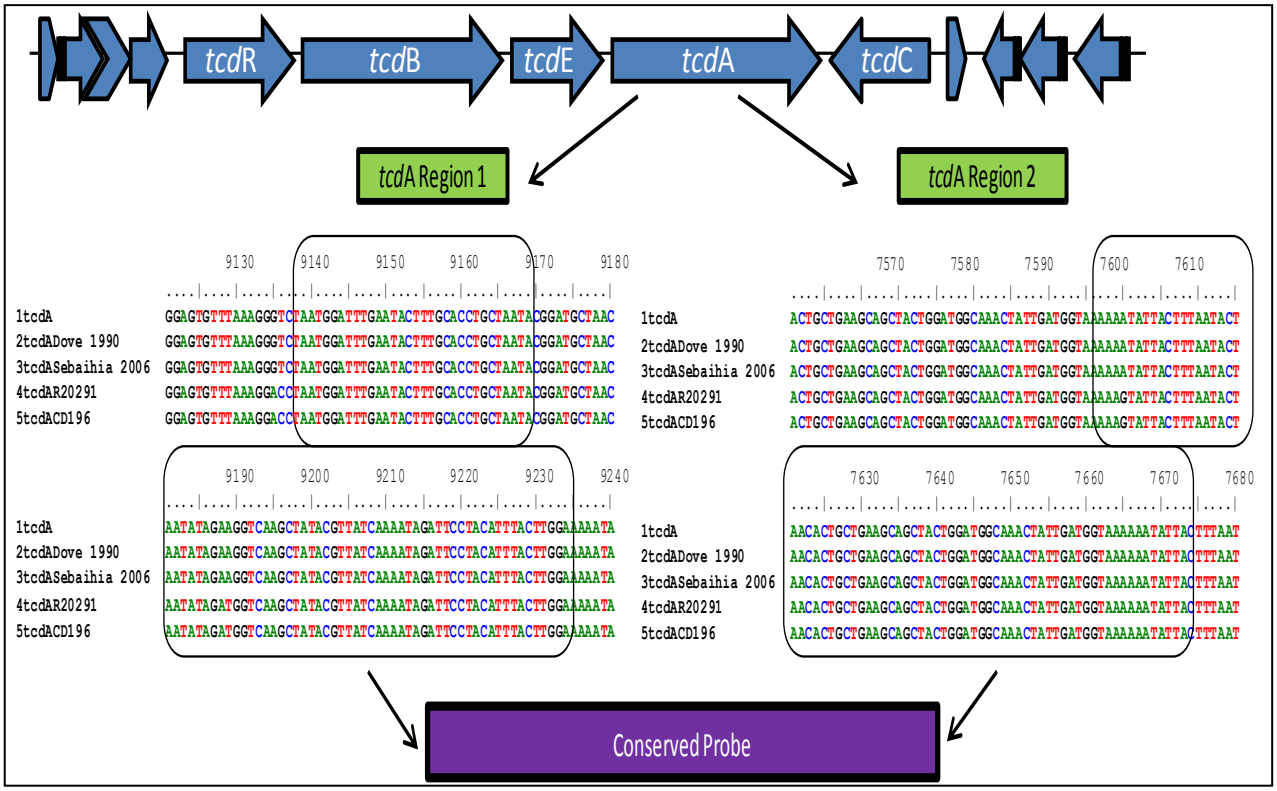


Figure 3.4: Schematic diagram showing deduction of conserved regions

Sequences of toxins A and B respectively obtained from the Genbank database were assessed for conserved regions using EBI ClustalW. The Bioedit *v*7.1.3 (Hall, 2001) method of viewing MSA allows conserved residues to be colour coded according to those used in the EBI program ClustalW. The guanine reside (G) is shown as **black**, cytosine (C) as **blue,** adenosine (A) as **green** and thymine (T) as **red**.

Table 3.5: Conserved regions of toxins identified via MSA

Regions within toxin A and toxin B of *C. difficile* were identified by ClustalX and the MSA analysis software Jalview™. The size of these regions were determined are listed above. Two regions in toxin A and three regions in toxin B were conserved across all the sequences collated from GenBank.

C. difficile Toxin	Region Length (bp)	Final Conserved Nucleotide Regions	References
Toxin A: 1	50	ATGGATTTGAATACTTTGCACCTGCTAATACGGATGCAA CAACATAGAA	Sebaihia *et al.*, 2006 Lemee *et al.*, 2005 Letournier *et al.*, 2003 Braun *et al.*, 2000 Sambol *et al.*, 2000 Kato *et al.*, 1998 Hundsberger *et al.*, 1997 Eichel-Streiber, 1995 Sauerborn & Eichel – Streiber, 1990 Dove *et al.*, 1990 Wren *et al.*, 1990
Toxin A: 2	76	AAAATATTACTTTAATACTAACACTGCTGTTGCAGTTACT GGATGGCAAACTATTAATGGTAAAAAATACTACTTT	
Toxin B: 1	51	TTGGCAAATAAGCTATCTTTTAACTTTAGTGATAAACAA GATGTACCTGTA	
Toxin B: 2	41	CATATTCTGGTATATTAAATTTCAATAATAAAATTTACTA T	
Toxin B: 3	52	TTTGAGGGAGAATCAATAAACTATACTGGTTGGTTAGAT TTAGATGAAAAGA	
Toxin B extra region	39	TCAAGACTCTATTATAG TAAGTGCAAATCAATATGAAGT	

3.5.2. Probe Design

Probes for the MAMEF assay were designed to recognise sequences within each conserved region (~50 nucleotides; Table 3.5) and were configured to take account of the requirements of the assay as described in Section 1.7. Thus the anchor probes were 17 nucleotides in length while the fluorescent detector probes were 22 nucleotides in length, and a 5 nucleotide gap between the anchor and fluorescent probes was incorporated to fit the MAMEF assay requirements. The potential capture and fluorescent detector probes designed from the regions can be seen in Tables 3.6 and 3.7.

Table 3.6: Conserved regions of toxins identified via MSA

Probes regions within toxin A of *C. difficile* were designed to recognise nucleotide sequences within conserved regions of toxin A, and incorporated features needed for use in a future MAMEF assay. Regions were divided at natural break within the nucleotide sequence- at GC or CC points. The entire conserved region and the capture and detector probes designed from them is shown. The colour coding of bases indicates areas of probe under design. Bases shown in **purple** indicate the capture probe under design and its corresponding detector sequence from the conserved region. Bases shown in **green** indicate the capture probe under design and its corresponding detector sequence from the conserved region. Bases shown in **blue** indicate the capture probe under design and its corresponding detector sequence from the conserved region. Bases shown in **red** indicate the capture probe under design and its corresponding detector sequence from the conserved region.

Clostridium difficile Toxin Region	Probe Regions (50 nt in total)						
	Entire Conserved Region	Capture Probe (17 nt)				Detector Probe Region (22 bp)	Remaining region after
		1	2	3	4		
1. Toxin A 50 bp Region	ATGGATTTGAATACTTTGCACCTGCTAATACGGATGCTAACAACATAGAA	GGATTTGAATACTTTGC	TTTGAATACTTTGCACC	GAATACTTTGCACCTGC		ACCTGCTAATACGGATGCTAAC	AACATA GAA
						TGCTAATACGGATGCTAACAAC	ATAGAA
						TAATACGGATGCTAACAACATA	GAA
2. Toxin A 72 bp Region	AAAATATTACTTTAATACTAACACTGCTGTTGCAGTTACTGGATGGCAAACTATTAATGGTAAAAAATACTACTTT	TACTTTAATACTAACAC	TTTAATACTAACACTGC	ACTAACACTGCTGTTGC	TGCTGTTGCAGTTACTG	TGCTGTTGCAGTTACTGGATGG	CAAACTATTAA
						TGTTGCAGTTACTGGATGGCAA	ACTATTAATGG
						ATGGCAAACTATTAATGGTAAA	AAATACACTT
						GATGGCAAACTATTAATGGTAA	AAAATACTACT

Table 3.7: Conserved regions of toxins identified via MSA

Probes regions within toxin B of *C. difficile* were designed to recognise nucleotide sequences within conserved regions of toxin B, and incorporated features needed for use in a future MAMEF assay. Regions were divided at natural break within the nucleotide sequence- at GC or CC points. The entire conserved region and the capture and detector probes designed from them is shown. The colour coding of bases indicates areas of probe under design. Bases shown in **purple** indicate the capture probe under design and its corresponding detector sequence from the conserved region. Bases shown in **green** indicate the capture probe under design and its corresponding detector sequence from the conserved region. Bases shown in **blue** indicate the capture probe under design and its corresponding detector sequence from the conserved region. Bases shown in **red** indicate the capture probe under design and its corresponding detector sequence from the conserved region.

Clostridium difficile Toxin Region	Probe Regions (50 nt in total)						
	Entire Conserved Region	Capture Probe (17 nt)				Detector Probe Region (22 bp)	Remaining DNA Region
		1	2	3	4		
1. Toxin B 51 bp Region	TTGGCAAATAAGCTATCTTTTAACTTTAGTGATAAACAAGATGTACCTGTA	TTGGCAAATAAGCTATC	ATAAGCTATCTTTTAAC	TCTTTTAACTTTAGTG		TTTTAACTTTAGTGATAAACAA TTTAGTGATAAACAAGATGTAC ATAAACAAGATGTACCTGTA	GATGTACCTGT CTGTA
2. Toxin B 41 bp Region	CATATTCTGGTATATTAAATTTCAATAATAAAATTTACTAT	CTGGTATATTAAATTTC				AATAATAAAATTTACTAT	
3. Toxin B 52 bp Region	TTTGAGGGAGAATCAATAAACTATACTGGTTGGTTAGATTTAGATGAAAAGA	AGGGAGAATCAATAAAC	GAATCAATAAACTATAC	TCAATAAACTATACTGG	TAAACTATACTGGTTGG	TATACTGGTTGGTTAGATTTAG TGGTTGGTTAGATTTAGATGAA TTGGTTAGATTTAGATGAAAAG TTAGATTTAGATGAAAAGA	ATGAAAAGA AAGA A

3.5.3. *In silico* characterisation of probe properties

The anchor and fluorescent detector probes for each toxin were analysed for GC content and secondary structures. Those with secondary structures were discarded from further use. Sequences were then imported into BLASTn and those sequences with significant hits to species other than *C. difficile* were discarded (9 *tcdA* sequences and 14 *tcdB* sequences). Those which showed homology to *C. difficile* only were selected for commercial synthesis (Table 3.8).

Table 3.8: Final probes for commercial synthesis

The probes above were commercially synthesised for further experiments. (Eurofins MWG Operon, Ebersberg, Germany).

Clostridium difficile Toxin	Anchor (Capture) Probe F Primer 5' → 3' (17 nt)	Detector Probe F Primer 5' → 3' (22 nt)
Toxin A Probe from 50 bp Region	TTTGAATACTTTGCACC	TGCTAATACGGATGCTAACAAC
Toxin A Probe from 76 bp	TTTAATACTAACACTGC	TGTTGCAGTTACTGGATGGCAA
Toxin B Probe	TCAAGACTCTATTATAG	TAAGTGCAAATCAATATGAAGT

3.5.5. PCR Thermocycle Optimisation

The annealing temperatures of the synthetically designed probes were optimised for *tcdA* and *tcdB* binding and to enable the PCR DIG-labelling of the probes for subsequent use in genomic hybridisation macroarraying experiments. The anchor and fluorescent detector probes were all tested across a temperature gradient (+/- 10^{o}C) corresponding to the annealing temperatures shown in Table 3.3. The PCR reactions revealed that the annealing temperatures recommended by Eurofins MWG Operon (Ebersberg, Germany) were optimal for each reaction (Figure 3.5; 3.6; 3.7).

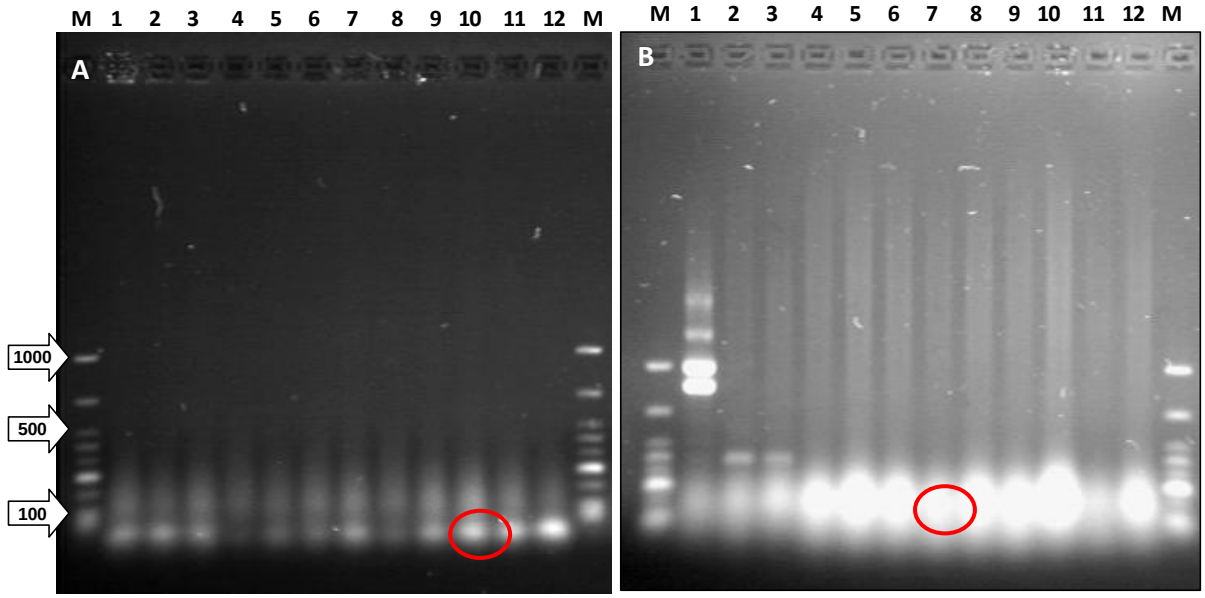


Figure 3.5: Anchor and detector gradient PCRs for *tcdA* 50bp probe

(A) This gel shows the optimal temperature of PCR annealing over a temperature gradient of 55.6^{o}C (+/- 10^{o}C). The gel also shows the PCR reaction for the anchor probe worked generating a product. (B) This gel shows the optimal temperature of PCR annealing over a temperature gradient of 63.5^{o}C (+/- 10^{o}C). The gel also shows the PCR reaction for the anchor probe worked generating a product. The optimal annealing temperatures for the anchor and fluorescent detector probes were determined as 55.6^{o}C and 66.9^{o}C. Marker (M) used is 1000 bp. PCR products at the optimal annealing temperatures are circled in **red**.

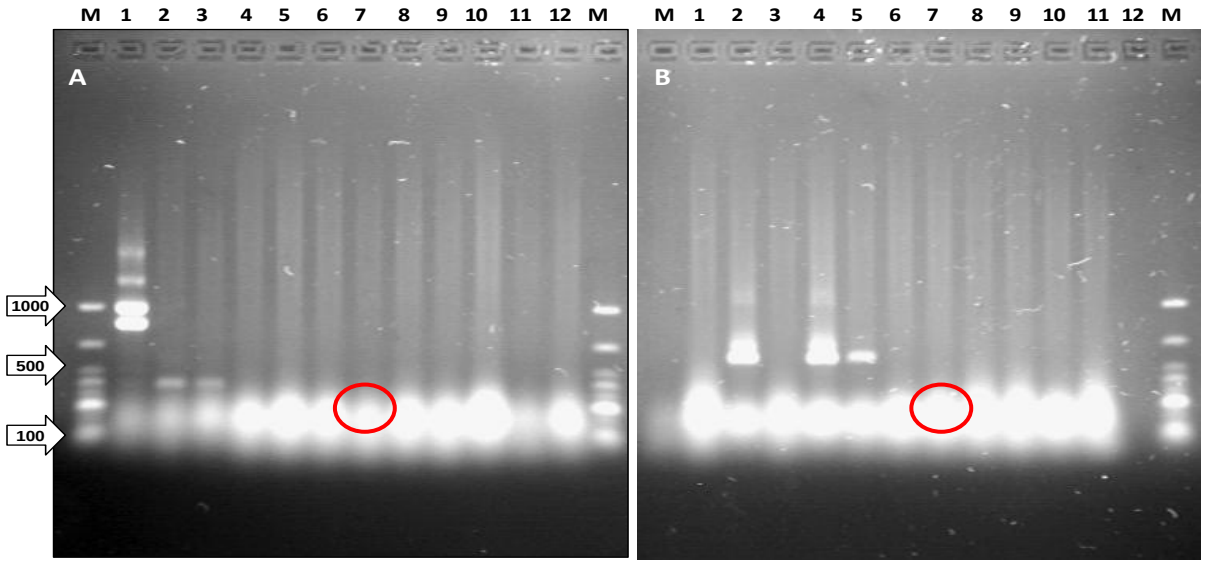


Figure 3.6: Anchor and detector gradient PCRs for *tcdA* 76 bp probe

(A) This gel shows the optimal temperature of PCR annealing over a temperature gradient of 37.3^{o}C (+/- 10^{o}C). The gel also shows the PCR reaction for the anchor probe worked generating a product. (B) This gel shows the optimal temperature of PCR annealing over a temperature gradient of 63.5^{o}C (+/- 10^{o}C). The gel also shows the PCR reaction for the anchor probe worked generating a product. The optimal annealing temperatures for the anchor and fluorescent detector probes were determined as 37.3^{o}C and 66.9^{o}C. Marker (M) used is 1000 bp. The amplicons produced from PCR at the optimal annealing temperatures are circled in **red**.

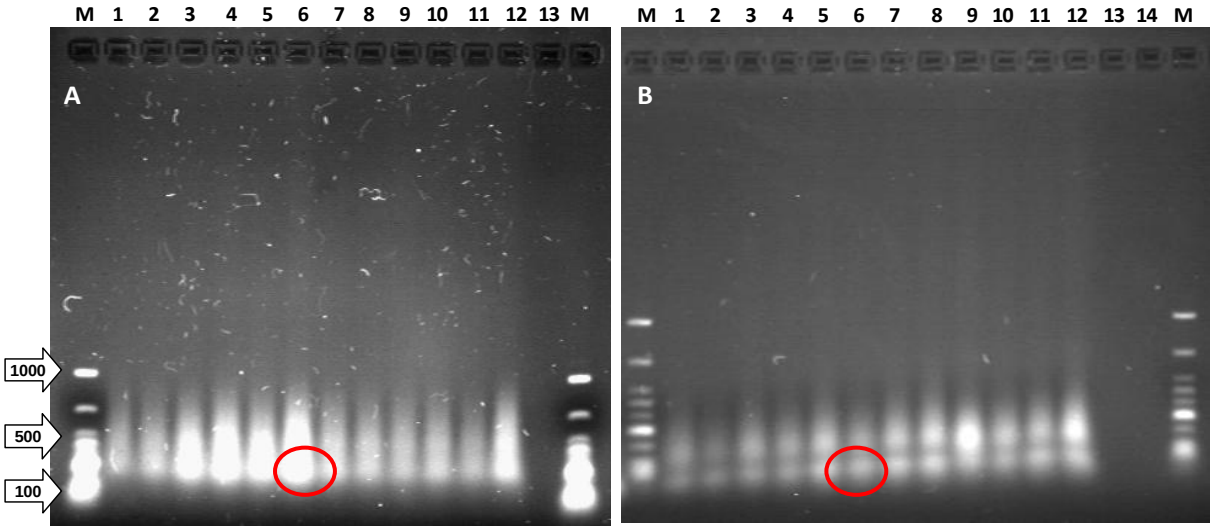


Figure 3.7: Anchor and detector gradient PCRs for *tcdB*

(A) This gel shows the optimal temperature of PCR annealing over a temperature gradient of 55.6^{o}C (+/- 10^{o}C). The gel also shows the PCR reaction for the anchor probe worked generating a product. (B) This gel shows the optimal temperature of PCR annealing over a temperature gradient of 63.5^{o}C (+/- 10^{o}C). The gel also shows the PCR reaction for the anchor probe worked generating a product. The optimal annealing temperatures for the anchor and fluorescent detector probes were determined as 55.6^{o}C and 66.9^{o}C. Marker (M) used is 1000 bp. PCR products at the optimal annealing temperatures are circled in **red**.

3.5.6. DNA Hybridisation dot blots

DNA hybridisation was used as a rapid method to screen all isolates in our collection against the designed probes.

3.5.6.1. Sensitivity of DIG-labelled probes

The sensitivities of each DIG-labelled DNA probe were assessed to determine the optimal concentration for use in further dot blot experiments. The higher the concentration of probe the more vivid the blot once hybridisation has occurred. The probe with the highest sensitivity and strongest chemiluminescent signal was produced by the *tcdA* 76 bp Anchor probe, which gave a signal at a dilution of 10 ng/ µl (Figure 3.8). This is important for capturing target DNA within a future MAMEF based assay. On the basis of the sensitivity results (Figure 3.8) the concentration of probe used for subsequent macroarray screening studies was 60 ng/ µl.

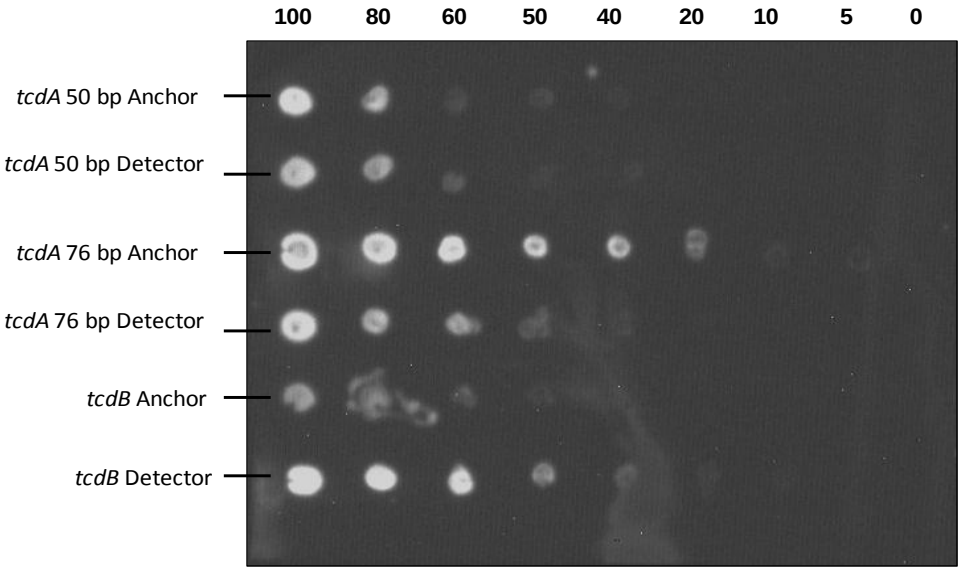


Figure 3.8: Sensitivities of DIG-labelled Probes

The probes (Table 3.7) were tested for sensitivity by performing a serial dilution of each probe (0-100 ng/µl). The probe concentration which showed the strongest signal was used for further macroarraying experiments. On the basis of the sensitivity results (Figure 3.7) the concentration of probe used for subsequent macroarray screening studies was 60 ng/ µl.

3.5.6.2. Macro-arraying genomic DNA onto a positively charged nylon membrane of *C. difficile isolates*

Genomic DNA from the *C. difficile* isolates tested were macroarrayed as shown in Table 3.9. As expected each strain containing a copy of the toxin A and B gene sequences gave a positive signal, the strength of which varied between isolates, likely due to an artefact of experimental procedure. The results do not reflect the toxin A and B gene copy numbers within the *C. difficile* genome. Indeed Wren *et al.* (1990) postulated that there was a single copy of toxin A within the *C. difficile* genome.

To confirm the specificity of the probes variant isolates of *C. difficile* (*tcdA*$^{-}$*tcdB*$^{+}$) lacking either the toxin A (ribotypes 017; 047: *tcdA*$^{-}$*tcdB*$^{+}$) or toxin B (Toxinotype XIa; XIb; DS1684: *tcdA*$^{-}$*tcdB*$^{-}$) gene sequences were included in the panel. The DNA from these isolates did not bind to the probes.

Table 3.9: gDNA from *C. difficile* isolates macroarrayed onto the nylon membrane.

Bromo-phenol blue + Lambda phage DNA	DS1759 (001)	DS1747 (001)	R8652 (001)	DS1750 (001)	12727 (001)	11204 (001)	R12824 (001)	R9399 (001)	R17752 (001)
DS1813 (027)	DS1801 (027)	R20291 (027)	DS1807 (027)	R26390 (027)	R24626 (027)	R10459 (106)	DS1798 (106)	DS1787 (106)	DS1771 (106)
R28614 (106)	VPI 10463	IX	DS1752 (012)	CD630 (012)	DS1723 (078)	R19058 (078)	R24498 (056)	DS2008 (056)	R26796 (056)
R27039 (002)	R25577 (002)	DS1748 (002)	R27038 (005)	DS1721 (005)	R25028 (005)	DS1742 (014)	R13400 (014)	R30061 (014)	R22537 (014)
R15552 (023)	DS1665 (023)	R30359 (333)	DS1724 (020)	R19168 (046)	R20408 (045)	R18091 (017)	R10542 (047)	R18045 (047)	R7771 (110)
R17978 (110)	R22680 (017)	R2139 (017)	R19222 (017)	R9557 (017)	R13695 (017)	XIa	XIb	DS1684 (010)	-

gDNA from the collection of *C. difficile* isolates tested was arrayed onto the positively charged membrane in the format shown above. Isolates were organised according to PCR ribotype. Toxin production per ribotype is shown in Table 3.1.

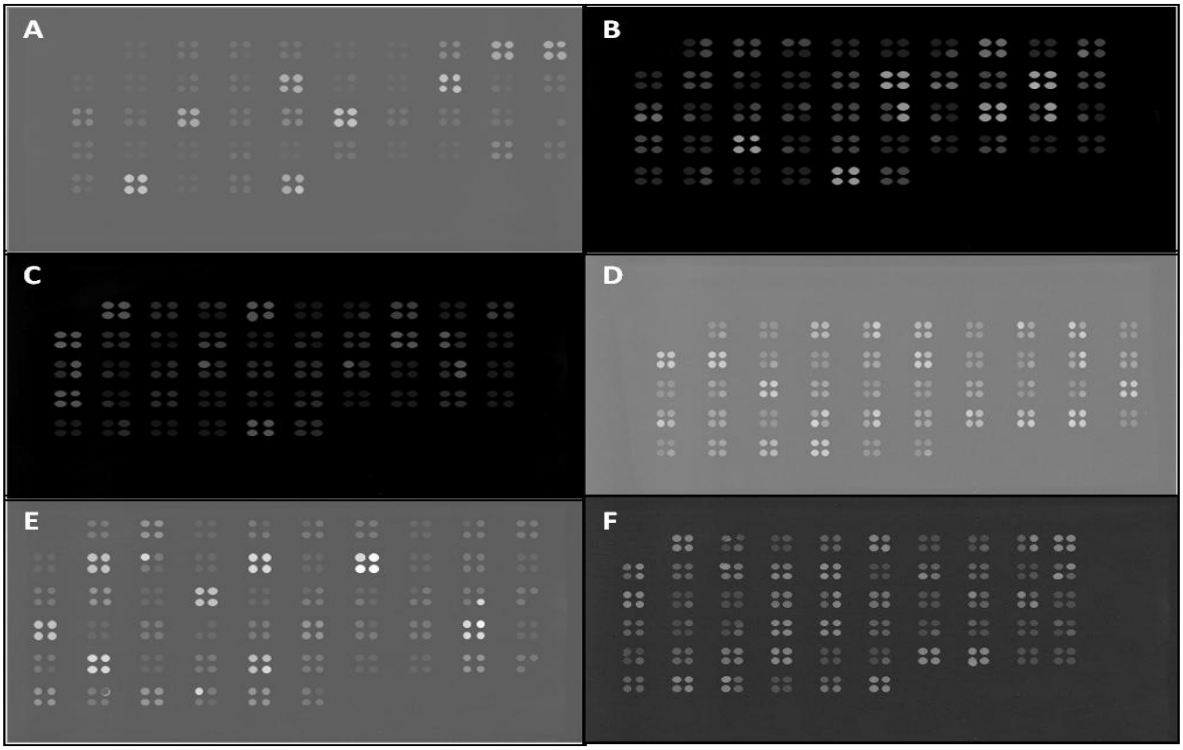


Figure 3.9: gDNA Macroarrays of *C. difficile* isolates

The macroarrays were performed against *C. difficile* isolates in our collection (58 isolates) using each probe as listed in Table 3.7. (A) *tcdA*50 anchor probe vs. *C. difficile* isolates, (B) *tcdA*50 detector probe vs. *C. difficile* isolates, (C) *tcdA*76 anchor probe vs. *C. difficile* isolates, (D) *tcdA*76 detector probe vs. *C. difficile* isolates, (E) *tcdB* anchor probe vs. *C. difficile* isolates, (F) *tcdB* detector probe vs. *C. difficile* isolates.

3.5.6.3. Dot blot of species unrelated to and closely related to *C. difficile*

To further confirm the specificity of the probes genomic DNA from other bacterial species, both close and distant relatives were subject to hybridisation analysis (Figure 3.10). Dots were spotted by hand in the order listed in Table 3.2. The probes did not bind to the bacterial species unrelated to *C. difficile* further indicating that the probes were highly specific to toxins A and B of *C. difficile*. Species of the *Clostridium* genus, including species of the LCT family closely related to *C. difficile* (which possess toxins closely related to *tcdB*) did not shown any probe hybridisation which further validates the specificity of the probes.

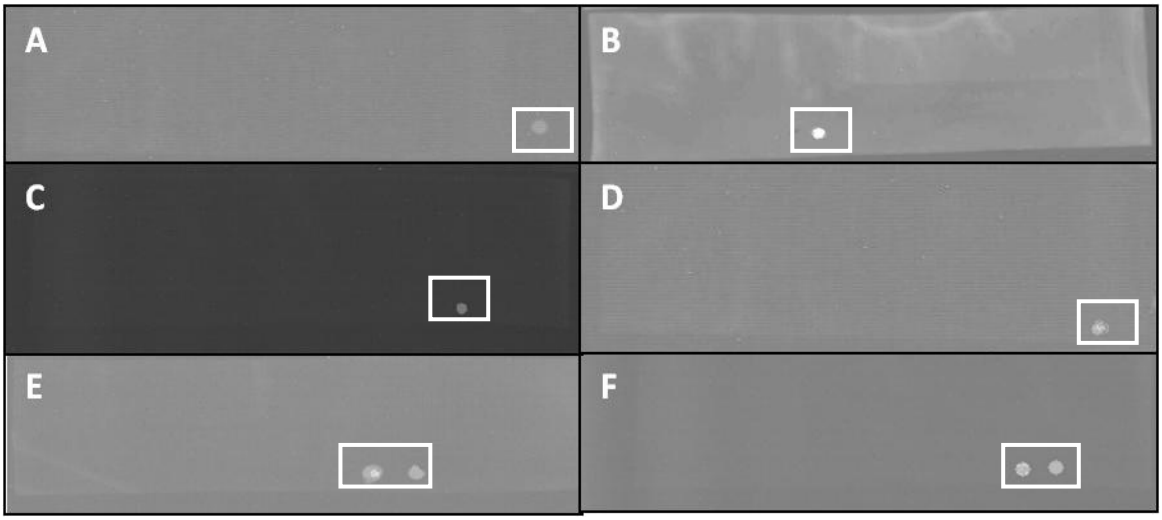


Figure 3.10: Dot blot of related and unrelated bacterial species by DNA hybridisation

Species which were related and unrelated to *C. difficile* were tested against the probes for specificity. There was no hybridisation of the probes to the gDNA on the membrane. Positive control for all reactions is CD630 and a variant strain control R22680 ($tcdA^{-}tcdB^{+}$) was also included (highlighted by white square). (A) *tcdA*50 anchor probe, (B) *tcdA*50 detector probe, (C) *tcdA*76 anchor probe, (D) *tcdA*76 detector probe, (E) *tcdB* anchor probe, (F) *tcdB* detector probe. See p98 for list of all species used.

3.5.6.4. Dot blot of human metagenomic gut DNA

To further confirm the specificity of the designed probes, metagenomic DNA from human gut flora, from ten human volunteers were examined for hybridisation against the designed DNA probes (Table 3.3). Dots were spotted by hand in the order listed in Table 3.2. The probes did not bind any human metagenomic DNA samples giving a clear blot. Thus as a result of these experiments the *tcdA* 76bp anchor and detector probes, and the *tcdB* anchor and detector probes were used in the MAMEF- based detection assay.

3.6. DISCUSSION

The underlying hypothesis for this study is that gene sequences encoding biologically essential protein structures are under considerable selective pressure, and are thus forced to conserve their primary sequences (Adhya & Patra, 2012). Indeed this is likely to be the case for toxins A and B, the organisms' principle virulence factors. We also hypothesised that these conserved regions are unlikely to contain any of the DNA restriction sites specified and examined by Rupnik's restriction enzyme toxin typing system (1998).

To determine the conserved nature of the regions which were not digested by the restriction enzymes used by Rupnik (1998), bioinformatic analyses were performed using *C. difficile* toxin A and B sequences deposited in the Genbank database. We found homology, of varying degrees, between regions in toxin A and toxin B with each other. Indeed the homology between both toxins has been suggested to be the result of a gene duplication event (Von Eichel-Streiber *et al*., 1992) as they share ~66% sequence and functional homology (Voth & Ballard, 2005). This was reflected in our bioinformatic analyses and gives credence to the gene duplication theory postulated by Von Eichel-Streiber *et al.* (1992); however it appears that toxin B is more stable than toxin A in *C. difficile*. This suggests that toxin B may have evolved in *C. difficile* first, and toxin A was duplicated later. Toxin A is absent, or truncated, within variant strains of *C. difficile*; thus arguably the detection of toxin B within the *C. difficile* genome is of the most import when considering assay design as all toxigenic *C. difficile* isolates possess this toxin. The role of the two toxins of *C. difficile* in disease however has been widely debated (Lyras *et al.*, 2009; Kuehne *et al.*, 2010)

Bioinformatic analysis of these conserved regions also revealed homology of toxin B to *C. sordellii* lethal toxin (TcsL), and to the TpeL toxin of *C. perfringens*, which are part of the LCT family. Toxin B and TcsL share a sequence homology of 85% (Voth & Ballard) and thus presented a challenge in defining a toxin B specific DNA probe (Von Eichel-Streiber *et al.*, 1992; Green *et al.*, 1995). In the context of evolution it is likely to be indicative of a common toxin ancestor within the *Clostridium* family and may have arisen as a result of divergent evolution. Toxin A did share homology to the TcsL sequences from the LCT family; however this homology was not as high as with toxin B (85%).

From bioinformatical analyses the individual structural domains of toxins A and B also shared homology with toxins associated with bacterial pathogens of insects, including *Photorhabdus luminescens*, and *P. asymbiotica* mcf (makes caterpillars floppy) toxins. This association of *C. difficile's* primary toxins with mcf toxins are potentially indicative of an insect associated past, and this finding may shed light on the evolutionary path of *C. difficile* itself. Indeed the presence of other insect virulence factors and gene homologs in the *C. difficile* genome may reflect the possibility that *C. difficile* may have originally been an insect pathogen.

Alternatively the toxin genes themselves may have been acquired as a result of HGT from insects. This hypothesis can be supported by the fact that the PaLoc of *C. difficile*, where the toxin genes are located (Section 1.4.5), is positioned between two insertion sequences, and as such its presence is suggestive of horizontal acquisition. However importantly the PaLoc is currently not mobile and appears to have lost its mobility over its evolution, hence it is still designated as a "pathogenicity locus" (Braun *et al.,* 1996). Indeed the insect associated genes also suggest that *C. difficile* may have originated as a scavenging organism able to degrade organic matter and insects in a potential environmental niche habitat. This is explored further in Chapter 4.

While published nucleotide sequences are a useful source of information, they only represent a fraction of the diversity found in nature. Thus we adopted a rigorous screening process to ensure the specificity of our probes; using *C. difficile* isolates, related *Clostridium* species and metagenomic DNA extracts of human gut flora. Indeed the probes designed were specific for conserved regions within the areas defined by Rupnik (1998) for future use in a MAMEF based toxin detection assay. In the context of the MAMEF assay it would be beneficial to employ anchor probes which are highly sensitive to target DNA to facilitate efficient capture of target DNA. The inclusion of a second specific probe carrying the fluorescent detection tag at the 5' end further enhanced the specificity of the assay (Section 1.7.1).

Probes were firstly tested for specificity against a comprehensive panel of 58 *C. difficile* clinical isolates, representing a range of differing toxinotypes. We observed differences in the binding efficiency of our probes to target DNA, with the most sensitive probe being the *tcdA* (76 bp) anchor probe. Probes were further screened for specificity against isolates

from near neighbours within the *Clostridium* family, including those which possessed LCT genes. Finally they were subjected to a more stringent analysis by running them against 10 metagenomic DNA extracts from human gut flora which represents approximately 10^{12} per gram wet weight bacteria (Shoemaker *et al*., 2001).

In conclusion, the ability of the resulting designed probes to bind specifically to toxins A and B confirms their suitability for inclusion in an assay able to detect the presence of pathogens in human faeces (as described in Chapter 5). Probes were designed using the above method were unique to toxigenic *C. difficile* and conserved within toxigenic isolates. Moreover the potential association of *C. difficil*e toxins to insect associated pathogenic toxins is a link which will be explored further in Chapter 4. Indeed if the evolution of *C. difficile* and its toxins can be traced we can better understand the organism and explore methods to prevent the severe disease caused by its toxins.

CHAPTER 4

EVOLUTION OF *CLOSTRIDIUM DIFFICILE* AND INVERTEBRATE ASSOCIATION

4.1. ABSTRACT

Bioinformatic analysis of the genome of *C. difficile* revealed the presence of genes encoding proteins homologous to known insect virulence factors. These genes included a binary toxin, a putative chitinase gene and a set of genes from *Photorhabdus luminescens* and *P. asymbiotica* linked to virulence. Further analysis of the genes encoding *C. difficile tcdA* and *tcdB* revealed the presence of mcf 1 & 2 proteins which is known play a role in insect pathogenicity.

The presence of these homologs suggests that at some stage in its evolution *C. difficile* may have been an insect pathogen. To determine if the bacterium has the ability to infect insects, we employed an insect model in collaboration with the University of Bath. We attempted to infect the tobacco hawk moth caterpillar *Manduca sexta* with the spore and vegetative forms of *C. difficile*. At 25^{o}C there appeared to be no adverse effects upon the insects after injection and feeding. However at 37^{o}C an effect was observed on the injected insects at pupation. The injection of spores and vegetative cells appeared to assist more insects to pupate, whereas injection of the controls (water and sodium taurocholate) resulted in insect mortality. The insects fed spores and vegetative cells did not show any signs of infection. Furthermore when the vegetative and spore forms of *C. difficile* were tested with haemolymph obtained from *M. sexta* there was evidence of antimicrobial activity. To understand whether temperature affected the ability of *C. difficile* spores to germinate, we assessed the germination of *C. difficile* spores at 25^{o}C, 30^{o}C and 37^{o}C. Results suggest 37^{o}C is optimal for germination.

In addition to screening for insect pathology we also determined the ability of *C. difficile* to degrade chitinase. A commercially available chitinase assay kit was used to ascertain the ability of *C. difficile* to produce chitinase, revealing that chitinase appears to be released from spore preparations when spore germination is initiated. The bioinformatic and *in vivo* experimental analysis of *C. difficile* has revealed that while there are genes within the *C. difficile* genome that appear to encode proteins associated with insect pathogenicity, they may not be biologically active. Moreover this inactivity may in fact suggest that although these genes may have originated from invertebrate associated ancestors, they may have lost biological function throughout *C. difficile's* evolution.

4.2. INTRODUCTION

Biological evolution has created millions of diverse species on this planet due primarily to repeated differentiation of existing species (Bourtzis & Miller, 2003). At its fundamental level an organism evolves by altering its genetic make up. This can be achieved through the loss, modification or acquisition of genes (horizontal gene transfer) which in turn is driven by natural selection. Genetic adaptations such as the acquisition of virulence factors allow bacteria to access and survive in new environments (Shames *et al.,* 2009). To do so they must overcome physical, cellular and molecular barriers presented by the host, they must grow and replicate and avoid the defence mechanisms of the host. This interaction can be harmful, benign or even beneficial to the host (Ochman & Moran, 2001).

Bacteria which have adapted to infect humans are often opportunistic pathogens that have "host-jumped" from another species. There are numerous examples of human pathogens which have evolved from diseases of insects and domestic animals including *B. anthracis* and *Y. pestis* (Waterfield *et al.,* 2004; Scully & Bidochka, 2006). As a result of evolution, mammals and insects have developed similar immune response mechanisms to pathogens.

The mammalian and insect innate immune systems have comparable structural and functional homology; however their adaptive immune systems are different. The innate immune system includes cellular and humoral defences. As part of the cellular immune system insects display phagocytosis, nodulisation and encapsulation. This response comprises haemocytes which phagocytose invading bacteria (similar to mammalian macrophages) and produce antibacterial reactive oxygen species (ROS) which kill the invaders in a similar manner to human neutrophils. Thus it is not surprising that insects and humans possess homologous proteins involved in superoxide production (Bergin *et al.,* 2005).

Processes of the humoral system include melanisation, haemolymph (insect blood system) clotting and antimicrobial peptide production such as cecropins, defensins, attacins, lysozymes and heat shock proteins (Salzet, 2001; Seed & Dennis, 2008). Similarities between the insect and human immune systems supports the argument that pathogenicity mechanisms developed to enable bacteria to infect insects may also contribute to virulence in humans. An example of an insect pathogen which evolved to infect humans is *B. anthracis*. This pathogen, similar to *C. difficile*, is a Gram positive

spore forming bacillus which, unlike *C. difficile*, is able to grow aerobically. Examination of the genome of *B. anthracis* reveals evidence of loss of gene function which is thought to have occurred during its evolution from an entomopathogenic ancestor of the *B. cereus* group into a mammalian pathogen (Ivanova *et al.*, 2003).

The *B. cereus* group share a high degree of genetically similar genes (80-100% sequence identity) and comprise *B. cereus* (primarily a soil commensal), *B. anthracis* (a mammalian pathogen) and *B. thuringiensis* (an insect pathogen). Some scientists have proposed that these should be classified as a single species; however due to the distinct pathogenic behaviour between species they have all maintained their individual status (Helgason *et al.*, 2000).

B. anthracis is the causative agent of anthrax. The plasmids of *B. anthracis* pXO1 and pXO2 confer toxin production. pXO1 carries the anthrax toxin genes *pag*A, *lef* and *cya*, and pXO2 carries the *cap* genes encoding antiphagocytic factors (Read *et al.*, 2003). Both *B. cereus* and *B. anthracis* possess 14 or 15 genes functioning as degraders of chitin, chitosan, starch and glycogen which are all associated with insects. Common virulence factors of *B. anthracis* and *B. cereus* include homologs of viral enhancin genes and an immune inhibitor gene (*inh*A) which selectively cleaves insect antimicrobial peptides (Read *et al.*, 2003).

The presence of plasmids encoding insecticidal toxins is the principal means of differentiating *B. thuringiensis* from other members of the group (Helgason *et al.*, 2000). *B. thuringiensis* is mainly an insect pathogen which produces crystal protein toxins (*cry*) at the sporulation stage of growth. These insecticidal toxins are used as a method of biocontrol and cause breakdown of the insect gut (Hofte & Whiteley, 1989). Vegetative insecticidal toxins (VIP) are produced at the vegetative stage of growth, however not all strains of *B. thuringiensis* produce this toxin. The VIP genes of *B. cereus*- specifically *B. thuringiensis* - are of interest as they are directly associated with insects. VIP genes do not show any homology to the δ- endotoxins of *B. thuringiensis*, but they do target insects which are not susceptible to the δ- endotoxins (Bhalla *et al.*, 2005). All these factors suggest a common insect associated ancestor for the bacillus group through virulence factor-containing plasmid acquisition.

Gram negative *Photorhabdus* species also produce insecticidal toxins. The *Photorhabdus* genus includes *P. luminescens*, *P. temperata* and *P. asymbiotica*. The latter is associated with human wounds and is not opportunistic but has links to soft tissue disease and has been thought to be transmitted via an invertebrate vector (spiders) (Gerrard *et al.*, 2004). The two former however are associated with nematodes. Nematodes contain *P. luminescens* and *P. temperata* within their gut. When they encounter an insect host in the soil, they bite the insect injecting the bacteria into the gut. In order to survive the *Photorhabdus* multiplies within the insect blood system (haemolymph) and feeds on the insect tissue, releasing toxins and proteases into the corpse (Ffrench-Constant *et al.*, 2003; Waterfield *et al.*, 2004).

Mcf 1 (Makes Caterpillars Floppy) toxin is released by *P. luminescens*. When expressed in a recombinant *E. coli* vector and injected into caterpillars, mcf 1 causes rapid degradation of the insect gut making the insect "floppy" (Daborn *et al.*, 2002). *Mcf1* is an 8.8 kb gene encoding a predicted protein 2929 aa long and has 20% homology to residues in the membrane translocation domain of *C. difficile* toxin B (Daborn *et al.*, 2002). Mcf 1 is also encoded adjacent to a *pheV*-tRNA gene located on the same site as the viral enhancin gene in *Y. pestis* suggesting conserved mechanisms of HGT (Waterfield *et al.*, 2004). Mcf 2 is encoded within *P. asymbiotica* and is 2993 aa in length.

Insecticidal toxin complexes (Tc) are produced by *Photorhabdus* after host death and are delivered directly into the insect gut (haemocoel) by nematodes. The toxins are produced in four forms; Tca, Tcb, Tcc and Tcd (Bowen *et al.*, 1998; Heerman and Fuchs, 2008). Other toxin complexes similar to the insecticidal toxin complexes include *Serratia entomophilia*'s *sep*ABC genes that cause "amber disease" in larvae of New Zealand grass grubs (Hurst *et al.*, 2000). The *sep* genes are encoded on a 120 kb plasmid pADAP which also encodes a prophage-like locus for an "anti-feeding effect". Prophage–like loci are also present in the genomes of *P. luminescens* and *P. asymbiotica* ATCC 43949, each encoding effector proteins. Some of these proteins display sequence similarity to the Mcf toxins of *Photorhabdus*, TcdA of *C. difficile* and YopT from *Yersinia enterocolitica* (Dove *et al.*, 1990; Yang *et al.*, 2006). These acquired virulence factors are classed as *Photorhabdus* virulence cassettes.

P. asymbiotica is an emerging human pathogen which, unlike the other *Photorhabdus* species, carries plasmids. Plasmids include pPAU1 of the North American strain and pPAA1 and pPAA2 of other strains which have yet to be fully sequenced. Within the *P. asymbiotica* ATCC 43949 genome there is a low copy number of the insecticidal toxin complexes in the Tca and Tcd encoding gene islands. This gene loss has been suggested as the reason for the absence of oral toxicity to model insects (Wilkinson *et al.,* 2009).

A rapid virulence annotation (RVA) technique has been utilised to screen the cosmid library of *P. asymbiotica* ATCC43949 against a range of vertebrate and invertebrate targets. The *P. asymbiotica* cosmid library (40000 bp insert size) was cloned into *E. coli* and screened against a murine macrophage cell line, the nematode *Caenorhabditis elegans*, the protozoan *Acanthamoeba polyphaga* and the caterpillars of the tobacco horn worm (*M. sexta*) and the greater wax moth (*Galleria mellonella*). From the screen 21 *P. asymbiotica* gene regions displayed toxicity factors common to one or more of the test models (Waterfield *et al.,* 2008). Any regions associated with human pathogenesis were determined by comparison with the insect pathogen *P. luminescens*. Interestingly the cellular immune response of insects is also inhibited by *P. asymbiotica*, all of which indicates that *P. asymbiotica* may have very recently evolved from a *Photorhabdus* ancestor to be capable of infecting both insects and humans (Waterfield *et al.,* 2008).

Insects also possess physical barriers as a defence from pathogenic bacteria. Their hard exoskeleton and gut lining is made from chitin (Kramer & Muthukrishnan, 1998). Chitin is a polymer of N-acetylglucosamine with similarities to cellulose (Bowen *et al.,* 1991). Indeed some pathogenic bacteria possess an enzymatic gene called a chitinase which "melts" the exoskeleton of insects through degradation of chitin to a lower molecular weight. Insects periodically shed their chitin layers using such chitinase enzymes, and some bacteria have acquired such enzymes through evolution to be able to combat insect defences. It is highly probable that these bacteria have co-evolved within an insect vector. Interestingly *C. difficile* possesses a putative chitinase gene which is biologically functional in its spore form, but not in its vegetative form. This chitinase is thought to be associated with the spore coat protein CotE (Lawley *et al.,* 2009; Permpoonpattna *et al.,* 2011).

To determine if *C. difficile* contains homologs to genes known to play a role in insect virulence we probed the NCBI database of published *C. difficile* genomes for homologs of the enhancin genes, VIP genes, mcf 1, TC toxins, RVA genes and the chitinase genes described above. We also determined the ability of clinical isolates of *C. difficile* to infect a live insect model.

4.2.1. Model Insects

Model systems are often used to study human infectious diseases. Most mammalian research uses primates, rodents and felines as hosts for microbial diseases; however there are many ethical concerns over the suffering endured by these animals (Scully & Bidochka, 2006). This has led some researchers to evaluate insects as an alternative model for the study of human pathogenic bacteria. There are many benefits to using insects as model hosts. They can be easily maintained, used in large numbers and do not require legal permission for experimental use which leads to reduced costs and labour time (Scully & Bidochka, 2006).

An example of a successful infection model is the fruit fly *Drosophila melanogaster* which has been used extensively to characterise the pathogenicity of a range of bacterial species including *Penicillium*, *P. aeruginosa*, *P. entomophila, Wolbachia* species, *Erwinia carotovora* and *Serratia marcescens* (Vallet-Gely *et al.,* 2008). The wax moth *G. mellonella* is also increasingly being used as a bacterial infection model (Seed & Dennis, 2008). The moth is a pest of bee hives, feeding upon pollen and destroying combs. Bacteria studied using this model include *B. cereus, P. aeruginosa*, *Proteus mirabilis* and fungal pathogens such as *Aspergillus (*Kavanagh and Reeves, 2004).

A further bacterial infection model, and the one used in this study, is the tobacco horn worm *M. sexta* the life cycle of which is shown in Figure 4.1 (Manduca project, 2001). The large size of the *Manduca* allows for easy dissection, bacterial recovery and general handling (Silva *et al.,* 2002; Daborn *et al.,* 2002). It is possible to infect these insects via direct injection into the hemocoel and by feeding which has led to it being employed as the primary model with which to study *Photorhabdus* species (Silva *et al.,* 2002). For these reasons we utilized this model to test our clinical isolates of *C. difficile,* supplied courtesy of Dr. N. Waterfield at the University of Bath.

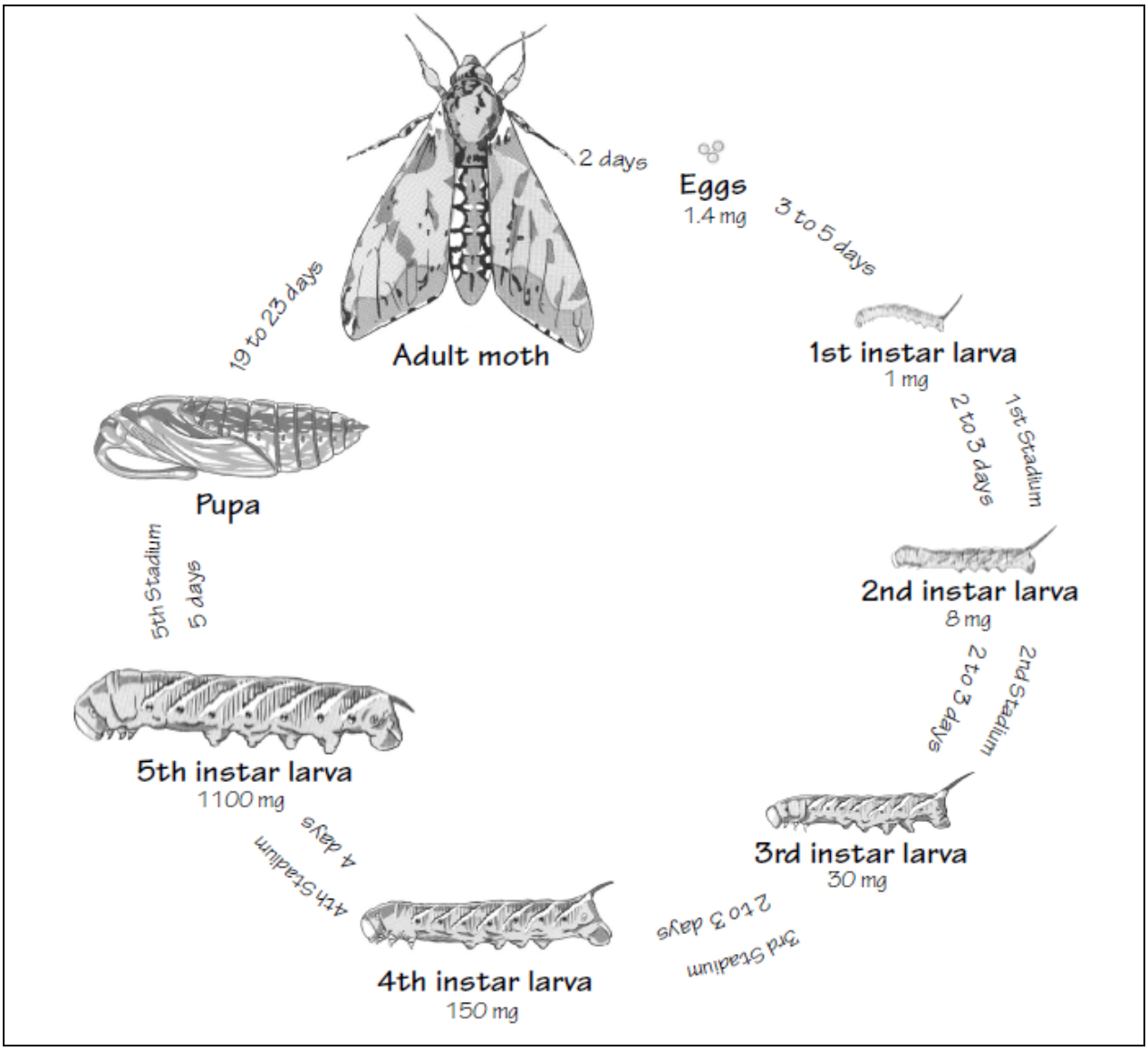


Figure 4.1: Lifecycle of *Manduca sexta* caterpillar

The larvae of the *Manduca* undergo several stages in their lifecycle.

Image taken from [insected.arizona.edu/manduca/PDFs/Posters.pdf @2001 Manduca Project]

4.2.2. *M. Sexta* Biology

Insects have three body parts: the head, the thorax and the abdomen. The insect *M. sexta* has a lifecycle as shown in Figure 4.2. The eggs of the insect are spherical and are approximately 1.5 mm in diameter for 3-5 days (Villenueva, 2007). The hatched larvae/caterpillars are cylindrical and have five pairs of legs (Figure 4.2). The main feature of the larvae is the red horn at the dorsal end segment from which the *M. sexta* derives its

name as the "tobacco hornworm". Each stage of growth is known as an instar and there are five instars in total. In the environment *Manduca* larvae live 9-15 inches deep underground in loose soil or leaf litter. After fifth instar stage the larvae pupate for 3-5 days and then hatch to form adult tobacco horn moths (Manduca Project, 2001).

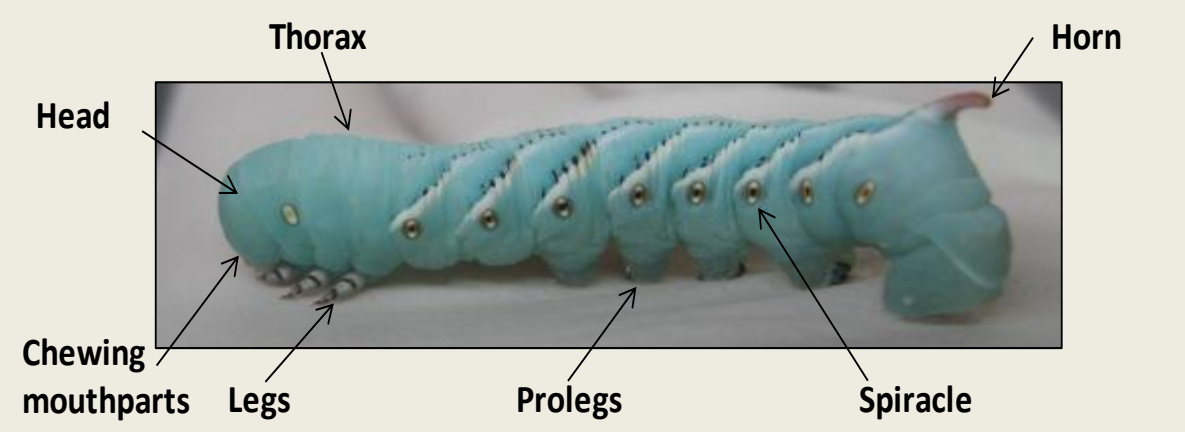


Figure 4.2: Larvae of *Manduca sexta*

The sections and major parts of the *M. sexta* larvae are shown here. The red horn and the characteristic lateral white stripes can be seen. The larvae have six legs, and prolegs which allow it to stay on the tobacco leaf.

4.2.3. *M. sexta* gut biology

The midgut of *M. sexta* is a simple tube running through the body of the insect allowing rapid passage and processing of large quantities of plant material (Brinkman *et al.*, 2008). The midgut itself is highly alkaline (pH 9) which enables the insect to extract nutrients from its food; however this environment may only allow for certain bacterial communities to survive. The bacterial composition of the *M. sexta* gut has been characterised. Commensal bacteria of the *M. sexta* include *Staphylococcus*, *Micrococcus*, *Corynebacterium*, *Bacillus* and *Paenibacillus* species (Van der Hoeven *et al.*, 2008). Often, breeding colonies of *M. sexta* are susceptible to epidemic infections thus antibiotics are added to suppress infection (Hoffman *et al.*, 1966).

4.2.4. *M. sexta* haemolymph

The haemolymph of *M. sexta* contains hemolin, peptidoglycan recognition proteins and C-type lectins (Jiang, 2008). Challenge of *Manduca* with bacteria induces synthesis of effector proteins and antimicrobial peptides which are secreted into the insect plasma. The

hemolin is a 48 kDa protein containing four immunoglobulin domains, and is unregulated in response to bacterial challenge. The C- type (calcium dependent) lectins are involved in binding bacteria and haemocyte aggregation (Rolf & Reynolds, 2009). The haemolymph also comprises a blue biliprotein known as insecticyanin (Holden *et al.*, 1987). The insecticyanin holoprotein is produced during the larval stage of growth and confers camouflage of the *Manduca.*

4.3. RESEARCH AIMS

The aims of this chapter are to:

- Analyse the genome of *C. difficile* for the presence of homologs to genes known to express insect virulence factors
- Determine the feasibility of infecting an insect model with *C. difficile* with a view to use the model to test the ability of our probes to detect *C. difficile.*
- Determine the ability of *C. difficile* to produce biologically active chitinase

4.4. MATERIALS & METHODS

4.4.1. Bioinformatic analysis of *C. difficile* virulence gene sequences using BLASTp

The amino acid (protein) sequences of Toxin A and B individual structural domains (glucosyltransferase domain, cysteine protease domain, transmembrane domain, receptor binding domain) were analysed to determine if they shared any sequence identity with amino acid sequences from other organisms stored within the GenBank sequence database [www. ncbi.nlm.nih.gov/Protein]. Amino acid sequence identity was determined using the Basic Local Alignment Search Tool protein homology search facility: BLASTp [http://blast.ncbi.nlm.nih.gov/Blast.cgi]. Sequences showing over 80% query coverage, and in >50% sequence identity, were considered to have significant homology (Altschul *et al.*, 1997). The query coverage shows how long the query sequence inputted into BLAST is covered by the one found (subject) by BLAST. The maximum identity refers to the percentage alignment of the Blast input (query) sequence to its matched (subject) sequence and indicates the maximum percentage of identical nucleotides or amino acids within the noted alignment length (http://blast.ncbi.nlm.nih.gov).

4.4.2. Bioinformatic analysis of the *C. difficile* genome for homologs to genes known to play a role in insect virulence

A total of 5 independent *C. difficile* genomes were analysed for the presence of homologs to the following virulence factors:

Table 4.1: Table of genes examined in this study

Genes were accessed and obtained from the NCBI Genbank database. The protein sequences were subjected to protein BLAST analysis and nucleotide sequences were examined via BLASTN against the non redundant NCBI database.

Bacterium Species	Gene Examined	Bioinformatic BLAST Sequence Analysis Performed
C. difficile	*tcdA*: glucosyltransferase domain, cysteine protease domain, transmembrane domain and receptor binding domain	PROTEIN
C. difficile	*tcdB*: glucosyltransferase domain, cysteine protease domain, transmembrane domain and receptor binding domain	PROTEIN
C. difficile	*cdtA* *cdtB*	PROTEIN
C. difficile	Putative chitinase/peroxiredoxin	PROTEIN
***B. cereus* ATCC 14579** ***C. botulinum* B1 str. Okra,** ***C. perfringens* D str. JGS1721**	Enhancin	PROTEIN
***P. asymbiotica* ATCC 43949**	RVA 1- 21 (Waterfield *et al.,* 2008)	PROTEIN
P. luminescens	*mcf 1* *mcf 2* Insecticidal Toxin genes	NUCLEOTIDE

4.4.3. Multiple sequence Alignments

To identify regions common to mcf, TC toxins, toxin A and toxin B individual nucleotide sequences were analysed using ClustalW Multiple Sequence Alignment (MSA) program provided by European Bioinformatics Institute (Larkin *et al.*, 2007). The results of this analysis were displayed using Bioedit *v*7.1.3 (Hall, 2001) MSA viewer in which bases are colour coded to highlight conserved regions.

4.4.4. Insect infection studies

To determine if *C. difficile* is capable of infecting an insect host we exposed the tobacco hornworm *M. sexta* to a series of bacterial challenges using either spores or vegetative organisms.

4.4.4.1. *Manduca sexta*

The tobacco hornworms (*M. sexta*) used in this study were supplied by Dr. N. Waterfield of the University of Bath (Figure 4.3). The *M. sexta* larvae were at fifth instar growth stage fed on an artificial wheat-germ diet (Reynolds & Nottingham, 1985) and one feeding

block ($1cm^2$) was used per insect. Unless otherwise stated the media was supplemented with the antibiotic chloramphenicol and formaldehyde to prevent decomposition. The insects were kept individually in a transparent plastic sealed container punctured with air holes to allow ventilation and maintained at 25^oC, in a DEFRA approved facility at the University of Bath.

Figure 4.3: Live *Manduca sexta* used in this study

The *Manduca* at fifth instar growth stage. Before experiments were performed they were wiped with 70% ethanol and handled using gloves.

4.4.4.2. Bacterial Strains & Growth conditions

The two isolates of *C. difficile* selected for the insect infection studies were DS1813, a hypervirulent 027 ribotype and DS1748 a non epidemic 002 ribotype. Spores and vegetative forms of each isolate were produced as previously described (Chapter 2).

4.4.4.3. Preliminary Oral Infection of *M. sexta* with *C. difficile* spores

A total of 90 *Manduca* at 2^{nd} instar were used in this preliminary study. 45 *M. sexta* previously fed on antibiotic-containing wheat germ were divided into three groups of 15. Each individual *Manduca* was fed a 1 cm^2 food block inoculated with 100 µl *C. difficile* spores of strain DS1813 at a concentration of 1×10^7 cfu/ml. For the second group each insect was fed a 1 cm^2 food block inoculated with 100 µl *C. difficile* spores of strain DS1748 at a concentration of 1×10^7 cfu/ml. As a control 15 *Manduca* were also fed food with sterile deionised water. *Manduca* were allowed to feed on the spore-containing antibiotic food for the next 3 days. This experiment was repeated using antibiotic free food disks. To determine if the presence of Sodium taurocholate enhanced the pathogenicity of *C. difficile* spores for *Manduca,* 100 µl of the bile salt sodium taurocholate (0.1%) was added to the 1 cm^2 antibiotic free food blocks. As a control *M. sexta* were fed food with

sdw. Following feeding the insects were examined for signs of infection/morbidity after 3 days.

4.4.4.4. Direct Injection of *Manduca sexta* with *C. difficile* at 25 ^{o}C

4.4.4.4.1 Spore challenge

Fifth-instar *M. sexta* larvae were injected with 50 µl of a stock spore concentration of 1.33 x 10^9 / ml spores of DS1813 and DS1748. This is a total of 6.65 x 10^7 spores. Replicates of 10 were used in this study. Injections were performed directly into the insect midgut using a 100 µl disposable syringe needle (Fisher Scientific, UK). Post injection, larvae were fed with wheat-germ and were monitored for symptoms of toxicity over time. This approach was repeated using DS1748 spores, and with sdw as a control.

4.4.4.4.2. Vegetative challenge

Experiment conducted as per above. Vegetative cultures of *C. difficile* strains DS1813, and DS1748 were grown anaerobically for 18 h, and larvae were injected with a total of 50 µl of a vegetative cell concentration of 2 x 10^9 cfu/ml cells. This is a total of 1 x 10^7 cfu injected into each insect.

4.4.4.5. Direct Injection of *M. sexta* with *C. difficile* at 37^{o}C

Experiments were performed as above however, insects were incubated at 37^{o}C.

4.4.4.6. Oral Infection of *M. sexta* with *C. difficile* spores

A total of 100 *M. sexta* 2nd instar larvae were divided into five groups of 20 fed on wheat-germ with no antibiotics. This experiment was repeated using antibiotic containing food discs. Each individual *Manduca* was fed a 1 cm^2 food block inoculated with 100 µl *C. difficile* spores of strain DS1813 at a concentration of 1 x 10^6 cfu/ml. For the second group each insect was fed a 1 cm^2 food block inoculated with 100 µl *C. difficile* spores of strain DS1748 at a concentration of 1 x 10^6 cfu/ml (Figure 5.4). The final group of 20 insects served as control and were fed a 1 cm^2 food block inoculated with 100 µl sdw.

The insects were monitored each morning for 5 days for symptoms of toxicity and faecal matter was collected from each insect and retained for analysis. Signs of infection included abnormal colouring of the larvae from blue-green to brown, unusual faecal matter production and death. At the end of the experiment all insects were killed by

freezing at -20^{o}C, subsequently dissected and the midgut was homogenized in sdw using a VortexGenie (Fisher Scientific, UK). To determine the number of recoverable bacteria a dilution series of faecal samples from each day was prepared with sdw and serial dilutions were then plated in duplicate onto CDMN agar. Plates were incubated for 48 h at 37°C anaerobically and individual colonies counted.

Figure 4.4: *Manduca sexta* feeding on wheat-germ based food block

Manduca were placed into individual containers with the wheat germ food block as the food source. Lids were applied to the containers to secure the insects, and each lid had ventilation holes.

4.4.4.7. The effect of temperature on the ability of *C. difficile* spores to germinate

All of the infection studies undertaken above with *M. sexta* were performed at 25^{o}C. To determine the ability of the *C. difficile* isolates used in our infection studies to germinate and replicate at this and other temperatures we incubated individual strains at 25^{o}C, 30^{o}C and 37^{o}C degrees. Spore germination was assessed by adding a 1 ml spore suspension (1 x 10^{7} spores/ ml) to 9 ml BHI broth in a 15 ml heat resistant falcon tube (Fisher Scientific, UK) and incubated at the relevant temperature. At 1 h intervals 1 ml of spores were removed from broth culture and the optical density (OD_{450}) was determined (Ultrospec 1100 *pro* UV/Visible spectrophotometer). The number of viable vegetative organisms and spores in each sample was determined by plating serial dilutions of heat shocked (80^{o}C for 10 min) and unheated samples onto BHI agar containing 0.1% sodium taurocholate. The plates were then incubated at the relevant germination temperature.

Experiments conducted at 37^{o}C were performed using a Bug Box Plus anaerobic cabinet (Ruskinn Technology Ltd, Bridgend, Wales). For experiments undertaken at 25^{o}C and 30^{o}C we employed a 3.4L anaerobic jar (Oxoid, Basingstoke, Hampshire, UK) with an anaerobic gas generating kit (Oxoid, Basingstoke, Hampshire, UK). The anaerobic jar was then placed into the incubator at the relevant temperature.

4.4.4.8. Effect of *Manduca* haemolymph on *C. difficile* spores and vegetative cells

Fifth instar *M. sexta* were killed by chilling on ice for 20 min, the head subsequently removed and haemolymph extracted. Each experiment was performed in triplicate. *C. difficile* spores of strains DS1813 (7 x 10^5 cfu/ml) and DS1748 (8.33 x 10^5 cfu/ml) were inoculated into individual Eppendorfs containing 200 µl of *M. sexta* haemolymph. A single colony *C. difficile* (strains DS1813 and DS1748) was inoculated into an Eppendorf containing 200 µl haemolymph. The inoculated haemolymph was incubated anaerobically for 48 h. Viable counts of vegetative cells and spores were conducted after incubation as described in Chapter 2 using CDMN selective agar.

4.4.4.9. Detection of Chitinase activity in spores and vegetative cells of *C. difficile*

Chitinase activity was determined as per Permpoonpattna *et al.* (2011). A pre-supplied Chitinase assay kit (Sigma CS0980) was used and the reactions performed (n =6) in a microtitre plate (Fisher Scientific, UK). The concentration of *C. difficile* vegetative cells used was 1.33 x 10^6 cfu/ml, and for the spore suspensions 2.67 x 10^6 cfu/ml were used in the reactions and *C. difficile* strains DS1813 and DS1748 were tested for chitinase activity. Spores of *C. difficile* were tested for chitinase activity following germination with sodium taurocholate (ST) at a range of concentrations- 0%, 3% and 5%. Spores were added to ST mixed with PBS and incubated anaerobically for 30 min. The chitinase activity was assessed following this.

The Substrate solution comprised 4-Nitrophenyl N-acetyl-β-D-glucosaminide (1 mg/ml) which was dissolved in 1 ml Assay buffer (A4855). Subsequently the Substrate and Standard solutions were calibrated to 37^{o}C. The positive control used was the chitinase control enzyme (chitinase from *Trichoderma viride*), which was mixed with Substrate solution. For each subsequent test 90 µl of substrate solution was mixed in with 10 µl *C. difficile* sample. Reaction time was 30 min at 37°C. After the reaction was stopped (200µl of 0.04g/ml sodium carbonate) and then the OD_{405} of supernatants was measured using a plate reader (Dynex, Worthing, West Sussex, UK).

4.4.4.10. Statistical Analysis

Statistical analysis was performed using GraphPad Prism version 5.04 for Windows, (GraphPad Software, La Jolla California USA, www.graphpad.com). Statistical significant differences were tested for using one way analysis of variance (ANOVA) at the 95%

confidence interval in conjunction with Dunnett's post test. A P value of <0.05 was considered significant (Bowker & Randerson, 2006).

4.5. RESULTS

4.5.1. Bioinformatic analysis of known *C. difficile* virulence genes using BLAST

4.5.1.1. Toxin A Analysis

BLAST comparisons of the four individual structural protein domains of toxin A revealed significant query coverage (99-100%) and maximum sequence identity of >41% to domains within bacteria of the LCT family. These included *C. difficile tcdB*, *C. sordellii tcsL*, *C. novyi tcnA* and *C. perfringens* toxins *tpeL* and *tcpA* (Table 4.2). Hits with species other than those in the LCT family include *Vibrio parahaemolyticus* and *P. fluorescens* putative toxin A within the glucosyltransferase domain. In the cysteine protease domain, maximum sequence identity to *Vibrio cholerae* RTX toxin was found to be 24%. The receptor binding domain of *tcdA* has homology to *Streptococcus sobrinus, S. oralis and C. perfringens* cell wall binding proteins. Interestingly only the membrane translocation domain of *tcdA* had identity (98%) to an insect associated bacterium- *P. luminescens subsp. laumondii.*

Table 4.2: Table depicting homology hits to structural domains of toxin A

The homology hits against each individual structural domain of *C. difficile* toxin A were deduced by protein BLAST. The percentage of sequence query coverage, and the percentage of maximum identity of each bacterial gene is displayed in correlation to each structural domain (*query coverage/maximum identity). Hits of > 50% sequence identity were considered significant. The full alignments from the protein BLAST [http://blast.ncbi.nlm.nih.gov/Blast.cgi] are detailed in the Appendix (p227).

Bacterial gene screened for homology	Homology with *C. difficile* Toxin A domain (*query coverage/maximum sequence identity) (%)			
	Glucosyl-transferase (GT) domain	Cysteine protease domain	Membrane Translocation domain	Receptor Binding Domain
C. difficile tcdB	*100/49	99/59	100/60	99/42
C. sordellii tcsL	100/52	99/57	100/60	99/41
C. perfringens tpeL	100/45	99/52	99/46	-
***C. novyi* α toxin**	99/33	98/34	96/38	99/50
***C. perfringens tcp*A**	99/45	99/52	99/46	-
***Citrobacter rodentium* lymphocyte inhibitory factor**	81/24	-	-	-
***Pseudomonas fluorescens* Putative toxin A**	81/25	-	-	-
***P. luminescens subsp. laumondii* Hypothetical protein**	-	-	98/22	-
***Yersinia mollaretii* Chemotaxis protein**	-	99/25	98/25	-
***Yersinia mollaretii* RTX toxin**	-	83/27	88/27	-
***Vibrio cholerae* RTX toxin**	-	83/24	-	-
***Streptococcus salivarius* Hypothetical protein**	-	-	-	99/24
***S. oralis* cell wall binding repeat domain protein**	-	-	-	99/31
***C. perfringens* cell wall binding repeat domain protein**	-	-	-	99/29

4.5.1.2. Toxin B Analysis

BLAST comparisons of the four individual structural protein domains of toxin B revealed significant query coverage (99-100%) and maximum sequence identity of >40% to domains within bacteria of the LCT family. These included *C. difficile tcdA*, *C. sordellii tcsL*, and *C. perfringens* toxins *tpeL* and *tcpA* (Table 4.3). The GT domain of *tcdB* had sequence identity of 21% to *P. fluorescens* toxin A and mcf toxins, while the cysteine protease domain had sequence identity of 25% to the insect pathogen *P. luminescens subsp. laumondii* hypothetical protein. There was 27% homology of the cysteine protease domain to *Y. mollaretii* RTX toxin, *P. asymbiotica* RTX toxin and *Vibrio cholerae* RTX toxin. Interestingly the membrane translocation domain of *tcdB* had sequence identity (24-33%) to the mcf toxins of *P. luminescens subsp. laumondii, P. asymbiotica, P. luminescens,* and *Xenorhabdus bovienii.* The receptor binding domain, however, was dramatically different showing 31% and 29% sequence identity to *Ruminococcus albus* and *S. oralis.*

Table 4.3: Table depicting homology hits to structural domains of toxin B

The homology hits against each individual structural domain of *C. difficile* toxin B were deduced by protein BLAST. The percentage of sequence query coverage, and the percentage of maximum identity of each bacterial gene is displayed in correlation to each structural domain (*query coverage/maximum identity). Hits of > 50% sequence identity were considered significant. The full alignments from the protein BLAST [http://blast.ncbi.nlm.nih.gov/Blast.cgi] are detailed in the Appendix (p227).

Bacterial gene screened for homology	Homology with *C. difficile* Toxin B (% Homology)			
	Glucosyl-transferase (GT) domain	Cysteine protease domain	Membrane Translocation domain	Receptor Binding Domain
C. difficile tcdA	*100/51	100/57	100/60	99/42
C. sordellii tcsL	100/75	100/74	100/92	100/74
C. perfringens tpeL	100/44	100/51	100/44	-
***C. novyi* α toxin**	99/34	98/36	96/37	99/34
C. perfringens tcpA	96/44	100/51	100/44	-
***Pseudomonas fluorescens* Putative toxin A**	82/21	-	-	-
***P. luminescens subsp. laumondii* (hyp. Protein)/ mcf toxin**	-	92/25	100/24	-
***P. asymbiotica* mcf toxin**	-	-	100/24	
***Yersinia mollaretii* RTX toxin**	-	93/27	88/33	-
***P. asymbiotica subsp. asymbiotica* RTX toxin**	-	83/27	-	-
***Xenorhabdus bovienii* Putative mcf toxin**	-	-	98/24	-
***P. luminescens* mcf2 toxin**	-	-	98/23	-
***Vibrio cholerae* RTX toxin**	-	84/27	-	-
***Pseudomonas fluorescens* mcf toxin**	82/21	-	-	-
Pseudomonas fluorescens fitD	-	-	98/26	-
***Ruminococcus albus* cell wall binding repeat domain protein**	-	-	-	99/31
***S. oralis* cell wall binding repeat domain protein**	-	-	-	100/29

4.5.2. Bioinformatic analysis of the *C. difficile* genome for homologs to genes known to play a role in insect virulence

The genome of *C. difficile* has potential links to genes associated with insect virulence. Genes in *C. difficile* were screened for insect associated gene homologs using protein BLAST (Altschul *et al.*, 1997). These genes included the binary toxin and the putative chitinase/peroxiredoxin gene. The genome of *C. difficile* was also screened for homologs to the enhancin genes from *B. anthracis*, *B. cereus*, *C. botulinum* and *C. perfringens*, which would further potentiate an evolutionary link with insects.

4.5.3. Bioinformatic analysis of binary toxin

The binary toxin of *C. difficile* is related to other ADP-ribosylating binary toxins of species that are directly linked to insect pathogenicity. These bacteria include the VIP2 toxin of *Bacillus thuringiensis* and, to a lesser extent, *C. perfringens* Iota toxin and *C. spiroforme*. BLAST (protein) analysis of the *C. difficile* binary toxins revealed links to the ADP-ribosylating toxins from other species, as expected (Figure 4.5). Links to the ISP2a protein of bacteria *Brevibacillus laterosporus* were found, which is known to cause toxicity to nematodes and beetles (Oliveiera *et al.*, 2004). There were no links to species associated with insects other than to the VIP genes of *B. thuringiensis*.

Accession	Description	Max score	Total score	Query coverage	E value
ZP_05401993.1	C2 toxin,component I [Clostridium difficile QCD-23m63]	215	215	100%	1e-54
ZP_05272668.1	C2 toxin,component I [Clostridium difficile QCD-66c26] >ref|ZP_05323063.1|	215	215	100%	1e-54
ZP_01802067.1	hypothetical protein CdifQ_04003036 [Clostridium difficile QCD-32g58]	215	215	100%	1e-54
2WN6_A	Chain A, Structural Basis For Substrate Recognition In The Enzymatic Compo	215	215	100%	1e-54
AAB67304.2	ADP-ribosyltransferase enzymatic component [Clostridium difficile CD196]	214	214	100%	3e-54
AAZ16242.1	binary toxin [Clostridium difficile] >gb|AAZ16243.1| binary toxin [Clostridium	211	211	100%	3e-53
CAB58425.1	Clostridium difficile binary toxin A; actin-specific ADP-ribosyltransferase	211	211	100%	3e-53
CAB58426.1	Clostridium difficile binary toxin A; actin-specific ADP-ribosyltransferase	207	207	100%	4e-52
ZP_02632017.1	iota toxin component Ia [Clostridium perfringens E str. JGS1987] >emb|CAA5	191	191	100%	2e-47
1GIQ_A	Chain A, Crystal Structure Of The Enzymatic Componet Of Iota-Toxin From C	191	191	100%	2e-47
CAA66611.1	Sa component [Clostridium spiroforme]	186	186	100%	5e-46
CAI40768.1	Isp2a protein [Brevibacillus laterosporus]	63.2	63.2	99%	1e-08
CAI43279.1	Isp2b protein [Brevibacillus laterosporus]	84.7	120	98%	3e-15
ZP_04105842.1	hypothetical protein bthur0008_59670 [Bacillus thuringiensis serovar berliner	81.6	117	98%	2e-14
ABR68092.1	Vip2A(BR) [Bacillus thuringiensis]	56.6	56.6	98%	8e-07
ZP_04075450.1	hypothetical protein bthur0013_57960 [Bacillus thuringiensis IBL 200] >gb|AA	53.5	53.5	98%	7e-06
1QS1_A	Chain A, Crystal Structure Of Vegetative Insecticidal Protein2 (Vip2) >pdb|1Q	53.5	53.5	98%	7e-06
1QS2_A	Chain A, Crystal Structure Of Vip2 With Nad	53.5	53.5	98%	7e-06
YP_002650774.1	C2 toxin,component I [Clostridium botulinum] >ref|YP_003034265.1| C2 toxi	62.8	62.8	95%	1e-08
BAA09942.1	C2 toxin (component-I) [Clostridium botulinum] >dbj|BAA32536.1| C2 toxin (	62.8	62.8	95%	1e-08
ACH42758.1	vegetative insecticidal protein [Bacillus thuringiensis]	69.3	103	86%	1e-10
ACH42759.1	vegetative insecticidal protein [Bacillus thuringiensis]	57.4	91.3	86%	5e-07

Figure 4.5: Protein BLAST homology hits to *C. difficile* Binary Toxin

The bacterial genes with high homology to *C. difficile* binary toxin are displayed as per BLAST format. The accession number directly links to the Genbank Accession for the bacterial gene, and the gene description is displayed. The maximum score is defined by the score of the highest scoring pair homology and is similar to the E- value. The E- value is defined as the Expect value and the lower the E value the more significant the homology hit. The total score corresponds to the total scores of the high scoring pair from the database sequence. Query coverage directly relates to the percentage length coverage of the query. The higher the percentage coverage, the higher the homology (Altschul *et al.*, 1997). Hits of > 80% query coverage were considered significant.

4.5.4. Bioinformatic analysis of putative *C. difficile* chitinase gene

The putative chitinase gene of *C. difficile* belongs to a family of chitinases known as the GH18 glycosylhydrolases, which hydrolyse chitin and have been identified in bacteria, fungi and insects. BLASTp analysis of the chitinase gene revealed high (100%) query coverage homology hits to peroxiredoxin/chitinase genes within other *C. difficile* strains, suggesting that this gene may be conserved (Figure 4.6). This gene however has links to *Clostridium bartletti*, an organism which can be isolated from human faeces (sequence query coverage 83%).

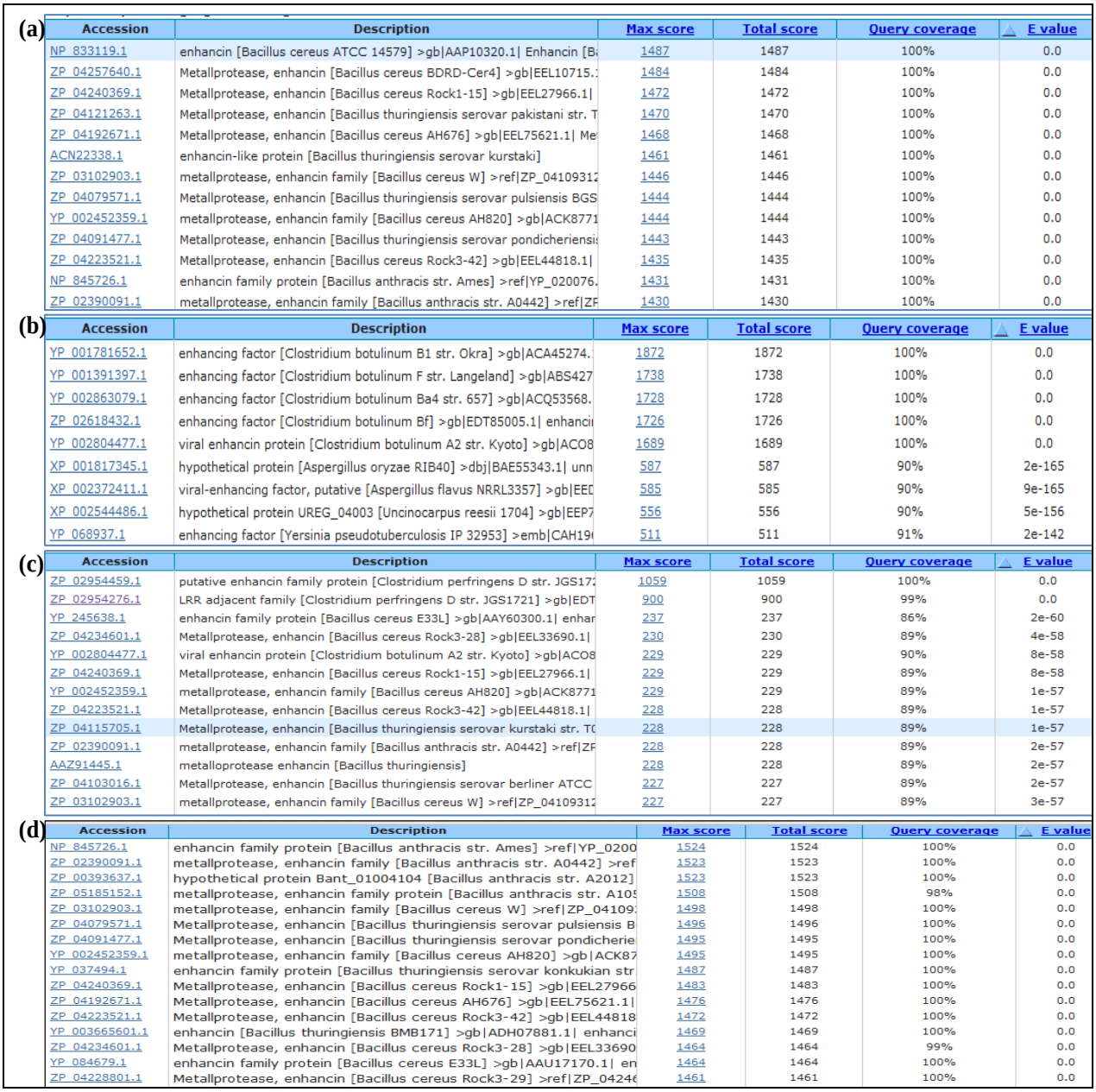
Accession	Description	Max score	Total score	Query coverage	E value
YP_001087934.1	putative bifunctional protein: peroxiredoxin/chitinase [Clostridium difficile 630	1479	1479	100%	0.0
ZP_01804190.1	hypothetical protein CdifQ_04001599 [Clostridium difficile QCD-32g58] >ref|Z	1473	1473	99%	0.0
ZP_05350571.1	putative bifunctional protein: peroxiredoxin/chitinase [Clostridium difficile ATC	1471	1471	99%	0.0
ZP_05329436.1	putative bifunctional protein: peroxiredoxin/chitinase [Clostridium difficile QCI	1471	1471	99%	0.0
ZP_05400853.1	putative bifunctional protein: peroxiredoxin/chitinase [Clostridium difficile QCI	1434	1434	99%	0.0
ZP_02212445.1	hypothetical protein CLOBAR_02062 [Clostridium bartlettii DSM 16795] >gb|E	539	887	83%	4e-151

Figure 4.6: Protein BLAST homology hits to *C. difficile* putative peroxiredoxin/chitinase

The bacterial genes with high homology to *C. difficile* the putative peroxiredoxin/ chitinase are displayed as per BLAST format. The accession number directly links to the Genbank Accession for the bacterial gene, and the gene description is displayed. The maximum score is defined by the score of the highest scoring pair homology and is similar to the E- value. The E- value is defined as the Expect value and the lower the E value the more significant the homology hit. The total score corresponds to the total scores of the high scoring pair from the database sequence. Query coverage directly relates to the percentage length coverage of the query. The higher the percentage coverage, the higher the homology (Altschul *et al.*, 1997). Hits of > 80% query coverage were considered significant.

4.5.5. Bioinformatic analysis of enhancin genes

To establish if genes within the *C. difficile* genome are homologous to the enhancin genes commonly associated with insects, the enhancin gene sequences from *B. cereus*, *C. botulinum subsp. Okra*, *C. perfringens,* and *B. anthracis* were inputted into BLASTp. The enhancin gene from *B. cereus* had no homology to *C. difficile*, but had 100% query coverage to the enhancin proteins within *B. thuringiensis* (Figure 4.7). The enhancin protein from *C. botulinum*, however, had 100% homology to other *C. botulinum* subspecies, and 90% to *Y. pestis* enhancing factors. *C. perfringens* D str. JGS1721 has 100% homology to enhancin from *C. botulinum* species, *B. cereus* species and the enhancin from *B. anthracis* strain Ames.



(a)

Accession	Description	Max score	Total score	Query coverage	E value
NP_833119.1	enhancin [Bacillus cereus ATCC 14579] >gb|AAP10320.1| Enhancin [Ba	1487	1487	100%	0.0
ZP_04257640.1	Metallprotease, enhancin [Bacillus cereus BDRD-Cer4] >gb|EEL10715.1	1484	1484	100%	0.0
ZP_04240369.1	Metallprotease, enhancin [Bacillus cereus Rock1-15] >gb|EEL27966.1|	1472	1472	100%	0.0
ZP_04121263.1	Metallprotease, enhancin [Bacillus thuringiensis serovar pakistani str. T	1470	1470	100%	0.0
ZP_04192671.1	Metallprotease, enhancin [Bacillus cereus AH676] >gb|EEL75621.1| Me	1468	1468	100%	0.0
ACN22338.1	enhancin-like protein [Bacillus thuringiensis serovar kurstaki]	1461	1461	100%	0.0
ZP_03102903.1	metallprotease, enhancin family [Bacillus cereus W] >ref|ZP_0410931	1446	1446	100%	0.0
ZP_04079571.1	Metallprotease, enhancin [Bacillus thuringiensis serovar pulsiensis BGS	1444	1444	100%	0.0
YP_002452359.1	metallprotease, enhancin family [Bacillus cereus AH820] >gb|ACK8771	1444	1444	100%	0.0
ZP_04091477.1	Metallprotease, enhancin [Bacillus thuringiensis serovar pondicheriensis	1443	1443	100%	0.0
ZP_04223521.1	Metallprotease, enhancin [Bacillus cereus Rock3-42] >gb|EEL44818.1|	1435	1435	100%	0.0
NP_845726.1	enhancin family protein [Bacillus anthracis str. Ames] >ref|YP_020076.	1431	1431	100%	0.0
ZP_02390091.1	metallprotease, enhancin family [Bacillus anthracis str. A0442] >ref|ZP	1430	1430	100%	0.0

(b)

Accession	Description	Max score	Total score	Query coverage	E value
YP_001781652.1	enhancing factor [Clostridium botulinum B1 str. Okra] >gb|ACA45274.1	1872	1872	100%	0.0
YP_001391397.1	enhancing factor [Clostridium botulinum F str. Langeland] >gb|ABS427	1738	1738	100%	0.0
YP_002863079.1	enhancing factor [Clostridium botulinum Ba4 str. 657] >gb|ACQ53568.	1728	1728	100%	0.0
ZP_02618432.1	enhancing factor [Clostridium botulinum Bf] >gb|EDT85005.1| enhanci	1726	1726	100%	0.0
YP_002804477.1	viral enhancin protein [Clostridium botulinum A2 str. Kyoto] >gb|ACO8	1689	1689	100%	0.0
XP_001817345.1	hypothetical protein [Aspergillus oryzae RIB40] >dbj|BAE55343.1| unn	587	587	90%	2e-165
XP_002372411.1	viral-enhancing factor, putative [Aspergillus flavus NRRL3357] >gb|EED	585	585	90%	9e-165
XP_002544486.1	hypothetical protein UREG_04003 [Uncinocarpus reesii 1704] >gb|EEP7	556	556	90%	5e-156
YP_068937.1	enhancing factor [Yersinia pseudotuberculosis IP 32953] >emb|CAH19	511	511	91%	2e-142

(c)

Accession	Description	Max score	Total score	Query coverage	E value
ZP_02954459.1	putative enhancin family protein [Clostridium perfringens D str. JGS172	1059	1059	100%	0.0
ZP_02954276.1	LRR adjacent family [Clostridium perfringens D str. JGS1721] >gb|EDT	900	900	99%	0.0
YP_245638.1	enhancin family protein [Bacillus cereus E33L] >gb|AAY60300.1| enhar	237	237	86%	2e-60
ZP_04234601.1	Metallprotease, enhancin [Bacillus cereus Rock3-28] >gb|EEL33690.1|	230	230	89%	4e-58
YP_002804477.1	viral enhancin protein [Clostridium botulinum A2 str. Kyoto] >gb|ACO8	229	229	90%	8e-58
ZP_04240369.1	Metallprotease, enhancin [Bacillus cereus Rock1-15] >gb|EEL27966.1|	229	229	89%	8e-58
YP_002452359.1	metallprotease, enhancin family [Bacillus cereus AH820] >gb|ACK8771	229	229	89%	1e-57
ZP_04223521.1	Metallprotease, enhancin [Bacillus cereus Rock3-42] >gb|EEL44818.1|	228	228	89%	1e-57
ZP_04115705.1	Metallprotease, enhancin [Bacillus thuringiensis serovar kurstaki str. T0	228	228	89%	1e-57
ZP_02390091.1	metallprotease, enhancin family [Bacillus anthracis str. A0442] >ref|ZP	228	228	89%	2e-57
AAZ91445.1	metalloprotease enhancin [Bacillus thuringiensis]	228	228	89%	2e-57
ZP_04103016.1	Metallprotease, enhancin [Bacillus thuringiensis serovar berliner ATCC	227	227	89%	2e-57
ZP_03102903.1	metallprotease, enhancin family [Bacillus cereus W] >ref|ZP_0410931	227	227	89%	3e-57

(d)

Accession	Description	Max score	Total score	Query coverage	E value
NP_845726.1	enhancin family protein [Bacillus anthracis str. Ames] >ref|YP_0200	1524	1524	100%	0.0
ZP_02390091.1	metallprotease, enhancin family [Bacillus anthracis str. A0442] >ref	1523	1523	100%	0.0
ZP_00393637.1	hypothetical protein Bant_01004104 [Bacillus anthracis str. A2012]	1523	1523	100%	0.0
ZP_05185152.1	metallprotease, enhancin family protein [Bacillus anthracis str. A105	1508	1508	98%	0.0
ZP_03102903.1	metallprotease, enhancin family [Bacillus cereus W] >ref|ZP_04109	1498	1498	100%	0.0
ZP_04079571.1	Metallprotease, enhancin [Bacillus thuringiensis serovar pulsiensis B	1496	1496	100%	0.0
ZP_04091477.1	Metallprotease, enhancin [Bacillus thuringiensis serovar pondicherie	1495	1495	100%	0.0
YP_002452359.1	metallprotease, enhancin family [Bacillus cereus AH820] >gb|ACK87	1495	1495	100%	0.0
YP_037494.1	enhancin family protein [Bacillus thuringiensis serovar konkukian str	1487	1487	100%	0.0
ZP_04240369.1	Metallprotease, enhancin [Bacillus cereus Rock1-15] >gb|EEL27966	1483	1483	100%	0.0
ZP_04192671.1	Metallprotease, enhancin [Bacillus cereus AH676] >gb|EEL75621.1|	1476	1476	100%	0.0
ZP_04223521.1	Metallprotease, enhancin [Bacillus cereus Rock3-42] >gb|EEL44818	1472	1472	100%	0.0
YP_003665601.1	enhancin [Bacillus thuringiensis BMB171] >gb|ADH07881.1| enhanci	1469	1469	100%	0.0
ZP_04234601.1	Metallprotease, enhancin [Bacillus cereus Rock3-28] >gb|EEL33690	1464	1464	99%	0.0
YP_084679.1	enhancin family protein [Bacillus cereus E33L] >gb|AAU17170.1| en	1464	1464	100%	0.0
ZP_04228801.1	Metallprotease, enhancin [Bacillus cereus Rock3-29] >ref|ZP_04246	1461	1461	100%	0.0

Figure 4.7: Protein BLAST homology hits to enhancin genes

(a) *B. cereus* ATCC 14579 (b) *C. botulinum* B1 str. Okra (c) *C. perfringens* D str. JGS1721 (d) *B. anthracis* Ames. A range of bacteria known to possess expressed enhancin genes were analysed for homologs to *C. difficile*. Results are displayed as per BLAST format. The accession number directly links to the Genbank Accession for the bacterial gene, and the gene description is displayed. The maximum score is defined by the score of the highest scoring pair homology and is similar to the E- value. The E- value is defined as the Expect value and the lower the E value the more significant the homology hit. The total score corresponds to the total scores of the high scoring pair from the database sequence. Query coverage directly relates to the percentage length coverage of the query. The higher the percentage coverage, the higher the homology hit (Altschul *et al.*, 1997). Hits of > 80% query coverage were considered significant.

4.5.6. Bioinformatical analysis of invertebrate associated virulence factors

4.5.6.1. Screening for homology with Mcf toxin genes against *C. difficile* toxin genes

From Table 4.2 and 4.3 it was revealed that the membrane translocation domains of TcdA and TcdB had high (99-100%) homology to *P. luminescens* and *P. asymbiotica* Mcf genes. Hence the amino acid sequences of these toxins were inputted into Bioedit *v*7.1.3 (Hall, 2001) and analysed using an MSA viewer using colour codes equivalent to those used in ClustalW (Thompson *et al.*, 1997; Figure 4.8). Mcf 1 from *P. luminescens* had 18-19% sequence identity to TcdA and TcdB respectively. BLAST analysis revealed 17-18% sequence identity of TcdB to the mcf from *P. asymbiotica.* This was expected as 20% sequence identity of the mcf toxins to TcdB has been described previously (Dove *et al.*, 1990).

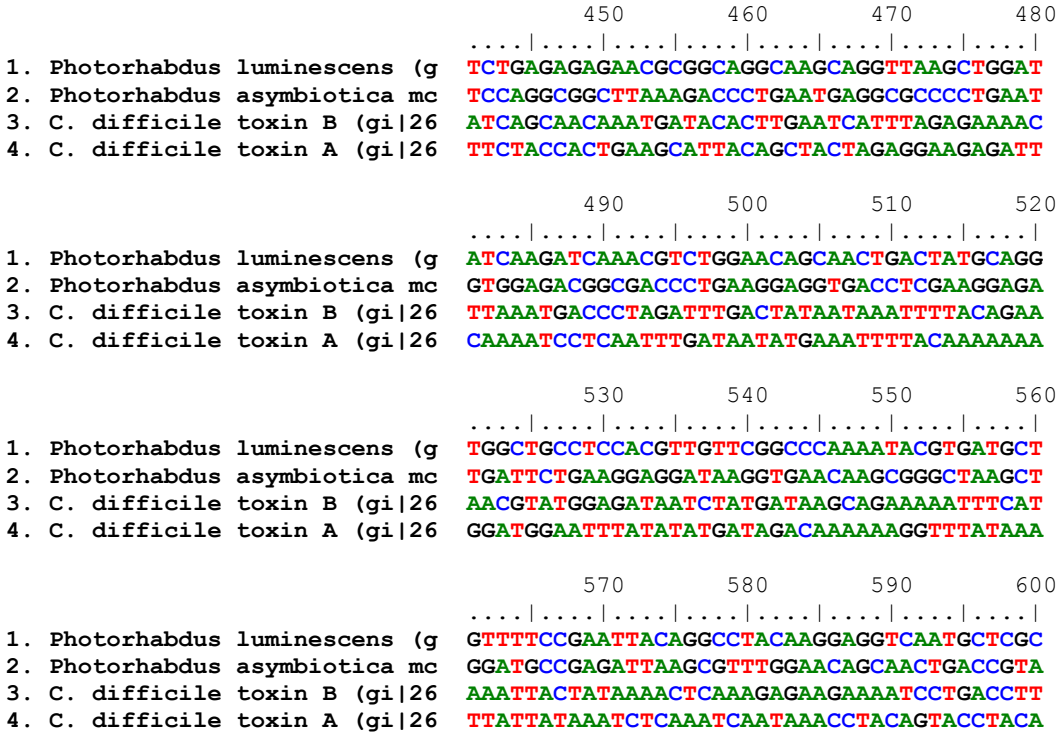


Figure 4.8: MSA of mcf toxin of *P. luminescens* and *P. asymbiotica* against toxins A and B of *C. difficile*

The nucleotide sequences of *tcdA* and *tcdB* of *C. difficile*, and the mcf toxin of *P. luminescens* and *P. asymbiotica*, were imported into EBI ClustalW for alignments. The Bioedit *v*7.1.3 (Hall, 2001) method of viewing MSA allows conserved residues to be colour coded according to those used in the EBI program ClustalW. The guanine reside (G) is shown as **black**, cytosine (C) as **blue,** adenosine (A) as **green** and thymine (T) as **red**.

4.5.6.2. Screening for homology with Insecticidal toxin genes against *C. difficile* toxin genes

As the mcf toxins from *P. luminescens* shared 20% homology with *tcdB* of *C. difficile*, it was of interest to screen the insecticidal toxins from *P. luminescens* for homology to TcdA or TcdB. Hence the nucleotide sequences of the insecticidal toxins TccA, TccB, TccC, TcaC, and TcdA of *P. luminescens* and *P. asymbiotica* TccC were inputted into ClustalW and analysed using an MSA viewer (Figure 4.9). Bioinformatic analysis of the insecticidal toxins revealed no significant homology to any *C. difficile* toxins. Significant homology at the genome level is considered to be 20% or above.

Figure 4.9: MSA of Insecticidal Toxin Complexes of *P. luminescens* and *P. asymbiotica* against toxins A and B of *C. difficile* The nucleotide sequences of *tcdA* and *tcdB* of *C. difficile, tccA, tccB1, tccC, tcaC,* and *tcdA1* of *P. luminescens* and *P. asymbiotica tccC* were imported into EBI ClustalW for alignments. The Bioedit *v*7.1.3 (Hall, 2001) method of viewing MSA allows conserved residues to be colour coded according to those used in the EBI program ClustalW. The guanine reside (G) is shown as **black**, cytosine (C) as **blue,** adenosine (A) as **green** and thymine (T) as **red**.

4.5.6.3. Screening for homology with RVA genes from *P. asymbiotica*

To determine of there were genes within *C. difficile* that were linked to the *Photorhabdus* virulence gene set, the 21 RVA genes from the cosmid library of *P. asymbiotica* ATCC43949 (associated with cross-species virulence) were inputted into BLAST for potential homology hits to genes in the *C. difficile* genome. BLAST analysis revealed no homology to any known *C. difficile* toxins.

4.5.7. Insect infection studies

To determine if *C. difficile* is capable of infecting an insect host, we exposed the tobacco hornworm *M. sexta* to a series of bacterial challenges using either spores or vegetative organisms.

4.5.7.1. Direct injection of *Manduca sexta* with *C. difficile* at 25^{o}C

Attempts to infect *M. sexta* with *C. difficile* spores were unsuccessful. Injected spores administered in combination with the germinant sodium taurocholate had no obvious effect on the health of the *M. sexta*. Indeed the *M. sexta* also survived challenge with vegetative cells showing no symptoms of toxicity or morbidity.

4.5.7.2. Direct Injection of *M. sexta* with *C. difficile* at 37^{o}C

Further attempts to infect *M. sexta* with *C. difficile* spores were unsuccessful at a higher temperature of 37^{o}C. There was no rapid acute infection and thus the *Manduca* followed their normal life cycle (Figure 4.1) and as expected started to pupate after 5 d at 5th instar. Furthermore we employed sodium taurocholate as a germinant in conjunction with spores; however no infection manifested. Interestingly *Manduca* which only received sodium taurocholate or water were unable to pupate correctly at 37^{o}C (Figure 4.10). Indeed the *M. sexta* also survived challenge using vegetative cells and showed no symptoms of toxicity or morbidity.

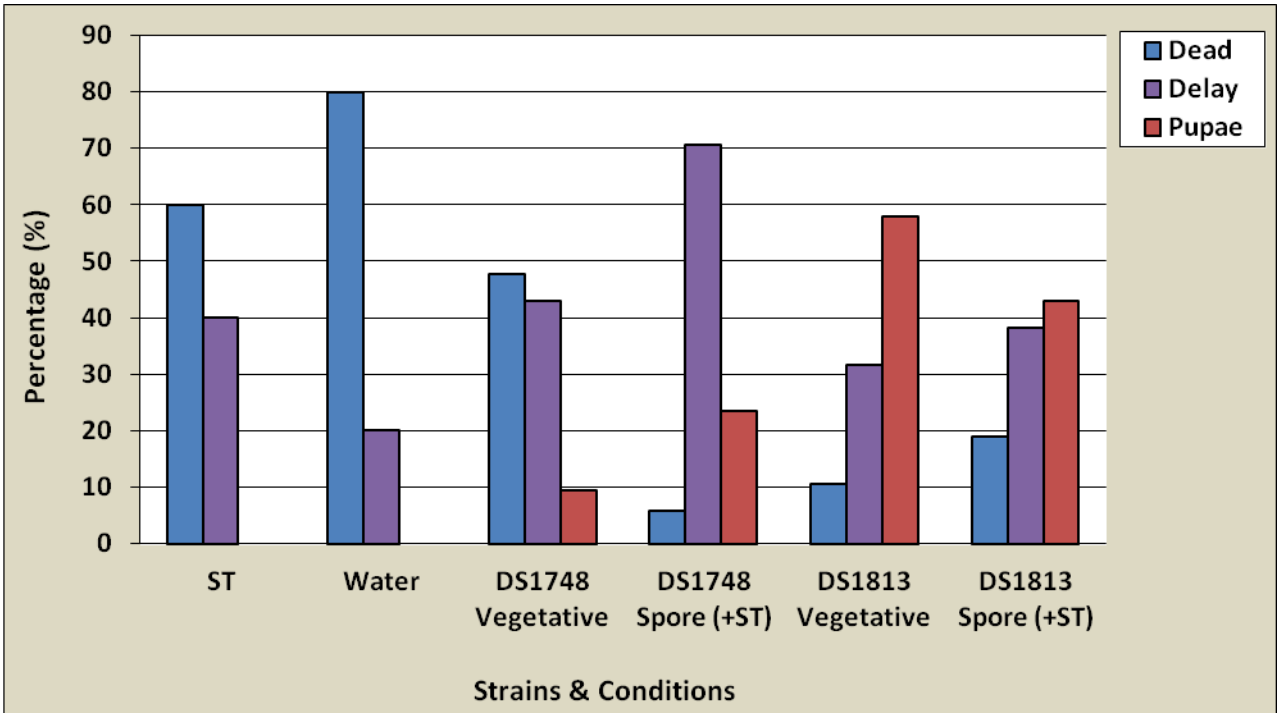


Figure 4.10: Results of direct injection of *C. difficile* in *M. sexta*

Insects were directly injected with *C. difficile* vegetative cells and spores of strains DS1813 and DS1748. Spores were injected in combination with ST to attempt to elicit an infective response. Controls used were ST and sterile water. In total 20 *Manduca* were tested in each sample group. The number of *Manduca* which died after injection with ST is shown in **blue**, the number of *Manduca* which survived and showed a subsequent delay in pupation; and thus their normal lifecycle are show in **purple** and finally those which were able to pupate normally after 5 days are shown in **red**. As these values are percentages of *Manduca* statistical analysis of the values was not conducted.

4.5.7.3. Effect of *Manduca* haemolymph on *C. difficile* spores and vegetative cells

C. difficile was tested against *M. sexta* haemolymph to determine if the response to direct injection with *C. difficile* was influenced by compounds within the insect haemolymph. Spores of strains DS1813 (7×10^5 cfu/ml) and DS1748 (8.33×10^5 cfu/ml) were inoculated into haemolymph and recovered on CDMN selective agar. In response to haemolymph 2.17×10^3 cfu/ml of strain DS1813 and 3.58×10^4 cfu/ml of strain DS1748 were recovered. Indeed this is indicative of potential antimicrobial compounds present with the haemolymph of *M. sexta*. Furthermore, there was no recovery of vegetative *C. difficile* from the inoculated haemolymph after incubation.

4.5.7.4. Oral Infection of *M. sexta* with *C. difficile* spores

To determine the feasibility of infecting *Manduca* via feeding, insects were fed spores with and without sodium taurocholate upon antibiotic containing and antibiotic-free wheat germ. After 3 d we saw no evidence of morbidity in any of our feeding studies suggesting that oral uptake does not initiate infection.

4.5.7.5. Potential for gastrointestinal colonisation of *M. sexta*

We tested *M. sexta* as a possible model to study the ability of *C. difficile* to colonise the gut. To determine if these differences in hydrophobicity affected the ability of the spores to adhere to the insect midgut and subsequently cause infection, *Manduca* were fed spores and the numbers recovered over a 5 d period. Two strains with different hydrophobic characteristics were used; DS1813 spores which are highly hydrophobic (77% RH) and DS1748 spores which are hydrophilic (14%).

In a subsequent series of experiments spores of each strain of *C. difficile* were fed to *Manduca* on antibiotic containing or antibiotic-free wheat germ with and without taurocholate. Faecal matter was cultured to ascertain the number of spores that passed through the insect, or remained within the insect. The expectation was that successful infection would result in an increase in the number of cells recovered. Upon the fifth day of the study the *Manduca* were killed and dissected to remove the mid gut. This was then cultured to ascertain the number of bacteria remaining within the gut. Infection of the *Manduca* was unsuccessful as the number of bacteria recovered decreased over time and there was no evidence of an increase in the total number of *C. difficile,* suggesting that the organism had simply passed through the gut. Culture of the gut contents at 5 days revealed no difference in the numbers of bacteria recovered from each test group as shown in Figure 4.11. This was confirmed by use of one way ANOVA revealed no significant difference between the mean results from each test group ($P = 0.0769$) (Figure 4.11). The mean results for each test group were further compared to each other using Dunnett's multiple comparison post-hoc test, also revealing no significant difference in recovery.

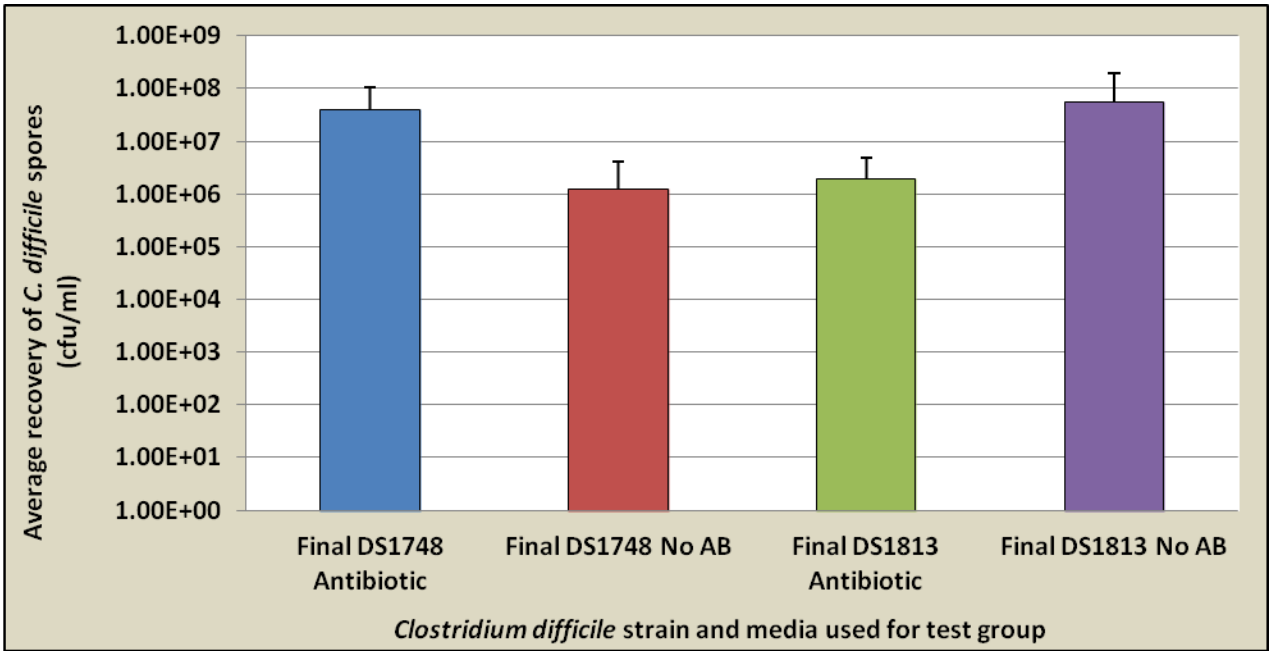
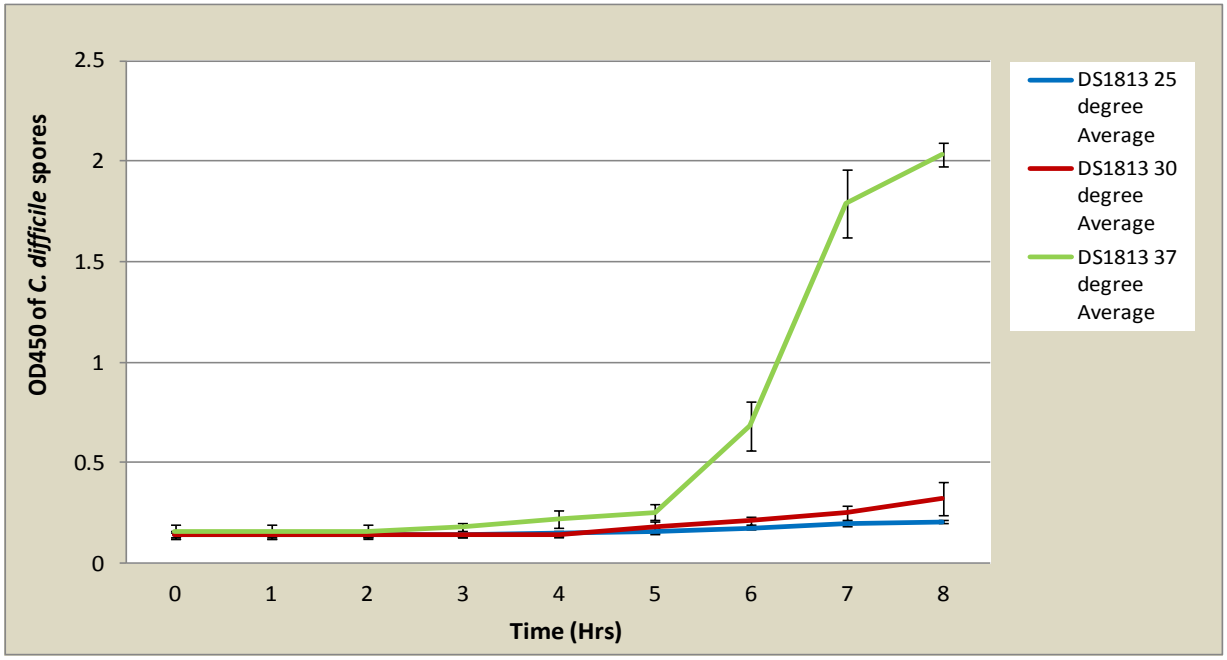


Figure 4.11: Average final recovery of *C. difficile* per *Manduca sexta* test group

The average recovery was determined by adding up the number of spores recovered and divided per insect (20). The final number of spores recovered appears to be equivalent to the number of spores initially fed (1 x 10^6) to the insects. The final recovered DS1748 spores on antibiotic containing wheat germ is shown in **blue**, final recovered DS1748 spores on antibiotic-free wheat germ is shown in **red**, final recovered DS1813 spores on antibiotic containing wheat germ is shown in **green**, and final recovered DS1748 spores on antibiotic free wheat germ is shown in **purple**. There is no statistical significance between the mean values, determined by use of one way ANOVA (P = 0.0769).

4.5.8. The effect of temperature on the ability of *C. difficile* spores to germinate

To determine if the temperature at which the insects were stored (25^{o}C) had an effect on the ability of *C. difficile* spores to germinate; we looked at the ability of spores to germinate at 25^{o}C, 30^{o}C and 37^{o}C. Results suggest *C. difficile* spore of strains DS1813 and DS1748 germinate better at 37^{o}C (Figure 4.12). At 25^{o}C there was a significant reduction in germination for DS1813, which appears to only germinate well at 37^{o}C. DS1748 appears to be able to germinate with varying ability across 25^{o}C, 30^{o}C and 37^{o}C.



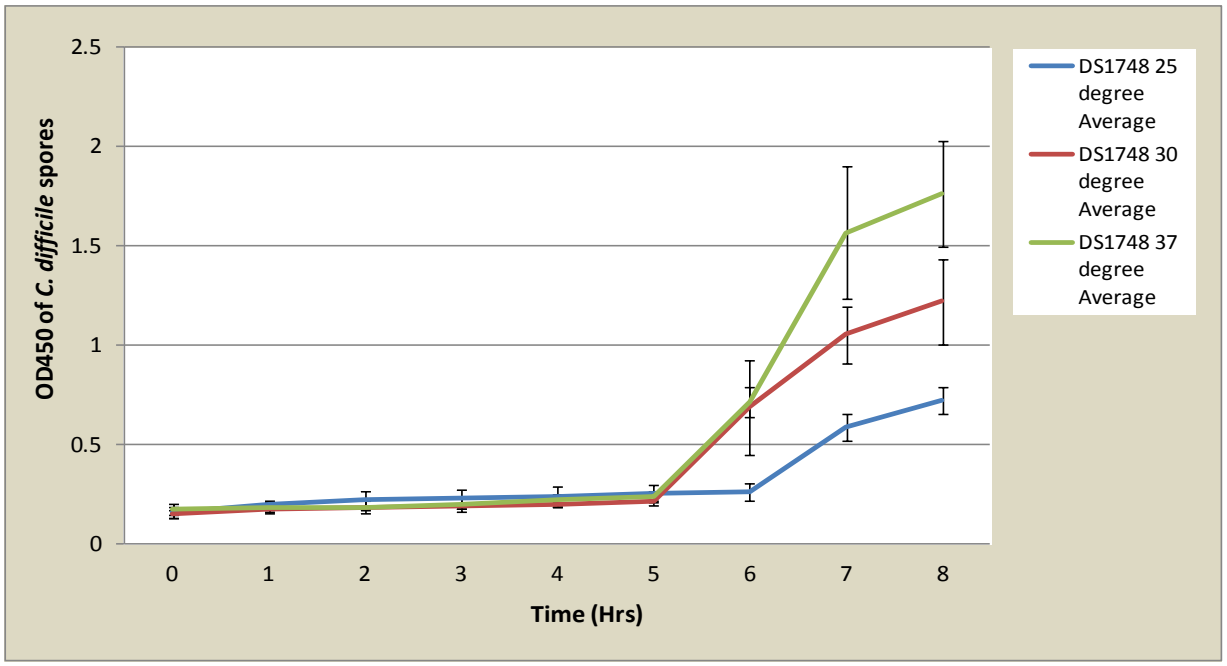


Figure 4.12: Germination graphs of DS1813 spores and DS1748 spores

Spores were subjected to three different temperatures and germination monitored. The optical density (450 nm) of the spores was measured each hour over a period of 8 hours. An increase in optical density indicates germination has occurred. The optical density of DS1813 was determined at 25°C **blue**, 30°C **red** and 37°C **green**. The optical density of DS1748 was determined at 25°C **blue**, 30°C **red** and 37°C **green**. One way ANOVA revealed highly significant differences in DS1813 germination at 37°C. Analysis of DS1748 revealed a significant differences in the germination abilities of spores at different temperatures (P = 0.002).

Statistical analysis via one way ANOVA revealed that there were highly significant differences in the germination ability of DS1813 spores between temperatures ($P < 0.001$). The most significant was at 37^{o}C as deduced via Dunnett's multiple comparison post hoc test. Analysis of DS1748 germination revealed significant differences between the test temperatures ($P = 0.002$).

4.5.9. Detection of Chitinase activity in spores of DS1813 & DS1748

To determine if the peroxiredoxin/chitinase gene in *C. difficile* was biologically functional, the chitinase activity of germinating and ungerminated spores of *C. difficile* strains DS1813 and DS1748 and their vegetative cells were assessed using the Chitinase Assay kit (Sigma Aldrich, Dorset, UK). Vegetative cells of both strains failed to demonstrate chitinase activity. The same was true for ungerminated spores (Figure 4.13). However germinating spores (treated with the germinant sodium taurocholate) showed evidence of chitinase activity similar to that recently reported by (Permpoonpattna *et al.*, 2011). Statistical analysis using one way ANOVA ($P < 0.0001$) revealed a significant difference when comparing the mean results between each strain with and without addition of taurocholate. When comparing the effect of sodium taurocholate on the strains, the most significant results were from spores of strains DS1813 and DS1748 when germinating with 3% ST, and spores of strain DS1813 germinating with 5% ST when compared to the blank in a Dunnetts multiple comparison post-hoc test.

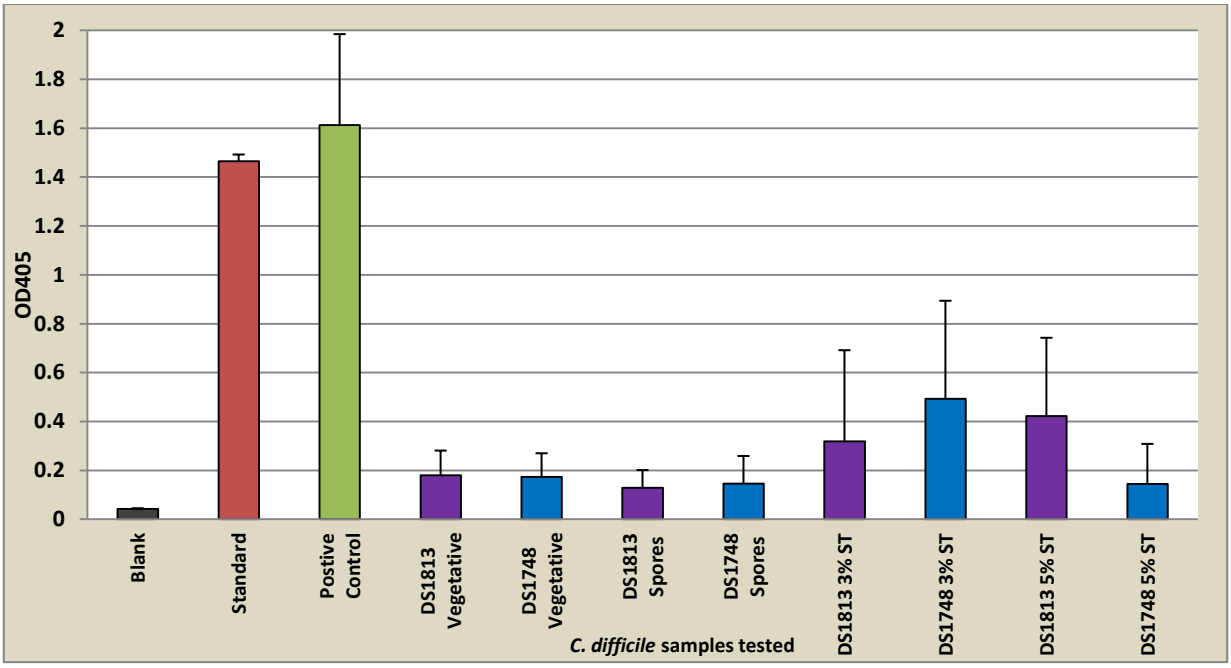


Figure 4.13: Chitinase activity of *C. difficile*

Vegetative cells and spores of *C. difficile* strains DS1813 and DS1748 were tested for chitinase activity using a chitinase assay kit (Sigma Aldrich). A blank solution was used to calibrate the assay, and a Standard marker for significant chitinase activity. The Positive control (Chitinase from *Trichoderma viride*) was also used to determine if the reaction was viable. Vegetative cells of both strains did not demonstrate chitinase activity upon testing. Untreated spores did not exhibit chitinase activity; however spores treated with the germinant ST appeared to show a marked increase in chitinase activity.

4.6. DISCUSSION

Bioinformatic analysis of the genome of *C. difficile* revealed the presence of genes encoding proteins homologous to known insect virulence factors. Analysis of the genes encoding *C. difficile tcdA* and *tcdB* revealed homology to the mcf proteins which are known play a role in insect pathogenicity. Furthermore individual structural protein domains of each toxin revealed significant structural homology to proteins associated with insect infection. For toxin A this included *P. fluorescens* putative toxin A and a *P. luminescens subsp. laumondii* hypothetical protein. The structure of toxin B also has significant association with *P. fluorescens* putative toxin A and the mcf toxins of *P. luminescens* and *P. asymbiotica*.

The putative chitinase/peroxiredoxin gene within the *C. difficile* genome has been suggested to be associated with the spore coat proteins and the ability of the spore to germinate (Permpoonpattna *et al.,* 2011). Thus the ability of selected *C. difficile* vegetative cells and germinating and ungerminating spores to produce a biologically functional chitinase was assessed. As expected we found that chitinase was released from germinating spores of both strains. This suggests that the spores may be releasing chitinase during germination to potentially enable macromolecular degradation of soil elements, such as fungi and dead insects, and thus allow nutrient uptake. Spores from other members of the clostridial family are usually saprophytic, residing in soil so it may be the case that the chitinase release from germinating spores is a relic of *C. difficile's* evolutionary history. Hence it could also be assumed that *C. difficile* may have divergently evolved alongside a saprophytic *Clostridium* into a human pathogen (Figure 1.11).

To establish whether *C. difficile* is able to infect insects we employed the *M. sexta* infection model host system. We found that *C. difficile* spores and vegetative cells were unable to elicit symptoms of infection upon either injected or oral delivery at 25^{o}C and 37^{o}C. While challenge at 37^{o}C with *C. difficile* had no obvious effects on the health of the insect, we did observe that the majority of insects followed their normal life cycle at this temperature in that they pupated after 5 days in contrast to unchallenged controls which were unable to pupate correctly. The presence of the bile salt germinant sodium taurocholate had no affect on infectivity.

This suggests that immune specific priming leading to developmental changes may have occurred following exposure to spores (Alberea & Schaub, 2008). In order to defend against infection, insects possess an innate immune system capable of pathogen recognition and elimination, releasing small effector molecules and antimicrobial peptides (Waterfield *et al.,* 2004). This may have occurred upon direct injection of *C. difficile* into the *Manduca* haemocoel, inducing an adaptive immune response. It would be of interest to examine which immune defence molecules are expressed in *Manduca* upon challenge with *C. difficile.*

Interestingly experiments using the haemolymph of the *Manduca* revealed that compounds within the haemolymph reduced the viability of *C. difficile* vegetative cells and spores after inoculation. Thus as there was a reduction in spore viability, and no vegetative *C. difficile* survived, credence can be given to the theory that potential antimicrobial peptides within the haemolymph are active. There was no neutralisation step involved in the experiment as the potential antimicrobial compounds have not been pinpointed. Indeed a further explanation of the lack of pathology is that *C. difficile* toxin A and B receptors are not present within the insect.

To determine if the incubation temperatures at which the experiments were performed influenced the ability of the spores to germinate, we assessed spore germination at the following range of temperatures: 25^{o}C, 30^{o}C and 37^{o}C. There appeared to be a significant increase in germination ability for both DS1813 ($P < 0.001$) and DS1748 strains ($P = 0.002$) at 37^{o}C. Thus incubation of the *Manduca* at 37^{o}C should have elicited an infection; however as infection did not occur we can conclude that temperature had no effect on the ability of *C. difficile* to infect the *Manduca*.

We also attempted utilise *M. sexta* as a potential model for the colonisation of *C. difficile*. Two strains with differing hydrophobic characteristics were tested; the highly hydrophobic DS1813 and hydrophilic DS1748. We established that there was no difference in the passage or adherence ability of either strain through the *Manduca*. Indeed there was also no difference between antibiotic treated and non-antibiotic treated food upon the colonisation abilities of *C. difficile.* There was no replication of *C. difficile* in the *Manduca*, and thus we have concluded that it is a poor model with which to study *C. difficile* colonisation.

Thus we have analysed the genome of *C. difficile* for insect associated genes and homologs, and have found genes which are similar to known insect virulence factors. We have attempted to establish an insect infection model for *C. difficile* but were unsuccessful. The origins of *C. difficile* and its evolution have yet to be established. While we cannot comment upon the evolutionary origins of the bacterium itself, we have found that toxin B of *C. difficile* is structurally similar to the insect associated mcf toxins.

Toxin A does not share this homology. This may suggest that toxin B may have originated as the first infective toxin from an insect associated ancestor, and thus the variant toxin A may have arisen as a duplication event. In contrast to toxin B which is structurally stable, toxin A has variations within its gene coding sequences and thus this structural instability is also indicative of gene duplication (Rupnik *et al.*, 1999).

CHAPTER 5
RAPID METHODS OF DETECTING TOXIN A AND TOXIN B USING MAMEF

5.1. ABSTRACT

To enable the development of a rapid sensor, DNA probes specific for conserved regions within toxins A and B were identified as described in Chapter 3. The probes with the highest sensitivity were incorporated into a rapid MAMEF based assay and were subsequently assessed for their ability to detect *C. difficile* (vegetative and spores) in a range of organic matrices which included milk and faecal matter. Using this platform we were able to detect as few as 100 spores in a 500 µl suspension of faecal matter within 40 seconds. This prototype biosensor had the potential to be developed into a real time beside diagnostic assay.

5.2. INTRODUCTION

Microwave accelerated metal enhanced fluorescence (MAMEF) is a technique which utilises low power microwave heating and a phenomenon known as metal enhanced fluorescence (MEF). The underlying principal of the technology is based on the selective heating of water, whereby the aqueous medium used in the assay is heated via microwave power to a higher temperature than plasmonic particles (electrons) within the metal surface. In this case silver is used as plasmons can be easily excited using this metal (plasmon resonance effect). Therefore a temperature gradient between the cold metal surface and the warm aqueous solution is formed, which allows DNA or other biomarkers to be recognised, bound and subsequently for fluorescence to occur (Geddes, 2010). Non radiative energy transfer occurs between fluorophores and plasmon electrons in a non continuous film. The metal surface becomes a mirror which intensifies the fluorophore signal (Geddes & Lakowicz, 2002; Das *et al.*, 2002). The assay does not require prior purification of organic material and so it highly rapid and is able to detect targets with organic material without quenching of the signal. This means a signal can be detected easily in raw organic material.

Conventional methods currently used to detect *C. difficile* infection as described in Section 1.6. Although these methods require long turnover times and are laborious they are used because of their sensitivity and specificity. It is important to develop a rapid detection method for *C. difficile* to prevent transfer of micro-organisms between patients and to enable healthcare professionals to provide an efficient treatment regimen for the patient. The prevention of cross infection between patients would reduce treatment times and costs to the NHS.

Currently it costs approximately £400 per day for each infected patient, with the estimated length of stay for patients with *C. difficile* being between 14-27 days. Thus the estimated cost of treating a patient with *C. difficile* in Wales would be £8 613, and given that over 800 patients with *C. difficile*-related infections are seen per year by the University Hospital Wales, Cardiff, this represents a considerable avoidable cost burden (Cardiff & Vale NHS Trust, 2010).

The laboratory tests currently used to detect *C. difficile* include toxinogenic assays, PCR methods, antigen detection (GDH) and ELISA based assays as described in Section 1.6. The challenges when using these methods include the time taken to conduct the tests, and their lack of sensitivity. Symptoms of *C. difficile* infection normally manifest within 24-48 hours of infection (as described Section 1.2), and thus there is a pressing need to diagnose *C. difficile* during the early, subclinical phase of infection to enable treatment and to minimise cross transmission within the hospital environment.

A biosensor capable of detecting the presence of *C. difficile* in clinical samples in <60 seconds without the need for complex pre-processing would dramatically reduce the time required to obtain confirmation of the presence of the organism compared to current laboratory assays. Ideally such an assay would be able to detect *C. difficile* with high sensitivity and specificity, whilst being inexpensive to run and low maintenance.

A technology approach which offers the potential to develop such a rapid sensor is MAMEF. The approach has previously been employed to detect a range of pathogens including *B. anthracis*, *Salmonella typhimurium* and *C. trachomatis* (Aslan *et al.*, 2008; Zhang *et al*., 2011, Tennant *et al*., 2011). The technology breaks open vegetative bacteria and spores releasing target DNA which is rapidly detected by pathogen specific probes. In the case of *B. anthracis*, a Gram positive spore former similar to *C. difficile*, pathogen specific DNA targets were detected within 60 seconds of the start of the assay.

5.3. RESEARCH AIMS

The aims of this chapter are to:

- To determine the ability of conserved probes identified from *C. difficile* Toxin A and Toxin B (as *per* Chapter 3) to detect target DNA using a MAMEF detector platform
- To determine microwave conditions which optimise the release of target DNA from *C. difficile* spores and vegetative bacteria
- Determine the ability of the sensor platform to identify *C. difficile* target DNA within a range of organic matrices
- Determine the sensitivity of the assay using known concentrations of *C. difficile* spores and vegetative cells spiked into organic matrices

5.4. MATERIALS & METHODS

5.4.1. Bacterial Strains, biological fluids and genomic DNA

Genomic DNA was isolated from *C. difficile* strain CD630 (supplied by NTCC, HPA, UK) using Chelex 100 resin (BioRad Laboratories, UK) as detailed in Chapter 3. Whole milk (3.5g fat content) was purchased from Whole Foods (Harbour East, Baltimore, MD, USA). PBS (0.01 M phosphate buffer, 0.0027 M potassium chloride and 0.137 M sodium chloride) was made by dissolving 1 tablet in 200 ml diH_2O (Sigma Aldrich, USA). Human faecal matter was provided by a healthy volunteer.

5.4.2. Anchor and Fluorescent Probes and target DNA

Probes specific for conserved regions of toxin A and toxin B of *C. difficile* were designed and validated as described in Chapter 3. Probes were modified as follows to enable incorporation into the MAMEF detection platform (Figure 5.1); a thiol group was added to the 5' region of the anchor probe to enable binding of the DNA to the surface of the silver island film (SiF) while the capture probe was labelled with an Alexa® (Invitrogen, USA) fluorophore (Aslan *et al.,* 2008; Zhang *et al.,* 2011). The Alexa fluorophore 488 (green) was used to label toxin A and the Alexa fluorophore 594 (red) was used to label toxin B. These fluorophores were chosen as their emissions were at wider ends of the fluorescent spectrum and thus would yield two different colours upon excitation.

	Toxin A	**Toxin B**
Anchor Probe	Thiol-TTTTT-TTTAATACTAACACTGC	Thiol- TTTTTT-CAAGACTCTATTATAG
Capture Probe	Alexa-488 – TGTTGCAGTTACTGGATGGCAA	Alexa-594-TAAGTGCAAATCAATATGAAG
Synthetic oligonucleotides	AAATTATGATTGTGACGTAATCCCA ATACAACGTCAATGACCTACCGTT	AGTTCTGAGATAATATCTAATCCC AAT ATTCACGTTTAGTTATACTTG

Table 5.1: Table of Oligonucleotides

The anchor probes for both toxins have a run of 5 consecutive Ts (thymine bases) included to increase the flexibility of the probes in detecting target DNA once bound to the silver surface. The capture probes have an Alexa ® group added to aid fluorescent detection. The target region negative strand was also synthesised to aid assay development.

The nucleotide sequences of the anchor and capture probes from each toxin are shown in Table 5.1. The common regions in anchor probes were preceded by a run of 5 consecutive Ts (thymine bases) included to increase the flexibility of the probe design once bound to the silver surface, and enhanced by the inclusion of a C at the terminal of the probe to which a thiol group was subsequently added (Aslan *et al.*, 2008). The negative strand of the target region was also synthesised to bind to the capture and anchor probes (Invitrogen, USA) (Table 5.1).

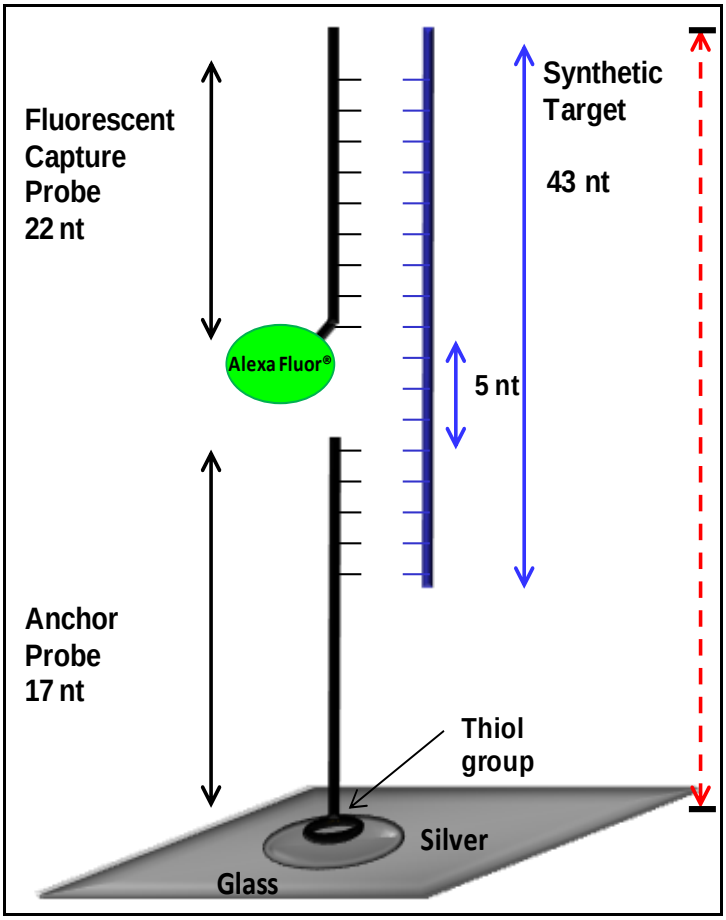


Figure 5.1: Configuration of 3 piece DNA detection assay

This configuration as used for the design of both toxin A and toxin B probes. The anchor probe is fixed to the SiF by the addition of the thiol group and is 17 oligonucleotides in length. The fluorescent probe is 22 oligonucleotides in length and has the Alexa® attached to the 3' end. When a target DNA strand binds to the anchor DNA, it should also bind to the fluorescent probe causing release of plasmons which under laser light induce fluorescence. Inadequate binding does not induce fluorescence. The 5 nucleotide spacer region is included to aid the fluorescent reaction to occur.

5.4.3. Formation of gold lysing triangles to microwave *C. difficile*

To focus the microwave power in the microwave, gold lysing triangles were used. Gold lysing triangles were provided by the Institute of Fluorescence (University of Maryland). Glass microscope slides (Starfrost®, LightLabs, USA) were covered with a paper mask (12.5 mm in size and 1 mm gap size) leaving a triangle bow-tie region exposed. Gold was deposited onto the glass microscope slides using a BOC Edwards 306 Auto vacuum E-beam evaporation deposition unit, in equilateral gold triangles of 12.5 mm and 100 nm thick at 3.0×10^{-6} Torr (Tennant *et al.*, 2011). Two layers of self-adhesive silicon isolators (Sigma Aldrich, USA) (diameter = 2.5 mm) were placed on top of the bow-tie regions, creating a lysing chamber (Zhang *et al.*, 2011) as shown in Figure 5.2.

5.4.4. The release of *C. difficile* DNA from vegetative cells and spores using microwave irradiation

To optimise DNA release the following approaches were investigated:

(i) Boiling *C. difficile* (1×10^5 cfu/ml) was suspended into PBS buffer into 15 ml Falcon tubes (Fisher Scientific, USA) and boiled in a water bath on a hot plate (Fisher Scientific, USA) for 3 h. After each 30 min, a 100 µl sample was taken from the boiling spores and cooled on ice. The DNA released from *C. difficile* was examined using gel electrophoresis.

(ii) Gold Triangles *C. difficile* (1×10^5 cfu/ml) was suspended into PBS buffer and pipetted at a volume of 500 µl into the gold tie lysing chamber and exposed to a 15 s microwave pulse at 80% power in a GE microwave Model No. JE2160BF01, kW 1.65 (M/W). The microwaved solution was examined for the presence of viable organisms. Further experiments involved spiking milk and human faecal matter (both diluted with PBS buffer) with varying concentrations of *C. difficile* and then performing microwave radiation. Each experiment was conducted in triplicate.

5.4.5. Bacterial quantification after focussed microwave irradiation

(i) Vegetative cells: The control samples (4×10^6 cfu/ml) were serially diluted using Miles Misra (1938) drop count method (as described in Chapter 2). Briefly 20 µl of bacterial sample was serially diluted in 1:10 ratio in 180 µl BHI broth and drop counts performed on BHI agar and incubated at 37^{o}C in a 3.4L anaerobic jar (Oxoid, Basingstoke, Hampshire, UK) with an anaerobic gas generating kit (Oxoid, Basingstoke,

Hampshire, UK). Post lysis the number of remaining viable bacteria (cfu/ml) was enumerated using the Miles Misra method as above (1938).

(ii) Spores: The control samples (1.33 x 10^7 cfu/ml) were serially diluted using Miles Misra (1938) drop count method as described in Chapter 2. Briefly 20 µl of the spore sample was serially diluted in 1:10 ratio in 180 µl BHI broth and drop counts performed on BHI agar supplemented with the spore germinant sodium taurocholate (0.1%). Plates were incubated at 37^{o}C in a 3.4L anaerobic jar (Oxoid, Basingstoke, Hampshire, UK) with an anaerobic gas generating kit (Oxoid, Basingstoke, Hampshire, UK). Before DNA release the sample was heated at 80^{o}C for 10 min to ensure removal of any vegetative cells. After lysis the number of remaining viable spores (cfu/ml) was enumerated using the Miles- Misra method as above (1938).

5.4.6. Gel electrophoresis of samples

Post microwave irradiation the sample was centrifuged at 3000 g for 15 min (Heraeus Primo R; Fisher Scientific, Waltham, MA, USA) and the supernatant resuspended in a ratio of 1:2 ethanol (Sigma Aldrich, USA). The supernatant was removed and resuspended in 100 µl sdw. The samples separated on 2% (w/v) agarose (Sigma Aldrich, USA) gels prepared in 1 X TAE buffer (Sigma Aldrich, USA) with 5 µl Ethidium Bromide (Sigma Aldrich, USA) and were run at 70 V using a BioRad electrophoresis Power Pac (Bio Rad Laboratories). Pre-lysis samples were included as a control.

5.4.7. Silver Island Film formation on glass substrate

Silver island Films (SiFs) were prepared on Silane-prep ™ glass slides (Sigma Aldrich, USA) to enable silver adhesion, as follows. A solution of silver nitrate was made by adding 0.5 g in 60 ml dH_2O to 200 ml of freshly prepared 5% (w/v) sodium hydroxide solution and 2 ml ammonium hydroxide. The solution was continuously stirred at RT, and then cooled to 5^{o}C in an ice bath. The slides were soaked in the solution and 15 ml of fresh D-glucose was added (0.72 g in 15 ml dH_2O). The temperature of the solution was raised to 40 ^{o}C and as the colour of the mixture turned from yellow/green to yellow-brown, the slides were removed from the mixture, washed with dH_2O, and sonicated using an MSE Soniprep 150 sonicator (Sanyo) for 1 min at RT (Tennant *et al.,* 2011). SiFs used in this study were between an OD_{450} of 0.4 - 0.5 (Figure 5.2).

5.4.8. Preparation of MAMEF assay platform for detection of *C. difficile* DNA

Glass slides with SiFs deposited were coated with self-adhesive silicon isolators containing oval wells (2.0 mm=diameter, 632 mm=length, 619 mm=width) (Sigma Aldrich, USA). The thiolated anchor probe was decapped to remove the thiol groups and enable binding to the SiF. This was achieved by diluting 40 µM anchor probe into 100 µl of 1 M Tris- EDTA (TE) buffer and adding 9 µl to 250 mM of 20 µl dithiothreitol (DTT) (38.5 mg DTT into 1 ml TE buffer). The mixture was incubated at RT for 60 min. The decapped anchor probe (1 µM) was diluted into 4 ml TE buffer and 100 µl of anchor probe was added to and incubated in each oval well of the SiF for 75 min. After incubation the anchor was removed and 50 µl of 1 µM Alexa-Fluor® probe was added to 50 µl of DNA from the target organism. The SiFs containing the bound probes and DNA was then incubated using MAMEF for 25 s in a microwave cavity at 20% microwave power, GE microwave Model No. JE2160BF01, 1.65 kiloWatts. In the presence of target DNA the three piece assay is complete and enhanced fluorescence can be observed. The sample was then removed from the well after MAMEF and the well washed with 100 µl TE buffer 3 times. Each experiment was conducted in triplicate.

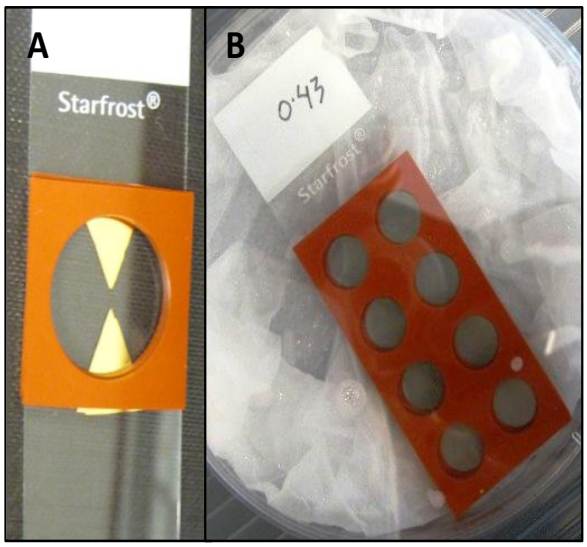


Figure 5.2: Gold lysing triangle on Starfrost ® slide and Silver island Film

(A) Gold triangle with silicone isolators added to create lysing chamber. In this lysing chamber 500 µl volume can be added and irradiated. (B) The SiF here is at an OD_{450} of 0.43 and has multi-well silicone isolators added. Each well can hold an individual DNA assay reaction; hence multiple repeats can be conducted. The SiF must be kept moist to prevent the silver from being oxidised.

5.4.9. Detection and fluorescence spectroscopy

The presence of target DNA was confirmed by the generation of a fluorescent signal following excitation with laser light. The following fluorophores were employed (details in Table 5.2). Fluorescence was emitted by the DNA MAMEF capture assay and measured using a diode laser (Figure 5.3) and a Fibre Optic Spectrometer (HD2000) (Ocean Optics, Inc) by collecting the emission intensity (I) through a notch filter.

Alexa Fluor® Dye	Absorption (nm)	Emission (nm)	Fluorescent Colour	*C. difficile* Toxin	Laser Notch Filter (nm)
Alexa 488	**495**	**519**	**GREEN**	**Toxin A**	**473**
Alexa 594	**590**	**617**	**RED**	**Toxin B**	**610**

Table 5.2: Table of fluorescence characteristics

Alexa Fluor dyes used to detect *C. difficile* toxins A and B. The absorbance and their fluorescence emission (nm) coupled with the laser filter is used to give the fluorescence intensity measurement.

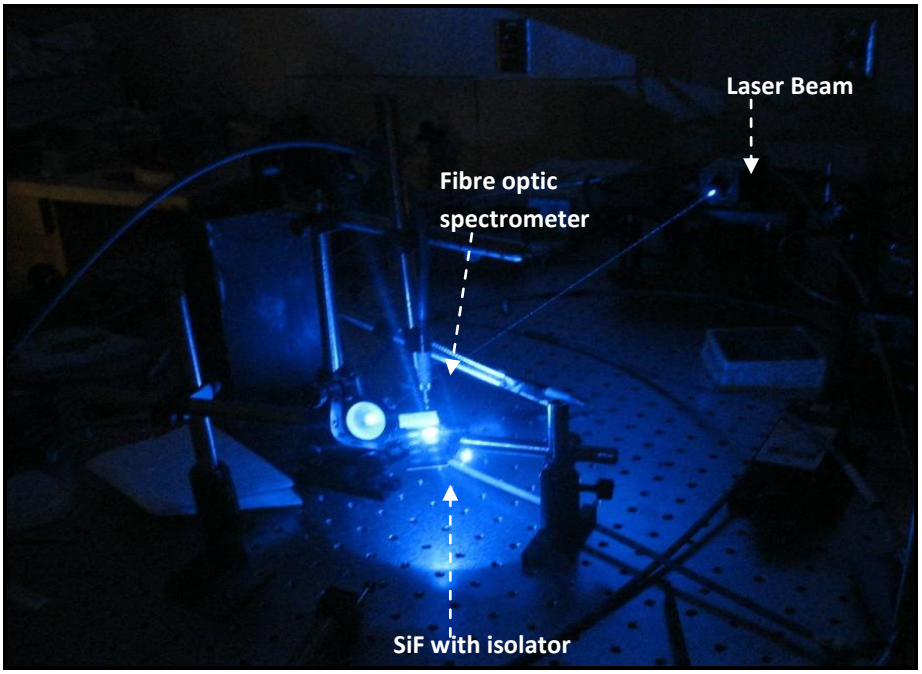


Figure 5.3: Laser platform used to excite the samples

The laser beam is directed into each well using glass prisms, and fluorescence is detected using a notch filter and a Fibre optic spectrometer.

5.4.10. Statistical Analysis

Statistical analysis was performed using GraphPad Prism version 5.04 for Windows, (GraphPad Software, La Jolla California USA, www.graphpad.com). Three replicates (n=3) were performed for each detection experiment. Statistical significant differences were tested for using one way analysis of variance (ANOVA) at the 95% confidence interval in conjunction with a Kruskall-Wallis test. Dunnett's post test was employed to compare sample data to control data. A P value of <0.05 was considered significant (Bowker & Randerson, 2006).

5.5. RESULTS

5.5.1. Detection of synthetic DNA targets by MAMEF

The ability of the toxin A and B probes to detect synthetic target DNA (43 nucleotides in length) was determined. Target DNA was diluted in TE buffer to give the following range of concentrations; 1 nM, 10 nM, 100 nM, 500 nM, 1000 nM and 100 μM with TE buffer serving as a control. The lowest concentration of DNA detected by each probe was 1 nM (Figure 5.4 A & C). Targets were readily detected by MAMEF as can be seen in Figure 5.4 (B & D). As the concentration of the synthetic target increases the fluorescent intensity increases. Taking the concentration of 10 nM as an example, the intensity for Toxin A is 20 AU and Toxin B is 45 AU. For toxin A probes the excitation is 495 nm, and emission: 519 nm, and toxin B excitation is 590 nm and emission is 617 nm.

The difference in signal intensity seen between toxin A and toxin B may be due to several factors; including the optical density of the SiF used, the concentration and specificity of the anchor and capture probes and the excitation wavelength of the laser used to produce fluorescent emission. For example the concentration of anchor probe to toxin B in this experiment was ten-fold higher that for toxin A.

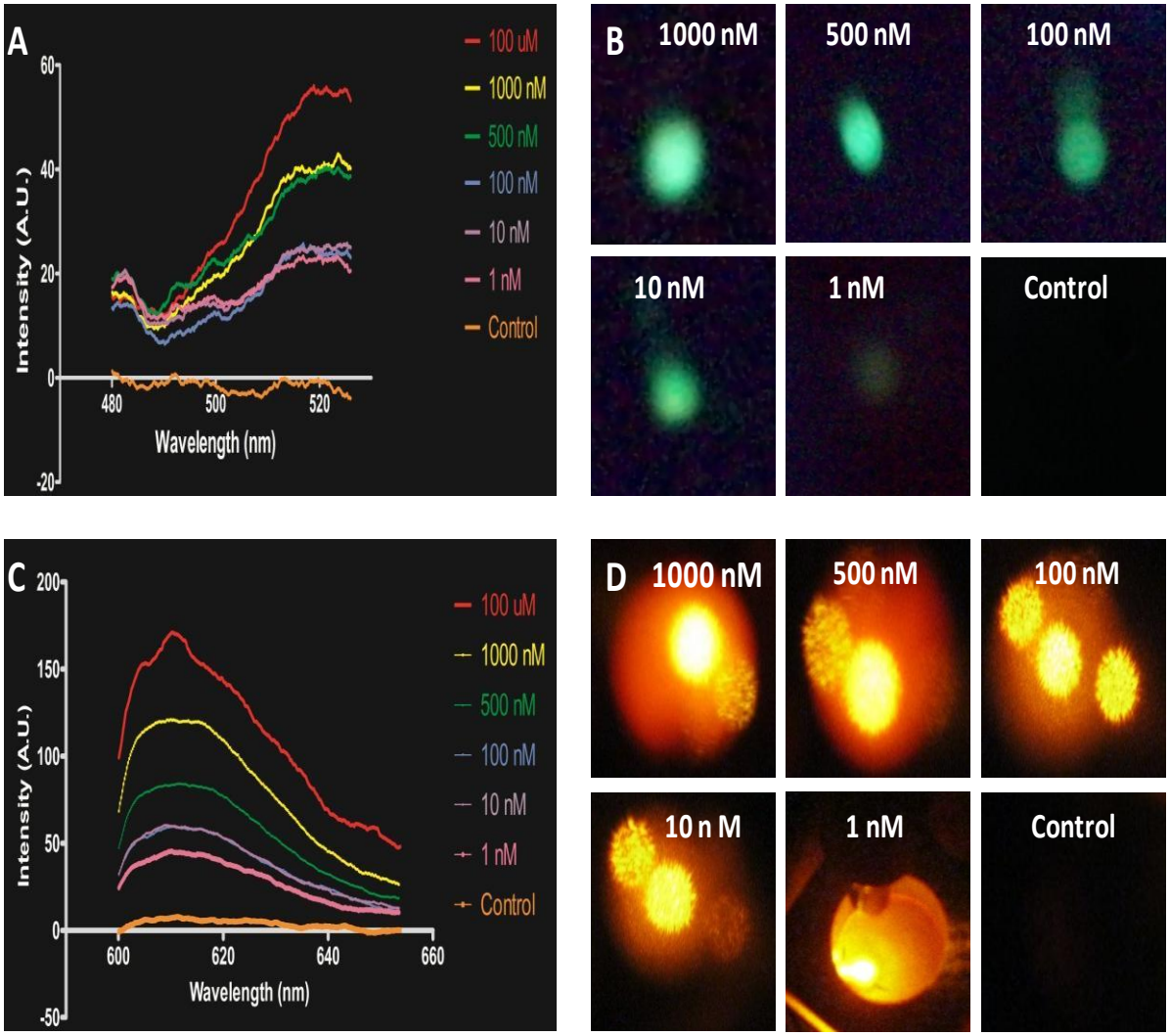


Figure 5.4: Detection of various concentrations of synthetic oligonucleotides in TE buffer by MAMEF
(A) This graph shows detection of toxin A synthetic oligonucleotide. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases. (B) The fluorescent signal produced at each concentration. Excitation: 495 nm, Emission: 519 nm. (C) The graph shows detection of toxin B synthetic oligonucleotide. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases. (D) The fluorescent signal produced at each concentration. Excitation: 590 nm, Emission: 617 nm**.**

5.5.2. Detection of synthetic DNA in whole milk by MAMEF

To determine the ability of these probes to detect toxins A and B in the presence of biological material, target DNA was mixed with whole milk. The whole milk was diluted by 10% with PBS and the synthetic target DNA was diluted in the milk to final concentrations of 1 nM, 10 nM, 100 nM, 500 nM and 1000 nM. As can be seen in Figure 5.5 (B & D) the limit of sensitivity was similar to that seen with PBS at 1nM for both Toxin A and B. Interestingly reducing the concentration of Toxin B anchor probe from 10

µM to 1 µM did not affect the sensitivity of the assay. As before there was an increase in fluorescent intensity as the concentration of the target oligonucleotide increased although the intensity of the signal was less (~5 AU) than that seen in the presence of PBS when comparing the results from experiments using 1 µM anchor (Figure 5.4 A; Figure 5.5 A & C). At the concentration of 10 nM the intensity in the presence of milk for Toxin A is ~20 AU and Toxin B is 25 AU.

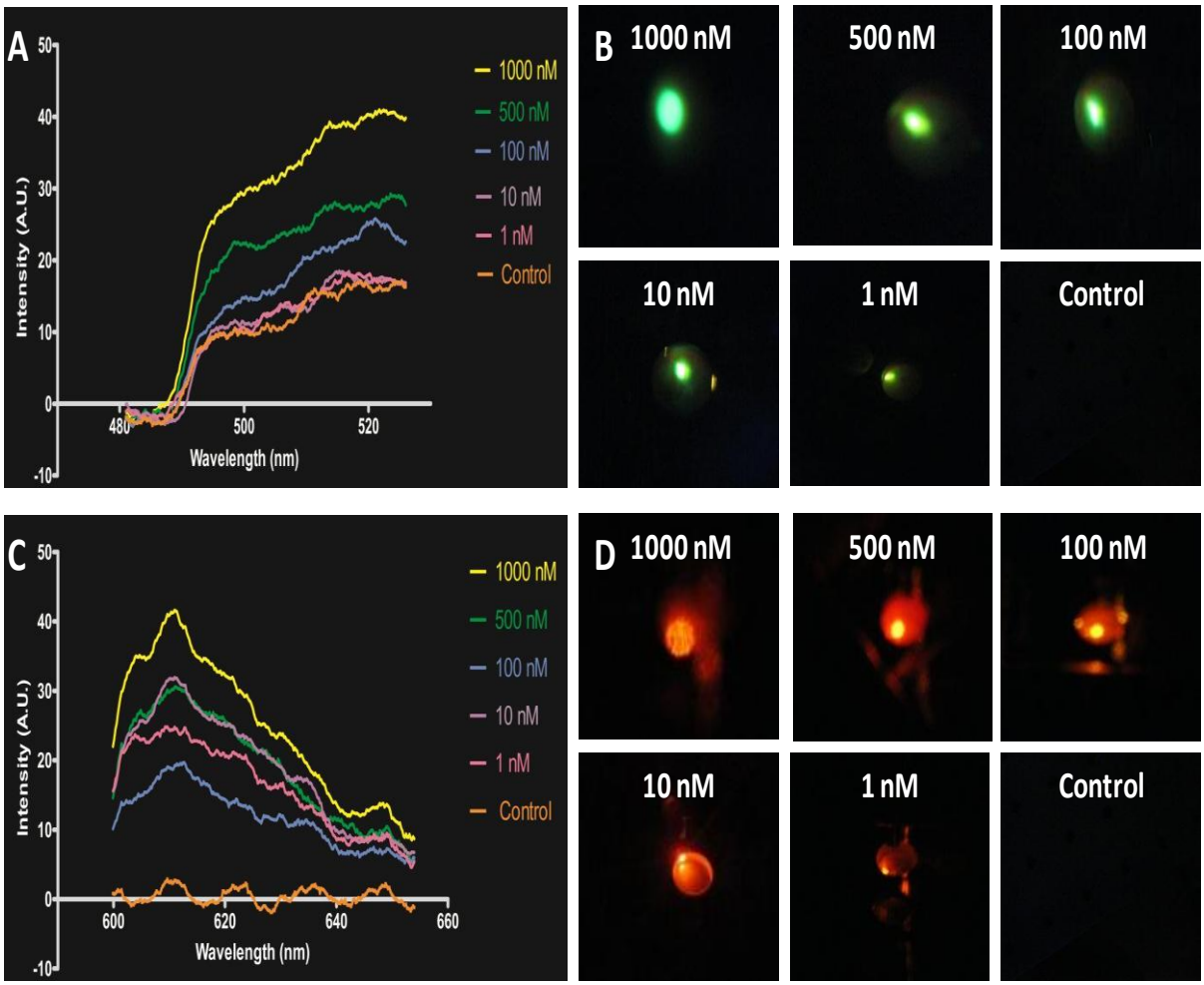


Figure 5.5: Detection of various concentrations of synthetic oligonucleotides in whole milk by MAMEF

(A) This graph shows detection of toxin A synthetic oligonucleotide in whole milk. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases. The signal has not been quenched by the milk. (B) The fluorescent signal produced at each concentration. Excitation: 495 nm, Emission: 519 nm. (C) The graph shows detection of toxin B synthetic oligonucleotide in the presence of whole milk. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases. The signal is strong with no evidence of quenching. (D) The fluorescent signal produced at each concentration. Excitation: 590 nm, Emission: 617 nm.

5.5.3. Detection of synthetic DNA in human faecal matter by MAMEF

The aim of this study is to detect *C. difficile* directly from human faecal samples. Therefore experiments were performed where target DNA was mixed with human faecal matter. Faecal matter was diluted by 50% with PBS and the synthetic target DNA was then added to give the following range of concentrations; 1 nM, 10 nM, 100 nM, 500 nM and 1000 nM. The limit of sensitivity was 1 nM for toxin A and 1 nM for toxin B. As expected an increase in fluorescent intensity was observed as the concentration of target oligonucleotide increased (Figure 5.6 A & B). The signal intensity however was less than observed in the presence of PBS and whole milk for both toxins A and B (Figure 5.4 A; Figure 5.5 A & C; Figure 5.6 A & B). At the concentration of 10 nM the intensity in the presence of milk for Toxin A is ~10 AU and Toxin B is 12 AU which are similar.

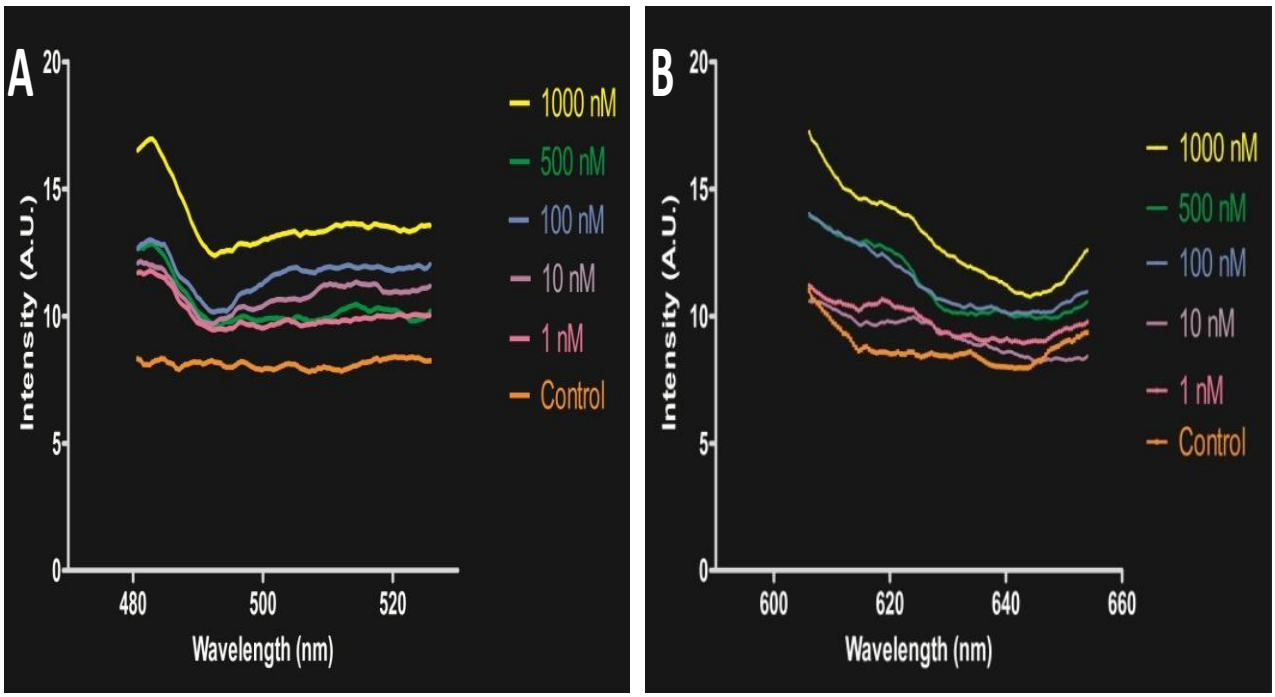


Figure 5.6: Detection of various concentrations of synthetic oligonucleotides in faecal matter by MAMEF

(A) This graph shows detection of toxin A synthetic oligonucleotide in the presence of human faecal matter. It can be seen that there is still a strong signal. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases. (B) The first graph shows detection of toxin B synthetic oligonucleotide. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases.

5.5.4. Statistical analysis of synthetic DNA detection within organic matrices

Statistical analysis of differences in synthetic target DNA detection of toxin A was performed using one way ANOVA, which revealed a P value of 0.0534. Thus as $P > 0.05$, this indicates there is no significant difference between the organic matrices used (TE buffer, Milk and faecal matter). This was further confirmed using the Kruskall-Wallis test where $P = 0.2588$.

One way ANOVA results of toxin B synthetic target DNA detection revealed a P value of 0.107. As $P > 0.05$ there is no significant difference between the organic matrices used (TE buffer, milk and faecal matter). This was further confirmed using the Kruskall-Wallis test where $P = 0.158$. Also there was a significant difference between the control sample of toxin A and B and the test concentrations as determined by Dunnetts multiple comparison of variance test (Graphpad Prism 5).

5.5.5. Detection of genomic DNA from *C. difficile* Strain 630

To establish if the toxin A and B probes could detect targets within gDNA, we employed DNA from *C. difficile* strain CD630. Genomic DNA from CD630 was microwaved in PBS buffer for a shorter time of 8 seconds at 70% power to determine if target DNA could be detected. A control of unmicrowaved gDNA in PBS was employed. The toxin probes were able to detect gDNA at concentrations of 0.010 µg before microwave treatment, and 0.010 µg and 0.005 µg after microwave treatment (Figure 5.7 A & B). The limit of sensitivity was 0.005 µg gDNA for both toxin A and toxin B. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases (Figure 5.7 A & B). The signal intensity however was similar to synthetic target intensity observed in the presence of PBS, and greater than that observed when the synthetic targets were tested in the presence of whole milk and faecal matter (Figure 5.4 A; Figure 5.5 A & C; Figure 5.6 A & B; Figure 5.6).

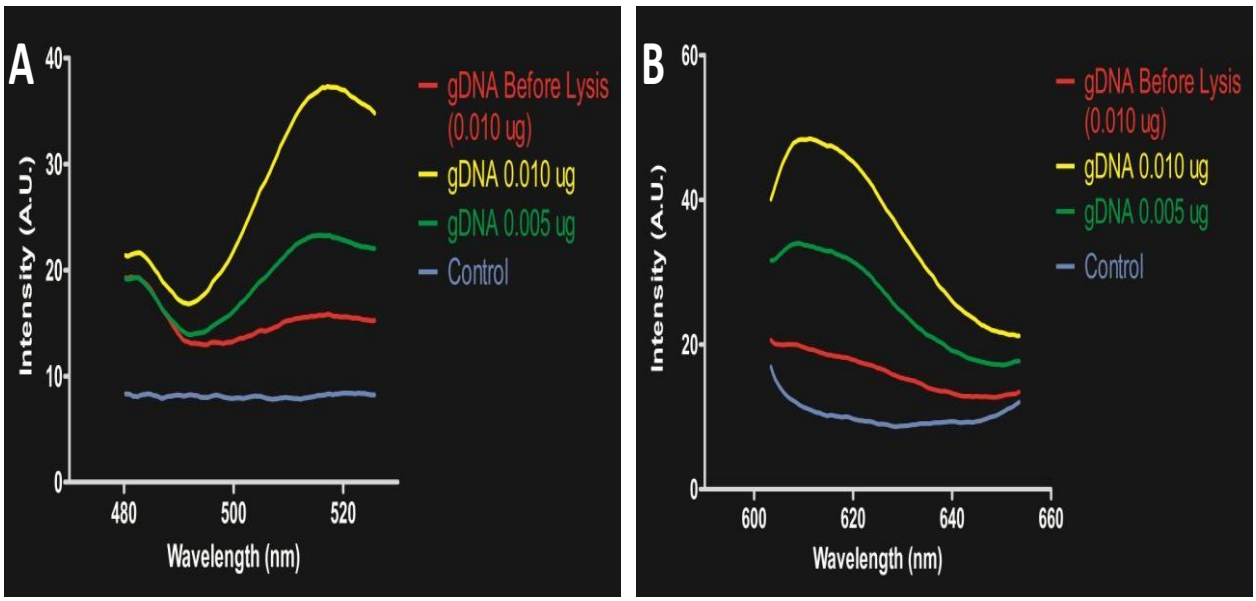


Figure 5.7: Detection of target DNA within various concentrations of genomic DNA by MAMEF

(A) This graph shows detection of toxin A synthetic oligonucleotide in the presence of human faecal matter. It can be seen that there is still a strong signal. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases. (B) The first graph shows detection of toxin B synthetic oligonucleotide. There is an increase in fluorescent intensity as the concentration of the oligonucleotide increases.

5.5.6. Release of target DNA from a *C. difficile* spore preparation via boiling

A boiling based method was developed to release DNA from *C. difficile*. The release of DNA from spores was assessed by mixing spores from a ~80% preparation with either PBS or diH_20, followed by boiling for 3 hours with samples removed every 30 min to determine if DNA has been released using gel electrophoresis. As can be seen in Figure 5.8, there appeared to be no DNA release from water-treated spores. However in contrast, spores suspended in PBS produced a smear upon gel electrophoresis which was thought to represent fragmented DNA. Therefore PBS was used as the diluent for samples used in further experiments.

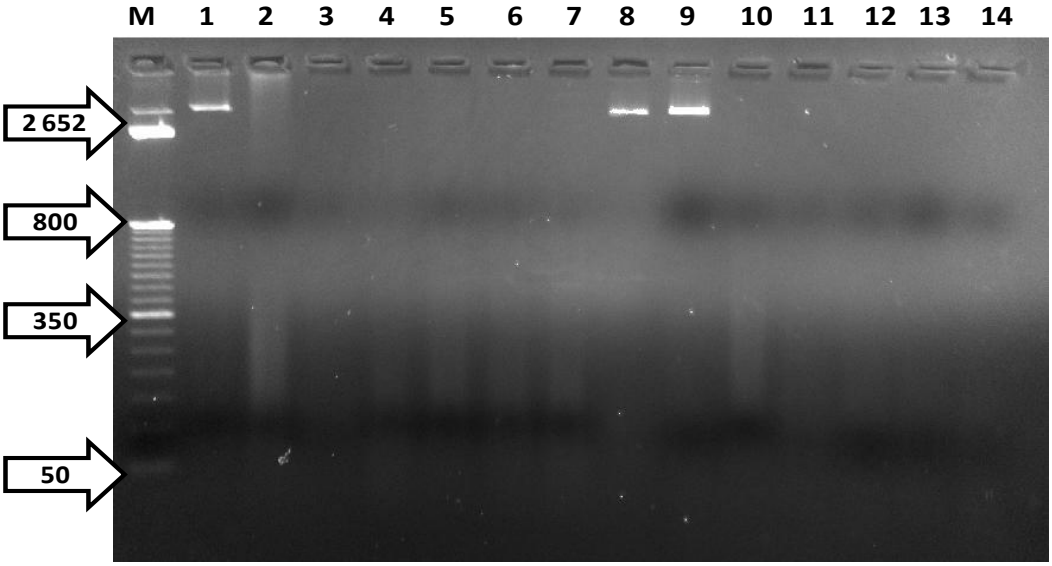


Figure 5.8: Gel electrophoresis of boiled *C. difficile* spores and vegetative cell mixed spore preparation
C. difficile spores and vegetative cells at a concentration of 1 x 10^5 cfu/ml were boiled in water and PBS buffer to ascertain DNA release. A sample was taken every half an hour for 3 h. **Lanes 1-7: Boiling in PBS:** Lane 1: shows the control sample of spores without boiling or centrifugation, Lane 2: Spore DNA release after 30 min boiling, Lane 3: Spore DNA release after 60 min boiling, Lane 4: Spore DNA release after 90 min boiling, Lane 5: Spore DNA release after 120 min boiling, Lane 6: Spore DNA release after 1500 min boiling, Lane 7: S Spore DNA release after 180 min boiling. **Lanes 5-14: Boiling in Water:** Lane 8: shows the control sample of spores without boiling or centrifugation, Lane 9: Spore DNA release after 30 min boiling, Lane 10: Spore DNA release after 60 min boiling, Lane 11: Spore DNA release after 90 min boiling; Lane 12: Spore DNA release after 120 min boiling; Lane 13: Spore DNA release after 150 min boiling, Lane 14: Spore DNA release after 180 min boiling.

5.5.7. Release of DNA from a *C. difficile* spore preparation using focussed microwave irradiation

We sought to develop a method which would rapidly release target DNA using microwave radiation. Gold triangles were used to assist in breaking open bacteria when exposed to microwave irradiation. This method has previously been used with *B. anthracis* to release DNA from spores (Aslan *et al.*, 2008). Rapid heating occurs at the 1 mm gap between two adjacent triangles where the microwaves are focussed thus reducing the need for complex DNA extraction methods. A total of 500 µl of vegetative bacteria and spores were tested at a range of microwave powers and times until an optimal period of irradiation was identified (15 s in the microwave cavity at 80% power). Each sample was run under gel electrophoresis to establish DNA release (Figure 5.9).

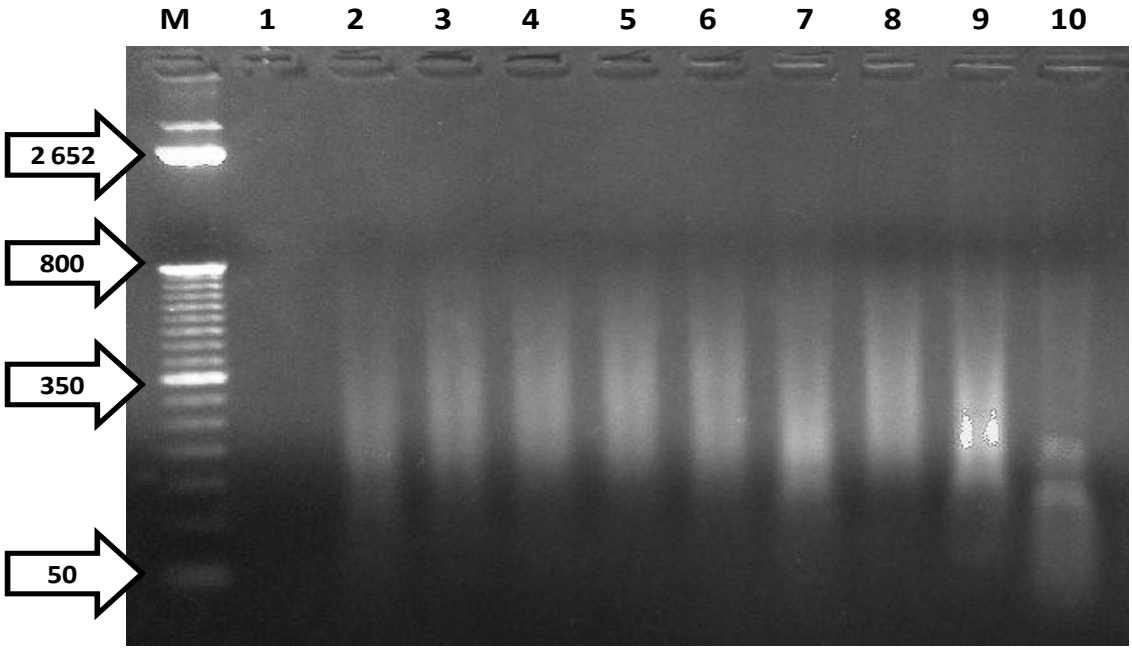


Figure 5.9: Gel electrophoresis of *C. difficile* spores and vegetative cell mixed preparations

C. difficile spores and vegetative cells at a concentration of 1 x 10^5 cfu/ml were lysed using the gold lysing triangles. The power used was 80% microwave power. The 1 kb marker (M) was used. Lane 1 represents the control which was not lysed. Lane 2: lysis at 8 s, Lane 3: 9 s, Lane 4: 10 s, Lane 5: 11 s, Lane 6: 12 s, Lane 7: 13 s, Lane 8: 14 s, Lane 9: 15 s and Lane 10: 16 s. The lane which showed the best lysis with the strongest banding (between 50 -200 bp) was Lane 9 (15 seconds). This means that the DNA had not been broken down to a level where it could not be detected via MAMEF, unlike in Lane 10.

5.5.8. Bacterial quantification following focussed microwave irradiation

The number of viable vegetative bacteria and spores in our spore preparations pre- and post- microwave treatment were determined (Figure 5.10). We observed a 4 x 10^3 cfu/ml log reduction in vegetative counts following microwave treatment. The corresponding reduction in viable spores was 1.33 x 10^4 cfu/ml.

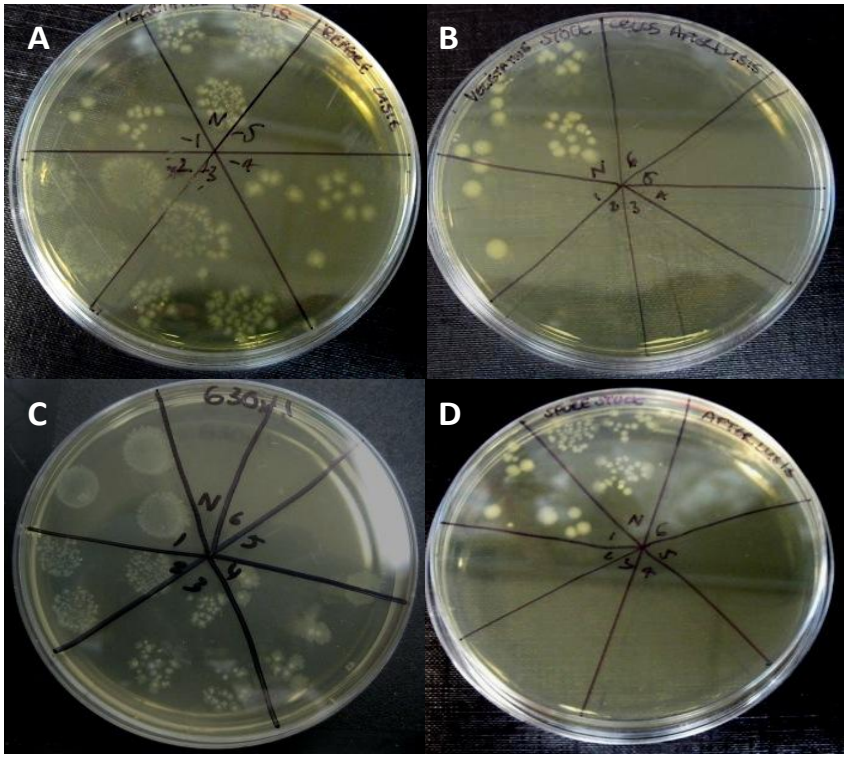


Figure 5.10: Viable counts of *C. difficile* spores and vegetative cells before and after lysis in PBS.
(A) *C. difficile* vegetative cells were lysed by microwave irradiation at a concentration of 4 x 10^6 cfu/ml (B) After lysis the concentration of vegetative cells was 6.67 x 10^2 cfu/ml. (C) Spores were lysed at a concentration of 1.33 x 10^7 cfu/ml. (D) After lysis 3.67 x 10^3 were left.

5.5.9. Detection of *C. difficile* target DNA in PBS

Using the conditions described above (80% microwave power) spores and vegetative cells of *C. difficile* CD630 were suspended in PBS and microwaved for 15 s. Following treatment, vegetative bacteria were diluted to give the following range of final concentrations in each assay well; 1000 cfu, 100 cfu and 10 cfu of vegetative bacteria. Following treatment, spores were diluted to give the following range of final concentrations in each assay well; 10000 cfu, 1000 cfu, 100 cfu and 10 cfu. PBS buffer was used as a control (Figure 5.11 A & B).

The probes readily detected DNA released from vegetative cells in PBS, yielding a signal of approximately 20 Intensity units at 1000 cfu. The limit of sensitivity was 10 cfu for both toxin A and toxin B. There is an increase in fluorescent intensity as the concentration of vegetative cells increases (Figure 5.11 A). However the signal intensity was higher for the toxin B probe at 60 Intensity Units at 1000 cfu (Figure 5.11 B). As mentioned previously this may be due to a range of factors including anchor concentration, optical density of the SiF and efficiency of the fluorophore emission upon binding to target DNA.

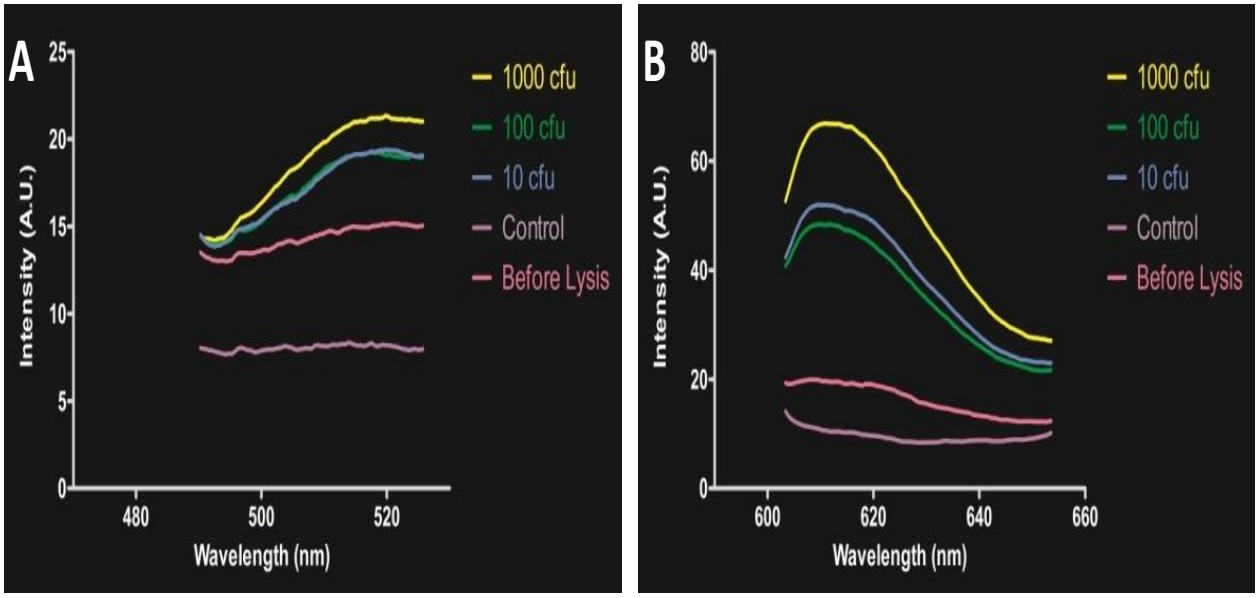


Figure 5.11: Detection of DNA from microwaved vegetative *C. difficile* in PBS using MAMEF

(A) Detection of toxin A from vegetative *C. difficile* lysed in PBS. (B) Detection of toxin B from vegetative *C. difficile* lysed in PBS. As the concentration of cfu increases the fluorescent intensity also increases.

The probes readily detected DNA released from the spore preparation in PBS, yielding a strong signal of approximately 15 Intensity units at 1000 cfu. The limit of sensitivity was 10 cfu for both toxin A and toxin B. There is an increase in fluorescent intensity as the concentration of spore preparation increases (Figure 5.12 A). The signal intensity is similar for the toxin B probe at 18 Intensity Units at 1000 cfu (Figure 5.12 B).

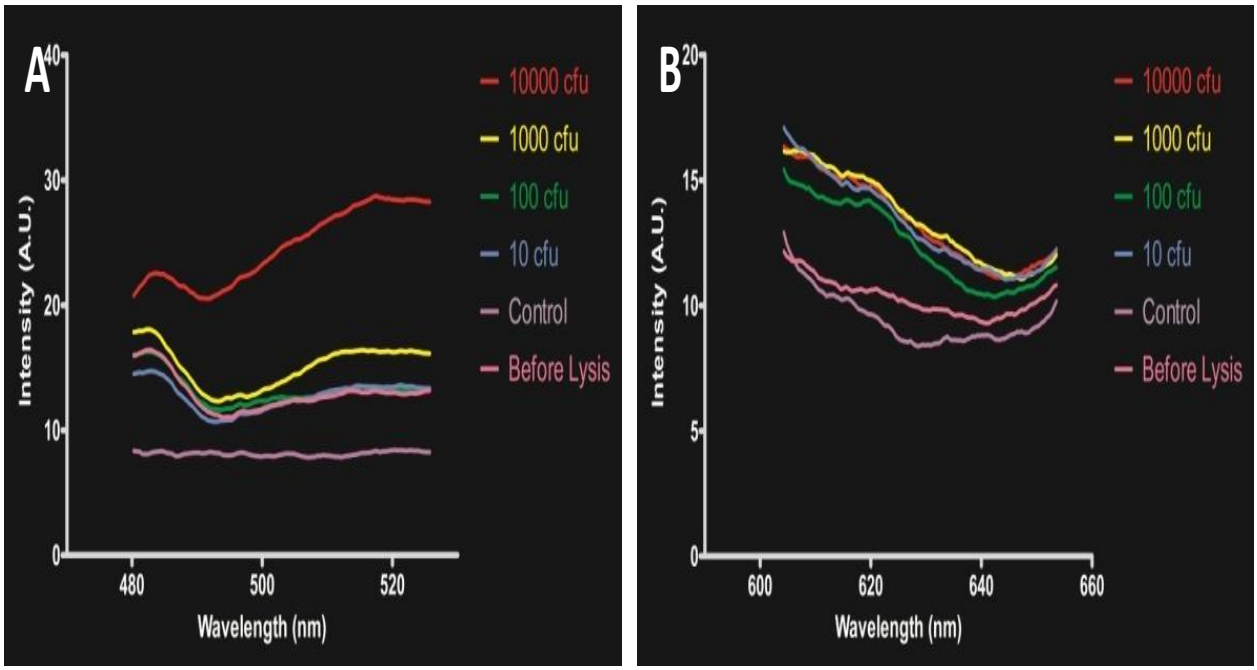


Figure 5.12: Detection of DNA from microwaved *C. difficile* spore preparation in PBS using MAMEF
(A) Detection of toxin A from spores of *C. difficile* lysed in PBS. A concentration of 10000 cfu of spores were used for toxin and B studies. (B) Detection of toxin B from spores of *C. difficile* lysed in PBS. As the concentration of cfu increases the fluorescent intensity also increases.

5.5.10: Detection of *C. difficile* target DNA in whole milk

Spores of *C. difficile* CD630 were suspended in milk and microwaved for 15s using 80% microwave power. Following treatment, spores were diluted to give the following range of final concentrations in each assay well; 1000 cfu, 100 cfu and 10 cfu (Figure 5.13). The probes readily detected DNA released from spores in milk and the intensities appear to be similar as can be seen in Figure 5.13 A & B. The limit of sensitivity was 10 cfu for both toxin A and toxin B. For the concentration of 1000 cfu the signal intensity was ~ 5AU less than in the presence of PBS for both toxins A and B (Figure 5.13 A & B). However the signal is higher for the toxin B probe when detecting the spores (25 Intensity Units at 1000 cfu) in Figure 5.13 B compared to the same concentration in toxin A.

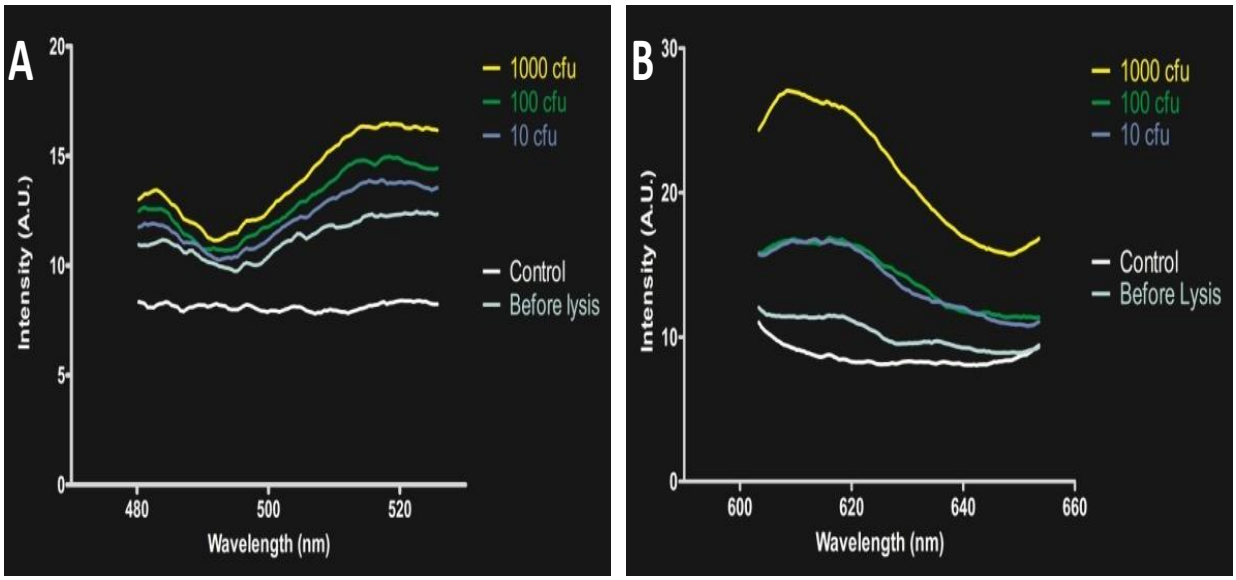


Figure 5.13: Detection of DNA from microwaved *C. difficile* in whole milk using MAMEF

(A) Detection of toxin A from spores of *C. difficile* lysed in whole milk. (B) Detection of toxin B from spores of *C. difficile* lysed in whole milk. As the concentration of cfu increases the fluorescent intensity also increases

5.5.11. Detection of *C. difficile* target DNA in Human faecal matter

Spores of *C. difficile* CD630 were suspended in human faecal matter and microwaved for 15s using 80% microwave power. Following treatment, spores were diluted to give the following range of final concentrations in each assay well; 1000 cfu, 100 cfu and 10 cfu (Figure 5.14). The probes readily detected DNA released from spores in faeces as can be seen in Figure 5.14 A & B. The limit of sensitivity was 10 cfu for both toxin A and toxin B. The signal intensity was slightly higher than observed in the presence of PBS and milk

at 25 AU for toxin A and 20 AU for toxin B (Figure 5.13 A & B; Figure 5.14 A & B). In this experiment toxin A emitted a higher fluorescent signal than toxin B in response to the target.

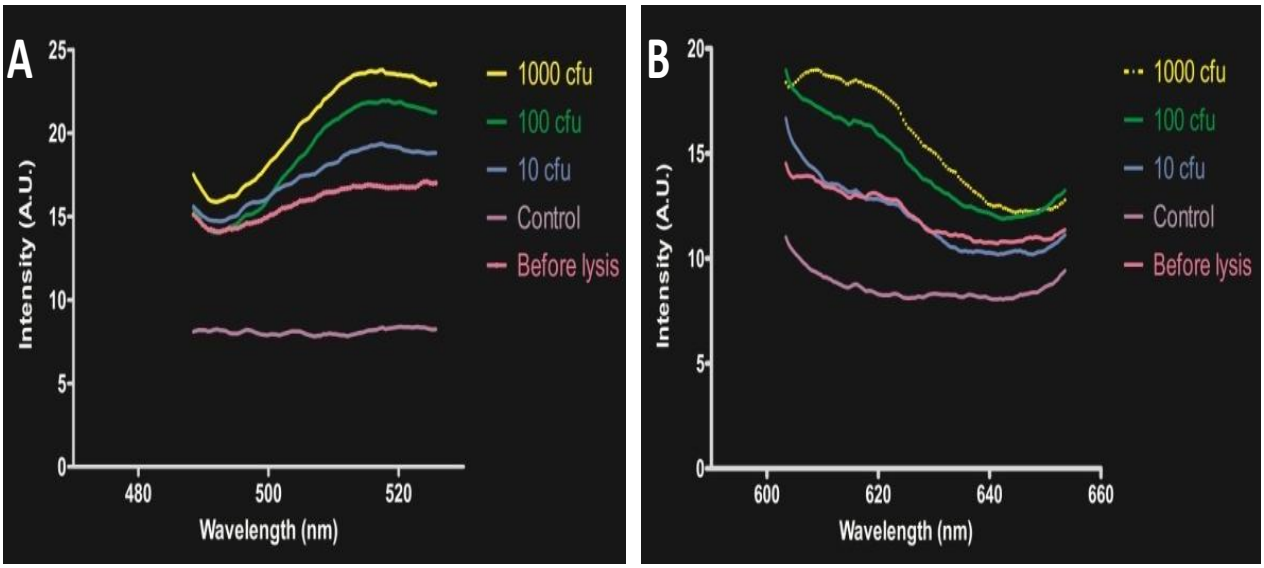


Figure 5.14: Detection of DNA from lysed *C. difficile* in varying biological media using MAMEF
(A) Detection of toxin A from spores of *C. difficile* microwaved in human faecal matter. (B) Detection of toxin B from spores of *C. difficile* microwaved in human faecal matter. For each assay it is clear that as the cfu increases the fluorescent intensity increases.

5.5.12. Statistical analysis of *C. difficile* DNA detection within organic matrices

Statistical analysis of toxin A target DNA detection was performed using one way ANOVA, in conjunction with a Kruskall-Wallis test revealing a P value of 0.069. Thus as P >0.05, this indicates there is no significant difference between the organic matrices used (TE buffer, Milk and faecal matter). One way ANOVA of toxin B target DNA detection revealed a P value of 0.008. As P < 0.05 there is a significant difference in detection between the organic matrices used (TE buffer, Milk and faecal matter). This was further confirmed using the Kruskall-Wallis test where P = 0.018. There is no significant difference in detection between spore concentrations used; however there was a significant difference between the control samples of tcdA and B and the test concentrations as determined by Dunnetts multiple comparison of variance test (Graphpad Prism 5).

5.6. DISCUSSION

We have developed a MAMEF-based assay capable of detecting conserved regions of the genes encoding toxins A and B of *C. difficile*. We were able to detect as few as 10 cfu of *C. difficile* in a faecal suspension within 40 seconds. To put this into context, this represents a high level of sensitivity as in human infection approximately 10^6- 10^8 spores are released. This level of sensitivity and speed compares favourably with all of the currently available *C. difficile* detection methods as described in Section 1.6. Indeed, there is presently no assay capable of directly detecting the presence of *C. difficile* within faecal samples. As far as we are aware, our assay is also the only system capable of detecting both toxins A and B.

At present *C. difficile* assays which employ a DNA based detection system utilise PCR to amplify specific DNA targets within 2- 8 hours (Zheng *et al.,* 2004; Sloan *et al.,* 2008; Goldenberg *et al.,* 2009). This reaction is enzyme-based, and thus requires prior purification to remove enzyme inhibitors which prevent the amplification of target DNA. This can result in false negative reactions as described in Section 1.6.2 (Bickley & Hopkins, 1999). Other factors leading to false negative results include poor DNA extraction methods and the presence of exogenous DNases and RNases within a reaction (Ballagi-Pordany & Belak, 1996; Woodford & Johnson, 2004). False positive reactions can occur due to carry over contamination from previously amplified targets and exogenous target DNA present in reagents and sterile culture material (Kwok & Higuchi, 1989). Thus it would not be possible to directly detect target DNA from a faecal sample without prior purification when using PCR.

In contrast the MAMEF DNA detection system does not require purification and is sensitive enough to detect DNA without the need for an amplification step, further reducing the time required to achieve a result. An advantage of this assay is the construction of the hybridisation platform which allows efficient binding of target DNA through the incorporation of a flexible region within the anchor DNA. This comprises of a region of 5 thymidines which do not bind to target DNA, but instead facilitate the flexibility of the anchor enhancing its ability to capture target DNA in the solution (Aslan *et al.,* 2010). DNA targets can be detected rapidly and efficiently within biological matrices such as blood (Aslan *et al.,* 2010; Tennant *et al.,* 2011). This study however is the first report of MAMEF based DNA target detection directly from milk and faeces.

A further advantage of the prototype assays we have developed is its ability to detect the genes encoding both toxins A and B. Current methods for diagnosis of CDAD are described in Section 1.6. Strains of *C. difficile* with an intact toxin B encoding gene but variations in the genes encoding toxin A also exist. Hence PCR and RT-PCR DNA based assays such as BD Gene Ohm and Xpert Cepheid are configured to detect only toxin B (Deshpande *et al.,* 2011). However the role of toxins A and B in disease has been widely debated, with some suggesting only toxin B is essential for virulence and other suggesting both toxins are capable of causing fulminant disease (Lyras *et al.,* 2009; Kuehne *et al.,* 2010). This highlights the importance of a diagnostic system able to detect both toxins A and B in an infected patient.

The principle component of any nucleic acid based detection system is DNA. Thus it is essential that any DNA based approach includes a method to liberate DNA from the target microorganism. We developed a microwave based method which released DNA from a mixture of *C. difficile* vegetative bacteria and spores. The mechanism by which DNA is released is currently unclear; it has been proposed that spores of bacteria release DNA and biological components upon exposure to the electric fields generated by focussed microwaves (Kim *et al.,* 2009). A further possibility is that internal pressure generated by the microwave irradiation of water in the spore core results in the disruption of the spore the release of biological components (Vaid & Bishop, 1998).

We found that microorganisms suspended in PBS yielded DNA while those suspended in water did not. This may be attributed to the presence of salts within the PBS aiding the potential disruption of vegetative bacteria and spores by elevating the boiling point of the solution (Joffe, 1945). Indeed the release of ionic components such as Ca^{2+}DPA may also contribute to this phenomenon. Previous analysis of DPA release from microwaved spores has indicated that stable complexes are formed between DPA and other ionic components which may affect the boiling point of the solution (Celandroni *et al.,* 2004). The exact mechanism of microwave heating has yet to be determined; it is thought that microwaves can cause ions to accelerate and collide with other molecules causing dipoles and thus an electric field by which disruption of protein structure of microorganisms can occur (Banik *et al.,* 2003).

Our assay is able to detect DNA targets within genomic DNA derived from *C. difficile*. In these preliminary experiments we used preparations containing both vegetative cells and spores as a model of human infection (Section 1.2). It is thus likely that the probes are detecting DNA derived from vegetative cells, rather than DNA liberated from broken spores. Given that the MAMEF technology has the ability to detect a single copy number of the target DNA, future studies will focus on the production of ultra pure spore preparations to determine whether spore DNA or vegetative cell DNA is being detected (Aslan *et al.,* 2008; Lawley *et al.,* 2009).

With regards the sensitivity and specificity of our assay we determined the ability of our probes to detect synthetic and genomic DNA in a variety of organic matrices. Each probe detected 1 nM target DNA (the lowest concentration tested) in PBS, milk, and human faecal matter. We also examined the ability of our probes to detect DNA from microwaved spore and vegetative cell preparations in PBS, milk and faecal matter. Our preliminary findings indicate a limit of sensitivity of 10 cfu, the lowest concentration of cfu tested. To further improve this assay the lower limit of detection below a concentration of 10 cfu should be ascertained using pure spore preparations. Target DNA from unmicrowaved spore preparations was also detected. This suggests DNA from vegetative cells in the preparation, or genomic DNA on the surface of the spores, is being detected by the probes.

To realise the potential of our promising preliminary data, future studies must focus on optimisation of the MAMEF platform. Thus the microwaves and laser platform would need to be incorporated into a hand held device capable of being used at the bedside. Also the SiFs used in the majority of experiments varied in their optical densities (between 0.4-0.5) due to the silver adhering unevenly to the silicone glass surface (Aslan *et al.,* 2010). Thus an improvement of the system would be to use SiFs with the exact same optical densities across the silicone glass slide surface. This would allow an even fluorescence intensity to be emitted when a target is detected. Furthermore, optimisation of the amount of anchor DNA bound to the SiF upon the fluorescent intensity would increase the sensitivity of the assay.

We anticipate that a future assay will be multiplexed and thus able to detect the presence of both toxins A and B in a sample. Further studies would involve examining a wide range infected *C. difficile* patient samples and carrying out clinical trials.

We believe our prototype *C. difficile* detection assay has the potential to be developed into a real-time assay capable of identifying individuals infected with *C. difficile* within 40 seconds. Such an assay would fulfil the pressing need for a rapid and sensitive detection system able to detect both toxin A and B encoding genes of *C. difficile*. In the future we believe this assay would be used as a means of screening patients upon admission to hospital to enable appropriate treatment regimens and clinical management decisions to be made.

CHAPTER 6

GENERAL DISCUSSION

6.1. GENERAL DISCUSSION

The aim of this research was to examine the virulence factors which contributed to the prevalence of the hospital bacterium *C. difficile*. It was also important to better understand the organism for which this assay was designed by not only exploring its genetic characteristics, but by further exploring its physical characteristics to enable a future assay to be designed optimally. We also designed probes specifically able to detect *C. difficile* toxins A and B within human faecal matter using a MAMEF based technological assay.

The resistant spore form of *C. difficile* facilitates its persistence in the hospital environment, and enables it to initiate infection by mechanisms which have yet to be elucidated. As part of our research we sought to determine which characteristics of the *C. difficile* spore contributed to the organisms' virulence. To determine if spore structure contributed to surface binding we characterised the abilities of spores representing a range of clinical isolates of differing ribotypes to adhere to organic and inorganic surfaces. Using the MATH hydrophobicity assay of Rosenberg *et al.* (1980) we identified an association between ribotype and spore hydrophobicity in that spores which demonstrated the highest percentage hydrophobicity were predominantly of hypervirulent and epidemic ribotypes.

We then demonstrated that spores of different clinical ribotypes were better able to adhere to stainless steel, a material common to the hospital environment. This provided potential clues as to how hypervirulent isolates are transmitted, survive and spread within the hospital environment. Spores which exhibited high hydrophilic and hydrophobic properties were further investigated for adherence to organic material in the form of human adenocarcinoma cells (Caco 2 and HT 29). It was observed that spores with high hydrophobicity were better able to adhere to the adenocarcinoma cells in contrast to spores with hydrophilic characteristics. Thus we found that the surface hydrophobicity of spores influenced their ability to adhere to organic and inorganic surfaces.

To determine if variations in spore adherence were associated with spore structure we characterised the architecture of selected spores by electron microscopy. Spores of the isolate DS1813, with the greatest hydrophobicity, were found to possess a distinct exosporium similar to that shown surrounding spores of strain CD630 (Lawley *et al.*, 2009). The presence, or absence, of this layer is thought to contribute to the degree with

which spores of *Bacillus* spp adhere to surfaces (Koshikawa *et al.,* 1989). The exosporium of *C. difficile* appears to plays an essential role in the attachment of the spore to the gut lining and thus is an important step in the process of bacterial colonisation (Panessa-Warren *et al.,* 1997). We also observed that the exosporial layer did not surround spores of every isolate examined, suggesting that there is no direct correlation between spore hydrophobicity and the presence of an exosporium. Another reason for examining spore adherence was to establish whether *C. difficile* spores could adhere to materials used within the context of a MAMEF based technological detection assay.

Following attachment to the cells lining the gut, the spore form of *C. difficile* must germinate to initiate infection. In the context of this study we observed that certain strains, such as R20291 (BI/NAP1/027), exhibited varying germination ability when grown in different laboratory culture media which suggests that strains differ with regards to the factors which trigger their germination. Indeed we isolated a variant of 027 capable of significant germination in the absence of the co-germinant sodium taurocholate. While the basis of these differences in relative germination in synthetic laboratory media are intriguing, their relevance to *in vivo* conditions has yet to be determined. Overall our data supports the current hypotheses that spores of *C. difficile* regulate their germination by sensing ideal locations to attach and germinate within the gut. This may be a result of bile salts, such as sodium taurocholate which has a positive effect on germination, and chenodeoxycholate which has an inhibitory effect on germination (Sorg & Sonenshein, 2008; Giel *et al.,* 2010).

The key virulence factors of *C. difficile* which are expressed following attachment and germination are its two toxins. Toxin A and toxin B gene sequences of *C. difficile* were examined using bioinformatic analysis for conserved areas to understand the organism's evolution as well as to determine regions suitable for probe design. Bioinformatic analysis of both toxins revealed significant homology to each others' structural domains, which supports the theory that the toxins may have originally evolved as a duplication of the other (Von Eichel-Streiber, 1992). The original toxin ancestor has yet to be elucidated, and there is some debate as to which toxin was duplicated. Even so from our studies we hypothesise that the *tcdA* gene, which is more susceptible to HGT and had less areas of conservation, may have been a duplication of the more stable *tcdB*.

We also identified homology with virulence genes associated with insect pathogenicity. The majority of the structural domains of toxin B shared significant homology with the mcf toxin from *P. luminescens* and a number of other proteins associated with insect infection. Indeed this homology suggests that toxin B may have originated as a toxin from an insect associated ancestor.

To better understand if *C. difficile* has an association with insects, we employed the insect *Manduca sexta* as a model infection system. *C. difficile* spores and vegetative cells were unable to elicit symptoms of infection upon either injected or oral delivery at 25^{o}C and 37^{o}C. An interesting observation was that while injection with spores at 37^{o}C had no obvious effects on the health of the insects they did pupate in contrast to the unchallenged controls, suggesting that exposure to spores had an effect on insect development (Alberea & Schaub, 2008).

The failure of spores to initiate infection could be due to the fact that the insect lacks the necessary germination triggers. In an attempt to trigger germination we exposed insects to a mixture of spores supplemented with the bile salt sodium taurocholate; however no germination or subsequent infection manifested. This result may be an artefact of the insect system whereby the anaerobic environments (haemolymph/haemocoel; Wyatt & Meyer, 1956) were not able to sustain *C. difficile*. Further to this it may be likely that the insects do not possess the toxin receptors needed for pathology. Indeed the insect haemocoel may also contain antimicrobial factors which could account for the failure to initiate infection in the *M. sexta*.

To determine if this indeed the case the vegetative and spore forms of *C. difficile* were suspended in *M. sexta* haemolymph. A reduction in the number of viable spores was observed supporting the hypothesis that the haemolymph is antimicrobial, possibly as a consequence of the production of antimicrobial peptides (Waterfield *et al.*, 2004). Attempts to develop a gut colonisation model were also unsuccessful due to the fact that there was no difference in the passage or adherence ability of either strain tested. On the basis of these results, we have concluded that *M. sexta* is a poor model with which to study *C. difficile* pathogenicity and colonisation.

C. difficile has also been found to possess chitinase/ peroxiredoxin genes within its spore coat proteins which are associated spore germination (Permpoonpattna *et al.,* 2011). To determine if this chitinase was indeed a relic of an insect infective past, we analysed spores of the widest differing hydrophobicity and found that chitinase was indeed released from germinating spores of both strains. This suggests that spores may be releasing chitinase during germination to potentially enable macromolecular degradation of soil elements, such as fungi and dead insects, and thus allow nutrient uptake, but also indicates that the presence of insect associated genes within the *C. difficile* genome may be associated with the evolution of the whole genome of *C. difficile*. Hence the common ancestor to *C. difficile* may indeed be an insect associated pathogen.

Conserved sequences within the gene encoding regions of *C. difficile* toxins A and B were identified to design probes for use in the MAMEF assay. Candidate probes specific to conserved regions within toxins A and toxin B were synthesised and examined for specificity using a rigorous screening process against 58 clinical isolates of *C. difficile,* unrelated species, species from the LCT family and 10 human metagenomic gut DNA extracts. The probes used in this study were highly specific, gave no false positives or negatives and were incorporated into a prototype MAMEF based detection system. Using this system we were able to detect 10 cfu of *C. difficile* in a faecal suspension within 40 seconds. Indeed as described previously (Section 1.6) there are currently no assays available which are able to rapidly detect *tcdA* and *tcdB* within human faecal suspensions near to this time frame. We have also proven that *C. difficile* can be directly detected in other complex organic matrices such as milk. This prototype biosensor has the potential to be developed into a real time beside diagnostic assay.

6.2. CONCLUSIONS

The aim of this thesis was to determine the contribution of *C. difficile's* virulence factors to its pathogenicity, while simultaneously designing an assay capable of detecting *C. difficile* toxins A and B using the MAMEF technological platform. The ability of *C. difficile* spores to germinate in the presence of different laboratory culture media was examined and we suggested that strains differed with regards to their germination triggers. Furthermore we observed that spores produced from a range of clinical isolates varied in their ability to adhere to organic and inorganic surfaces. This adherence ability was associated with spore surface hydrophobicity, with the most hydrophobic spores being amongst the most virulent ribotypes. These results suggest that spore adherence is closely associated with the spread and virulence of certain strains.

From bioinformatic analysis it appears that the toxin genes of *C. difficile* appear to have evolved as a duplication of each other, with toxin A being a duplicate of toxin B. There was also homology of the *C. difficile* toxin genes to genes linked to insect pathogenicity, suggesting that C. *difficile* may have associated with insects throughout its evolution. Whether the genome acquired the toxins through HGT or evolved from an insect toxin ancestor is still the subject of debate.

Finally conserved DNA signatures within toxins A and B of *C. difficile* were identified and analysed for specificity and sensitivity to *C. difficile*. Probes were engineered into the MAMEF platform and the assay was able to detect *C. difficile* at a lower detection limit of 10 cfu in human faecal suspensions within 40 seconds.

6.3. FUTURE WORK

To build on the research described above we propose that the following studies be considered:

1. Examine the abilities of a larger collection of *C. difficile* clinical isolates to germinate in different laboratory culture media to establish if there is an association between ribotype and germination ability.

2. Characterise the ability of spores of a representative collection of *C. difficile* clinical isolates to adhere to human adenocarcinoma cells to determine if there is a definitive association between virulence and adherence.

3. Determine the role of the exosporium in the pathogenicity of *C. difficile.*

4. Determine the basis of the antimicrobial activity of *M. sexta* haemolymph.

5. Determine if direct injection of *C. difficile* into the haemolymph of other insects, such as cockroaches or *Galleria,* elicits a similar response. Indeed this would add support to our hypothesis that *C. difficile* evolved from an insect associated ancestor.

6. Determine if the toxins of *C. difficile* had an effect upon the insect gut. This study can be conducted by adding the toxins to insect gut tissue culture. Any resulting pathology would indicate the presence of *C. difficile* toxin receptors on the surface of these cells.

7. Currently the MAMEF assay detects two individual toxins in separate reactions; thus the assay would be benefit to be made into a single assay.

8. Develop the prototype MAMEF based *C. difficile* detection assay.

9. File a patent for the toxin A and B probe sequences and explore the commercialisation of this assay.

6.4. FUTURE BENEFITS AND APPLICATIONS

The potential benefits of an assay capable of detecting both toxins of *C. difficile* in faecal suspensions are numerous. Indeed current assays are designed to detect only one of the two *C. difficile* toxins, have low sensitivities and often take hours to yield a result (Section 1.6). In the context of hospital diagnosis, a variety of tests are used for CDI diagnosis, and there is currently no universal standard assay capable of rapidly (<60 mins) and accurately detecting the presence of both toxins. Thus an assay capable of specifically detecting both toxins A and B, without the need for prior purification or amplification, within raw human faecal samples in 40 seconds would be of tremendous benefit to the healthcare industry.

Indeed such an assay would negate the need for the use of multiple non-specific, non-rapid assays, and has the potential to provide results at the patient bedside. This rapid diagnosis would allow the treatment regimen for patients to be tailored accordingly, preventing cross infection and further disease transmission. Currently within the NHS (UK) it takes approximately three days to diagnose a person with suspected toxigenic *C. difficile*, and it costs ~ £600 to keep a person in hospital; hence it would cost £1800 per patient. Patients with infection can expect to spend up to 27 days in hospital and in Wales it currently costs £8 000 to treat each individual case of *C. difficile* (Cardiff & Vale NHS Trusts, 2011). Thus any measure which would reduce the costs incurred by the NHS with regards to diagnosing *C. difficile* infection, and save lives in the process, would be of use.

Another benefit of the *C. difficile* detection assay is its use of conserved sequences which are unlikely to change as a result of selective pressure. Indeed these conserved sequences are likely to remain conserved within toxigenic *C. difficile* and thus these probe sequences could potentially be applied to any platform technology- not only the MAMEF technological platform.

CHAPTER 7

REFERENCES

Abigail, H. D., April, K. R. & Clifford, C. S. (2011) Super toxins from a super bug: structure and function of *Clostridium difficile* toxins. *Biochemical Journal,* 436**,** 517-526.

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A. & Carniel, E.(1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences,* 96**,** 14043.

Adhya, T. K. & Patra, M. (2012) Darwin and microbial evolution. *Nature at Work: Ongoing Saga of Evolution***,** 103-118.

Åkerlund, D, T., Persson, I., Unemo, M., Noren, T., Svenungsson, B., Wullt, M. & Burman, L. G. (2008) Increased sporulation rate of epidemic *Clostridium difficile* type 027/NAP1. *Journal of Clinical Microbiology,* 46**,** 1530-1533.

Aktories K. & Barbieri, J. T. (2005) Bacterial cytotoxins: targeting eukaryotic switches. *Nature Reviews Microbiology,* 3**,** 397-410.

Aktories K. & Just, I. (2005) *Clostridial* rho-inhibiting protein toxins. *Bacterial Virulence Factors and Rho GTPases***,** 113-145.

Aktories, K. & Wilkins, T. C. (2000) *Clostridium difficile,* Springer.

Albera, U. M. P. V. G. & Schaubb, G. A. (2008) The innate immune system of mammals and insects. *Trends in innate immunity,* 15**,** 21-44.

Albesa-Jove, D., Bertrand, T., Carpenter, E. P., Swain, G. V., Lim, J., Zhang, J., Haire, L. F., Vasisht, N., Braun, V. & Lange, A (2010) Four distinct structural domains in *Clostridium difficile* Toxin B Visualized using SAXS. *Journal of Molecular Biology*.

Alcala, L., Marin, M., Madrid, M., Dominguez-Garcia, E., Catalan, P., Pelaez, M. T., Sanchez-Somolinos, M. & Bouza, E. (2010) Comparison of ImmunoCard toxins A&B and the new semi automated Vidas *Clostridium difficile* toxin A&B tests for diagnosis of *C. difficile* infection. *Journal of Clinical Microbiology,* 48**,** 1014.

Alonso, R., Martin, A., Pelaez, T., Marin, M., Rodriguez-Creixems, M. & Bouze, E. (2005) An improved protocol for pulsed-field gel electrophoresis typing of *Clostridium difficile*. *Journal of Medical Microbiology,* 54**,** 155.

Alouf, J. E. & Popoff, M. R. (2006) *The Comprehensive Sourcebook of Bacterial Protein Toxins,* Elsevier.

Al Saif, N. & Brazier, J. S. (1996) The distribution of *Clostridium difficile* in the environment of South Wales. *Journal of Medical Microbiology,* 45**,** 133-137.

Altman, D. G. & Bland, J. M. (1994) Diagnostic tests 1: Sensitivity and specificity. *BMJ: British Medical Journal,* 308**,** 1552.

Altman, D. G. & Bland, J. M. (1994) Diagnostic tests 2: Predictive values. *BMJ: British Medical Journal,* 309**,** 102.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389.

Alvarez-Perez, S., Blanco, J., Bouza, E., Alba, P., GIibert, X., Maldonado, J., & Garcia, M. E. (2009) Prevalence of *Clostridium difficile* in diarrhoeic and non-diarrhoeic piglets. *Veterinary Microbiology,* 10**,** 1-16.

Amimoto, K., Noro, T., Oishi, E. & Shimizu, M. (2007) A novel toxin homologous to large clostridial cytotoxins found in culture supernatant of *Clostridium perfringens* type C. *Microbiology,* 153**,** 1198.

Arena, I., Bravo, A., Soberan, M. & Gamez, I. (2010) Role of alkaline phosphatase from *Manduca sexta* in the mechanism of action of *Bacillus thuringiensis* Cry1Ab toxin. *Journal of Biological Chemistry,* 285**,** 12497.

Arizona, T. U. O. (2001) The Manduca Project.

Arnold, A., Pope, C., Bray, S., Riley, P., Breathnach, A., Krishna, S. & Planche, T. (2010) Prospective assessment of two-stage testing for *Clostridium difficile*. *Journal of Hospital Infection,* 76**,** 18-22.

Aslan, K., Leonenko, Z., Lakowicz, J. R. & Geddes, C. D. (2005) Annealed silver-island films for applications in metal-enhanced fluorescence: Interpretation in terms of radiating plasmons. *Journal of Fluorescence,* 15**,** 643-654.

Aslan, K., Gryczynski, I., Malicka, J., Matveeva, E., Lakowicz, J. R. & Geddes, C. D. (2005) Metal-enhanced fluorescence: an emerging tool in biotechnology. *Current Opinion in Biotechnology,* 16**,** 55–62.

Aslan, K., Zhang, Y., Hibbs, S., Baillie, L., Previte, M. J. R. & Geddes, C. D. (2007) Microwave-accelerated metal-enhanced fluorescence: application to detection of genomic and exosporium anthrax DNA in< 30 seconds. *The Analyst,* 132**,** 1130-1138.

Aslan, K. & Geddes, C. D. (2008) A review of an ultrafast and sensitive bioassay platform technology: Microwave-accelerated Metal-enhanced Fluorescence. *Plasmonics,* 3**,** 89-101

Aslan, K., Previte, M. J. R., Zhang, Y., Gallagher, T., Baillie, L., & Geddes (2008) Extraction and detection of DNA from *Bacillus anthracis* spores and the vegetative cells within 1 min. *Analytical Chemistry,* 80**,** 4125-4132.

Aslan, K., Baillie, L. & Geddes, C. D. (2010) Ultra fast and sensitive detection of biological threat agents using microwaves, nanoparticles and luminescence. *Journal of Medical Chemical, Biological and Radiological Defense,* 8**,** 1-21.

Aspinall, S. T. & Hutchinson, D. N. (1992) New selective medium for *isolating Clostridium difficile* from faeces. *Journal of clinical pathology,* 45**,** 812.

Ballagi-Pordany, A. & Belak, S. (1996) The use of mimics as internal standards to avoid false negatives in diagnostic PCR. *Molecular and cellular probes,* 10**,** 159-164.

Bakri, M. M., Brown, D. J., Butcher, J. P. & Sutherland, A. D. (2009) *Clostridium difficile* in ready-to-eat salads, Scotland. *Emerging Infectious Diseases,* 15**,** 817.

Banik, S., Bandyopadhyay, S. & Ganguly, S. (2003) Bioeffects of microwave: brief review. *Bioresource Technology,* 87**,** 155-159.

Banno, Y., Kobayashi, T., Watanabe, K., Uneo, K. & Y., Nozawa. (1981) Two toxins (D-1 and D-2) of *Clostridium difficile* causing antibiotic-associated colitis: purification and some characterization. *Biochemistry International,* 2**,** 629-635.

Banno, Y., Kobayashi, T., Kono, H., Watanabe, K., Uneo, K. & Y., Nozawa. (1984) Biochemical characterization and biologic actions of two toxins (D-1 and D-2) from *Clostridium difficile*. *Review of Infectious Diseases,* 6**,** S11- S20.

Barbut, F. & Petit, J. C. (2001) Epidemiology of *Clostridium difficile*-associated infections. *Clinical Microbiology and Infection,* 7**,** 405-410.

Barbut, F., Delmee, M., Brazier, J. S., Petit, J. C., Poxton, I. R., Rupnik, M., Lalande, V., Schneider, C., Mastrantonio, P. & Alonso, R. (2003) A European survey of diagnostic methods and testing protocols for *Clostridium diffi*cile. *Clinical Microbiology and Infection,* 9**,** 989-996.

Barroso, L. A., Moncrief, J., Lyerly, D. M. & Wilkins, T. D. (1994) Mutagenesis of the *Clostridium difficile* toxin B gene and effect on cytotoxic activity. *Microbial Pathogenesis,* 16**,** 297-303.

Barth, H., Aktories, K., Popoff, M. R. & Stiles, B. G. (2004) Binary Bacterial Toxins: Biochemistry, biology, and applications of common *Clostridium* and *Bacillus* proteins. *Microbiology and Molecular Biology Reviews,* 68**,** 373.

Barth, H. & Stiles, B. G. (2008) Binary actin-ADP-ribosylating toxins and their use as molecular Trojan horses for drug delivery into eukaryotic cells. *Current medicinal chemistry,* 15**,** 459-469.

Bartlett, J. G., Chang, T. W. & Onderdonk, A. B. (1978) Will the real *Clostridium* species responsible for antibiotic associated colitis please step forward? *The Lancet***,** 338.

Bartlett, J. G. (2006) Narrative review: the new epidemic of *Clostridium difficile*-associated enteric disease. *Annals of internal medicine,* 145**,** 758.

Baverud, V. (2002) *Clostridium difficile* infections in animals with special reference to the horse. A review. *Veterinary quarterly,* 24**,** 203-219

Belanger, S. D., Boissinot, M., Clairoux, N., Picard, F. J. & Bergeron, M. G. (2003) Rapid detection of *Clostridium difficile* in feces by Real-Time PCR. *Journal of Clinical Microbiology,* 41**,** 730-734

Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., Ouellette, B. F., Rapp, B. A. & Wheeler, D. L. (1999) GenBank. *Nucleic Acids Research,* 27**,** 12.

Bergin, D., Reeves, E. P., Renwick, J., Wientjes, F. B. & Kavanagh, K. (2005) Superoxide production in *Galleria mellonella* hemocytes: identification of proteins homologous to the NADPH oxidase complex of human neutrophils. *Infection and Immunity,* 73**,** 4161.

Bhalla, R., Dalal, M., Panguluri, S. K., Jagadish, B., Mandaokar, A. D., Singh, A. K. & Kumar, P. A. (2005) Isolation, characterization and expression of a novel vegetative insecticidal protein gene of *Bacillus thuringiensis*. *FEMS Microbiology Letters,* 243**,** 467-472.

Bickley, J. & Hopkins, D. (1999) Inhibitors and enhancers of PCR. *Analytical molecular biology. Quality and validation,* 6**,** 81-102.

Bidet, P., Lalande, V., Salauze, B., Burghoffer, B., Avesani, V., Delmee, M., Rossier, A., Barbut, F. & Petit, J. C. (2000) Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. *Journal of Clinical Microbiology,* 38**,** 2484.

Biller, P., Shank, B., Lind, L., Brennan, M., Tkatch, L., Killgore, G., Thompson, A. & Mcdonald, L. C. (2007) Moxifloxacin therapy as a risk factor for *Clostridium difficile*–Associated Disease during an outbreak: Attempts to control a new epidemic strain. *Infection Control and Hospital Epidemiology,* 28**,** 198-201.

Bongaerts, G. P. A. & Lyerly, D. M. (1994) Role of toxins A and B in the pathogenesis of *Clostridium difficile* disease. *Microbial Pathogenesis,* 17**,** 1-12.

Borriello, S. P. (Ed.) (1990) *Clinical and Molecular Aspects of Anaerobes,* Wrightson Biomedical Publishing.

Boutzis, K. & Miller, T. A. (2006) *Insect symbiosis,* CRC.

Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M. & von Eichel-Streiber, C. (1996) Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene,* 181**,** 29-38.

Braun, V., Mehlig, M., Moos, M., Rupnik, M., Kalt, B., Mahony, D. E. & Eichel-Streiber, C. V. (2000) A chimeric ribozyme in *Clostridium difficile* combines features of group I introns and insertion elements. *Molecular Microbiology* 36**,** 1447-1459.

Brinkmann, N., Martens, R. & Tebbe, C. C. (2008) Origin and diversity of metabolically active gut bacteria from laboratory-bred larvae of *Manduca sexta* (*Sphingidae, Lepidoptera, Insecta*). *Applied and Environmental Microbiology,* 74**,** 7189-7196.

Bruggeman, H. (2005) Genomics of clostridial pathogens: implication of extrachromosomal elements in pathogenicity. *Current opinion in microbiology,* 8**,** 601-605.

Bruggemann, H. & Gottschalk, G. (Eds.) (2009) *Clostridia: Molecular biology in the post genomic era*. Caister Academic Press

Buggy, B. P., Hawkins, C. C. & Fekety, R. (1985) Effect of adding sodium taurocholate to selective media on the recovery of *Clostridium difficile* from environmental surfaces. *Journal of Clinical Microbiology,* 21**,** 636-637.

Bulmer, M. (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics,* 129**,** 897

Burdon, D. W. & George, R. H. (1978) Pseudomembranous colitis. *The Lancet***,** 445.

Burns, D. A., Heap, J. T. & Minton, N. P. (2010) SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. *Journal of Bacteriology,* 192**,** 657.

Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G. & Fairweather, N. (2001) Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Molecular Microbiology,* 40**,** 1187-1199.

Calabi, E. & Fairweather, N. (2002) Patterns of sequence conservation in the S-layer proteins and related sequences in *Clostridium difficile*. *Journal of Bacteriology,* 184**,** 3886.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brussow, H. (2003) Prophage genomics. *Microbiology and Molecular Biology Reviews,* 67**,** 238.

Cardiff & Vale NHS Trust, (2011) Cardiff and Vale UHB *C. difficile* success. [http://www.cardiffandvaleuhb.wales.nhs.uk/sitesplus/documents/864/Cdiff.pdf] Accessed 04/2011

Carman, R. J., Wickham, K. N., Chen, L., Lawrence, A. M., Boone, J. H., Wilkins, T. D., Kerkering, T. M. & Lyerly, D. M. (2012) Glutamate dehydrogenase (GDH) is highly conserved among *Clostridium difficile* ribotypes. *Journal of Clinical Microbiology*.

Carroll, K. C. & Loeffelholz, M. (2011) Conventional versus molecular methods for the detection of *Clostridium difficile*. *Journal of Clinical Microbiology,* 49**,** S49-S52.

Chaudary, A., Nath, G., Tikoo, A. & Sanyal, S. C. (1999) Enteropathogenicity and antimicrobial susceptibility of new *Escherichia* spp. *Journal of Diarrhoeal Diseases Research,* 17**,** 85-87.

Cohen, S. H., Tang, Y. J. & Silva JR. J, (2000) Analysis of the pathogenicity locus in *Clostridium difficile* strains. *Journal of Infectious Diseases,* 181**,** 659-663.

Curry, S. R., Marsh, J. W., Muti, C. A., O'Leary, M. M., Pasculle, A. W. & Harrison, L. H. (2007) tcdC genotypes associated with severe TcdC truncation in an epidemic clone and other strains of *Clostridium difficile*? *Journal of Clinical Microbiology,* 45**,** 215-221.

Curry, S. R., Schlackman, J. L., Hamilton, T. M., Henderson, T. K., Brown, N. T., Marsh, H, J. W., Shutt, K. A., Brooks, M. M., Pasculle, A. W. & Muto, C. A. (2011) Perirectal swab surveillance for *Clostridium difficile* by use of selective broth pre-amplification and real-time PCR detection of tcdB. *Journal of Clinical Microbiology,* 49**,** 3788-3793.

Dhalluin, A., Bougeois, I., Pestel-Caron, M., Camiade, E., Raux, G., Courtin, P., Chapot-Chartier, M. P. & Pons, J. L. (2005) Acd, a peptidoglycan hydrolase of *Clostridium difficile* with N-acetylglucosaminidase activity. *Microbiology,* 151**,** 2343.

Dawson, L. F., Valiente, E. & Wren, B. W. (2009) *Clostridium difficile*—A continually evolving and problematic pathogen. *Infection, Genetics and Evolution*.

Dawson, L. F., Stabler, R. A. & Wren, B. W. (2008) Assessing the role of p-cresol tolerance in *Clostridium difficile*. *Journal of Medical Microbiology,* 57**,** 745.

Davies, H. A. & Borriello, S. P. (1990) Detection of capsule in strains of *Clostridium difficile* of varying virulence and toxigenicity. *Microbial Pathogenesis,* 9**,** 141.

De Maagd, R. A., Bravo, A., Berry, C., Crickmore, N. & Schnepf, H. E. (2003) Structure, diversity and evolution of protein toxins from spore forming entomopathogenic bacteria *Annual Reviews in Genetics,* 37**,** 409-433.

Debast, S. B., Van Leengoed, L., Goorhuis, A., Harmanus, C., Kuipers, E. J. & Bergwerff, A. A. (2009) *Clostridium difficil*e PCR ribotype 078 toxinotype V found in diarrhoeal pigs identical to isolates from affected humans. *Environmental Microbiology,* 11**,** 505.

Delmee, M., Van Broeck, J., Simon, A., Janssens, M. & Avesani, V. (2005) Laboratory diagnosis of *Clostridium difficile*-associated diarrhoea: a plea for culture. *Journal of Medical Microbiology,* 54**,** 187-191.

Denecker, G., Declercq, W., Geuijen, C. A. W., Boland, A., Benabdillah, R., Van Gurp, M., Sory, M. P., Vandenabeele, P. & Cornelis, G. R. (2001) *Yersinia enterocolitica* YopP-induced apoptosis of macrophages involves the apoptotic signalling cascade upstream of bid. *Journal of biological chemistry,* 276**,** 19706.

Deneve, C., Delomenie, C., Barc, M.-C., Collignon, A. & Janoir, C. (2008) Antibiotics involved in *Clostridium difficile*-associated disease increase colonization factor gene expression. *Journal of Medical Microbiology,* 57**,** 732–738.

Department of Health & Health Protection Agency (2008) *Clostridium difficile* infection: How to deal with the problem. DH Publishers.

Dieffenback, C. W. & Dveksler, G. S. (Eds.) (2003) *PCR Primer: A Laboratory Manual,* New York, Cold Spring Harbour Press.

Dieckmann, U. (2002) *Adaptive dynamics of infectious diseases: in pursuit of virulence management,* Cambridge Univ Pr.

Dineen, S. S., Villapakkam, A. C., Nordman, J. T. & Sonenshein, A. L. (2007) Repression of *Clostridium difficile* toxin gene expression by CodY. *Molecular Microbiology,* 66**,** 206-219.

Dineen, S. S., Mcbride, S. M. & Sonenshein, A. L. (2010) Integration of metabolism and virulence by *Clostridium difficile* CodY. *Journal of Bacteriology,* 192**,** 5350.

Doing, K. M., Hintz, M. S., Keefe, C., Horne, S., Levasseur, S. & Kulikowski, M. L. (2010) Reevaluation of the Premier *Clostridium difficile* toxin A and B immunoassay with comparison to glutamate dehydrogenase common antigen testing evaluating Bartels

cytotoxin and Prodesse ProGastro™ Cd polymerase chain reaction as confirmatory procedures. *Diagnostic Microbiology & Infectious Disease,* 66**,** 129-134.

Driks, A. (2004) The Bacillus spore coat. *Phytopathology,* 94**,** 1249-1251.

Driks, A. (2009) The *Bacillus anthracis* spore. *Molecular Aspects of Medicine*.

Drudy, D., Harnedy, N., Fanning, S., Hannan, M. & Kyne, L. (2007) Emergence and control of fluoroquinolone-resistant, toxin A-negative, toxin B-positive *Clostridium difficile. Infection Control and Hospital Epidemiology,* 28**,** 932-940.

Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M. & Charles, J. F. (2003) The genome sequence of the entomopathogenic bacterium Photorhabdus luminescens. *Nature biotechnology,* 21**,** 1307-1313.

Dupont, H. L., Garey, K., Caeiro, J.-P. & Jiang, Z.-D. (2008) New advances in *Clostridium difficile* infection: changing epidemiology, diagnosis, treatment and control. *Current Opinion in Infectious Diseases,* 21**,** 500-507.

Dupuy, B., Govind, R., Antunes, A. & Matamouros, S. (2008) *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *Journal of Medical Microbiology,* 57**,** 685.

Durre, P. (2005) *Handbook on Clostridia,* Taylor & Francis.

Eastwood, K., Else, P., Charlett, A. & Wilcox, M. (2009) Comparison of nine commercially available Clostridium difficile toxin detection assays, a real-time PCR assay for *C. difficile tcd*B, and glutamate dehydrogenase detection assay to cytotoxin testing and cytotoxigenic culture methods. *Journal of Clinical Microbiology,* 47**,** 3211.

Egerer, M., Giesemann, T., Jank, T., Satchell, K. J. F. & Aktories, K. (2007) Auto-catalytic cleavage of *Clostridium difficile* toxins A and B depends on cysteine protease activity. *Journal of biological chemistry,* 282**,** 25314.

Elliott, T., Worthington, T., Osman, H. & Gill, M. (2007) *Lecture Notes: Medical Microbiology & Infection,* Blackwell Publishing.

Eltringham, I. J. (2009) Reflection and Reaction: Diagnosis of *Clostridium difficile* infection by toxin detection kits. *Lancet Infectious Diseases,* 9**,** 141-142.

Eveillard, M., Fourel, V., Bare, M. C., Kerneis, S., Coconnier, M. H., Karjalainen, T., Bourlioux, P. & Servin, A. L. (1993) Identification and characterization of adhesive

factors of Clostridium difficile involved in adhesion to human colonic enterocyte-like Caco-2 and mucus-secreting HT29 cells in culture. *Molecular Microbiology,* 7**,** 371-381

Eurosurveillance (2009) Eurosurveillance statistics.

Faust, C., Ye, B. & Song, K.-P. (1998) The enzymatic domain of *Clostridium difficile* Toxin A is located within its N-Terminal region. *Biochemical and Biophysical Research Communications,* 251**,** 100-105

Fawley, W. N., Underwood, S., Freeman, J., Baines, S. D., Saxton, K., Stephenson, K., Jr., R. C. O. & Wilcox, M. H. (2007) Efficacy of hospital cleaning agents and germicides against epidemic *Clostridium difficile* strains. *Infection Control and Hospital Epidemiology,* 28**,** 920-925.

Fawley, W. N., Freemen, J., Smith, C., Harmanus, C., Van Den Berg, R. J., Kuijper, E. J. & Wilcox, M. H. (2008) Use of highly discriminatory fingerprinting to analyze clusters of *Clostridium difficile* infection cases due to epidemic ribotype 027 Strains? *Journal of Clinical Microbiology,* 46**,** 954-960.

Fenner, L., Widmer, A. F., Goy, G., Rudin, S. & Frei, R. (2008) Rapid and reliable diagnostic algorithm for detection of *Clostridium difficile*? *Journal of Clinical Microbiology,* 46**,** 328-330.

Ferretti, J. J., Gilpin, M. L. & Russell, R. R. (1987) Nucleotide sequence of a glucosyltransferase gene from *Streptococcus sobrinus* MFe28. *Journal of Bacteriology,* 169**,** 4271.

Frazier, K. S., Herron, A. J., Hines, M. E., Gaskin, J. M. & Altman, N. H. (1993) Diagnosis of enteritis and enterotoxemia due to *Clostridium difficil*e in captive ostriches (*Struthio camelus*). *Journal of Veterinary Diagnostic Investigation,* 5**,** 623.

Freeman, J., Baines, S. D., Jabes, D. & Wilcox, M. H. (2005) Comparison of the efficacy of ramoplanin and vancomycin in both in vitro and in vivo models of clindamycin-induced *Clostridium difficile* infection. *Journal of Antimicrobial Chemotherapy,* 56**,** 717–725.

Gal, M., Northey, G. & Brazier, J. S. (2005) A modified pulsed-field gel electrophoresis (PFGE) protocol for subtyping previously non-PFGE typeable isolates of *Clostridium difficile* polymerase chain reaction ribotype 001. *Journal of Hospital Infection,* 61**,** 231-236.

Geddes, C. D. & Lakowicz, J. R. (2002) Editorial: Metal-enhanced fluorescence. *Journal of Fluorescence,* 12**,** 121-129.

George, R. H., Symonds, J. M., Dimock, F., Brown, J. D., Arabi, Y., Shinagawa, N., Keighley, M. R. B., Alexander-Williams, J. & Burdon, D. W. (1978) Identification of *Clostridium difficile* as a cause of pseudomembranous colitis. *British Medical Journal***,** 695.

Genth, H., Dreger, S. C., Huelsenbeck, J. & Just, I. (2008) *Clostridium difficile* toxins: More than mere inhibitors of Rho proteins. *The International Journal of Biochemistry & Cell Biology,* 40**,** 592-597.

Gerding, D. N., Johnson, S., Peterson, L. R., Mulligan, M. E. & Silva JR, J. (1995) *Clostridium difficile*-associated diarrhea and colitis. *Infect Control Hosp Epidemiology,* 16**,** 459-77.

Gerding, D. N. (2009) *Clostridium difficile* 30 years on: what has, or has not, changed and why? *International Journal of Antimicrobial Agents,* 33**,** S2-S8.

Geric, B., Carman, R. J., Rupnik, M., Genheimer, C. W., Sambol, S. P., Lyerly, D. M., Gerding, D. N. & Johnson, S. (2006) Binary toxin-producing, large clostridial toxin-negative *Clostridium difficile* strains are enterotoxic but do not cause disease in hamsters. *The Journal of infectious diseases,* 193**,** 1143-1150.

Geric Stare, B. & Rupnik, M. (2010) *Clostridium difficile* toxinotype XI (AB-) exhibits unique arrangement of PaLoc and its upstream region. *Anaerobe*

Gerrard, J., Waterfield, N. & Vohra, R. (2004) Human infection with *Photorhabdus asymbiotica*: an emerging bacterial pathogen. *Microbes and infection,* 6**,** 229-237.

Giannasca, P. J. & Warny, M. (2004) Active and passive immunization against *Clostridium difficile* diarrhea and colitis. *Vaccine,* 22**,** 848-856.

Giel, J. L., Sorg, J. A., Sononeshein, A. L. & Zhu, J. (2010) Metabolism of bile salts in mice influences spore germination in *Clostridium difficile*.

Giesemann, T., Jank, T., Gerhard, R., Maier, E., Just, I., Benz, R. & Aktories, K. (2006) Cholesterol-dependent pore formation of *Clostridium difficile* toxin A. *Journal of biological chemistry,* 281**,** 10808.

Giesemann, T., Egerer, M., Jank, T. & Aktories, K. (2008) Processing of *Clostridium difficile toxins. J Med Microbiol,* 57**,** 690-696.

Gilligan, P. H. (2008) Is a two-step glutamate dehyrogenase antigen-cytotoxicity neutralization assay algorithm superior to the premier Toxin A and B enzyme immunoassay for laboratory detection of *Clostridium difficile*? *Journal of Clinical Microbiology,* 46**,** 1523.

Goh, S., Riley, T. V. & Chang, B. J. (2005) Isolation and characterization of temperate bacteriophages of *Clostridium difficile*. *Applied and Environmental Microbiology,* 71**,** 1079.

Goldenberg, S. D., Cliff, P. R., Smith, S., Milner, M. & French, G. L. (2009) Two-step glutamate dehydrogenase antigen real-time polymerase chain reaction assay for detection of toxigenic *Clostridium difficile*. *Journal of Hospital Infection*.

Goldenberg, S. D. & French, G. L. (2011) Diagnostic testing for Clostridium difficile: a comprehensive survey of laboratories in England. *Journal of Hospital Infection*.

Goldenberg, S. D., Gumban, M., Hall, A., Patel, A. & French, G. L. (2011) Lack of effect of strain type on detection of toxigenic Clostridium difficile by glutamate dehydrogenase and polymerase chain reaction. *Diagnostic microbiology and infectious disease,* 70**,** 417-419.

Goulding, D., Thompson, H., Emerson, J., Fairweather, N. F., Dougan, G. & Douce, G. R. (2009) Distinctive profiles of infection and pathology in hamsters infected with *Clostridium difficile* strains 630 and B1. *Infection and Immunity,* 77**,** 5478.

Govind, R., Fralick, J. A. & Rolfe, R. D. (2006) Genomic organization and molecular characterization of *Clostridium difficile* bacteriophage {phi} CD119. *Journal of Bacteriology,* 188**,** 2568.

Govind, R., Vediyappan, G., Rolfe, R. D., Dupuy, B. & Fralick, J. A. (2009) Bacteriophage-Mediated Toxin Gene Regulation in *Clostridium difficile*. *Journal of virology,* 83**,** 12037.

Green, R. H. (1974) The association of viral activation with penicillin toxicity in guinea pigs and hamsters. *The Yale Journal of Biology and Medicine,* 47**,** 166.

Green, G. A., Schue, V. & Monteil, H. (1995) Cloning and characterization of the cytotoxin L-encoding gene of *Clostridium sordellii*: homology with *Clostridium difficile* cytotoxin B. *Gene,* 161**,** 57-61.

Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D. W., Fung, R., Golubchik, T., Harding, R. M., Jeffery, K. J. M. & Jolley, K. A. (2010) Multilocus Sequence Typing of *Clostridium difficile*. *Journal of Clinical Microbiology,* 48**,** 770.

Griffith, F. (1923) The influence of immune serum on the biological properties of pneumococci. *Great Britain Rep. on Public Health and Med. Subjects, Ministry of Health,* 18**,** 1-13.

Gulke, I., Pfeifer, G., Liese, J., Fritz, M., Hofmann, F., Aktories, K. & Barth, H. (2001) Characterization of the enzymatic component of the ADP-ribosyltransferase toxin CDTa from *Clostridium difficile*. *Infection and Immunity,* 69**,** 6004.

Gumerlock, P. H., Tang, Y. J., Weiss, J. B. & Silva JR, J. (1993) Specific detection of toxigenic strains of *Clostridium difficile* in stool specimens. *Journal of Clinical Microbiology,* 31**,** 507.

Hacek, D. M., Ogle, A. M., Fisher, A., Robisek, A. & Peterson, L. R. (2010) Significant impact of terminal room cleaning with bleach on reducing nosocomial *Clostridium difficile*. *American Journal of Infection Control*.

Hacker, J., Hochhut, B., Middendorf, B., Schneider, G., Buchrieser, C., Gottschalk, G. & Dobrindt, U. (2004) Pathogenomics of mobile genetic elements of toxigenic bacteria. *International Journal of Medical Microbiology,* 293**,** 453-461.

Hacker, J., Hentschel, U. & Dobrindt, U. (2003) Prokaryotic chromosomes and disease. *Science,* 301**,** 790.

Hall, I. C. & O'Toole, E. (1935) Intestinal flora in newborn infants with description of a new pathogenic anaerobe. *American Journal of Disease in Children* 19**,** 309-402.

Hall, T. (2001) BioEdit software *v*7.3.1. http://www.mbio.ncsu.edu/bioedit/bioedit.html [Accessed 2012]

Hammitt, M. C., BueschelL, D. M., Keel, M. K., Glock, R. D., Cuneo, P., Deyoung, D. W., Reggairdo, C., Trinh, H. T. & Songer, J. G. (2008) A possible role for *Clostridium difficile* in the aetiology of calf enteritis. *Veterinary Microbiology,* 127**,** 343-352.

Haraldsen, J. D. & Sonenshein, A. L. (2003) Efficient sporulation in *Clostridium difficile* requires disruption of the sσK gene. *Molecular Microbiology,* 48**,** 811-821.

Hatheway, C. L. (1990) Toxigenic clostridia. *Clinical Microbiology Reviews,* 3**,** 66.

Heermann, R. & Fuchs, T. (2008) Comparative analysis of the *Photorhabdus luminescens* and the *Yersinia enterocolitica* genomes: uncovering candidate genes involved in insect pathogenicity. *BMC genomics,* 9**,** 40.

Hellickson, L. A. & Owens, K. L. (2007) Cross-contamination of *Clostridium difficile* spores on bed linen during laundering. *AJIC: American Journal of Infection Control,* 35**,** 32-33.

Henriques, A. O. & Moran, C. P. (2007) Structure, Assembly, and Function of the Spore Surface Layers. *Annual Review of Microbiology,* 61, 555.

Hernandez-Martinez, P., Naseri, B., Navarro, A., Cerrillo, G., Escriche, B., Ferra, J. & Herrero, S. (2010) Increase in midgut microbiota load induces an apparent immune priming and increases tolerance to *Bacillus thuri*ngiensis. *Environmental Microbiology,* 12**,** 2730-2737.

Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. (2009) Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*.

Hidalgo, I. J., Raub, T. J. & Borchardt, R. T. (1989) Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology,* 96**,** 736.

Hinkson, P. L., Dinardo, C., Deceiro, D., Klinger, J. D. & JR., R. H. B. (2008) Tolevamer, an anionic polymer, neutralizes toxins produced by the BI/027 strains of *Clostridium difficile*. *Antimicrobial Agents and Chemotherapy,* 52**,** 2190–2195.

Hoffman, J. D., Lawson, F. R. & Yamamoto, R. (1966) Tobacco hornworms. *Insect colonization and mass production***,** 479-486.

Hofte, H. & Whiteley, H. R. (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiology and Molecular Biology Reviews,* 53**,** 242.

Holden, H. M., Rypniewski, W. R., Law, J. H. & Rayment, I. (1987) The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* L. at 2.6 A resolution. *The EMBO journal,* 6**,** 1565.

Hu, M. Y., Katchar, K., Kyne, L., Maroo, S., Tummala, S., Dreisbach, V., Xu, H., Leffler, D. A. & Kelly, C. P. (2009) Prospective derivation and validation of a clinical prediction rule for recurrent *Clostridium difficile* infection. *Gastroenterology,* 136**,** 1206-12

Igimi, S., Ryu, C. H., Park, S. H., Sasaki, Y., Sasaki, T. & Kumagai, S. (2008) Transfer of conjugative plasmid pAM 1 from *Lactococcus lactis* to mouse intestinal bacteria. *Letters in applied microbiology,* 23**,** 31-35.

Integrated DNA Technologies (2009) Oligoanalyser 3.1.

Igarashi, T., Yamamoto, A. & Goto, N. (2000) PCR for detection and identification of *Streptococcus sobrinus*. *Journal of Medical Microbiology,* 49**,** 1069.

Ishikawa, H. (1989) Biochemical and molecular aspects of endosymbiosis in insects. *International review of cytology,* 116**,** 1-46.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N. & Lapidus, A. (2003) Genome sequence of **Bacillus cereus** and comparative analysis with **Bacillus anthracis**. *Nature,* 423**,** 87-91.

Jank, T., Giesemann, T. & Aktories, K. (2007) Rho-glucosylating *Clostridium difficile* toxins A and B: new insights into structure and function. *Glycobiology,* 17**,** 15R - 22R.

Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. J. (2005) *Immunobiology: the immune system in health and disease,* Garland Science.

Janoir, C., Pechine, S., Grosdidier, C. & Collignon, A. (2007) Cwp84, a surface-associated protein of *Clostridium difficile,* is a cysteine protease with degrading activity on extracellular matrix proteins. *Journal of Bacteriology,* 189**,** 7174.

Janvilisri, T., Scaria, J., Thompson, A. D., Nicholson, A., Limbago, B. M., Arroyo, L. G., Songer, J. G., Grohn, Y. T. & Chang, Y. F. (2009) Microarray identification of *Clostridium difficile* core components and divergent regions associated with host origin. *Journal of Bacteriology,* 191**,** 3881.

Jhung, G, M. A., Thompson, A. D., Killgore, G. E., Zukowski, W. E., Songer, G., Warny, M., Johnson, S., Gerding, D. N., Mcdonald, L. C. & Limbag, B. M. (2008) Toxinotype V *Clostridium difficile* in humans and food animals. *Emerging Infectious Diseases,* 14**,** 1039-1045

Jiang, H. (2008) The biochemical basis of antimicrobial responses in *Manduca sexta*. *Insect Science,* 15**,** 53-66.

Jobstl, M., Heuberger, S., Indra, A., Nepf, R., Kofer, J. & Wagner, M. (2009) *Clostridium difficile* in raw products of animal origin. *International Journal of Food Microbiology*.

Just, I. & Gerhard, R. (2004) Large clostridial cytotoxins. *Reviews of physiology, biochemistry and pharmacology,* 152**,** 23-47.

Kane, S. V., Sandborn, W. J., Rufo, P. A., Zholudev, A., Boone, J., Lyerly, D., Camilleri, M. & Hanauer, S. B. (2003) Fecal lactoferrin is a sensitive and specific marker in identifying intestinal inflammation. *American Journal of Gastroenterology,* 98**,** 1309-1314.

Kang, J. K., Hwang, J. S., Nam, H. J., Ahn, K. J., Seok, H., Kim, S. K., Yun, E. Y., Pothoulakis, C., Lamont, J. T. & Kim, H. (2011) The insect peptide Coprisin prevents *Clostridium difficile*-mediated acute inflammation and mucosal damage through selective antimicrobial activity. *Antimicrobial Agents and Chemotherapy,* 55**,** 4850-4857.

Karginov, F. V. & Hannon, G. J. (2010) The CRISPR system: small RNA-guided defence in bacteria and archaea. *Molecular cell,* 37**,** 7-19.

Karlsson, S., Burman, L. G. & Akerlund, T. (1999) Suppression of toxin production in *Clostridium difficile* VPI 10463 by amino acids. *Microbiology,* 145**,** 1683.

Kato, H., Yokoyama, T. & Arakawa, Y. (2005) Typing by sequencing the slpA gene of *Clostridium difficile* strains causing multiple outbreaks in Japan. *Journal of Medical Microbiology,* 54**,** 167.

Kavanagh, K. & Reeves, E. P. (2004) Exploiting the potential of insects for in vivo pathogenicity testing of microbial pathogens. *FEMS Microbiology Reviews,* 28**,** 101-112.

Kawada, M., Chen, C. C., Arihiro, A., Nagatani, K., Watanabe, T. & Mizoguchi, E. (2008) Chitinase 3-like-1 enhances bacterial adhesion to colonic epithelial cells through the interaction with bacterial chitin-binding protein. *Laboratory Investigation,* 88**,** 883-895.

Keighley, M. R. B., Alexander-Williams, J., Arabi, Y., Youngs, D., Burdon, D. W., Shinagawa, N., Thompson, H., Bentley, S. & George, R. H. (1978) Diarrhoea and pseudomembranous colitis after gastrointestinal operations. *The Lancet***,** 1164 - 1167.

Killgore, G., Thompson, A., Johnson, S., Brazier, J., Kuijper, E., Pepin, J., Frost, E. H., SavelkoulL, P., Nicholson, B. & Van Den Berg, R. J. (2008) Comparison of seven techniques for typing international epidemic strains of *Clostridium difficile*: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing. *Journal of Clinical Microbiology,* 46**,** 431.

Kim, S. Y., Shin, S. J., Song, C. H., Jo, E. K., Kim, H. J. & Park, J. K. (2009) Destruction of *Bacillus licheniformis* spores by microwave irradiation. *Journal of Applied Microbiology,* 106**,** 877-885.

Klaassen, C. H. W., Van Haren, H. A. & Horrevorts, A. M. (2002) Molecular fingerprinting of *Clostridium difficile* isolates: pulsed-field gel electrophoresis versus amplified fragment length polymorphism. *Journal of Clinical Microbiology,* 40**,** 101.

Klapproth, J. M. A., Sasaki, M., Sherman, M., Babbin, B., Donnenberg, M. S., Fernades, P. J., Scaletsky, I. C. A., Kalman, D., Nusrat, A. & Williams, I. R. (2005) *Citrobacter rodentium* lifA/efa1 is essential for colonic colonization and crypt cell hyperplasia in vivo. *Infection and Immunity,* 73**,** 1441.

Knight, P. J. K., Crickmore, N. & Ellar, D. J. (1994) The receptor for *Bacillus thuringiensis* CrylA (c) delta endotoxin in the brush border membrane of the lepidopteran *Manduca sexta* is aminopeptidase N. *Molecular Microbiology,* 11**,** 429-436.

Knetsch, C. W., Bakker, D., De Boer, R. F., Sanders, I., Hofs, S., Kooistra-Smid, A. M. D., Corver, J., Eastwood, K., Wilcox, M. H. & Kuijper, E. J. (2011) Comparison of real-time PCR techniques to cytotoxigenic culture methods for diagnosing Clostridium difficile infection. *Journal of Clinical Microbiology,* 49**,** 227.

Kononen, E., Rasinpera, M., Virolainen, A., Mentula, S. & Lyytikainen, O. (2009) Diagnostic trends in *Clostridium difficile* detection in Finnish microbiology laboratories. *Anaerobe*.

Kramer, K. J. & Muthukrishnan, S. (1998) Insect chitinases: molecular biology and potential use as biopesticides. *Insect Biochemistry and Molecular Biology,* 27**,** 887-900.

Kuehne, S. A., Cartman, S. T., Heap, J. T., Kelly, M. L., Cockayne, A. & Minton, N. P. (2010) The role of toxin A and toxin B in Clostridium difficile infection. *Nature,* 467**,** 711-713.

Kuijpers, E. J., Coignard, B., Brazier, J. S., Suetens, C., Drudy, D., Wiuff, C., Pituch, H., Reichert, P., Schneider, F. & Widmer, A. F. (2008) Update of *Clostridium difficile*-associated disease due to PCR ribotype 027 in Europe. *Euro Surveill,* 12**,** E1 - 2.

Kuijpers, E. J. & Surawicz, C. M. (2008) *Clostridium difficile* infection. *The Lancet,* 371**,** 1486-1488.

Kvach, E. J., Ferguson, D., Riska, P. F. & Landry, M. L. (2010) Comparison of BD GeneOhm Cdiff Real-Time PCR Assay with a Two-Step Algorithm and a Toxin A/B Enzyme-Linked Immunosorbent Assay for Diagnosis of Toxigenic *Clostridium difficile* Infection. *Journal of Clinical Microbiology,* 48**,** 109.

Kwok, S. & Higuchi, R. (1989) Avoiding false positives with PCR. *Nature,* 339**,** 237-238.

Kyne, L. (2010) *Clostridium difficile*--Beyond Antibiotics. *The New England Journal of Medicine,* 362**,** 264.

Larkin M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) ClustalW and ClustalX version 2. Bioinformatics (2007) 23(21): 2947-2948.

Larson, A. M., Fung, A. M. & Fang, F. C. (2010) Evaluation of tcdB real-time PCR in a three-step diagnostic algorithm for detection of toxigenic *Clostridium difficile*. *Journal of Clinical Microbiology,* 48**,** 124.

Larson, H. E., Price, A. B., Honour, P. & Borriello, S. P. (1978) *Clostridium difficile* and the aetiology of pseudomembranous colitis. *The Lancet***,** 1063 - 1066.

Lawley, T. D., Croucher, N. J., Yu, L., Clare, S., Sebaihia, M., Goulding, D., Pickard, D. J., Parkhill, J., Choudhary, J. & Dougan, G. (2009) Proteomic and genomic characterization of highly infectious *Clostridium difficile* 630 spores. *J. Bacteriol.,* 191**,** 5377-5386.

Lemee, L., Dhalluin, A., Pestel-Caron, M., Lemeland, J. F. & Pons, J. L. (2004) Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. *Journal of Clinical Microbiology,* 42**,** 2609.

Lemee, L., Bourgeoi, I., Ruffin, E., Collignon, A., Lemeland, J.-F. & Pons, J.-L. (2005) Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of *Clostridium difficile*. *Microbiology* 151**,** 3171–3180

Lemieux, J., Woody, S. & Camilli, A. (2008) Roles of the sortases of *Streptococcus pneumoniae* in assembly of the RlrA pilus. *Journal of Bacteriology,* 190**,** 6002.
Lesk, A. M. (2005) *Introduction to Bioinformatics,* Oxford University Press.

Letournier, O., Ottone, S., Delauzun, V., Bastide, M.-C. & Foussadier, A. (2003) Molecular cloning, overexpression in *Escherichia coli*, and purification of 6 his-tagged C-terminal domain of *Clostridium difficile* toxins A and B. *Protein Expression and Purification,* 31**,** 276–285

Leverrier, P., Dimova, D., Pichereau, V., Auffray, Y., Boyaval, P. & Jan, G. (2003) Susceptibility and adaptive response to bile salts in *Propionibacterium freudenreichii*: physiological and proteomic analysis. *Applied and Environmental Microbiology,* 69**,** 3809.

Lopez, R., Garcia, E., Garcia, P. & Garcia, J. L. (2004) Cell wall hydrolases. *The pneumococcus***,** 75–88.

Lowy, I., Molrine, D. C., Leav, B. A., Blair, B. M., Baxter, R., Gerding, D. N., Nichol, G., Thomas JR, W. D., Leney, M. & Sloan, S. (2010) Treatment with monoclonal antibodies against Clostridium difficile toxins. *New England Journal of Medicine,* 362**,** 197.

Luna, V. A., King, D. S., Gulledge, J., Cannons, A. C., Amuso, P. T. & Cattani, J. (2007) Susceptibility of *Bacillus anthracis*, *Bacillus cereus*, *Bacillus mycoides*, *Bacillus pseudomycoides* and *Bacillus thuringiensis* to 24 antimicrobials using Sensititre(R)

automated microbroth dilution and Etest(R) agar gradient diffusion methods. *J. Antimicrob. Chemother.,* 60**,** 555-567.

Lyerly, D. M., Krivan, H. C. & Wilkins, T. D. (1988) *Clostridium difficile*: Its disease and toxins. *Clinical Microbiology Reviews,* 1**,** 1-18.

Lyerly, D. M., Barroso, L. A. & Wilkins, T. D. (1991) Identification of the latex test-reactive protein of *Clostridium difficile* as glutamate dehydrogenase. *Journal of Clinical Microbiology,* 29**,** 2639.

Lyras, D., O'Connor, J. R., Howarth, P. M., Sambol, S. P., Carter, G. P., Phumoonna, T., Poon, R., Adams, A., Vedantam, G., Johnson, S., Gerding, D. N. & Rood, J. I. (2009) Toxin B is essential for virulence of *Clostridium difficile*. *Nature,* 458**,** 1176-1179.

Macconnachie, A. A., Fox, R., Kennedy, D. R. & Seaton, R. A. (2009) Faecal transplant for recurrent Clostridium difficile-associated diarrhoea: a UK case series. *QJM,* 102**,** 781.

Mahillon, J. & Chandler, M. (1998) Insertion sequences. *American Society for Microbiology,* 62 725–774.

Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, M., Nakano, M. & Yamashita, A. (2003) Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of *V cholerae*. *The Lancet,* 361**,** 743-749.

Mani, N. & Dupuy, B. (2001) Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proceedings of the National Academy of Sciences,* 98**,** 5844.

Marsh, J. W., O'Learly, M. M., Shutt, K. A., Pasculle, A. W., Johnson, S., Gerding, D. N., Muto, C. A. & Harrison, L. H. (2006) Multilocus variable-number tandem-repeat analysis for investigation of *Clostridium difficile* transmission in hospitals. *Journal of Clinical Microbiology,* 44**,** 2558-2566.

Marsh, J. W., O'Leary, M. M., Shutt, K. A., Sambol, S. P., Johnson, S., Gerding, D. N. & Harrison, L. H. (2010) Multilocus Variable-Number Tandem-Repeat Analysis and Multilocus Sequence Typing Reveal Genetic Relationships among *Clostridium difficile* Isolates Genotyped by Restriction Endonuclease Analysis. *Journal of Clinical Microbiology,* 48**,** 412.

Matamouros, S., England, P. & Dupuy, B. (2007) *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Molecular Microbiology,* 64**,** 1274-1288.

Mehlig, M., Moos, M., Braun, V., Kalt, B., Mahony, D. E. & Eichel-Streiber, C. V. (2001) Variant toxin B and a functional toxin A produced by *Clostridium dfficile* C34. *FEMS Microbiology Letters* 198 171-176.

Mejean, V. & Claverys, J. P. (1993) DNA processing during entry in transformation of *Streptococcus pneumoniae*. *Journal of biological chemistry,* 268**,** 5594.

McFarland, L. V. (2005) Alternative treatments for *Clostridium difficile* disease: what really works? *Journal of Medical Microbiology,* 54**,** 101-111.

Mclemore, J. B., Tainter, F. H. & Mcelreath, S. D. (2000) Bacteria associated with butterfly stain of Chilean tepa. *Current microbiology,* 40**,** 86-90.

Miles, A. A., Misra, S. S. & Irwin, J. O. (1938) The estimation of the bactericidal power of the blood. *The Journal of Hygiene***,** 732-749.

Miller, M. A. (2007) Clinical management of *Clostridium difficile*–Associated Disease. *Clinical Infectious Diseases,* 45**,** 122-128.

Miller, K. & Fraser, T. (2009) Q: What is the role of probiotics in the treatment of acute *Clostridium difficile*-associated diarrhea? *Cleveland Clinic Journal of Medicine,* 76**,** 391.

Moncrief, J. S., Barroso, L. A. & Wilkins, T. D. (1997) Positive regulation of *Clostridium difficile* toxins. *Infection and Immunity,* 65**,** 1105.

Morita, H., Toh, H., Fukida, S., Horikawa, H., Oshima, K., Suzuki, T., Murakami, M., Hisamatsu, S., Kato, Y. & Takizawa, T. (2008) Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. *DNA research*.

Morrison, M. & Miron, J. (2006) Adhesion to cellulose by *Ruminococcus albus*: a combination of cellulosomes and Pil-proteins? 1. *FEMS Microbiology Letters,* 185**,** 109-115.

Mt, Y. A., Msc, N. S., Kabir, A. M. A., Takagi, A. & Koga, Y. (1998) Lactic acid-mediated suppression of *Helicobacter pylori* by the oral administration of *Lactobacillus salivarius* as a probiotic in a gnotobiotic murine model. *American Journal of Gastroenterology,* 93**,** 2097-2101.

Mullany, P., Pallen, M., Wilks, M., Stephen, J. R. & Tabaqchali, S. (1996) A group II intron in a conjugative transposon from the gram-positive bacterium, *Clostridium difficile*. *Gene,* 174**,** 145-150.

Mulligan, M. E., Citron, D., Gabay, E., Kirby, B. D., George, W. L. & Finegold, S. M. (1984) Alterations in human fecal flora, including in growth of *Clostridium difficile*, related to cefoxitin therapy. *Antimicrobial Agents and Chemotherapy,* 26**,** 343-346.

Murdoch, C. W. & Campana, R. J. (1983) Bacterial species associated with wetwood of elm. *Phytopathology,* 73**,** 1270-1273.

Murdoch, D. A. & Mitchelmore, I. J. (1989) Isolation of *Peptostreptococcus heliotrinreducens* from human polymicrobial abscesses. *Letters in applied microbiology,* 9**,** 223-225.

Muto, A. & Osawa, S. (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. *Proceedings of the National Academy of Sciences,* 84**,** 166.

Mutters, R., Nonnenmacher, C., Susin, C., Albrecht, U., Kropatsch, R. & Schumacher, S. (2009) Quantitative detection of *Clostridium difficile* in hospital environmental samples by real-time polymerase chain reaction. *Journal of Hospital Infection,* 71**,** 43-48.

Nagahama, M., Ohkubo, A., Oda, M., Kobayashi, K., Amimoto, K., Miyamoto, K. & Sakurai, J. (2011) *Clostridium perfringens* TpeL glycosylates the Rac and Ras subfamily proteins. *Infection and Immunity,* 79**,** 905.

NFSTC (2009) Chelex 100 Extraction. USA.

Ni Eidhin, D. B., O'Brien, J. B., Mccabe, M. S., Athie-Morales, V. & Kellerher, D. P. (2008) Active immunization of hamsters against *Clostridium difficile* infection using surface-layer protein. *FEMS Immunology & Medical Microbiology,* 52**,** 207-218.

Novak-Weekely, S. M., Marlowe, E. M., Miller, J. M., Cumpio, J., Nomura, J. H., Vance, P. H. & Wiessfeld, A. (2010) *Clostridium difficile* testing in the clinical laboratory by use of multiple testing algorithms. *Journal of Clinical Microbiology,* 48**,** 889.

Numan, S. C., Veldkamp, P., Kuijper, E. J., van den Berg, R. J., & van Dissel, J. T. (2007). *Clostridium difficile*-associated diarrhoea: bovine anti-*Clostridium difficile* whey protein to help aid the prevention of relapses. *Gut, 56,* 888–889.

O'Brien, M. & Colwell, R. R. (1987) A rapid test for chitinase activity that uses 4 methylumbelliferyl-N-acetyl-beta-D-glucosaminide. *Applied and Environmental Microbiology,* 53**,** 1718.

Office of National Statistics Online. 2010 [http://www.statistics.gov.uk/hub/index.html] Accession Date: 08/2011.

Pamer, E. G. (2007) Immune responses to commensal and environmental microbes. *Nature immunology,* 8**,** 1173-1178.

Panessa-Warren, B. J., Tortora, G. T. & Warren, J. B. (2007) High Resolution FESEM and TEM reveal bacterial spore attachment. *Microscopy and Microanalysis,* 13**,** 251-266.

Paredes, C. J., Alasker, K. V. & Papoutsakis, E. T. (2005) A comparative genomic view of Clostridial sporulation and physiology. *Nature Reviews: Microbiology,* 3**,** 969-978.

Paredes-Sabja, D., Bond, C., Carman, R. J., Setlow, P. & Sarker, M. R. (2008) Germination of spores of *Clostridium difficile* strains, including isolates from a hospital outbreak of *Clostridium difficile*-associated disease (CDAD). *Microbiology,* 154**,** 2241.

Pechine, S., Janoir, C., Boureau, H., Gleizes, A., Tsapis, N., Hoys, S., Fattal, E. & Collignon, A. (2007) Diminished intestinal colonization by Clostridium difficile and immune response in mice after mucosal immunization with surface proteins of Clostridium difficile. *Vaccine,* 25**,** 3946-3954.

Pepin, J., Saheb, N., Coulombe, M. A., Alary, M. E., Corriveau, M. P., Authier, S., Leblanc, M., Rivard, G., Bettez, M., Primeau, V., Nguyen, M., Jacob, C. E. & Lanthier, L. (2005) Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*–Associated Diarrhoea: A cohort study during an epidemic in Quebec. *Clinical Infectious Diseases,* 41**,** 1254–1260.

Perelle, S., Gibert, M., Bourlioux, P., Cothier, G. & Popoff, M. R. (1997) Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infection and Immunity,* 65**,** 1402.

Perez, J., Springthorpe, V. S. & Sattar, S. A. (2005) Activity of selected oxidizing microbicides against the spores of *Clostridium difficile*: Relevance to environmental control. *AJIC: American Journal of Infection Control,* 33**,** 320-325

Permpoonpattana, P., Tolls, E. H., Nadem, R., Tan, S., Brisson, A. & Cutting, S. M. (2011) Surface layers of *Clostridium difficile* endospores. *J. Bacteriol.***,** JB.05182-11.

Peterson, L. R., Manson, R. U., Paule, S. M., Hacek, D. M., Robicsek, A., Thomson, R. B. & Kaul, K. L. (2007) Detection of toxigenic *Clostridium difficile* in stool samples by Real-Time Polymerase Chain Reaction for the diagnosis of *C. difficile*-Associated Diarrhoea. *Clinical Infectious Diseases,* 45**,** 115

Piggot, P. J. & Losick, R. (2002) Sporulation genes and intercompartmental regulation. *Bacillus subtilis and its Closest Relatives***,** 483–517.

Piggot, P. J. & Hilbert, D. W. (2004) Sporulation of *Bacillus subtilis*. *Current opinion in microbiology,* 7**,** 579-586.

Pinto, M., Appay, S,, Simon-Assmann, P., Chevalier, G., Dracopoli, N., Fogh, J., and Zweibaum, A. (1982) Enterocytic differentiation of cultured human colon cancer cells by replacement of glucose by galactose in the medium. *Btol Celt* 44:193-196.

Pinto, M., Robine-Leon, S., Appay, M.D., Kedinger, M., Triadou, N., Dussaulx, E., Lacroix, B., Simon-Assmann, P., Ha¡en, K., Fogh, J. and Zweibaum, A. (1983) Enterocyte-like differentiation and polarization of the human colon carcinoma cell line Caco-2 in culture. Biol. Cell 47, 323-330.

Planche, T., Achaizu, A., Holliman, R., Riley, P., Poloniecki, J., Breathnack, A. & Krishna, S. (2008) Diagnosis of *Clostridium difficile* infection by toxin detection kits: a systematic review. *The Lancet Infectious Diseases,* 8**,** 777-784.

Planche, T. & Wilcox, M. (2011) Reference assays for *Clostridium difficile* infection: one or two gold standards? *Journal of clinical pathology,* 64**,** 1-5.

Planche, T. & Wilcox, M. (2010) Two-step testing for *C. difficile*: no answers yet. *The Journal of hospital infection,* 75**,** 325.

Poilane, I., Karjalainen, T., Barc, M. C., Bourlioux, P. & Collignon, A. (1998) Protease activity of *Clostridium difficile* strains. *Canadian journal of microbiology,* 44**,** 157-161.

Poutanen, S. M. & Simor, A. E. (2004) *Clostridium difficile*-associated diarrhea in adults. *Canadian Medical Association Journal,* 171**,** 51-58.

Poxton, I. R., McCoubrey, J. & Blair, G. (2001) The pathogenicity of *Clostridium difficile*. *Clinical Microbiology & Infection,* 7**,** 421-427.

Priest, F. G., Barker, M., Baillie, L. W. J., Holmes, E. C. & Maiden, M. C. J. (2004) Population structure and evolution of the Bacillus cereus group. *Journal of Bacteriology,* 186**,** 7959.

Primrose, S. B. and Twyman, R. M. (2007) Basic Techniques. In: *Principles of Gene Manipulation and Genomics*. 7th Edition, Blackwell Publishing. pp 15-35.

Pruitt, R. N., Chambers, M. G., Ng, K. K. S., Ohi, M. D. & Lacy, D. B. (2010) Structural organization of the functional domains of Clostridium difficile toxins A and B. *Proceedings of the National Academy of Sciences,* 107**,** 13467.

Qa'Dan, M., Spyres, L. M. & Ballard, J. D. (2000) pH-Induced Conformational changes in *Clostridium difficile* Toxin B. *Infection and Immunity,* 68**,** 2470.

Quinn, C. D., Sefers, S. E., Babiker, W., He, Y., Alcabasa, R., Stratton, C. W., Carroll, K. C. & Tang, Y. W. (2010) C. Diff Quik Chek Complete Enzyme Immunoassay Provides a Reliable First-Line Method for Detection of *Clostridium difficile* in Stool Specimens. *Journal of Clinical Microbiology,* 48**,** 603.

Read, T. D., Peterson, S. N., Tourasse, N., Baillie, L. W., Paulsen, I. T., Nelson, K. E., Tettelin, H., Fouts, D. E., Eisen, J. A. & Gills, S. R. (2003) The genome sequence of *Bacillus anthracis Ames* and comparison to closely related bacteria. *Nature,* 423**,** 81-86.

Reinert, D. J., Jank, T., Aktories, K. & Schulz, G. E. (2005) Structural Basis for the function of *Clostridium difficile* Toxin B. *Journal of Molecular Biology,* 351**,** 973–981.

Reineke, J., Tenzer, S., Rupnik, M., Koschincki, A., Hasselmayer, O., Schrattenholz, A., Schild, H. & Von Eichel-Streiber, C. (2007) Autocatalytic cleavage of *Clostridium difficile* toxin B. *Nature,* 446**,** 415-419.

Reynolds, S. E. & Nottingham, S. F. (1985) Effects of temperature on growth and efficiency of food utilization in fifth-instar caterpillars of the tobacco hornworm, *Manduca sexta. Journal of insect physiology,* 31**,** 129-134.

Reynolds, C. B., Emerson, J. E., de la Riva, L., P., F. R. and Fairweather, N. F. (2010). Variation and conservation: The regulation of *Clostridium difficile* cell wall protein CwpV. 3rd Interational Clostridium difficile symposium, Bled, Slovenia.

Reynolds, C. B., Emerson, J. E., De La Riva, L., Fagan, R. P. & Fairweather, N. F. (2011) The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS pathogens,* 7**,** e1002024.

Rifkin, G. D., Fekety, F. R. & Silva, J. (1977) Antibiotic-induced colitis implication of a toxin neutralised by *Clostridium sordellii*. *The Lancet***,** 1103-1107.

Rinttila, T., Kassinen, A., Malinen, E., Krogius, L. & Palva, A. (2004) Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *Journal of Applied Microbiology,* 97**,** 1166-1177.

Rodou, A., Ankrah, D. O. & Stathopoulos, C. (2010) Toxins and secretion systems of *Photorhabdus luminescens*. *Toxins,* 2**,** 1250-1264.

Rolff, J. & Reynolds, S. (2009) *Insect infection and immunity: evolution, ecology, and mechanisms,* Oxford University Press, USA.

Rozen, S., and Skaletsky, H. J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ, pp 365-386

Rupnik, M., Braun, V., Soehn, F., Janc, M., Hofstetter, M., Laufenberg-Feldmann, R. & von Eichel-Streiber, C. (1997) Characterization of polymorphisms in the toxin A and B genes of *Clostridium difficile*. *FEMS Microbiology Letters,* 148**,** 197-202.

Rupnik, M., Avesani, V., Janc, M., von Eichel-Streiber, C. & Delmee, M. (1998) A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. *Journal of Clinical Microbiology,* 36**,** 2240-2247.

Rupnik, M. (2001) How to detect *Clostridium difficile* variant strains in a routine laboratory. *Clinical Microbiology & Infection,* 7**,** 417-420

Rupnik, M. (2008) Heterogeneity of large clostridial toxins: importance of *Clostridium difficile* toxinotypes. *FEMS Microbiology Reviews,* 32**,** 541–555.

Rupnik, M., Wilcox, M. H. & Gerding, D. N. (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nature Reviews: Microbiology,* 7**,** 526-536.

Ryle, A. P., Sanger, F., Smith, L. F. & Kitai, R. (1955) The disulphide bonds of insulin. *Biochemical Journal,* 60**,** 541.

Saah, A. J. & Hoover, D. R. (1997) "Sensitivity" and "Specificity" reconsidered: The meaning of these terms in analytical and diagnostic settings. *Annals of internal medicine,* 126**,** 91.

Sadd, B. M. & Schmid-Hempel, P. (2006) Insect immunity shows specificity in protection upon secondary pathogen exposure. *Current Biology,* 16**,** 1206-1210.

Salyers, A. A. & Whitt, D. D. (2002) Bacterial Pathogenesis: A Molecular Approach. 2nd Edition ed., ASM Press.

Salzet, M. (2001) Vertebrate innate immunity resembles a mosaic of invertebrate immune responses. *Trends in Immunology,* 22**,** 285-288.

Sambol, S. P., Merrigan, M. M., Lyerly, D., Gerding, D. N., Johnson, S. & Barbeiri, J. T. (2000) Toxin gene analysis of a variant strain of *Clostridium difficile* that causes human clinical disease. *Infection and Immunity,* 68**,** 5480-5487.

Schmidt, H. & Hensel, M. (2004) Pathogenicity islands in bacterial pathogenesis. *Clinical Microbiology Reviews,* 17**,** 14.

Schwan, C., Stecher, B., Tzivelekidis, T., Van Han, M., Rohde, M., Hardt, W. D., Wehland, J. & Aktories, K. (2009) *Clostridium difficile* Toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria.

Scully, L. R. & Bidochka, M. J. (2006) Developing insect models for the study of current and emerging human pathogens. *FEMS Microbiology Letters,* 263**,** 1-9.

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeno-Tarraga, A. M. & Wang, H. (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics,* 38**,** 779-786.

Sebaihia, M., Peck, M. W., Minton, N. P., Thomson, N. R., Holden, M. T. G., Mitchell, W. J., Carter, A. T., Bentley, S. D., Mason, D. R. & Crossman, L. (2007) Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the Clostridial genomes. *Genome research,* 17**,** 1082.

Seed, K. D. & Dennis, J. J. (2008) Development of *Galleria mellonella* as an alternative infection model for the *Burkholderia cepacia* complex. *Infection and Immunity,* 76**,** 1267.

Shames, S. R., Auweter, S. D. & Finlay, B. B. (2009) Co-evolution and exploitation of host cell signalling pathways by bacterial pathogens. *The International Journal of Biochemistry & Cell Biology,* 41**,** 380-389.

She, R. C., Durrant, R. J. & Petti, C. A. (2009) Evaluation of enzyme immunoassays to detect *Clostridium difficile* toxin from anaerobic stool culture. *American journal of clinical pathology,* 131**,** 81.

Sheahan, K. L., Cordero, C. L. & Fullner Satchell, K. J. (2004) Identification of a domain within the multifunctional *Vibrio cholerae* RTX toxin that covalently cross-links actin. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 9798.

Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S. & Hayashi, H. (2002) Complete genome sequence of *Clostridium perfringens,* an anaerobic flesh-eater. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 996.

Schink, B., Ward, J. C. & Zeikus, J. G. (1981) Microbiology of wetwood: role of anaerobic bacterial populations in living trees. *Microbiology,* 123**,** 313.

Shoemaker, N. B., Vlamakis, H., Hayes, K. & Salyers, A. A. (2001) Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Applied and Environmental Microbiology,* 67**,** 561.

Siani, H., Groen, H., Maillard, J. Y. and Baillie, L. (2010). Characterisation of variant *C. difficile* morphotypes - Microscopic examination and proteome analysis. 3rd International *Clostridium difficile* symposium, Bled, Slovenia

Singh, V. K., Gvindarajan, R., Naik, S. & Kumar, A. (2000) The effect of hairpin structure on PCR Amplification efficiency. *Mol Biol Today,* 1**,** 657-659.

Sloan, L. M., Duresko, B. J., Gustafson, D. R. & Rosenblatt, J. E. (2008) Comparison of Real-Time PCR for detection of the tcdC gene with four toxin immunoassays and culture in diagnosis of *Clostridium difficile* Infection. *Journal of Clinical Microbiology,* 46**,** 1996-2001.

Small, J. D. (1968) Fatal enterocolitis in hamsters given lincomycin hydrochloride. *Laboratory of Animal Care,* 18**,** 411 - 420.

Snell, H., Ramos, M., Longo, S., John, M. & Hussain, Z. (2004) Performance of the TechLab C. DIFF CHEK-60 enzyme immunoassay (EIA) in combination with the C. *difficile* Tox A/B II EIA kit, the Triage *C. difficile* panel immunoassay, and a cytotoxin assay for diagnosis of *Clostridium difficile*-associated diarrhea. American Society of Microbiology.

Soehn, F., Wagenknecht-Wiesner, A., Leukel, P., Kohl, M., Weidmann, M., VonEichel-Streiber, C. & Braun, V. (1998) Genetic rearrangements in the pathogenicity locus of *Clostridium difficile* strain 8864-implications for transcription, expression and enzymatic activity of toxins A and B. *Molecular Genetics and Genomics,* 258**,** 222-232.

Sorek, R., Kunin, V. & Hugenholtz, P. (2008) CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews*.

Sorg, J. A. & Sonenshein, A. L. (2008a) Bile salts and glycine as cogerminants for *Clostridium difficile* spores? *Journal of Bacteriology,* 190**,** 2505-2512.

Sorg, J. A. & Sonenshein, A. L. (2008b) Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination. *Journal of Bacteriology*.

Spigaglia, P. & Mastrantonio, P. (2002) Molecular analysis of the pathogenicity locus and polymorphism in the putative negative regulator of toxin production (TcdC) among *Clostridium difficile* clinical isolates. American Society of Microbiology.

Stabler, R. A., Gerding, D. N., Songer, J. G., Drudy, D., Brazier, J. S., Trinh, H. T., Witney, A. A., Hinds, J. & Wren, B. W. (2006) Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. *Journal of Bacteriology,* 188**,** 7297-7305.

Stabler, R. A., Dawson, L. F., Phua, L. T. H. & Wren, B. W. (2008) Comparative analysis of BI/NAP1/027 hypervirulent strains reveals novel toxin B-encoding gene (tcdB) sequences. *Journal of Medical Microbiology,* 57**,** 771–775.

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T. D., Sebaihia, M., Quail, M. A. & Rose, G. (2009) Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol,* 10**,** R102.

Staneck, J. L., Weckbach, L. S., Allen, S. D., Siders, J. A., Gilligan, P. H., Coppitt, G., Kraft, J. A. & Willis, D. H. (1996) Multicenter evaluation of four methods for *Clostridium difficile* detection: ImmunoCard C. *difficile*, cytotoxin assay, culture, and latex agglutination. *Journal of Clinical Microbiology,* 34**,** 2718-2721.

Stare, B. G., Delmee, M. & Rupnik, M. (2007) Variant forms of the binary toxin CDT locus and tcdC gene in *Clostridium difficile* strains. *Journal of Medical Microbiology,* 56**,** 329.

Stephen, A. M. & Cummings, J. H. (1980) The microbial contribution to human faecal mass. *Journal of Medical Microbiology,* 13**,** 45.

Stevenson, J. P. (1966) The normal bacterial flora of the alimentary canal of laboratory stocks of the desert locust, *Schistocerca gregaria Forskål*. *Journal of Invertebrate Pathology,* 8**,** 205-211.

Stubbs, S. L. J., Brazier, J. S., O'Neill, G. L. & Duerden, B. I. (1999) PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 different PCR ribotypes. *Journal of Clinical Microbiology,* 37**,** 461.

Swindells, J., Brenwald, N., Reading, N. & Oppenheim, B. (2010) An evaluation of diagnostic tests for *Clostridium difficile* infection. *Journal of Clinical Microbiology*.

Sullivan, N. M., Pellet, S. & Wilkins, T. D. (1982) Purification and characterization of Toxins A and B of *Clostridium difficile*. *Infection and Immunity,* 35**,** 1032-1040

Surawicz, C. M., Mcfarland, L. V., Elmer, G. & Chinn, J. (1989) Treatment of recurrent *Clostridium difficil*e colitis with vancomycin and *Saccharomyces boulardii*. *American Journal of Gastroenterology,* 84**,** 1285-7.

Taha, S., Johansson, O., Rivera Jonsson, S., Heimer, D. & Krovacek, K. (2007) Toxin production by and adhesive properties of *Clostridium difficile* isolated from humans and horses with antibiotic-associated diarrhea. *Comparative immunology, microbiology and infectious diseases,* 30**,** 163-174.

Tan, K. S., Wee, B. Y. & Song, K. P. (2001) Evidence for holin function of tcdE gene in the pathogenicity of *Clostridium difficile*. *Journal of Medical Microbiology,* 50**,** 613.

Tan, M. W. (2002) Cross-species infections and their analysis. *Annual Reviews in Microbiology,* 56**,** 539-565.

Tasteyre, A., Barc, M. C., Karjalainen, T., Dodson, P., Hyde, S., Bourlioux, P. & Borriello, P. (2000) A *Clostridium difficile* gene encoding flagellin. *Microbiology,* 146**,** 957.

Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H. & Karjalainen N, T. (2001) Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infection and Immunity,* 69**,** 7937.

Taylor, N. S., Thorne, G. M. & Bartlett, J. G. (1981) Comparison of two toxins produced by *Clostridium difficile*. *Infection and Immunity,* 34**,** 1036-1043.

Tedesco, F. J., Barton, R. W. & Alpers, D. H. (1974) Clindamycin-associated colitis. A prospective study. *Annals of internal medicine,* 81**,** 429.

Tennant, S. M., Zhang, Y., Galen, J. E., Geddes, C. D. & Levine, M. M. (2011) Ultra-Fast and sensitive detection of non-typhoidal *Salmonella* using Microwave-Accelerated Metal-Enhanced Fluorescence (MAMEF). *PLoS ONE,* 6**,** e18700.

Tenover, F. C., Novak-Weekley, S., Woods, C. W., Peterson, L. R., Davis, T., Schreckenberger, P., Fang, F. C., Dascal, A., Gerding, D. N. & Nomura, J. H. (2010) Impact of strain type on detection of toxigenic Clostridium difficile: comparison of molecular diagnostic and enzyme immunoassay approaches. *Journal of Clinical Microbiology,* 48**,** 3719-3724.

Tenover, F. C., Baron, E. J., Peterson, L. R. & Persing, D. H. (2011) Laboratory diagnosis of *Clostridium difficile* infection:: Can molecular amplification methods move us out of uncertainty? *The Journal of Molecular Diagnostics*.

Thomas, C. M. & Nielsen, K. M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology,* 3**,** 711-721.

Ticehurst, J. R., Aird, D. Z., Dam, L. M., Bork, A. P., Hargrove, J. T. & Carroll, K. C. (2006) Effective detection of toxigenic *Clostridium difficile* by a two-step algorithm including tests for antigen and cytotoxin. Am Soc Microbiol.

Turgeon, D. K., Novicki, T. J., Quick, J., Carlson, L. D., Millers, P., Ulness, B., CentA., Ashley, R., Larson, A. & Coyle, M. (2003) Six rapid tests for direct detection of *Clostridium difficile* and its toxins in fecal samples compared with the fibroblast cytotoxicity assay. *Journal of Clinical Microbiology,* 41**,** 667-670.

Turnbull, P. C. B. (1999) Definitive identification of *Bacillus anthracis*—a review. *Journal of Applied Microbiology,* 87**,** 237-240.

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Raplh, S., Rombauts, S. & Salamov, A. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science,* 313**,** 1596.

Underwood, S., Guan, S., Vijayasubhash, V., Baines, S. D., Graham, L., Lewis, R. J., Wilcox, M. H. & Stephenson, K. (2009) Characterization of the Sporulation Initiation Pathway of *Clostridium difficile* and Its Role in Toxin Production. *Journal of Bacteriology,* 191**,** 7296.

Vaid, A. & Bishop, A. H. (1998) The destruction by microwave radiation of bacterial endospores and amplification of the released DNA. *Journal of Applied Microbiology,* 85**,** 115-122.

Vallet-Gely, I., Lemaitre, B. & Boccard, F. (2008) Bacterial strategies to overcome insect defences. *Nature Reviews Microbiology,* 6**,** 302-313.

Van den Berg, R. J., Kuijper, E. J., Coppenraet, L. & Claas, E. C. J. (2006) Rapid diagnosis of toxinogenic *Clostridium difficile* in faecal samples with internally controlled real-time PCR. *Clinical Microbiology & Infection,* 12**,** 184-186.

Van den Berg, R. J., Vaessen, N., Endtz, H. P., Schulin, T., Van der Vorm, E. R. & Kuijper, E. J. (2007) Evaluation of real-time PCR and conventional diagnostic methods for the detection of *Clostridium difficile*-associated diarrhoea in a prospective multicentre study. *Journal of Medical Microbiology,* 56**,** 36.

Van Der Hoeven, R., Betrabet, G. & Forst, S. (2008) Characterization of the gut bacterial community in *Manduca sexta* and effect of antibiotics on bacterial diversity and nematode reproduction. *FEMS Microbiology Letters,* 286**,** 249-256.

Vlisidou, I., Dowling, A. J., Evans, I. R., Waterfield, N. & Wood, W. (2009) Drosophila embryos as model systems for monitoring bacterial infection in real time. *PLoS pathogens,* 5**,** e1000518.

Von Eichel-Streiber, C. V. & Sauerborn, M. (1990) Nucleotide sequence of *Clostridium difficile* toxin A. *Nucleic Acids Research,* 18**,** 1629-1630.

Von Eichel-Streiber, C., Laufenberg-Feldmann, R., Sartingen, S., Schulze, J. & Sauerborn, M. (1992) Comparative sequence analysis of the *Clostridium difficile* toxins A and B. *Molecular Genetics and Genomics,* 233**,** 260-268.

Von Eichel-Streiber, C., Boquet, P., Sauerborn, M. & Thelestam, M. (1996) Large Clostridial cytotoxins—a family of glycosyltransferases modifying small GTP-binding proteins. *Trends in Microbiology,* 4**,** 375-382.

Voth, D. E. & Ballard, J. D. (2005) *Clostridium difficile* Toxins: Mechanism of action and role in disease. *Clinical Microbiology Reviews,* 18**,** 247-263

Waligora, A.-J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A. & Karjalainen, T. (2001) Characterization of a cell surface protein of *Clostridium difficil*e with adhesive properties. *Infect. Immun.,* 69**,** 2144-2153.

Walsh, P. S., Metzger, D. A. & Higuchi, R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques,* 10.

Wasserman, W. W. & Sandelin, A. (2004) Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics,* 5**,** 276-287

Waterfield, N. R., Wren, B. W. & Ffrench-Constant, R. H. (2004) Invertebrates as a source of emerging human pathogens. *Nature Reviews Microbiology,* 2**,** 833-841

Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench" *Bioinformatics*

Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C. & Mattick, J. S. (2002) Extracellular DNA required for bacterial biofilm formation. *Science,* 295**,** 1487.

Wauters, G., Janssens, M., Steigerwalt, A. G. & Brenner, D. (1988) *Yersinia mollaretii* sp. nov. and *Yersinia bercovieri* sp. nov., formerly called *Yersinia enterocolitica* biogroups 3A and 3B. *International Journal of Systematic and Evolutionary Microbiology,* 38**,** 424.

Wilcox, M. H., Fawley, W. N. & Parnell, P. (2000) Value of lysozyme agar incorporation and alkaline thioglycollate exposure for the environmental recovery of *Clostridium difficile*. *Journal of Hospital Infection,* 44**,** 65-69.

Wilcox, M. H. (2007) Evidence for low risk of *Clostridium difficile* infection associated with tigecycline. *Clinical Microbiology & Infection,* 13**,** 949.

Wilcox, M. H. & Eastwood, K. A. (2009) Evaluation Report: *Clostridium difficile* detection assays. IN NHS (Ed.)

Wilkins, T. D. (2000) *Clostridium difficile,* Springer Verlag.

Wilkinson, P., Waterfield, N., Crossman, L., Corton, C., Sanchez-Contreras, M., Vlisidou, I., Barron, A., Bignell, A., Clark, L. & Ormond, D. (2009) Comparative genomics of the emerging human pathogen *Photorhabdus asymbiotica* with the insect pathogen *Photorhabdus luminescens*. *BMC genomics,* 10**,** 302.

Williams, O. M. & Spencer, R. C. (2009) The management of *Clostridium difficile* infection. *British Medical Bulletin*.

Williams, G. J., Denyer, S. P., Hosein, I. K., Hill, D. W. & Maillard, J. Y. (2009) Limitations of the efficacy of surface disinfection in the healthcare setting. *Infection Control and Hospital Epidemiology,* 30**,** 570-573.

Willis, A. T. (1969) *Clostridia of wound infection*, Butterworths, London.

Wilson, K. H., Kennedy, M. J. & Fekety, F. R. (1982) Use of sodium taurocholate to enhance spore recovery on a medium selective for *Clostridium difficile*. *Journal of Clinical Microbiology,* 15**,** 443-446.

Wilson, K. H. (1983) Efficiency of various bile salt preparations for stimulation of *Clostridium difficil*e spore germination. *Journal of Clinical Microbiology,* 18**,** 1017-1019.

Wood, T. M., Wilson, C. A. & Stewart, C. S. (1982) Preparation of the cellulase from the cellulolytic anaerobic rumen bacterium *Ruminococcus albus* and its release from the bacterial cell wall. *Biochemical Journal,* 205**,** 129.

Woodford, N. & Johnson, A. P. (2004) *Genomics, Proteomics and Clinical Bacteriology: Methods and Reviews*, Humana Press.

Wren, B. W., Clayton, C. L., Castledine, N. B. & Tabaqchali, S. (1990) Identification of toxigenic Clostridium difficile strains by using a toxin A gene-specific probe. *Journal of Clinical Microbiology,* 28**,** 1808.

Wren, B. W., Heard, S. R., Al-Saleh, A. I. & Tabaqchali, S. (1993) Characterisation of *Clostridium difficile* strains by polymerase chain reaction with toxin A-and B-specific primers. *Journal of Medical Microbiology,* 38**,** 109.

Wren, M. W., Sivapalan, M., Kinson, R. & Shetty, N. R. (2009) Laboratory diagnosis of *Clostridium difficile* infection: An evaluation of tests for faecal toxin, glutamate dehydrogenase, lactoferrin and toxigenic culture in the diagnostic laboratory. *British journal of biomedical science,* 66**,** 1

Wyatt, G. R. & Meyer, W. L. (1959) The chemistry of insect haemolymph. *The Journal of General Physiology,* 42**,** 1005.

Yamakawa, K., Karasawa, T., Ikoma, S. & Nakamura, S. (1996) Enhancement of *Clostridium difficile* toxin production in biotin-limited conditions. Soc General Microbiol.

Yamakawa, K., Kamiya, S., Meng, X. Q., Karasawa, T. & Nakamura, S. (1994) Toxin production by *Clostridium difficile* in a defined medium with limited amino acids. Soc General Microbiol.

Yang, G., Dowling, A. J. & Gerike, U. (2006) *Photorhabdus* virulence cassettes confer injectable insecticidal activity against the wax moth. *Journal of Bacteriology,* 188**,** 2254.

Zaisz, N. H., Rupnik, M., Kuijper, E., Harmanus, C., Michielsen, D., Janssens, K. & Nubel, U. (2009) Typing *Clostridium difficile* strains based on tandem repeat sequences. *BMC Microbiology,* 9**,**

Zhang, Y., Agreda, P., Kelley, S., Gaydos, C. & Geddes, C. D. (2011) Development of a microwave-accelerated metal-enhanced fluorescence 40 second,< 100 cfu/mL point of care assay for the detection of *Chlamydia trachomatis*. *Biomedical Engineering, IEEE Transactions on***,** 1-1.

Zheng, L., Keller, S. F., Lyerly, D. M., Carman, R. J., Genheimer, C. W., Gleaves, C. A., Kohlhepp, S. J., Young, S., Perez, S. & Ye, K. (2004) Multicenter evaluation of a new screening test that detects *Clostridium difficil*e in fecal specimens. *Journal of Clinical Microbiology,* 42**,** 3837-3840

Zvelebil, M. & Baum, J. O. (2008) *Understanding Bioinformatics*, Taylor & Francis Group.

CHAPTER 8

APPENDIX

8.1. Bacterial Staining

8.1.1. Gram Staining

Colonies of *C. difficile* were spread onto a microscope slide using a sterile loop and emulsified using water. Slides were fixed under Bunsen flame and subsequently stained with crystal violet (Oxoid Ltd, Basingstoke, UK) for sixty seconds, washed with water and then flooded with Gram's iodine (Oxoid Ltd, Basingstoke, UK) for sixty seconds. These were then washed with water and flooded with 95% ethanol (Fisher Scientific, UK) for ten seconds, washed with water and counterstained with safranin (Oxoid Ltd, Basingstoke, UK) for thirty seconds, rinsed, blotted and allowed to air dry. The samples were viewed under light microscope under oil immersion at 1000x TM

8.1.2. Spore Staining

Spores from *C. difficile* were spread onto a glass microscope slide using a sterile cool loop. Once this was air dried the slide was fixed by passing it briefly through a Bunsen flame two or three times without exposing the dried film directly to the flame. The slides were placed over a screened boiling water bath and flooded with Malachite Green (Oxoid Ltd, Basingstoke, UK). Slides were then washed with water and counterstained with Safranin (Sigma Aldrich, UK) for 30 seconds and washed with water (Oxoid Ltd, Basingstoke, UK). The samples were viewed under light microscope under oil immersion at 1000x TM. Spores strained green and vegetative cells red.

8.2. Statistics

To enable interpretation of data statistical analyses were conducted. Statistical tests used in this thesis are described.

8.2.1. One way ANOVA:

One way analysis of variance (ANOVA) tests are performed on groups of variables and each group should represent separate responses to, or effects of, one factor. Thus the null hypothesis should be that the groups have the same mean (Bowker & Randerson, 2006).

8.2.2. Kruskall-Wallis test

This is a non- parametric test equivalent to one way ANOVA. The tests each group to determine if there is a different median. This is used as a final test post one way ANOVA if the assumptions of ANOVA are violated (Bowker & Randerson, 2006)

8.2.3. T- Test

A two sample t-test is used to compare the mean value of two independent random samples of normalised data (Bowker & Randerson, 2006). The non-parametric equivalent of a t test is the Wilcoxon's rank test to discern if the difference between two median values is not significantly different from zero (Bowker & Randerson, 2006).

8.2.4. P- values

The P value is a measure of probability of the likeliness of a hypothesis being true or the significance of certain data. The P value statistic ranges between 0 - 1. Thus if P = 0 then the biological event is impossible and if P = 1 then it is certain (Bowker & Randerson, 2006).

8.2.5. Multiple comparison tests

The Dunnett's multiple comparison test is used to compare samples with the control group (GraphPad Software, La Jolla California USA, www.graphpad.com). A Tukey-Kramer test is also used test if significant differences between means of pairs of data are unexpected after ANOVA is performed.

8.3. Solutions for Southern Hybridisation

8.3.1. Dig Easy Hyb Buffer

For the hybridisation of the DIG probe to the membrane, DIG Easy Hyb buffer (Roche Diagnostics, Charles Avenue, West Sussex, UK) was obtained from (Roche Diagnostics, Charles Avenue, West Sussex, UK). It is specially designed for use with DIG detection methods.

8.3.2. Maleic Acid Buffer

Maleic acid buffer is used as the basic buffer for mixing further solutions needed for dot blots. Maleic acid buffer was prepared by adding 0.1 M maleic acid (Sigma Aldrich, UK) to 0.15 M sodium chloride - NaCl (Sigma Aldrich, UK) with polished H_2O in a 1000 ml Duran bottle. The pH of the solution was calibrated to pH= 7.5 by adding Sodium hydroxide pellets (Sigma Aldrich, UK) (NaOH). The solution was mixed thoroughly and autoclaved at 121 ^oC, and left to cool before use.

8.3.3. Stock solution of 20x Saline-Sodium citrate buffer (SSC)

To wash the nylon membrane after hybridisation a stock solution of SSC buffer was made. SSC buffer at 20x concentration was prepared by adding 0.3 M tri-sodium citrate (Sigma Aldrich, UK) at pH =7 to 3 M NaCl (Sigma Aldrich, UK) with polished H_2O in a 1000 ml Duran bottle. The solution was mixed thoroughly before use.

8.3.4. SSC X 2

This low stringency wash solution was used to remove unspecifically bound probes from the membrane via washing. This has a high salt content and is washed at a low temperature (Roche Diagnostics, Charles Avenue, West Sussex, UK). This thus allows for a more efficient dot to be observed. SSC buffer at 2x concentration was prepared by adding 50 ml 20x SSC and 5 ml 10% SDS (Sodium Dodecyl sulphate, Sigma Aldrich, UK) to 500 ml polished H_2O. The solution was mixed thoroughly before use.

8.3.5. SSC X 0.5

This high stringency wash solution is used to remove any undesired binding with low homology. This has a low salt content and is washed at a high temperature (Roche Diagnostics, Charles Avenue, West Sussex, UK). SSC buffer at 0.5x concentration was prepared by adding 12.5 ml 20x SSC and 5 ml 10% SDS (Sodium Dodecyl Sulphate,

Sigma Aldrich, UK) to 500 ml polished H_2O. The solution was mixed thoroughly before use.

8.3.6. SSC X 0.1

This high stringency wash solution is also used to further remove any undesired binding with low homology in conjunction with 0.5x SSC buffer. This has a low salt content and is washed at a high temperature (Roche Diagnostics, Charles Avenue, West Sussex, UK). SSC buffer at 0.1x concentration was prepared by adding 2.5 ml 20x SSC and 5 ml 10% SDS (Sodium Dodecyl sulphate, Sigma Aldrich, UK) to 500 ml polished H_2O. The solution was mixed thoroughly before use.

8.3.7. Washing buffer

Washing buffer is used to assist in detection of any probe bound to the DNA and was prepared by adding 0.3% Tween® 20 (Sigma Aldrich, UK) to 1 L pre-prepared maleic acid buffer in a 1000 ml Duran bottle. The solution was mixed thoroughly before use.

8.3.8. Stock 10% Blocking solution

The blocking solution is used to minimise the background on the membrane during detection of the bound probe. Blocking solution was obtained from Roche (Roche Diagnostics, Charles Avenue, West Sussex, UK) and mixed in maleic acid buffer to a concentration of 10% (w/v) in a 500 ml Duran bottle. The powder was dissolved by microwaving 5x for 30 seconds each with 10 minutes mixing in between. The final stock solution was autoclaved at 121 ^{o}C, and left to cool. Aliquots of the stock solution were frozen at -20^{o}C until needed. This solution has to be used immediately after preparation to avoid spoiling.

8.3.9. 1% Blocking solution

The 1% blocking solution was made from 10% blocking solution by mixing into maleic acid buffer. The final solution was autoclaved at 121 ^{o}C, left to cool and used immediately.

8.3.10. Detection Buffer

To detect the bound probe on the membrane detection buffer was prepared by adding 100 mM Tris-HCl (pH 9.5) (Sigma Aldrich, UK) to 100 mM NaCl (Sigma Aldrich, UK) in a 500 ml Duran bottle. The solution was autoclaved at 121 ^{o}C and left to cool.

8.4. Bioinformatic analysis of known *C. difficile* virulence genes using BLAST

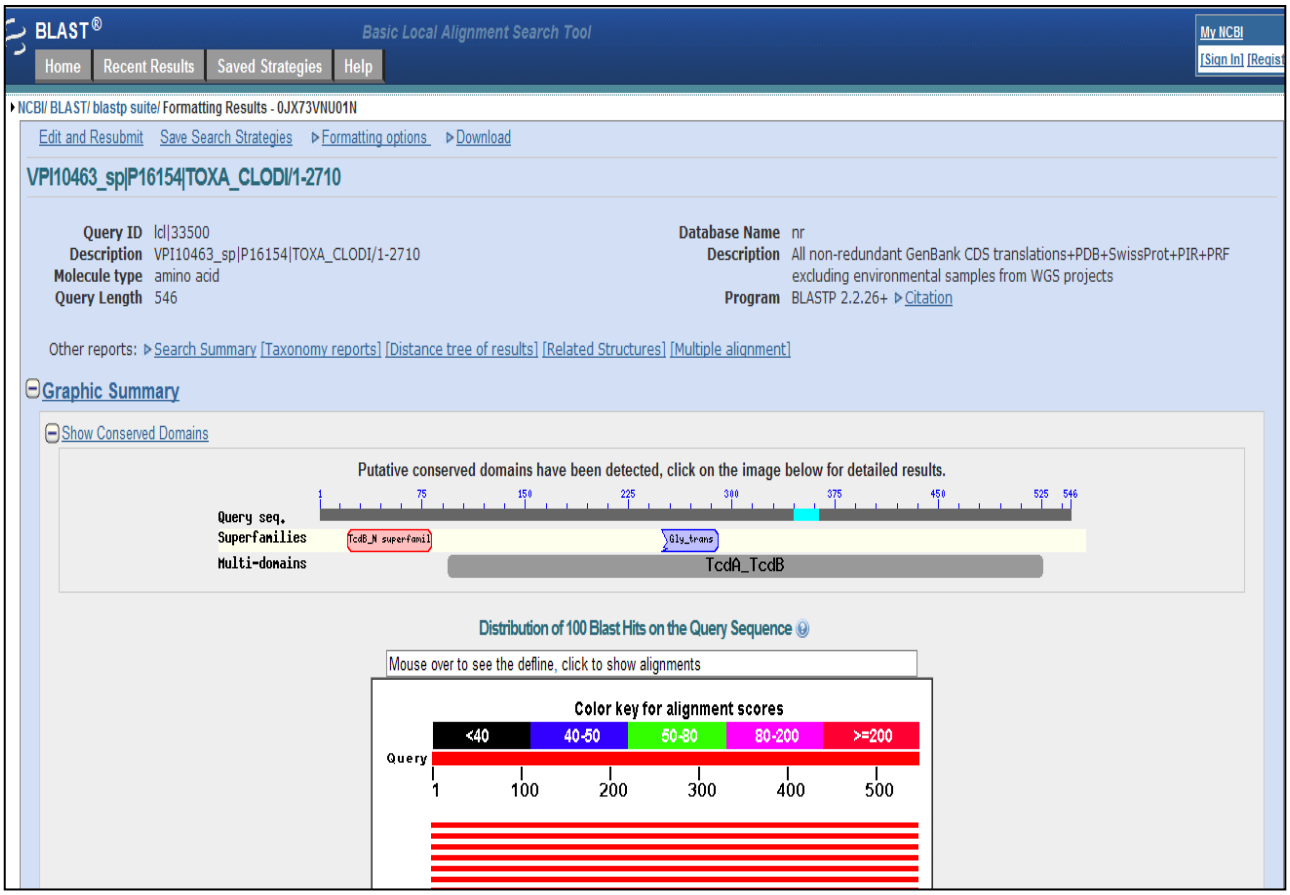
8.4.1. Toxin A protein BLAST alignments

[http://blast.ncbi.nlm.nih.gov/Blast.cgi]

8.4.1.1. Glucosyltransferase domain:

Analysis of the glucosyltransferase domain of *C. difficile* toxin A using protein BLAST revealed the following alignment:

Query length= 546 amino acids.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	1097	1097	100%	0.0	100%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| Tc	1097	1097	100%	0.0	100%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	1097	1097	100%	0.0	100%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	1096	1096	100%	0.0	100%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	1095	1095	100%	0.0	100%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	1093	1093	100%	0.0	99%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	1093	1093	100%	0.0	99%
AFN52237.1	TcdA [[Clostridium] difficile]	1092	1092	100%	0.0	99%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	1092	1092	100%	0.0	99%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	1091	1091	100%	0.0	99%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	1091	1091	100%	0.0	99%
AAC08437.1	truncated toxin A [[Clostridium] difficile]	1082	1082	100%	0.0	98%
CAC19893.1	truncated toxin A [[Clostridium] difficile]	1077	1077	100%	0.0	97%
CAC03681.1	toxin A [[Clostridium] difficile]	1073	1073	100%	0.0	98%
CAC79962.1	toxin B [[Clostridium] difficile]	539	539	100%	0.0	49%
CAC41640.1	TcdB-C34 Cluster1-2 [[Clostridium] difficile]	536	536	100%	0.0	49%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	563	563	100%	2e-177	51%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	561	561	100%	9e-177	52%
YP_001087135.1	tcdB gene product [Clostridium difficile 630] >ref|ZP_05349824.1|	560	560	100%	2e-176	51%
ZP_05328744.1	toxin B [Clostridium difficile QCD-63q42]	560	560	100%	2e-176	51%
EHJ40398.1	cell wall-binding repeat protein [Clostridium difficile 70-100-2010]	560	560	100%	2e-176	51%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	555	555	100%	9e-175	50%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	555	555	100%	1e-174	50%
EHJ31817.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	541	541	100%	9e-170	49%
CAC19891.1	toxin B [[Clostridium] difficile]	541	541	100%	1e-169	49%
AAG18011.1	cytotoxin B [[Clostridium] difficile]	539	539	100%	4e-169	49%
CAA80815.1	toxin B [[Clostridium] difficile]	538	538	100%	1e-168	49%
BAF46125.1	TpeL [Clostridium perfringens]	469	469	100%	6e-146	45%
ACF49258.1	TpeL [Clostridium perfringens A]	468	468	100%	6e-145	46%

ACF49258.1	TpeL [Clostridium perfringens A]	468	468	100%	6e-145	46%
2VKD_A	Chain A, Crystal Structure Of The Catalytic Domain Of Lethal Toxin	566	566	99%	0.0	53%
CAA71690.1	TcdA protein [[Clostridium] difficile]	1072	1072	99%	0.0	97%
ZP_02636804.1	TpeL [Clostridium perfringens B str. ATCC 3626] >gb|EDT23010.1|	465	465	99%	7e-144	45%
ZP_02865634.1	TcpA [Clostridium perfringens C str. JGS1495] >gb|EDS79391.1| T	464	464	99%	1e-143	45%
3SRZ_A	Chain A, Clostridium Difficile Toxin A (Tcda) Glucolsyltransferase D	1093	1093	99%	0.0	100%
2BVL_A	Chain A, Crystal Structure Of The Catalytic Domain Of Toxin B Fro	555	555	99%	0.0	50%
4DMV_A	Chain A, Crystal Structure Of The Gt Domain Of Clostridium Difficile	1092	1092	99%	0.0	100%
2BVM_A	Chain A, Crystal Structure Of The Catalytic Domain Of Toxin B Fro	553	553	99%	0.0	50%
CAA88565.1	alpha-toxin [Clostridium novyi]	252	252	99%	9e-69	33%
2VK9_A	Chain A, Crystal Structure Of The Catalytic Domain Of Alpha-Toxin	253	253	98%	2e-73	33%
ADH94633.1	TcdB [[Clostridium] difficile]	556	556	96%	3e-175	52%
ADH94634.1	TcdB [[Clostridium] difficile]	554	554	96%	3e-174	52%
ADH94636.1	TcdB [[Clostridium] difficile]	553	553	96%	3e-174	52%
ADH94635.1	TcdB [[Clostridium] difficile]	553	553	96%	4e-174	52%
ADH94624.1	TcdB [[Clostridium] difficile]	553	553	96%	4e-174	52%
ADH94629.1	TcdB [[Clostridium] difficile]	550	550	96%	6e-173	51%
ADH94632.1	TcdB [[Clostridium] difficile]	545	545	96%	2e-171	51%
ADH94627.1	TcdB [[Clostridium] difficile]	543	543	96%	2e-170	51%
ADH94631.1	TcdB [[Clostridium] difficile]	543	543	96%	2e-170	51%
ADH94630.1	TcdB [[Clostridium] difficile]	535	535	96%	2e-167	50%
ADH94626.1	TcdB [[Clostridium] difficile]	534	534	96%	4e-167	50%
ADH94623.1	TcdB [[Clostridium] difficile]	533	533	96%	1e-166	50%
ADH94628.1	TcdB [[Clostridium] difficile]	532	532	96%	1e-166	50%
ADH94625.1	TcdB [[Clostridium] difficile]	531	531	96%	2e-166	50%
ZP_05120822.1	Glycosyltransferase sugar-binding region containing DXD motif fam	58.2	58.2	94%	4e-06	22%
EGY29542.1	Glycosyltransferase [Candidatus Regiella insecticola R5.15]	96.3	96.3	86%	6e-18	24%

ADH94628.1	TcdB [[Clostridium] difficile]	532	532	96%	1e-166	50%
ADH94625.1	TcdB [[Clostridium] difficile]	531	531	96%	2e-166	50%
ZP_05120822.1	Glycosyltransferase sugar-binding region containing DXD motif fam	58.2	58.2	94%	4e-06	22%
EGY29542.1	Glycosyltransferase [Candidatus Regiella insecticola R5.15]	96.3	96.3	86%	6e-18	24%
EGY28465.1	Peptidase C80 protein [Candidatus Regiella insecticola R5.15]	100	100	86%	5e-19	25%
YP_346341.1	hypothetical protein [Pseudomonas fluorescens Pf0-1] >gb|ABA72	93.6	93.6	82%	6e-17	23%
EIK69977.1	putative toxin protein [Pseudomonas fluorescens Q8r1-96]	117	117	81%	1e-24	25%
YP_004351772.1	hypothetical protein PSEBR_a620 [Pseudomonas brassicacearum su	116	116	81%	3e-24	24%
YP_003366177.1	lifA gene product [Citrobacter rodentium ICC168] >gb|AAW52502.	73.6	73.6	81%	1e-10	24%
YP_005206075.1	cytotoxin mcf [Pseudomonas fluorescens F113] >gb|AEV60680.1|	116	116	81%	3e-24	25%
YP_004498542.1	glycosyltransferase family protein [Serratia sp. AS12] >ref|YP_004	84.7	84.7	76%	1e-14	26%
EIO58881.1	tcdA/TcdB catalytic glycosyltransferase domain protein [Escherich	62.8	62.8	73%	2e-07	23%
EHV09211.1	glycosyltransferase sugar-binding region containing DXD motif fami	62.8	62.8	73%	2e-07	24%
ZP_03263719.1	toxin B [Escherichia coli O157:H7 str. EC4042] >gb|EDZ84682.1| t	62.0	62.0	73%	4e-07	23%
ZP_02815165.1	toxin B [Escherichia coli O157:H7 str. EC869] >ref|ZP_05950388.1	62.0	62.0	73%	4e-07	23%
EFW63331.1	toxin B [Escherichia coli O157:H7 str. EC1212] >gb|EIP62080.1| tc	62.0	62.0	73%	4e-07	23%
ZP_02802532.1	toxin B [Escherichia coli O157:H7 str. EC4196] >gb|EDU31090.1| t	62.0	62.0	73%	4e-07	23%
YP_325655.1	putative cytotoxin [Escherichia coli O157:H7 EDL933] >ref|ZP_027	62.0	62.0	73%	4e-07	23%
NP_052665.1	toxin B [Escherichia coli O157:H7 str. Sakai] >dbj|BAA31815.1| To	61.6	61.6	73%	5e-07	23%
ZP_03085363.1	toxin B [Escherichia coli O157:H7 str. EC4024]	61.6	61.6	73%	5e-07	23%
YP_346342.1	hypothetical protein [Pseudomonas fluorescens Pf0-1] >gb|ABA72	120	120	71%	2e-25	26%
ZP_04642320.1	RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC 4396	97.1	97.1	70%	4e-18	27%
ZP_04629276.1	hypothetical protein yberc0001_36060 [Yersinia bercovieri ATCC 4	69.3	69.3	70%	1e-09	23%
ZP_06195543.1	adherence factor [Chlamydia muridarum Weiss]	72.4	72.4	69%	2e-10	23%
NP_296815.1	adherence factor [Chlamydia muridarum Nigg] >ref|ZP_06194616.1	72.4	72.4	69%	2e-10	23%
NP_296816.1	adherence factor [Chlamydia muridarum Nigg] >gb|AAF39293.1| ac	69.7	69.7	62%	2e-09	25%
ZP_06194619.1	adherence factor [Chlamydia muridarum Nigg] >ref|ZP_06195546.1	69.7	69.7	62%	2e-09	25%
EHW03958.1	glycosyltransferase sugar-binding region containing DXD motif fami	54.3	95.1	55%	6e-05	28%
ZP_06053303.1	toxin B [Grimontia hollisae CIP 101886] >gb|EEY72053.1| toxin B [(	60.8	60.8	50%	8e-07	24%
EHJ31823.1	hypothetical protein HMPREF1123_00959 [Clostridium difficile 050-I	523	523	47%	0.0	99%
EGY28951.1	Cytotoxin L/B [Candidatus Regiella insecticola R5.15]	68.9	68.9	39%	6e-10	29%
EHV29059.1	glycosyltransferase sugar-binding region containing DXD motif fami	57.0	57.0	39%	8e-06	27%
EFU97591.1	glycosyltransferase sugar-binding region containing DXD motif fami	57.0	57.0	39%	9e-06	27%

ZP_04629276.1	hypothetical protein yberc0001_36060 [Yersinia bercovieri ATCC 4	69.3	69.3	70%	1e-09	23%
ZP_06195543.1	adherence factor [Chlamydia muridarum Weiss]	72.4	72.4	69%	2e-10	23%
NP_296815.1	adherence factor [Chlamydia muridarum Nigg] >ref|ZP_06194616.1	72.4	72.4	69%	2e-10	23%
NP_296816.1	adherence factor [Chlamydia muridarum Nigg] >gb|AAF39293.1| ac	69.7	69.7	62%	2e-09	25%
ZP_06194619.1	adherence factor [Chlamydia muridarum Nigg] >ref|ZP_06195546.1	69.7	69.7	62%	2e-09	25%
EHW03958.1	glycosyltransferase sugar-binding region containing DXD motif fami	54.3	95.1	55%	6e-05	28%
ZP_06053303.1	toxin B [Grimontia hollisae CIP 101886] >gb|EEY72053.1| toxin B [(	60.8	60.8	50%	8e-07	24%
EHJ31823.1	hypothetical protein HMPREF1123_00959 [Clostridium difficile 050-I	523	523	47%	0.0	99%
EGY28951.1	Cytotoxin L/B [Candidatus Regiella insecticola R5.15]	68.9	68.9	39%	6e-10	29%
EHV29059.1	glycosyltransferase sugar-binding region containing DXD motif fami	57.0	57.0	39%	8e-06	27%
EFU97591.1	glycosyltransferase sugar-binding region containing DXD motif fami	57.0	57.0	39%	9e-06	27%
EHU73379.1	glycosyltransferase sugar-binding region containing DXD motif fami	57.0	57.0	39%	9e-06	27%
EJE84889.1	toxin B, partial [Escherichia coli O26:H11 str. CVM10021]	56.6	56.6	39%	1e-05	30%
EHW81443.1	glycosyltransferase sugar-binding region containing DXD motif fami	56.2	56.2	39%	2e-05	27%
EII08397.1	PF11996 family protein [Escherichia coli 5.0959]	55.1	55.1	39%	6e-05	30%
YP_002756600.1	toxin B [Escherichia coli] >gb|ACL51990.1| toxin B [Escherichia co	55.1	55.1	39%	6e-05	30%
YP_003377848.1	toxin B [Escherichia coli O26:H-] >gb|ADB20445.1| toxin B [Esche	55.1	55.1	39%	6e-05	30%
YP_003237860.1	putative adherence factor, Efa1 homolog [Escherichia coli O26:H1	55.1	55.1	39%	6e-05	30%
YP_128484.1	hypothetical protein PBPRA0243 [Photobacterium profundum SS9]	62.8	62.8	39%	1e-07	29%
CAA80819.1	alpha-toxin [Clostridium novyi]	128	128	39%	9e-31	38%
ZP_07395770.1	glycosyltransferase sugar-binding domain-containing hypothetical	65.1	65.1	37%	2e-08	29%
1807295A	toxin B	204	204	35%	5e-59	50%
CAA43302.1	toxin A [[Clostridium] difficile]	307	307	28%	5e-99	100%
EHJ31822.1	hypothetical protein HMPREF1123_00958 [Clostridium difficile 050-I	218	218	19%	1e-65	98%
1814453C	toxin A	177	177	16%	2e-50	100%
EHJ31821.1	hypothetical protein HMPREF1123_00957 [Clostridium difficile 050-I	169	169	15%	3e-47	98%
AAG18375.1	truncated toxin A [[Clostridium] difficile] >emb|CAB60776.1| TcdA	91.3	91.3	8%	2e-19	100%

Alignment information for genes shown in Table 4.2:

C. difficile tcdB

```
>emb|CAC79962.1|  toxin B [[Clostridium] difficile]
Length=560

 Score =  539 bits (1388),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 268/549 (49%), Positives = 391/549 (71%), Gaps = 4/549 (1%)

Query  1    MSLISKEELIKLA-YSIRPRENEYKTILTNLDEYNKLTTNNNENKYLQLKKLNESIDVFM  59
            MS +++++L K+A     R +E+EY   IL   L+EY+ ++ N      KYL+LK +N    D ++
Sbjct  1    MSSVNRKQLEKMANVRFRVQEDEYVAILDALEEYHNMSENTVVEKYLKLKDINSLTDTYI  60

Query  60   NKYKTSSRNRALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD  119
            + YK S RN+AL    K+ ++ E++ +KNSN +PVEKNLHF+WIGG+++D A+ YI QW D
Sbjct  61   DTYKKSGRNKALKKFKEYLVTEILELKNSNLTPVEKNLHFIWIGGQINDTAINYINQWKD  120

Query  120  INAEYNIKLWYDSEAFLVNTLKKAIVESSTTEALQLLEEEIQNPQFDNMKFYKKRMEFIY  179
            +N++YN+ ++YDS AFL+NTLKK I+ES++ + L+    E + +P+F++   F++KRM+ IY
Sbjct  121  VNSDYNVNVFYDSNAFLINTLKKTIIESASNDTLESFRENLNDPEFNHTAFFRKRMQIIY  180

Query  180  DRQKRFINYYKSQINKPTVPTIDDIIKSHLVSEYNRDETVLESYRTNSLRKINSNHGIDI  239
            D+Q+ FINYYK+Q   +      IDDI+K++L +EY++D    L +Y    SL K+  N G D+
Sbjct  181  DKQQNFINYYKAQKEENPDLIIDDIVKTYLSNEYSKDIDELNAYIEESLNKVTENSGNDV  240

Query  240  RANSLFTEQELLNIYSQELLNRGNLAAASDIVRLLALKNFGGVYLDVDMLPGIHSDLFKT  299
            R     F    E+ N+Y QEL+ R NLA ASDI+R+   LKN GGVYLDVDMLPGIH DLFK
Sbjct  241  RNFEEFKTGEVFNLYEQELVERWNLAGASDILRVAILKNIGGVYLDVDMLPGIHPDLFKD  300

Query  300  ISRPSSI--GLDRWEMIKLEAIMKYKKYINNYTSENFDKLDQQLKDNFKLIIESKSEKSE  357
            I++P S+    +D WE ++LEAIMKYK+YI  YTS++FD LD++++ +F+ ++ SKS+KSE
Sbjct  301  INKPDSVKTAVD-WEEMQLEAIMKYKEYIPEYTSKHFDTLDEEVQSSFESVLASKSDKSE  359

Query  358  IFSKLENLNVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLNPAI  417
            IF  L   + VS LE+K+AFA GS+I+QALIS + SY ++L+I+Q++NRY+ LN   L P I
Sbjct  360  IFLPLGGIEVSPLEVKVAFAKGSIIDQALISAKDSYCSDLLIKQIQNRYKILNDTLGPII  419

Query  418  ESDNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSGPGAYASAY  477
               N+F   T    F +SL    A  EN  F+ KI  YL+VGF PEA +TI+LSGP  YA AY
Sbjct  420  SQGNDFNTTMNNFGESLGAIANEENISFIAKIGSYLRVGFYPEANTTITLSGPTIYAGAY  479

Query  478  YDFINLQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQFEKYVRDY  537
             D +  +E +I+ ++ +S+L  F+FP+ N+SQ TEQE NSLW F++  AK QFE+Y ++Y
Sbjct  480  KDLLTFKEMSIDTSILSSELRNFEFPKVNISQATEQEKNSLWQFNEERAKIQFEEYKKNY  539

Query  538  TGGSLSEDN  546
              G+L ED+
Sbjct  540  FEGALGEDD  548
```

C. sordellii tcsL

```
>emb|CAA57959.1|  cytotoxin L [[Clostridium] sordellii]
Length=2364

 Score =  561 bits (1446),  Expect = 9e-177, Method: Compositional matrix adjust.
 Identities = 286/547 (52%), Positives = 395/547 (72%), Gaps = 1/547 (0%)

Query  1    MSLISKEELIKLAY-SIRPRENEYKTILTNLDEYNKLTTNNNENKYLQLKKLNESIDVFM  59
            M+L++K +L K+ Y    R +E+EY   IL   L+EY+ ++ ++    KYL+LK +N    D ++
Sbjct  1    MNLVNKAQLQKMVYVKFRIQEDEYVAILNALEEYHNMSESSVVEKYLKLKDINNLTDNYL  60

Query  60   NKYKTSSRNRALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD  119
            N YK S RN+AL    K+ +  EV+ +KN++ +PVEKNLHF+WIGG+++D A+ YI QW D
Sbjct  61   NTYKKSGRNKALKKFKEYLTMEVLELKNNSLTPVEKNLHFIWIGGQINDTAINYINQWKD  120

Query  120  INAEYNIKLWYDSEAFLVNTLKKAIVESSTTEALQLLEEEIQNPQFDNMKFYKKRMEFIY  179
            +N++Y +K++YDS AFL+NTLKK IVES+T    L+    E + +P+FD   KFY+KRME IY
Sbjct  121  VNSDYTVKVFYDSNAFLINTLKKTIVESATNNTLESFRENLNDPEFDYNKFYRKRMEIIY  180

Query  180  DRQKRFINYYKSQINKPTVPTIDDIIKSHLVSEYNRDETVLESYRTNSLRKINSNHGIDI  239
            D+QK FI+YYKSQI +      ID+IIK++L +EY++D    L   Y   SL KI +N+G DI
Sbjct  181  DKQKHFIDYYKSQIEENPEFIIDNIIKTYLSNEYSKDLEALNKYIEESLNKITANNGNDI  240

Query  240  RANSLFTEQELLNIYSQELLNRGNLAAASDIVRLLALKNFGGVYLDVDMLPGIHSDLFKT  299
            R     F +++L+ +Y+QEL+ R NLAAASDI+R+   LK  GGVYLDVD+LPGI  DLFK+
Sbjct  241  RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQPDLFKS  300

Query  300  ISRPSSIGLDRWEMIKLEAIMKYKKYINNYTSENFDKLDQQLKDNFKLIIESKSEKSEIF  359
            I++P SI     WEMIKLEAIMKYK+YI  YTS+NFD LD++++ +F+  + SKS+KSEIF
Sbjct  301  INKPDSITNISWEMIKLEAIMKYKEYIPGYTSKNFDMLDEEVQRSFESALSSKSDKSEIF  360

Query  360  SKLENLNVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLNPAIES  419
              L+++ VS LE+KIAFA  SVINQALIS + SY ++LVI Q+KNRY+ LN +LNP+I
Sbjct  361  LPLDDIKVSPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKNRYKILNDNLNPSINE  420

Query  420  DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSGPGAYASAYYD  479
              +F  T KIF D L + +  +N MF+ KI  YL+VGF P+ RSTI+LSGPG Y  AY D
Sbjct  421  GTDFNTTMKIFSDKLASISNEDNMMFMIKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD  480

Query  480  FINLQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQFEKYVRDYTG  539
            F   ++N+     L   +L   F+FP+   +SQLTEQEI SLWSF+QA AK QFE+Y + Y
Sbjct  481  LLMFKDNSTNIHLLEPELRNFEFPKTKISQLTEQEITSLWSFNQARAKSQFEEYKKGYFE  540

Query  540  GSLSEDN  546
            G+L ED+
Sbjct  541  GALGEDD  547
```

C. perfringens tpeL

```
>dbj|BAF46125.1|  TpeL [Clostridium perfringens]
Length=1651

 Score =  469 bits (1207),  Expect = 7e-146, Method: Compositional matrix adjust.
 Identities = 249/549 (45%), Positives = 364/549 (66%), Gaps = 6/549 (1%)

Query  1    MSLISKEELIKLAYSIRPRENEYKTILTNLDEYNKLTTNNNENK---YLQLKKLNESIDV  57
            M L+SKE+LI LA +  P+E EYK IL  LDEYN L  +  +N    YL+L +L++SID+
Sbjct  1    MGLMSKEQLIILAKNSSPKEGEYKKILELLDEYNLLNNSVEKNSIDLYLKLNELSKSIDI  60

Query  58   FMNKYKTSSRNRALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQW  117
            ++ KYK S RN AL  LK D+ KEVI IK++N  P+EKN+HFVW+GG +++I+++YI QW
Sbjct  61   YLKKYKNSKRNNALYQLKSDLTKEVIEIKDTNLKPLEKNIHFVWVGGMINNISIDYINQW  120

Query  118  ADINAEYNIKLWYDSEAFLVNTLKKAIVESSTTEALQLLEEEIQNPQFDNMKFYKKRMEF  177
             DIN++Y   +WYDSEA LVN LKKAI++SS  E L   E  + +  FD+ KFY++RME
Sbjct  121  KDINSDYETIIWYDSEALLVNILKKAIIDSSNKEVLTKYESVLNDNSFDSNKFYRERMEV  180

Query  178  IYDRQKRFINYYKSQINKPTVPTIDDIIKSHLVSEYNRDETVLESYRTNSLRKINSNHGI  237
            I+ +QK F NYY +        +++D+IK +L+ +Y + +  LE Y   S     +N
Sbjct  181  IFRKQKEFNNYYNTN--DNYTKSLNDVIKVYLIEKYLKTDEELEKYINESKEVFKANGAK  238

Query  238  DIRANSLFTEQELLNIYSQELLNRGNLAAASDIVRLLALKNFGGVYLDVDMLPGIHSDLF  297
            DIR   +  + EL +IY QELL R NLA+ASDI+R++ L   GG+YLDVD+LPGI   +F
Sbjct  239  DIREYDILDDVELKSIYEQELLMRFNLASASDIIRVIVLNKLGGIYLDVDVLPGIKKHIF  298

Query  298  KTISRPSSIGLDRWEMIKLEAIMKYKKYINNYTSENFDKLDQQLKDNFKLIIESKSEKSE  357
            K I++P++I  ++W+MI+LE IMKYK+YI  YT  +F  L   L++  +  +  K+ KS+
Sbjct  299  KDINKPTNISENKWQMIQLETIMKYKQYIKGYTENSFKNLPSDLQEMLQEKVVEKNLKSD  358

Query  358  IFSKLENLNVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLNPAI  417
            IF +L ++ +S+L+ KIAF  G + NQ LISK+ SY  NL+I Q+KNRY  +N+ L+ AI
Sbjct  359  IFQRLGDIFISELDTKIAFMFGKIANQVLISKKNSYSLNLIINQIKNRYNIINKCLSSAI  418

Query  418  ESDNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSGPGAYASAY  477
            E  +NF +T  IF   L N        F++K+  YL  G+MP+ R+T++SGPG Y +AY
Sbjct  419  EKGSNFNNTVDIFIQQL-NEFYVNEGFFVSKVMGYLGDGYMPDMRATLNISGPGIYTAAY  477

Query  478  YDFINLQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQFEKYVRDY  537
            YD +   E ++   +  DL  F+ P+  +SQ TEQEINS W+F+Q  ++ +++K V  Y
Sbjct  478  YDLLYFNERSLNPQILQEDLKYFEVPQALISQQTEQEINSSWIFNQVKSQIEYKKLVEKY  537

Query  538  TGGSLSEDN  546
            T  SLSE++
Sbjct  538  TNKSLSEND  546
```

C. novyi α toxin

```
>emb|CAA88565.1|  alpha-toxin [Clostridium novyi]
Length=2178

 Score =  252 bits (643),  Expect = 9e-69, Method: Compositional matrix adjust.
 Identities = 182/554 (33%), Positives = 299/554 (54%), Gaps = 19/554 (3%)

Query  3    LISKEELIKLA-YSIRPRENEYKTILTNLDEYNK-LTTNNNENKYLQLKKLNESIDVFMN  60
            LI++E+L+K+A   ++ +E EY  IL  L+ +N+ +   + +  Y +L KLNE +D +
Sbjct  2    LITREQLMKIASIPLKRKEPEYNLILDALENFNRDIEGTSVKEIYSKLSKLNELVDNYQT  61

Query  61   KYKTSSRNRALSNLKKDILKEVI-LIKNSNTSPV-EKNLHFVWIGGEVSDIALEYIKQWA  118
            KY +S RN AL N +   +  E+  LIKNS TS +  KNL F+WIGG +SD +LEY   W
Sbjct  62   KYPSSGRNLALENFRDSLYSELRELIKNSRTSTIASKNLSFIWIGGPISDQSLEYYNMWK  121

Query  119  DINAEYNIKLWYDSEAFLVNTLKKAIVESSTTEALQLLEEEIQNPQFDNMKFYKKRMEFI  178
              N +YNI+L+YD  + LVNTLK AI++ S+   ++  +  I +  + + KFY  RM+ I
Sbjct  122  MFNKDYNIRLFYDKNSLLVNTLKTAIIQESSKVIIEQNQSNILDGTYGHNKFYSDRMKLI  181

Query  179  YDRQKRFINYYKSQINKPTVPTIDDIIKSHLVSEYNRDETVLESYRTNSLRKINSNHGID  238
            Y  ++     Y+   N     ++DDII + L + +  D   L + + N+  K+ +    D
Sbjct  182  YRYKRELKMLYE---NMKQNNSVDDIIINFLSNYFKYDIGKLNNQKENNNNKMIAIGATD  238

Query  239  IRANSLFTEQELLNIYSQELLNRGNLAAASDIVRLLALKNFGGVYLDVDMLPGIHSDLFK  298
            I   ++ T + L + Y QEL+   NLAAASDI+R+  LK +GGVY D+D LPG++  LF
Sbjct  239  INTENILTNK-LKSYYYQELIQTNNLAAASDILRIAILKKYGGVYCDLDFLPGVNLSLFN  297

Query  299  TISRPSSIGLDRWEMIKLEAIMKYKKYINNYTSENFDKLDQQLKDNFKLIIESKSEKSEI  358
             IS+P+ +  + WE    EAI   KK +NNY  +  +++  ++K+     + +  + +++
Sbjct  298  DISKPNGMDSNYWEAAIFEAIANEKKLMNNYPYKYMEQVPSEIKERILSFVRNH-DINDL  356

Query  359  FSKLENLNVSDLEI-----KIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHL  413
              L ++ +S LEI     K A     +  N  +IS   S   N +I Q++NRY+ LN  +
Sbjct  357  ILPLGDIKISQLEILLSRLKAATGKKTFSNAFIISNNDSLTLNNLISQLENRYEILNSII  416

Query  414  NPAIE----SDNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSG  469
                +     D+     +++  ++   + + + S F  +I  YL   GF PE   ST+  SG
Sbjct  417  QEKFKICETYDSYINSVSELVLETTPKNLSMDGSSFYQQIIGYLSSGFKPEVNSTVFFSG  476

Query  470  PGAYASAYYDFINLQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQ  529
            P  Y+SA   D  +  +NT +  L + +    F+   N     T  E  S W      A+ +
Sbjct  477  PNIYSSATCDTYHFIKNTFD-MLSSQNQEIFEASNNLYFSKTHDEFKSSWLLRSNIAEKE  535

Query  530  FEKYVRDYTGGSLS  543
            F+K ++ Y G +L+
Sbjct  536  FQKLIKTYIGRTLN  549
```

*C. perfringens tcp*A

```
>ref|ZP_02865634.1|  TcpA [Clostridium perfringens C str. JGS1495]
 gb|EDS79391.1|  TcpA [Clostridium perfringens C str. JGS1495]
Length=1776

 Score =  464 bits (1195),  Expect = 1e-143, Method: Compositional matrix adjust.
 Identities = 247/546 (45%), Positives = 362/546 (66%), Gaps = 6/546 (1%)

Query  4    ISKEELIKLAYSIRPRENEYKTILTNLDEYNKLTTNNNENK---YLQLKKLNESIDVFMN  60
            +SKE+LI LA +  P+E EYK IL  LDEYN L  +  +N    YL+L +L++SID+++
Sbjct  1    MSKEQLIILAKNSSPKEGEYKKILELLDEYNLLNNSVEKNSIDLYLKLNELSKSIDIYLK  60

Query  61   KYKTSSRNRALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWADI  120
            KYK S RN AL  LK D+ KEVI IK++N  P+EKN+HFVW+GG +++I+++YI QW DI
Sbjct  61   KYKNSKRNNALYQLKSDLTKEVIEIKDTNLKPLEKNIHFVWVGGMINNISIDYINQWKDI  120

Query  121  NAEYNIKLWYDSEAFLVNTLKKAIVESSTTEALQLLEEEIQNPQFDNMKFYKKRMEFIYD  180
            N++Y  +WYDSEA LVN LKKAI++SS  E L  E  + +  FD+ KFY++RME I+
Sbjct  121  NSDYETIIWYDSEALLVNILKKAIIDSSNKEVLTKYESVLNDNSFDSNKFYRERMEVIFR  180

Query  181  RQKRFINYYKSQINKPTVPTIDDIIKSHLVSEYNRDETVLESYRTNSLRKINSNHGIDIR  240
            +QK F NYY +        +++D+IK +L+ +Y + +  LE Y   S      +N   DIR
Sbjct  181  KQKEFNNYYNTN--DNYTKSLNDVIKVYLIEKYLKTDEELEKYINESKEVFKANGAKDIR  238

Query  241  ANSLFTEQELLNIYSQELLNRGNLAAASDIVRLLALKNFGGVYLDVDMLPGIHSDLFKTI  300
               +  + EL +IY QELL R NLA+ASDI+R++ L   GG+YLDVD+LPGI   +FK I
Sbjct  239  EYDILDDVELKSIYEQELLMRFNLASASDIIRVIVLNKLGGIYLDVDVLPGIKKHIFKDI  298

Query  301  SRPSSIGLDRWEMIKLEAIMKYKKYINNYTSENFDKLDQQLKDNFKLIIESKSEKSEIFS  360
            ++P++I  ++W+MI+LE IMKYK+YI  YT  +F  L   L++  +  +  K+ KS+IF
Sbjct  299  NKPTNISENKWQMIQLETIMKYKQYIKGYTENSFKNLPSDLQEMLQEKVVEKNLKSDIFQ  358

Query  361  KLENLNVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLNPAIESD  420
            +L ++ +S+L+ KIAF  G + NQ LISK+ SY  NL+I Q+KNRY  +N+ L+ AIE
Sbjct  359  RLGDIFISELDTKIAFMFGKIANQVLISKKNSYSLNLIINQIKNRYNIINKCLSSAIEKG  418

Query  421  NNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSGPGAYASAYYDF  480
            +NF +T  IF   L N        F++K+  YL  G+MP+ R+T++SGPG Y +AYYD
Sbjct  419  SNFNNTVDIFIQQL-NEFYVNEGFFVSKVMGYLGDGYMPDMRATLNISGPGIYTAAYYDL  477

Query  481  INLQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQFEKYVRDYTGG  540
            +  E ++   +  DL  F+ P+  +SQ TEQEINS W+F+Q  ++ +++K V  YT
Sbjct  478  LYFNERSLNPQILQEDLKYFEVPQALISQQTEQEINSSWTFNQVKSQIEYKKLVEKYTNK  537

Query  541  SLSEDN  546
            SLSE++
Sbjct  538  SLSEND  543
```

Citrobacter rodentium lymphocyte inhibitory factor

```
>ref|YP_003366177.1|  lifA gene product [Citrobacter rodentium ICC168]
 gb|AAW52502.1|  lymphocyte inhibitory factor A [Citrobacter rodentium]
 emb|CBG89387.1|  lymphocyte inhibitory factor A [Citrobacter rodentium ICC168]
Length=3208

 GENE ID: 8711386 lifA | lymphocyte inhibitory factor A
[Citrobacter rodentium ICC168]  (10 or fewer PubMed links)

 Score = 73.6 bits (179),  Expect = 1e-10, Method: Compositional matrix adjust.
 Identities = 126/522 (24%), Positives = 225/522 (43%), Gaps = 89/522 (17%)

Query  26   ILTNLDEYNKLTTNNNENKYLQLKKLNESIDVFMNKYK-TSSRNRALSNLKKDILKEVIL  84
            I  ++ EYN LT  N+ N    L+K  + + V   +   T + N+ +  +  DI KE
Sbjct  234  IKKSIAEYNLLTEKNSRNGLKLLQKQADLLKVIKEEIPATETTNKNMETIIADIKKEY--  291

Query  85   IKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWADINAEYNIKLWYDSEAF----LVNTL  140
              S+T  +EKN+H +W+ G    +YIK +    E+   LW D+ AF       + L
Sbjct  292  --QSHTVDIEKNIHAIWVAGSPPESISDYIKTFLKTYKEFTYYLWVDNNAFGAAKFTSIL  349

Query  141  K--------KAIVESSTTEALQLL------EEEIQNPQFDNMKFYKKRMEFIYDRQKR--  184
            K       K I +++  +++  +        E+  N   D  K+ +K  E +YD  ++
Sbjct  350  KQIAFDLACKTIQQNTPQKSIDFIIEYNEIREKFNNSPSDQQKYLEKLRE-LYDSYQKTS  408

Query  185  ----------FI---------NYYKSQINKPTVPTIDDIIKSHLVSEYNRDETVLESY---  223
                      F+         N++    I K       D++  ++L +  N  E  ++SY
Sbjct  409  SPLKHMFNALFLENMIKLQDNFFNYCIMKGVTDINDELRVNYLKNVINLSEDDIDSYHKT  468

Query  224  ------RTNSL-RKINSNHG------IDIRANSLFTEQELLNIYSQELLNRGNLAAASDI  270
                  R   L  K+ S  G          D+++      ++ E  + Y  E+L R N  AASD+
Sbjct  469  ISDNKERVKRLIHKLQSEFGETRISIRDVKSLHSLSKSENNHNYQIEMLLRWNYPAASDL  528

Query  271  VRLLALKNFGGVYLDVDMLPGIHSDLFKTISRPSSIGLDRW-EMIKLEAIMK--YKKYIN  327
            +R+  LK  GG+Y D DM+P     +    I   ++ G +R+ E +KL   +     +++N
Sbjct  529  LRMYILKEHGGIYTDTDMMPAYSKQVIFKIMMETN-GDNRFLEDLKLRRAISDGVLRHVN  587

Query  328  --NYTSENFDKLDQQLKDNFKLIIE--SKSEKSEIFSKLENLNVSD-LEIKIAFALGS--  380
              N     N+D +     K+  K I+    SK  +  IF+K++      D + I    + L +
Sbjct  588  KQNIDEVNYDGISDADKNIIKKILAGISKIPEDNIFTKIDTKIPRDTMPILRRYHLWTDG  647

Query  381  ----VINQALISKQGSYLTNLVIEQVKNRYQFL--------NQHLNPAIESDNNFTDTIK  428
                +N  ++S + S + ++VI        Y+ L        NQ          +  ++F+ T K
Sbjct  648  WNIRGLNGFMLSHKDSEVVDVVIAGQNQAYREKIRDNVQNQIYFKQTDDLSSFSVTDK  707

Query  429  IFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSGP  470
            +    +        S+F    + + Q   +PEA ST+ +SGP
Sbjct  708  V--GGVLVKKYLSGSLF----SNFRQDTIIPEALSTLQISGP  743
```

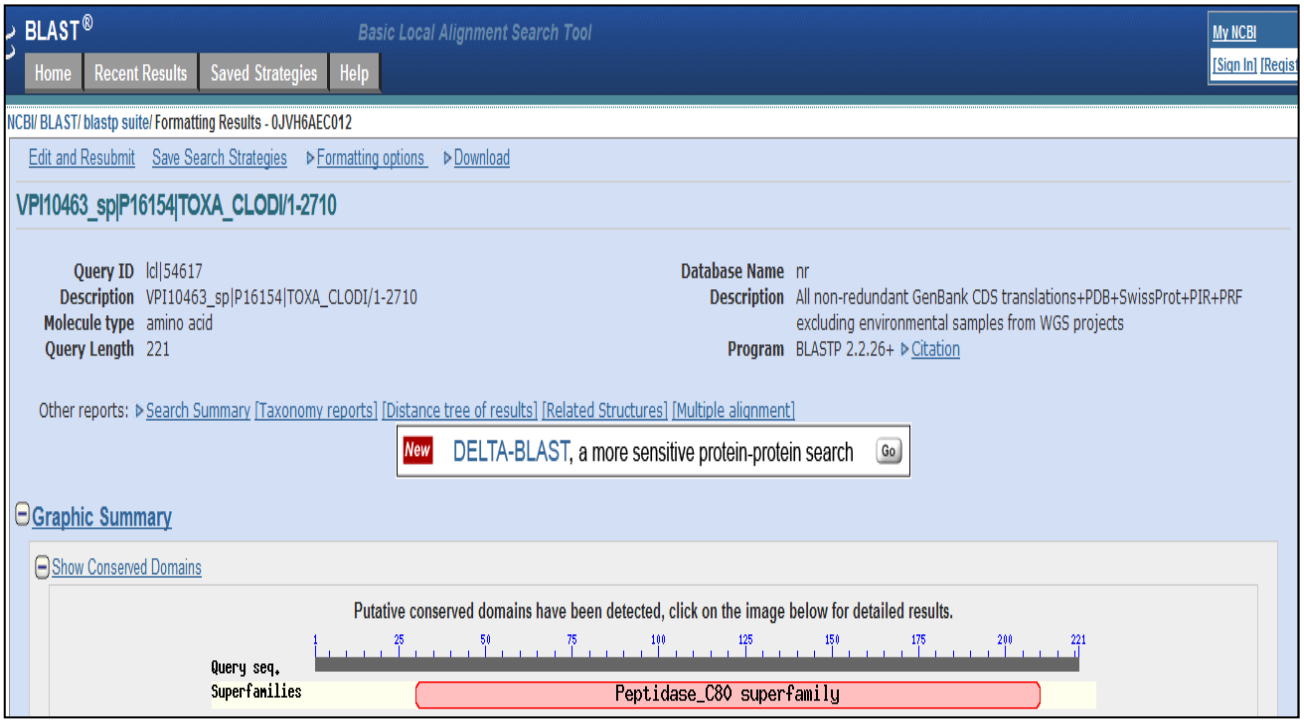
Pseudomonas fluorescens Putative toxin A

```
>gb|EIK69977.1|  putative toxin protein [Pseudomonas fluorescens Q8r1-96]
Length=2360

 Score =  117 bits (293),  Expect = 1e-24, Method: Composition-based stats.
 Identities = 116/473 (25%), Positives = 206/473 (44%), Gaps = 50/473 (11%)

Query  91   SPVEKNLHFVWIGGEVSDIALEYIKQWADINAE--YNIKLWYDSEAFLVNTLKKAIVESS  148
            + V K LHFVW+GG + +I  +Y+  W ++ A   Y++ LWYDS+A L     + IVE++
Sbjct  120  TEVPKTLHFVWLGGGLGNIQRDYLNVWKEVLARQGYSLNLWYDSDALLAWQTNRLIVEAA  179

Query  149  TTEA-LQLLEEEIQNPQFDNMKFYKKRMEFIYDRQKRFINYYKSQINKPTVPTIDDIIKS  207
              T+  LQ ++E I   +   +  Y +R   I  +Q+   +   + +   +    D+
Sbjct  180  KTDVFLQGVDERISEVELGGL--YNERT--IVLKQQMHAHISAAVVGGGSA---DEARMD  232

Query  208  HLVSEYNRDETVLESYRTNSLRKINSNHGIDIRANSLF-TEQELLNIYSQELLNRGNLAA  266
             L   Y +D  VL     ++ R +    G  +R  +      +L ++Y +E+  RGN+AA
Sbjct  233  LLSRAYGQDVQVLRKQLEDNRRSVLDMDGFKLRDIATGDVSLQLQDVYEREMCLRGNMAA  292

Query  267  ASDIVRLLALKNFGGVYLDVDMLPGIHSDLFKTISRPSSIGLDRWEMIKLEAIMKYKKYI  326
            ASD+VR   L   GG Y DVD LP     L +T+      G DR   + +  ++     +
Sbjct  293  ASDVVRAEVLYAEGGSYTDVDHLP----PLSQTLGAVDISGFDRNARLGVLQLL-----L  343

Query  327  NN---------YTSENFDKLDQQLKDNFKLIIESKSEKSEIFSKLENLNVSDLEIKIAFA  377
            NN          +S ++  +  +     +   +S+   SE+F++  +       ++
Sbjct  344  NNNPEWMPGRQASSSHYTHIGAEYFPALQAFAQSRPSLSEVFAQPADRLARPFALRALAM  403

Query  378  LGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLNPAIESDNNFTDTTKIFH--DSLF  435
              S+ N  L++  G  + N VIE+++  Y  ++      A + D   +D   +    + +F
Sbjct  404  QQSLSNAFLMAHPGCAVLNSVIERLRANYALIDASTRLAAQRDIALSDVAGMLRLVEEVF  463

Query  436  ---NSATAENSM--------FLTKIAPYLQVGFMPEARSTISLSGPGAYASAYYDFINLQ  484
               N      SM          +T IA +   G   E+  TI L+GPGA  D+
Sbjct  464  EKTNGPLTGLSMEDVPRAIALITAIATHFGDGIRFESEGTIYLTGPGAMRDGMVDYAKAH  523

Query  485  EN-TIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQFEKYVRD  536
             I ++L     I    P   ++  TE+E +  W  ++        EK+V +
Sbjct  524  LGAAIAESLSKEAAIA---PNGTVNGATEEEQDHSWKENETD----HEKWVSN  569
```

8.4.1.2. Cysteine Protease domain:

Analysis of the cysteine protease domain of *C. difficile* toxin A revealed the following alignments:

Query length=221 amino acids.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
3HO6_A	Chain A, Structure-Function Analysis Of Inositol Hexakisphosphate	448	448	100%	2e-157	100%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	448	448	100%	2e-142	100%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	448	448	100%	2e-142	100%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	447	447	100%	2e-141	100%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| Tc	447	447	100%	2e-141	100%
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	447	447	100%	2e-141	100%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	446	446	100%	3e-141	99%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	446	446	100%	5e-141	99%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	446	446	100%	7e-141	99%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	444	444	99%	5e-141	99%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	444	444	99%	6e-141	99%
AFN52237.1	TcdA [[Clostridium] difficile]	443	443	99%	5e-140	99%
3PA8_A	Chain A, Structure Of The C. Difficile Tcdb Cysteine Protease Dom	265	265	99%	9e-86	57%
CAC19891.1	toxin B [[Clostridium] difficile]	274	274	99%	9e-81	59%
ADH94630.1	TcdB [[Clostridium] difficile]	274	274	99%	1e-80	59%
ADH94626.1	TcdB [[Clostridium] difficile]	273	273	99%	2e-80	59%
CAA80815.1	toxin B [[Clostridium] difficile]	272	272	99%	3e-80	59%
EHJ31817.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	270	270	99%	2e-79	59%
AAG18011.1	cytotoxin B [[Clostridium] difficile]	270	270	99%	2e-79	59%
ADH94625.1	TcdB [[Clostridium] difficile]	270	270	99%	2e-79	59%
ADH94636.1	TcdB [[Clostridium] difficile]	268	268	99%	7e-79	58%
ADH94624.1	TcdB [[Clostridium] difficile]	268	268	99%	9e-79	58%
ADH94628.1	TcdB [[Clostridium] difficile]	268	268	99%	1e-78	58%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	268	268	99%	1e-78	57%
ADH94634.1	TcdB [[Clostridium] difficile]	268	268	99%	1e-78	57%
ADH94635.1	TcdB [[Clostridium] difficile]	268	268	99%	1e-78	57%
ADH94633.1	TcdB [[Clostridium] difficile]	266	266	99%	3e-78	58%
ZP_05328744.1	toxin B [Clostridium difficile QCD-63q42]	265	265	99%	1e-77	57%
YP_001087135.1	tcdB gene product [Clostridium difficile 630] >ref|ZP_05349824.1|	265	265	99%	1e-77	57%

ADH94629.1	TcdB [[Clostridium] difficile]	264	264	99%	2e-77	57%
ADH94623.1	TcdB [[Clostridium] difficile]	264	264	99%	2e-77	57%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	262	262	99%	1e-76	57%
ADH94627.1	TcdB [[Clostridium] difficile]	262	262	99%	1e-76	57%
ADH94632.1	TcdB [[Clostridium] difficile]	262	262	99%	1e-76	57%
ADH94631.1	TcdB [[Clostridium] difficile]	262	262	99%	1e-76	57%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	260	260	99%	5e-76	57%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	260	260	99%	6e-76	57%
BAF46125.1	TpeL [Clostridium perfringens]	211	211	99%	6e-59	52%
ZP_02865634.1	TcpA [Clostridium perfringens C str. JGS1495] >gb|EDS79391.1| T	211	211	99%	7e-59	52%
ZP_02636804.1	TpeL [Clostridium perfringens B str. ATCC 3626] >gb|EDT23010.1|	211	211	99%	7e-59	52%
ACF49258.1	TpeL [Clostridium perfringens A]	210	210	99%	1e-58	52%
ZP_04640347.1	Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC 439	82.4	82.4	99%	7e-15	25%
CAA88565.1	alpha-toxin [Clostridium novyi]	111	111	98%	1e-24	34%
ZP_04642320.1	RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC 4396	78.2	78.2	83%	2e-13	27%
ZP_00989505.1	Autotransporter adhesin [Vibrio splendidus 12B01] >gb|EAP95439.1	53.5	53.5	83%	1e-05	25%
ZP_04418869.1	RTX (Repeat in toxin) cytotoxin [Vibrio cholerae 12129(1)] >gb|EE	58.2	58.2	83%	4e-07	24%
EGQ99461.1	RTX toxin RtxA [Vibrio cholerae HE39]	54.3	54.3	83%	9e-06	24%
ZP_01982391.1	RTX toxin RtxA [Vibrio cholerae 623-39] >gb|EDL72938.1| RTX tox	54.3	54.3	83%	9e-06	24%
ZP_09365212.1	RTX repeat-containing cytotoxin [Vibrio ordalii ATCC 33509]	54.3	54.3	83%	9e-06	24%
ZP_04961062.1	RTX toxin RtxA [Vibrio cholerae AM-19226] >gb|EDN15712.1| RTX	54.3	54.3	83%	1e-05	24%
EJH63941.1	RTX toxin RtxA [Vibrio cholerae HE-45]	53.9	53.9	83%	1e-05	24%
3EEB_A	Chain A, Structure Of The V. Cholerae Rtx Cysteine Protease Doma	52.8	52.8	83%	1e-05	23%
EJH62273.1	RTX toxin RtxA [Vibrio cholerae HE-25]	53.9	53.9	83%	1e-05	24%
EGS69061.1	PGAP1-like family protein [Vibrio cholerae BJG-01]	53.9	53.9	83%	1e-05	24%
NP_762440.1	autotransporter adhesin [Vibrio vulnificus CMCP6] >gb|AAO07430.1	53.5	53.5	83%	2e-05	23%
NP_937086.1	RTX repeat-containing cytotoxin [Vibrio vulnificus YJ016] >dbj|BA	53.5	53.5	83%	2e-05	23%
EGR08431.1	RTX toxin RtxA [Vibrio cholerae HE48]	53.5	53.5	83%	2e-05	24%
ABX24512.1	hemolysin [Listonella anguillarum]	53.1	53.1	83%	2e-05	23%
ZP_04410971.1	RTX toxin RtxA [Vibrio cholerae TM 11079-80] >gb|EEO06653.1| R	53.1	53.1	83%	2e-05	24%
YP_004191172.1	autotransporter adhesin [Vibrio vulnificus MO6-24/O] >gb|ADV889	53.1	53.1	83%	2e-05	23%
EJH50736.1	RTX toxin RtxA domain protein [Vibrio cholerae HC-43B1]	52.4	52.4	83%	4e-05	24%

3FZY_A	Chain A, Crystal Structure Of Pre-Cleavage Form Of Cysteine Prote	48.5	48.5	83%	4e-04	22%
ZP_01975348.1	RTX [Vibrio cholerae B33] >gb|EAZ77024.1| RTX [Vibrio cholerae B	48.1	48.1	83%	0.001	23%
ZP_01970588.1	RTX toxin RtxA [Vibrio cholerae NCTC 8457] >gb|EAZ74142.1| RTX	48.1	48.1	83%	0.001	23%
YP_002810174.1	RTX toxin RtxA [Vibrio cholerae M66-2] >ref|ZP_07008277.1| RTX	48.1	48.1	83%	0.001	23%
ZP_05237921.1	RTX toxin RtxA [Vibrio cholerae MO10] >gb|EET22690.1| RTX toxin	48.1	48.1	83%	0.001	23%
ZP_01681473.1	RTX toxin RtxA [Vibrio cholerae V52] >gb|EAX61706.1| RTX toxin R	48.1	48.1	83%	0.001	23%
NP_231094.1	RTX toxin RtxA [Vibrio cholerae O1 biovar El Tor str. N16961] >gb|	48.1	48.1	83%	0.001	23%
ZP_01677789.1	RTX toxin RtxA [Vibrio cholerae 2740-80] >gb|EAX57825.1| RTX to	48.1	48.1	83%	0.001	23%
EJH31540.1	RTX toxin RtxA [Vibrio cholerae CP1041(14)]	48.1	48.1	83%	0.001	23%
ZP_05419777.1	RTX toxins and related Ca2+-binding proteins [Vibrio cholera CIRS	48.1	48.1	83%	0.001	23%
YP_005333215.1	RTX toxin RtxA [Vibrio cholerae IEC224] >gb|AFC58227.1| RTX tox	48.1	48.1	83%	0.001	23%
AAD21057.1	RtxA protein [Vibrio cholerae]	48.1	48.1	83%	0.001	23%
YP_005633914.1	RTX (Repeat in toxin) cytotoxin [Vibrio cholerae LMA3984-4] >gb|A	48.1	48.1	83%	0.001	23%
ZP_01956295.1	RTX toxin RtxA [Vibrio cholerae MZO-3] >gb|EAY41476.1| RTX tox	47.8	47.8	83%	0.001	23%
ZP_01979529.1	RTX toxin RtxA [Vibrio cholerae MZO-2] >gb|EDM53577.1| RTX tox	47.8	47.8	83%	0.001	23%
ZP_06028842.1	autotransporter adhesin [Vibrio cholerae INDRE 91/1] >gb|EEY491	47.4	47.4	83%	0.002	23%
YP_003712501.1	hypothetical protein [Xenorhabdus nematophila ATCC 19061] >emb	53.1	53.1	81%	2e-05	25%
CBA72128.1	LRR-MCF ORF3 [Arsenophonus nasoniae]	50.8	50.8	81%	1e-04	23%
YP_003468304.1	Mcf protein (fragment) [Xenorhabdus bovienii SS-2004] >emb|CBJ	47.4	47.4	81%	0.002	22%
CBA74898.1	N-terminal half of LRR-MCF ORF1. Contains LRR and RTX-like regior	49.7	49.7	78%	3e-04	22%
ZP_01811976.1	calmodulin-sensitive adenylate cyclase [Vibrionales bacterium SWA	50.1	50.1	77%	2e-04	28%
AAC08437.1	truncated toxin A [[Clostridium] difficile]	304	304	68%	1e-95	100%
CAA71690.1	TcdA protein [[Clostridium] difficile]	301	301	68%	1e-94	99%
CAC19893.1	truncated toxin A [[Clostridium] difficile]	301	301	68%	1e-94	99%
CBA72635.1	RTX-family protein [Arsenophonus nasoniae]	50.1	50.1	62%	2e-04	29%
ZP_04642268.1	hypothetical protein ymoll0001_31680 [Yersinia mollaretii ATCC 43	53.9	53.9	60%	1e-05	30%
ADX36376.1	putative RTX-toxin [Vibrio vulnificus]	49.3	49.3	56%	4e-04	26%
EHJ31823.1	hypothetical protein HMPREF1123_00959 [Clostridium difficile 050-1	157	157	34%	6e-43	99%
EHJ31824.1	hypothetical protein HMPREF1123_00960 [Clostridium difficile 050-1	135	135	29%	2e-33	98%

Alignment information for genes shown in Table 4.2:

C. difficile tcdB

```
>gb|ADH94630.1|  TcdB [[Clostridium] difficile]
Length=2329

 Score =  274 bits (700),  Expect = 1e-80, Method: Compositional matrix adjust.
 Identities = 129/220 (59%), Positives = 172/220 (78%), Gaps = 3/220 (1%)

Query  2    VDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNPKNS  61
            +DF++NT +DK YLL  KI S+   ++   + YVHYI+QLQGD ISYEA CNLF+KNP +S
Sbjct  530  LDFSQNTVIDKEYLLE-KISSST--KSSERGYVHYIVQLQGDKISYEAACNLFAKNPYDS  586

Query  62   IIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSEFAR  121
            I+ Q+N+ +S  +Y+ +     I E++KYRIP+R+ ++ K+K+T IGHGK EFNT  FA
Sbjct  587  ILFQKNIEDSEVAYYYNPTDSEIQEIDKYRIPDRISDRPKIKLTLIGHGKAEFNTDIFAG  646

Query  122  LSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDKITS  181
            L VDSLS+EI + +D  K DISPK++E+NLLGCNMFSY  NVEETYPGKLLL + DK++
Sbjct  647  LDVDSLSSEIETIIDLAKADISPKSIEINLLGCNMFSYSVNVEETYPGKLLLRVKDKVSE  706

Query  182  TLPDVNKNSITIGANQYEVRINSEGRKELLAHSGKWINKE  221
             +P ++++SI + ANQYEVRINSEGR+ELL HSG+WINKE
Sbjct  707  LMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKE  746


>gb|ADH94626.1|  TcdB [[Clostridium] difficile]
Length=2329

 Score =  273 bits (697),  Expect = 3e-80, Method: Compositional matrix adjust.
 Identities = 129/220 (59%), Positives = 171/220 (78%), Gaps = 3/220 (1%)

Query  2    VDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNPKNS  61
            +DF++NT  DK YLL  KI S+   ++   + YVHYI+QLQGD ISYEA CNLF+KNP +S
Sbjct  530  LDFSQNTVTDKEYLLE-KISSST--KSSERGYVHYIVQLQGDKISYEAACNLFAKNPYDS  586

Query  62   IIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSEFAR  121
            I+ Q+N+ +S  +Y+ +     I E++KYRIP+R+ ++ K+K+T IGHGK EFNT  FA
Sbjct  587  ILFQKNIEDSEVAYYYNPTDSEIQEIDKYRIPDRISDRPKIKLTLIGHGKAEFNTDIFAG  646

Query  122  LSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDKITS  181
            L VDSLS+EI + +D  K DISPK++E+NLLGCNMFSY  NVEETYPGKLLL + DK++
Sbjct  647  LDVDSLSSEIETIIDLAKADISPKSIEINLLGCNMFSYSVNVEETYPGKLLLRVKDKVSE  706

Query  182  TLPDVNKNSITIGANQYEVRINSEGRKELLAHSGKWINKE  221
             +P ++++SI + ANQYEVRINSEGR+ELL HSG+WINKE
Sbjct  707  LMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKE  746
```

C. sordellii tcsL

```
>emb|CAA57959.1|  cytotoxin L [[Clostridium] sordellii]
Length=2364

 Score =  260 bits (665),  Expect = 5e-76, Method: Compositional matrix adjust.
 Identities = 126/220 (57%), Positives = 168/220 (76%), Gaps = 3/220 (1%)

Query  2    VDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNPKNS  61
            +DF +NT LDK+Y+ + KI S+   +K Y+HYI+QLQGD ISYEA+CNLFSK+P +S
Sbjct  549  LDFAQNTVLDKDYV-SKKILSS--MKTRNKEYIHYIVQLQGDKISYEASCNLFSKDPYSS  605

Query  62   IIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSEFAR  121
            I+ Q+N+  S  +Y+       I E++KYRIP ++ NK  +K+TFIGHGK EFNT  FA
Sbjct  606  ILYQKNIEGSETAYYYYVADAEIKEIDKYRIPYQISNKRNIKLTFIGHGKSEFNTDIFAN  665

Query  122  LSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDKITS  181
            L VDSLS+EI + L+  K DISPK +E+NLLGCNMFSY  + EETYPGKLLL I D+++
Sbjct  666  LDVDSLSSEIETILNLAKADISPKYIEINLLGCNMFSYSISAEETYPGKLLLKIKDRVSE  725

Query  182  TLPDVNKNSITIGANQYEVRINSEGRKELLAHSGKWINKE  221
             +P ++++SIT+ ANQYEVRIN EG++E+L HSGKWINKE
Sbjct  726  LMPSISQDSITVSANQYEVRINEEGKREILDHSGKWINKE  765
```

C. perfringens tpeL

```
>dbj|BAF46125.1|  TpeL [Clostridium perfringens]
Length=1651

 Score =  211 bits (536),  Expect = 6e-59, Method: Compositional matrix adjust.
 Identities = 114/220 (52%), Positives = 151/220 (69%), Gaps = 4/220 (2%)

Query  2    VDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNPKNS  61
            ++FN+N  +DK  LLN +I SNN+     K Y+ YIIQLQGD +SYEA  NLF KNP NS
Sbjct  548  LNFNENKIIDKVELLN-RINSNNLINFDDKEYLRYIIQLQGDKVSYEAAINLFIKNPSNS  606

Query  62   IIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSEFAR  121
            I++Q   N    SY+ + + +SI  +    IPE LK K K+K+TFIGHG++EFNT ERFAS
Sbjct  607  ILVQEINN---ISYYFNSEYKSIDSIQFDNIPEILKGKNKIKLTFIGHGEEEFNTERFAS  663

Query  122  LSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDKITS  181
            L+V   S +I   LD IK + + K ++++LLGCNMFSY+ NVEETYPGKLL  ++D +
Sbjct  664  LTVKEFSKKIYKVLDMIKSNTNVKEIQIDLLGCNMFSYNINVEETYPGKLLKVVLDYVDK  723

Query  182  TLPDVNKNSITIGANQYEVRINSEGRKELLAHSGKWINKE  221
                 K  I I ANQYEVRIN +G+KELL+HSG+W++KE
Sbjct  724  IYNADIKPEIKISANQYEVRINKDGKKELLSHSGEWLSKE  763
```

C. novyi tcnα

```
>ref|ZP_02865634.1|  TcpA [Clostridium perfringens C str. JGS1495]
 gb|EDS79391.1|  TcpA [Clostridium perfringens C str. JGS1495]
Length=1776

 Score =  211 bits (536),  Expect = 7e-59, Method: Compositional matrix adjust.
 Identities = 114/220 (52%), Positives = 151/220 (69%), Gaps = 4/220 (2%)

Query  2    VDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNPKNS  61
            ++FN+N  +DK  LLN +I SNN+     K Y+ YIIQLQGD +SYEA  NLF KNP NS
Sbjct  545  LNFNENKIIDKVELLN-RINSNNLINFDDKEYLRYIIQLQGDKVSYEAAINLFIKNPSNS  603

Query  62   IIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSEFAR  121
            I++Q   N    SY+ + + +SI  +    IPE LK K K+K+TFIGHG++EFNT  FA
Sbjct  604  ILVQEINN---ISYYFNSEYKSIDSIQFDNIPEILKGKNKIKLTFIGHGEEEFNTERFAS  660

Query  122  LSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDKITS  181
            L+V   S +I   LD IK + + K ++++LLGCNMFSY+ NVEETYPGKLL  ++D +
Sbjct  661  LTVKEFSKKIYKVLDMIKSNTNVKEIQIDLLGCNMFSYNINVEETYPGKLLKVVLDYVDK  720

Query  182  TLPDVNKNSITIGANQYEVRINSEGRKELLAHSGKWINKE  221
                  K  I I ANQYEVRIN +G+KELL+HSG+W++KE
Sbjct  721  IYNADIKPEIKISANQYEVRINKDGKKELLSHSGEWLSKE  760
```

C. perfringens tcpA

```
>ref|ZP_02865634.1|  TcpA [Clostridium perfringens C str. JGS1495]
 gb|EDS79391.1|  TcpA [Clostridium perfringens C str. JGS1495]
Length=1776

 Score =  211 bits (536),  Expect = 7e-59, Method: Compositional matrix adjust.
 Identities = 114/220 (52%), Positives = 151/220 (69%), Gaps = 4/220 (2%)

Query  2    VDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNPKNS  61
            ++FN+N  +DK  LLN +I SNN+     K Y+ YIIQLQGD +SYEA  NLF KNP NS
Sbjct  545  LNFNENKIIDKVELLN-RINSNNLINFDDKEYLRYIIQLQGDKVSYEAAINLFIKNPSNS  603

Query  62   IIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSEFAR  121
            I++Q   N    SY+ + + +SI  +    IPE LK K K+K+TFIGHG++EFNT  FA
Sbjct  604  ILVQEINN---ISYYFNSEYKSIDSIQFDNIPEILKGKNKIKLTFIGHGEEEFNTERFAS  660

Query  122  LSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDKITS  181
            L+V   S +I   LD IK + + K ++++LLGCNMFSY+ NVEETYPGKLL  ++D +
Sbjct  661  LTVKEFSKKIYKVLDMIKSNTNVKEIQIDLLGCNMFSYNINVEETYPGKLLKVVLDYVDK  720

Query  182  TLPDVNKNSITIGANQYEVRINSEGRKELLAHSGKWINKE  221
                  K  I I ANQYEVRIN +G+KELL+HSG+W++KE
Sbjct  721  IYNADIKPEIKISANQYEVRINKDGKKELLSHSGEWLSKE  760
```

Yersinia mollaretii

Methyl accepting chemotaxis protein & RTX protein

```
>ref|ZP_04640347.1|  Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC
43969]
 gb|EEQ11141.1|  Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC
43969]
Length=1797

 Score = 82.4 bits (202),  Expect = 7e-15, Method: Compositional matrix adjust.
 Identities = 56/221 (25%), Positives = 115/221 (52%), Gaps = 10/221 (5%)

Query  2    VDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNPKNS  61
            +DF+   + D+ Y      +P+  +E  G+       I QLQGDD  +EA+  LF+K+   S
Sbjct  353  IDFSVWESPDQAYF--TSLPNTPLEPEGTHYEKTLIFQLQGDDTCFEASRALFNKHRYTS  410

Query  62   IIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSEFAR  121
             +Q    + A+ +   +  +  +     Y  P +L  + K+++T +GHG+ E +I+ F
Sbjct  411  EWLQLGDGKPAEVFTWGETYKKFV----YTSPLKLDKEGKIRITLVGHGETEGDITTFGG  466

Query  122  LSVDSLSNEISSFLDTI-KLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDKIT  180
            ++ ++L   +SS    +    +  K + +NL GC++ +    + +T PG+  L+I  K
Sbjct  467  MNAETLKGHLSSLFARLGSSSVLIKGITLNLTGCSLLNPKQPLADTLPGQ--LAIWLKQQ  524

Query  181  STLPDVNKNSITIGANQYEVRINSEGRKELLAHSGKWINKE  221
            + +  ++ ++ ++ A + ++ +   G+KE+  +   WINKE
Sbjct  525  AEILGLDDSNWSVNARENDLLVLENGKKEIRIND-HWINKE  564


>ref|ZP_04642320.1|  RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC
43969]
 gb|EEQ09150.1|  RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC
43969]
Length=1998

 Score = 78.2 bits (191),  Expect = 2e-13, Method: Compositional matrix adjust.
 Identities = 53/193 (27%), Positives = 100/193 (52%), Gaps = 24/193 (12%)

Query  37   IIQLQGDDISYEATCNLFSKNPKNSIIIQR---NMNESAKSYFLSDDGESILELNKYRI-  92
            IIQLQGD+  +++    F+K+PK S  +Q    N+N+   ++  SD         N Y+
Sbjct  556  IIQLQGDETCFKSAQAFFAKHPKQSEWLQLTDGNIND-VFTWSTSD--------NHYQYT  606

Query  93   --PERLKNKEKVKVTFIGHGKDEFNTSEFARLSVDSLSNEISSFLDTIKLDISPK--NVE  148
              P  L+    V++  +GHG    ++ +FA +S   +   +    D I    + K  N++
Sbjct  607  DKPLALEKSGAVRIILVGHG----DSKQFAGMSAVQVGATLGKVFDKISKQGATKFNNIK  662

Query  149  VNLLGCNMFSYDFNVEETYPGKLLLSIMDKITSTLPDVNKNSITIGANQYEVRINSEGRK  208
            +N+ GC++F     +E+T+PG+L + + ++   +     +++ A +Y +R    G+K
Sbjct  663  LNMAGCSLFDTYLPLEQTFPGQLAIWMNER--ADFWGLKPEQLSVVAYEYPLRAVENGKK  720

Query  209  ELLAHSGKWINKE  221
            E+    G+WINKE
Sbjct  721  EIFV-KGEWINKE  732
```

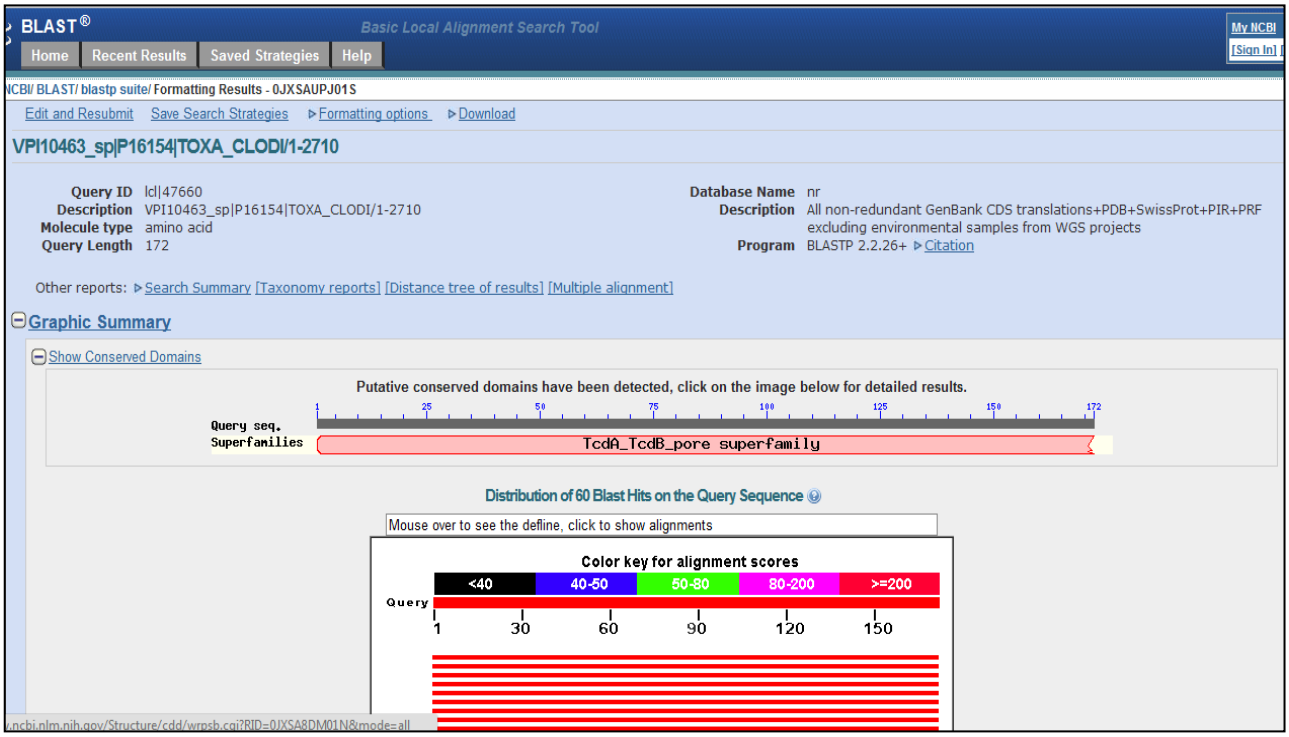
Vibrio cholerae RTX protein

```
>gb|EGQ99461.1|  RTX toxin RtxA [Vibrio cholerae HE39]
Length=4545

 Score = 54.3 bits (129),  Expect = 9e-06, Method: Composition-based stats.
 Identities = 46/192 (24%), Positives = 89/192 (46%), Gaps = 39/192 (20%)

Query  37    IIQLQGDDISYEATCNLFSKNPKNSIIIQRNMNESAKSYFLSDDGESILELNKYRI----  92
             I+Q++ DD+  +A  NL  K+P++S+++Q           L  DG         YR+
Sbjct  3463  IVQMENDDVVAKAAANLAGKHPESSVVVQ-----------LDSDG-------NYRVVYGD  3504

Query  93    PERLKNKEKVKVTFIGHGKD--EFNTSEFARLSVDSLSNEISSFLDTIKL--DISPKNVE  148
             P +L    K++   +GHG+D  E N +   +  S D L+ +++ F  +      +I+ K
Sbjct  3505  PSKLDG--KLRWQLVGHGRDHSESNNTRLSGYSADELAVKLAKFQQSFNQAENINNKPDH  3562

Query  149   VNLLGCNMFSYDFNVEETYPGKLLLSIMDKITSTLPDVNKNSITIGANQYEVRINSEGRK  208
             ++++GC++ S D    ++   G   ++ M        D N   + +      E+ ++  GRK
Sbjct  3563  ISIVGCSLVSDD---KQKGFGHQFINAM--------DANGLRVDVSVRSSELTVDEAGRK  3611

Query  209   ELLAHSGKWINK  220
                   +G W+ K
Sbjct  3612  HTKDANGDWVQK  3623


>ref|ZP_01982391.1|  RTX toxin RtxA [Vibrio cholerae 623-39]
 gb|EDL72938.1|  RTX toxin RtxA [Vibrio cholerae 623-39]
Length=4558

 Score = 54.3 bits (129),  Expect = 9e-06, Method: Composition-based stats.
 Identities = 46/192 (24%), Positives = 89/192 (46%), Gaps = 39/192 (20%)

Query  37    IIQLQGDDISYEATCNLFSKNPKNSIIIQRNMNESAKSYFLSDDGESILELNKYRI----  92
             I+Q++ DD+  +A  NL  K+P++S+++Q           L  DG         YR+
Sbjct  3476  IVQMENDDVVAKAAANLAGKHPESSVVVQ-----------LDSDG-------NYRVVYGD  3517

Query  93    PERLKNKEKVKVTFIGHGKD--EFNTSEFARLSVDSLSNEISSFLDTIKL--DISPKNVE  148
             P +L    K++   +GHG+D  E N +   +  S D L+ +++ F  +      +I+ K
Sbjct  3518  PSKLDG--KLRWQLVGHGRDHSESNNTRLSGYSADELAVKLAKFQQSFNQAENINNKPDH  3575

Query  149   VNLLGCNMFSYDFNVEETYPGKLLLSIMDKITSTLPDVNKNSITIGANQYEVRINSEGRK  208
             ++++GC++ S D    ++   G   ++ M        D N   + +      E+ ++  GRK
Sbjct  3576  ISIVGCSLVSDD---KQKGFGHQFINAM--------DANGLRVDVSVRSSELAVDEAGRK  3624

Query  209   ELLAHSGKWINK  220
                   +G W+ K
Sbjct  3625  HTKDANGDWVQK  3636
```

8.4.1.3. Membrane translocation domain

Analysis of the membrane translocation domain of *C. difficile* toxin A revealed the following alignments:

Query length=172 amino acids.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
EHJ31824.1	hypothetical protein HMPREF1123_00960 [Clostridium difficile 050-I	329	329	100%	1e-105	99%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	335	335	100%	3e-103	100%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	335	335	100%	4e-103	100%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	335	335	100%	5e-103	100%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	334	334	100%	8e-103	99%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	334	334	100%	8e-103	99%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	335	335	100%	8e-103	100%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	335	335	100%	9e-103	100%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| Tc	334	334	100%	1e-102	100%
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	334	334	100%	1e-102	100%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	334	334	100%	1e-102	100%
AFN52237.1	TcdA [[Clostridium] difficile]	333	333	100%	3e-102	99%
ADH94623.1	TcdB [[Clostridium] difficile]	222	222	100%	1e-63	60%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	222	222	100%	1e-63	60%
ADH94635.1	TcdB [[Clostridium] difficile]	222	222	100%	1e-63	60%
ADH94624.1	TcdB [[Clostridium] difficile]	222	222	100%	1e-63	60%
ADH94630.1	TcdB [[Clostridium] difficile]	222	222	100%	1e-63	60%
ADH94631.1	TcdB [[Clostridium] difficile]	221	221	100%	2e-63	60%
ADH94633.1	TcdB [[Clostridium] difficile]	221	221	100%	3e-63	60%
ADH94632.1	TcdB [[Clostridium] difficile]	221	221	100%	3e-63	60%
ADH94627.1	TcdB [[Clostridium] difficile]	221	221	100%	3e-63	60%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	221	221	100%	3e-63	60%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	221	221	100%	3e-63	60%

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
EHJ31824.1	hypothetical protein HMPREF1123_00960 [Clostridium difficile 050-1	329	329	100%	1e-105	99%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	335	335	100%	3e-103	100%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	335	335	100%	4e-103	100%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	335	335	100%	5e-103	100%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	334	334	100%	8e-103	99%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	334	334	100%	8e-103	99%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	335	335	100%	8e-103	100%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	335	335	100%	9e-103	100%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| Tc	334	334	100%	1e-102	100%
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	334	334	100%	1e-102	100%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	334	334	100%	1e-102	100%
AFN52237.1	TcdA [[Clostridium] difficile]	333	333	100%	3e-102	99%
ADH94623.1	TcdB [[Clostridium] difficile]	222	222	100%	1e-63	60%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	222	222	100%	1e-63	60%
ADH94635.1	TcdB [[Clostridium] difficile]	222	222	100%	1e-63	60%
ADH94624.1	TcdB [[Clostridium] difficile]	222	222	100%	1e-63	60%
ADH94630.1	TcdB [[Clostridium] difficile]	222	222	100%	1e-63	60%
ADH94631.1	TcdB [[Clostridium] difficile]	221	221	100%	2e-63	60%
ADH94633.1	TcdB [[Clostridium] difficile]	221	221	100%	3e-63	60%
ADH94632.1	TcdB [[Clostridium] difficile]	221	221	100%	3e-63	60%
ADH94627.1	TcdB [[Clostridium] difficile]	221	221	100%	3e-63	60%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	221	221	100%	3e-63	60%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	221	221	100%	3e-63	60%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	219	219	100%	2e-62	60%
EHJ31817.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	217	217	100%	1e-61	60%
AAG18011.1	cytotoxin B [[Clostridium] difficile]	217	217	100%	1e-61	60%
CAA80815.1	toxin B [[Clostridium] difficile]	217	217	100%	1e-61	60%
ADH94629.1	TcdB [[Clostridium] difficile]	217	217	100%	1e-61	60%
YP_001087135.1	tcdB gene product [Clostridium difficile 630] >ref|ZP_05349824.1|	217	217	100%	1e-61	60%
ADH94625.1	TcdB [[Clostridium] difficile]	216	216	100%	1e-61	60%
ADH94634.1	TcdB [[Clostridium] difficile]	216	216	100%	1e-61	60%
ADH94636.1	TcdB [[Clostridium] difficile]	216	216	100%	1e-61	60%
ZP_05328744.1	toxin B [Clostridium difficile QCD-63q42]	216	216	100%	2e-61	60%
ADH94628.1	TcdB [[Clostridium] difficile]	215	215	100%	6e-61	59%
CAC19891.1	toxin B [[Clostridium] difficile]	212	212	100%	4e-60	58%
ADH94626.1	TcdB [[Clostridium] difficile]	212	212	100%	4e-60	58%
BAF46125.1	TpeL [Clostridium perfringens]	152	152	99%	2e-39	46%
ZP_02865634.1	TcpA [Clostridium perfringens C str. JGS1495] >gb|EDS79391.1| T	152	152	99%	2e-39	46%
ZP_02636804.1	TpeL [Clostridium perfringens B str. ATCC 3626] >gb|EDT23010.1|	152	152	99%	2e-39	46%
ACF49258.1	TpeL [Clostridium perfringens A]	149	149	99%	3e-38	46%
ZP_04640347.1	Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC 439	49.7	49.7	98%	1e-04	25%
NP_930360.1	hypothetical protein plu3128 [Photorhabdus luminescens subsp. la	38.1	38.1	98%	0.85	22%
YP_003712268.1	hypothetical protein [Xenorhabdus nematophila ATCC 19061] >emb	37.0	37.0	98%	2.0	24%
ZP_01811976.1	calmodulin-sensitive adenylate cyclase [Vibrionales bacterium SWA	68.6	68.6	98%	1e-10	31%
YP_003712501.1	hypothetical protein [Xenorhabdus nematophila ATCC 19061] >emb	56.2	56.2	98%	9e-07	25%
YP_003468304.1	Mcf protein (fragment) [Xenorhabdus bovienii SS-2004] >emb|CBJ	47.0	47.0	98%	0.001	24%
CAA88565.1	alpha-toxin [Clostridium novyi]	112	112	96%	2e-25	38%
ADA82957.1	TpeL [Clostridium perfringens B]	134	134	91%	2e-33	45%
ADA82953.1	TpeL [Clostridium perfringens B]	134	134	91%	2e-33	45%
ZP_03319418.1	hypothetical protein PROVALCAL_02362 [Providencia alcalifaciens]	40.4	40.4	88%	0.17	25%
ZP_04642320.1	RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC 4396	50.1	50.1	88%	8e-05	27%
CBA72128.1	LRR-MCF ORF3 [Arsenophonus nasoniae]	40.0	40.0	87%	0.19	25%
ZP_05884844.1	cytotoxin Mcf [Vibrio coralliilyticus ATCC BAA-450] >gb|EEX34533.	42.0	42.0	73%	0.039	24%
XP_636034.1	hypothetical protein DDB_G0289839 [Dictyostelium discoideum AX4	35.8	35.8	64%	5.3	29%
NP_726796.2	wings apart-like, isoform B [Drosophila melanogaster] >gb|AAF457	35.4	35.4	51%	6.9	29%
NP_525042.2	wings apart-like, isoform A [Drosophila melanogaster] >sp|Q9W517	35.4	35.4	51%	7.1	29%
NP_875061.1	hypothetical protein [Prochlorococcus marinus subsp. marinus str.	35.4	35.4	45%	2.8	30%
YP_002604047.1	tpi gene product [Desulfobacterium autotrophicum HRM2] >gb|ACN	34.7	34.7	41%	9.8	30%
YP_001320103.1	hypothetical protein [Alkaliphilus metalliredigens QYMF] >gb|ABR48	33.9	33.9	33%	4.7	33%

ZP_08065205.1	choline binding protein E [Streptococcus peroris ATCC 700780] >g	176	977	99%	1e-42	30%
ZP_06979366.1	surface protein PspA [Streptococcus pneumoniae str. Canada MDR	168	912	99%	5e-42	32%
ZP_08523504.1	metallo-beta-lactamase domain protein [Streptococcus infantis SK	172	916	99%	1e-41	37%
ZP_05853491.1	putative cell wall binding repeat-containing domain protein [Blautia	171	1144	99%	8e-41	30%
ZP_08090001.1	glucan-binding domain-containing protein [Clostridium symbiosum V	155	1011	99%	9e-37	32%
ZP_09046953.1	hypothetical protein HMPREF1020_01032 [Clostridium sp. 7_3_54FA	155	1025	99%	1e-36	30%
ZP_10038893.1	metallo-beta-lactamase domain protein [Streptococcus sp. SK140]	155	962	99%	3e-36	29%
ZP_06342308.1	mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase [Bulleid	154	982	99%	1e-35	31%
ZP_08295167.1	cell wall-binding repeat protein [Actinomyces sp. oral taxon 170 st	150	850	99%	1e-35	29%
ZP_06344526.1	hypothetical protein CLOM621_05219 [Clostridium sp. M62/1] >gb|	152	1358	99%	1e-35	31%
ZP_07458344.1	choline binding protein E [Streptococcus sp. oral taxon 071 str. 73	153	807	99%	2e-35	32%
ZP_08602448.1	hypothetical protein HMPREF0993_01825 [Lachnospiraceae bacteri	154	1086	99%	2e-35	29%
EGL91250.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK25	152	806	99%	3e-35	32%
EIC78914.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK10	150	743	99%	2e-34	27%
ZP_08163928.1	NlpC/P60 family protein [Eggerthella sp. HGA1] >gb|EGC89895.1| N	148	781	99%	5e-34	25%
2QJ6_A	Chain A, Crystal Structure Analysis Of A 14 Repeat C-Terminal Fra	642	2249	99%	0.0	100%
ZP_08849318.1	hypothetical protein HMPREF9457_01027 [Dorea formicigenerans 4	165	899	99%	6e-39	28%
ZP_08108980.1	hypothetical protein HMPREF9475_03844 [Clostridium symbiosum W	152	996	99%	1e-35	30%
ZP_05331889.1	toxin A [Clostridium difficile QCD-63q42]	711	3032	99%	0.0	99%
ZP_05399187.1	toxin A [Clostridium difficile QCD-37x79]	647	2916	99%	0.0	94%
YP_694944.1	cell wall binding repeat-containing protein [Clostridium perfringens	240	1132	99%	8e-63	28%
ZP_09123063.1	hypothetical protein STRCR_1804 [Streptococcus criceti HS-6] >g	162	567	99%	1e-38	30%
ZP_08333659.1	hypothetical protein HMPREF0992_02583 [Lachnospiraceae bacteri	184	1288	99%	4e-45	31%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	477	2204	98%	3e-143	99%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	458	2110	98%	1e-136	99%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	400	1820	98%	2e-116	99%
EHJ36920.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	350	1666	97%	9e-112	86%
ZP_03288011.1	hypothetical protein CLONEX_00190 [Clostridium nexile DSM 1787]	227	1068	96%	4e-59	32%
2F6E_A	Chain A, Clostridium Difficile Toxin A C-Terminal Fragment 1 (Tcda-	247	1063	94%	6e-74	95%
AAZ40113.1	toxin A [[Clostridium] difficile] >gb|AAZ40136.1| toxin A [Clostridiu	213	1039	90%	8e-62	95%
AAZ40114.1	toxin A [[Clostridium] difficile] >gb|AAZ40116.1| toxin A [[Clostridi	224	1032	86%	7e-66	100%
AAZ40115.1	toxin A [[Clostridium] difficile] >gb|AAZ40127.1| toxin A [Clostridiu	213	1030	86%	9e-62	95%

Alignment information for genes shown in Table 4.2:

C. difficile tcdB

```
>gb|ADH94623.1|  TcdB [[Clostridium] difficile]
Length=2329

 Score =  222 bits (566),  Expect = 1e-63, Method: Compositional matrix adjust.
 Identities = 104/172 (60%), Positives = 151/172 (88%), Gaps = 0/172 (0%)

Query  1     QVNTLNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNA  60
             +VNTLNAAFFIQSLI+Y+S+K+ L++LS ++KVQ+YAQLFSTGLNTI D+ ++V L+S A
Sbjct  936   EVNTLNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTA  995

Query  61    VNDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLS  120
             +++TI++LPT++EG+P+++TI+DG++LGAAIKEL +  DPLL++E+EAK+G++A+N++ +
Sbjct  996   LDETIDLLPTLSEGLPVIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTAA  1055

Query  121   IAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNH  172
               A + S +GI +  +I L+P+AGISAGIPSLVNNELIL DKAT VV+YF+H
Sbjct  1056  TTAIITSSLGIASGFSILLVPLAGISAGIPSLVNNELILRDKATKVVDYFSH  1107


>ref|ZP_05270740.1|  toxin B [Clostridium difficile QCD-66c26]
 ref|ZP_05321143.1|  toxin B [Clostridium difficile CIP 107932]
 ref|ZP_05354976.1|  toxin B [Clostridium difficile QCD-76w55]
 ▶8 more sequence titles
 Length=2366

 Score =  222 bits (566),  Expect = 1e-63, Method: Compositional matrix adjust.
 Identities = 104/172 (60%), Positives = 151/172 (88%), Gaps = 0/172 (0%)

Query  1     QVNTLNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNA  60
             +VNTLNAAFFIQSLI+Y+S+K+ L++LS ++KVQ+YAQLFSTGLNTI D+ ++V L+S A
Sbjct  955   EVNTLNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTA  1014

Query  61    VNDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLS  120
             +++TI++LPT++EG+P+++TI+DG++LGAAIKEL +  DPLL++E+EAK+G++A+N++ +
Sbjct  1015  LDETIDLLPTLSEGLPVIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTAA  1074

Query  121   IAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNH  172
               A + S +GI +  +I L+P+AGISAGIPSLVNNELIL DKAT VV+YF+H
Sbjct  1075  TTAIITSSLGIASGFSILLVPLAGISAGIPSLVNNELILRDKATKVVDYFSH  1126
```

C. sordellii tcsL

```
>emb|CAA57959.1|  cytotoxin L [[Clostridium] sordellii]
Length=2364

 Score =  219 bits (558),  Expect = 2e-62, Method: Compositional matrix adjust.
 Identities = 103/172 (60%), Positives = 150/172 (87%), Gaps = 0/172 (0%)

Query  1     QVNTLNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNA  60
             +VNTLN+AFFIQSLI+Y++ K+ L++LS ++KVQ+YAQLFSTGLNTI D+ ++V L+S A
Sbjct  955   EVNTLNSAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTA  1014

Query  61    VNDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLS  120
             +++TI++LPT++EG+PI++TI+DG++LGAAIKEL + +DPLL++E+EAK+G++A+N++ +
Sbjct  1015  LDETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQEIEAKIGIMAVNLTAA  1074

Query  121   IAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNH  172
               A V S +GI +  +I L+P+AGISAGIPSLVNNELIL DKAT V++YF H
Sbjct  1075  STAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILQDKATKVIDYFKH  1126


>gb|EHJ31817.1|  cell wall-binding repeat protein [Clostridium difficile 050-P50-201
 gb|EHJ32980.1|  cell wall-binding repeat protein [Clostridium difficile 002-P50-2011]
Length=2367

 Score =  217 bits (552),  Expect = 1e-61, Method: Compositional matrix adjust.
 Identities = 104/172 (60%), Positives = 150/172 (87%), Gaps = 0/172 (0%)

Query  1     QVNTLNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNA  60
             +VNTLNAAFFIQSLI+Y+S+K+ L++LS ++KVQ+YAQLFSTGLNTI D+ ++V L+S A
Sbjct  956   EVNTLNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTA  1015

Query  61    VNDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLS  120
             +++TI++LPT++EG+PI++TI+DG++LGAAIKEL   DPLL++E+EAK+G++A+N++ +
Sbjct  1016  LDETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTTA  1075

Query  121   IAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNH  172
               A + S +GI +  +I L+P+AGISAGIPSLVNNEL+L DKAT VV+YF H
Sbjct  1076  TTAIITSSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKH  1127
```

C. perfringens tpeL

```
>ref|ZP_02636804.1|  TpeL [Clostridium perfringens B str. ATCC 3626]
 gb|EDT23010.1|  TpeL [Clostridium perfringens B str. ATCC 3626]
Length=1776

 Score =  152 bits (385),  Expect = 2e-39, Method: Compositional matrix adjust.
 Identities = 79/171 (46%), Positives = 117/171 (68%), Gaps = 0/171 (0%)

Query  2     VNTLNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNAV  61
             V+TLN +F IQS+IDY +     N LSTSVKVQ+Y Q+ +  L+ I D+  LV +I+ A
Sbjct  941   VSTLNTSFLIQSMIDYKAQNFDFNKLSTSVKVQIYCQITNISLSEIQDASNLVKIIAEAN  1000

Query  62    NDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLSI  121
                IN++PT+   IP+++TI+DGINL A I EL++  D LL+KEL A++G+++ NM+ +I
Sbjct  1001  EIEINLIPTLANAIPLITTIVDGINLIANIDELINTKDELLRKELAARIGIISSNMTAAI  1060

Query  122   AATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNH  172
             ++ +      G      L+PIAGIS+GIP+LVNN LIL +K+  +  YF H
Sbjct  1061  SSYILYFTEFGEVFNPLLVPIAGISSGIPTLVNNILILEEKSKEITEYFSH  1111


>gb|ACF49258.1|  TpeL [Clostridium perfringens A]
Length=1779

 Score =  149 bits (376),  Expect = 3e-38, Method: Compositional matrix adjust.
 Identities = 78/171 (46%), Positives = 117/171 (68%), Gaps = 0/171 (0%)

Query  2     VNTLNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNAV  61
             V+TLN +F I S+IDY ++    N LSTSVKVQ+Y Q+ +  L+ I D+  LV +I+ A
Sbjct  944   VSTLNTSFLIPSMIDYKAHNFDFNKLSTSVKVQIYCQITNISLSEIQDASNLVKIIAEAN  1003

Query  62    NDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLSI  121
                IN++PT+   IP+++TI+DGINL A I EL++  D L+KKEL A++G+++ NM+ +I
Sbjct  1004  EIEINLIPTLANAIPLITTIVDGINLIANIDELINTKDELIKKELAARIGVISSNMTAAI  1063

Query  122   AATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNH  172
             ++ +      G      L+PIAGIS+GIP+L+NN LIL +K+  +  YF H
Sbjct  1064  SSYILYFTEFGEVFNPLLVPIAGISSGIPTLINNILILEEKSKEITEYFSH  1114
```

C. novyi tcnα

```
>emb|CAA88565.1|  alpha-toxin [Clostridium novyi]
Length=2178

 Score =  112 bits (281),  Expect = 2e-25, Method: Compositional matrix adjust.
 Identities = 65/169 (38%), Positives = 106/169 (63%), Gaps = 3/169 (2%)

Query  5     LNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNAVNDT  64
             LN+A  +Q LIDY    ++L +++TS+KVQ YAQ+F   +  I ++ ++V +IS+A+N
Sbjct  965   LNSAMLMQLLIDYKPYTEILTNMNTSLKVQAYAQIFQLSIGAIQEATEIVTIISDALNAN  1024

Query  65    INVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLSIAAT  124
              N+L  +  G  + S I+DGINL AA+ EL +      +K +EAKVG+ +I   L  ++
Sbjct  1025  FNILSKLKVGSSVASVIIDGINLIAALTELKNVKTNFERKLIEAKVGMYSIGFILESSSL  1084

Query  125   VASIVGIGAEVTIF---LLPIAGISAGIPSLVNNELILHDKATSVVNYF  170
             ++ ++G  A   I     +P+AGI  G+PSLVNN L+L +K   +++YF
Sbjct  1085  ISGLLGATAVSEILGVISVPVAGILVGLPSLVNNILVLGEKYNQILDYF  1133
```

C. perfringens tcpA

```
>ref|ZP_02865634.1|  TcpA [Clostridium perfringens C str. JGS1495]
 gb|EDS79391.1|  TcpA [Clostridium perfringens C str. JGS1495]
Length=1776

 Score =  152 bits (385),  Expect = 2e-39, Method: Compositional matrix adjust.
 Identities = 79/171 (46%), Positives = 117/171 (68%), Gaps = 0/171 (0%)

Query  2     VNTLNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNAV  61
             V+TLN +F IQS+IDY +     N LSTSVKVQ+Y Q+ +  L+ I D+  LV +I+ A
Sbjct  941   VSTLNTSFLIQSMIDYKAQNFDFNKLSTSVKVQIYCQITNISLSEIQDASNLVKIIAEAN  1000

Query  62    NDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLSI  121
                IN++PT+   IP+++TI+DGINL A I EL++  D LL+KEL A++G+++ NM+ +I
Sbjct  1001  EIEINLIPTLANAIPLITTIVDGINLIANIDELINTKDELLRKELAARIGIISSNMTAAI  1060

Query  122   AATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNH  172
             ++ +      G      L+PIAGIS+GIP+LVNN LIL +K+  +  YF+H
Sbjct  1061  SSYILYFTEFGEVFNPLLVPIAGISSGIPTLVNNILILEEKSKEITEYFSH  1111
```

Photorhabdus luminescens subsp. *laumondii* hypothetical protein

```
>ref|NP_930360.1|  hypothetical protein plu3128 [Photorhabdus luminescens subsp.
laumondii TTO1]
 emb|CAE15502.1|  unnamed protein product [Photorhabdus luminescens subsp. laumondii
TTO1]
Length=2384

 GENE ID: 2803139 plu3128 | hypothetical protein
[Photorhabdus luminescens subsp. laumondii TTO1] (10 or fewer PubMed links)

 Score = 38.1 bits (87),  Expect = 0.87, Method: Composition-based stats.
 Identities = 44/199 (22%), Positives = 82/199 (41%), Gaps = 38/199 (19%)

Query  2     VNTLNAAFFIQSLIDYSSNKDVLN--------DLSTSVKVQLYAQLFSTGLNTIYDSIQL  53
             V+ LNA F IQSLI + ++K+  +         DL+T++KV  Y   L        + D  ++
Sbjct  1097  VDGLNAGFAIQSLIQWFTDKNRNDAARGIASPDLATALKVHSYLNLVQMAHGGVQDVTKV  1156

Query  54    VNLISNA-----------VNDTINVLP-TITEGIPIVSTILDGINLGAAIKELLDEHDPL  101
               L+  A           + D  ++L  T+ EG      +  G  +G    EL     + +
Sbjct  1157  TELVRTALRGEVVAAETSLKDFTSILGHTVNEG---AGVLFGGAMVGLDAYELAHAENDV  1213

Query  102   LKKELEAKV---------GVLAINMSLSIAATVASIVGIGAEVTIFLLPIAGISAGIPSL  152
               K     ++         G   +   L  A+T  +++G G  +      + G++ G  +L
Sbjct  1214  QKAVFGTQLAFDSASFVTGTAGVGAGLVGASTAGAVLGSGGVI------LGGLAVGFTAL  1267

Query  153   VNNELILHDKATSVVNYFN  171
                   ++ + A +V  YF+
Sbjct  1268  AQAFGVVAEDAKAVGRYFD  1286


>ref|ZP_10600388.1|  virulence surface antigen [Pseudomonas sp. GM102]
 gb|EJL93853.1|  virulence surface antigen [Pseudomonas sp. GM102]
Length=2360

 Score = 38.1 bits (87),  Expect = 1.0, Method: Composition-based stats.
 Identities = 38/180 (21%), Positives = 81/180 (45%), Gaps = 10/180 (6%)

Query  2     VNTLNAAFFIQSLIDYSSNKDVLN-DLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNA  60
             V+TLNA F IQ+L++     ++  +  L+ +V++  Y           + D   LV+L+  A
Sbjct  1160  VHTLNAGFAIQALMNALRGQEGPDRPLTLAVRLHAYVNYAQLVHGNVVDVAGLVSLVRQA  1219

Query  61    VND----TINVLPTITEGI-PIVSTILDGI----NLGAAIKELLDEHDPLLKKELEAKVG  111
             + +          V P +   + P V        G+    N+G  I +L       +    ++
Sbjct  1220  LVEEKLIAHTVAPVVRAAVGPSVGEATGGLLQLANVGFDIYQLATAQSDVERARFGTQLT  1279

Query  112   VLAINMSLSIAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFN  171
               + ++ LS+ A     +   GA +    + +AG++ G+ +L     ++ ++A  V  +F+
Sbjct  1280  FDSASLVLSVGAYAVGVTTAGAFLGGAAVILAGLAVGVAALAQGFAVIAEEAKQVGLFFD  1339
```

Yersinia mollaretii

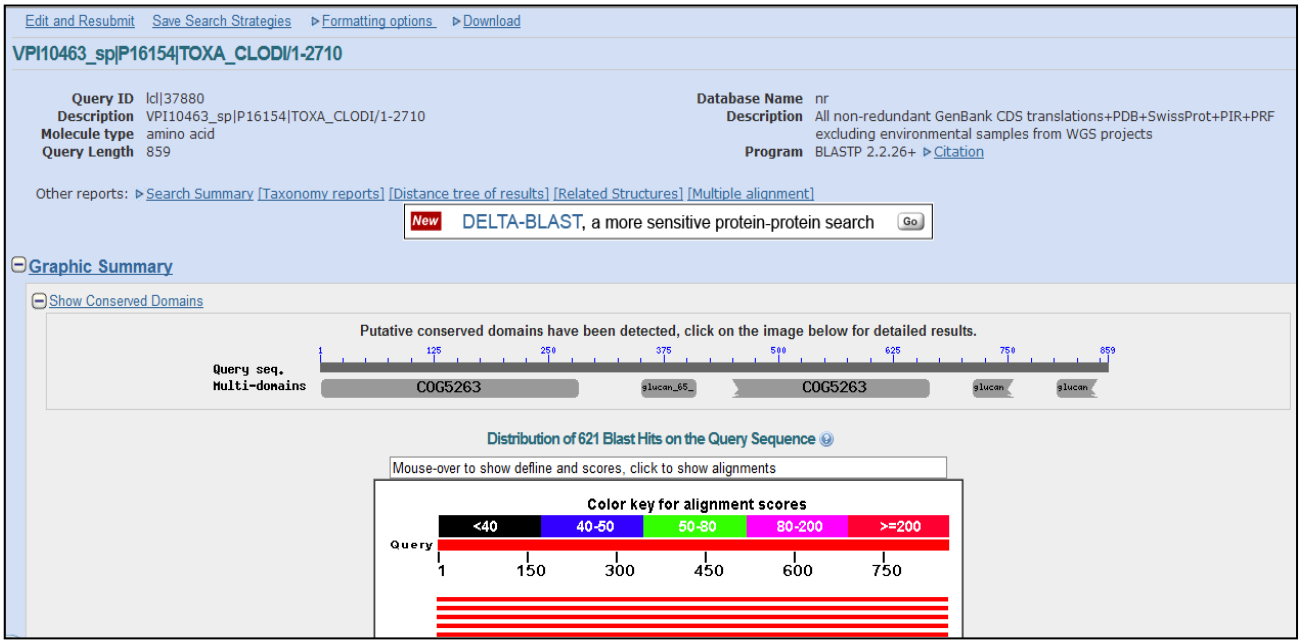
Methyl accepting chemotaxis protein & RTX protein

```
>ref|ZP_04642320.1|  RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC
43969]
 gb|EEQ09150.1|  RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC
43969]
Length=1998

 Score = 50.1 bits (118),  Expect = 8e-05, Method: Compositional matrix adjust.
 Identities = 46/169 (27%), Positives = 84/169 (50%), Gaps = 17/169 (10%)

Query  2     VNTLNAAFFIQSLIDYSSNKDVLND-LSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNA  60
             V+TLNAAF +Q+L+     +   +D LS  +++Q Y  L    +    D++ L  L++ A
Sbjct  925   VHTLNAAFLLQALMGQRPQQQGSSDALSWPMQLQNYVGLIQPTIGLAEDAVHLGGLVTQA  984

Query  61    VN--------DTINVL---PTITEGIP-IVSTILDGINLGAAIKELLDEHDPLLKKELEA  108
             +          T++ L    PT+   +P +   +LD  NL    I +L     P+
Sbjct  985   IAGVELKPLAQTLSALHASPTLGSVMPALPGLLLDAANLTGIIAQLATTDNPIEIAVAST  1044

Query  109   KVGVLAINMSLSIAATVASIV----GIGAEVTIFLLPIAGISAGIPSLV  153
              + +  +++AA V S +     G  A + +  +P+AGI+AG+P+LV
Sbjct  1045  NLTMATLTTGVNVAALVTSFIPAAAGASAVLGLVAVPLAGIAAGLPALV  1093


>ref|ZP_04640347.1|  Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC
43969]
 gb|EEQ11141.1|  Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC
43969]
Length=1797

 Score = 49.7 bits (117),  Expect = 1e-04, Method: Compositional matrix adjust.
 Identities = 46/184 (25%), Positives = 91/184 (49%), Gaps = 19/184 (10%)

Query  2    VNTLNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLIS---  58
             +TLNAAF +Q+L++ + +   +N LS  +++Q Y QL    L  ++D   + NL+
Sbjct  758  AHTLNAAFMLQTLMNINPSNGGINALSWPLQLQTYTQLAQNTLGLVHDVSAVANLVKLAS  817

Query  59   ----NAVNDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLA  114
                 ++   ++L T+  G  +V  +LD  N+     +L    DP+   E+   +  L
Sbjct  818  ATELKPLSAATSLLGTVAPG--VVGLLLDAANILGMSFQLSASTDPV---EINTTIANLT  872

Query  115  I-------NMSLSIAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVV  167
             +       N++  + +  A+   +  + +  +P+AGI+AG+P+LV N   L  S +
Sbjct  873  LSSLMVGTNIAALLTSLSAASAAVSGLLGMVAVPLAGIAAGLPALVGNYTTLAEQNKSAL  932

Query  168  NYFN  171
               F+
Sbjct  933  TAFD  936
```

8.4.1.4. Receptor binding domain

Analysis of the receptor binding domain of *C. difficile* toxin A revealed the following alignments:

Query coverage= 859 amino acids.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	1754	1944	100%	0.0	100%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	1750	1939	100%	0.0	99%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| To	1743	1929	100%	0.0	99%
AFN52237.1	TcdA [[Clostridium] difficile]	1707	1707	100%	0.0	97%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	1684	1833	100%	0.0	96%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	970	2493	100%	0.0	76%
ZP_05321147.1	toxin A [Clostridium difficile CIP 107932]	541	1643	100%	0.0	98%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	536	2427	100%	5e-164	98%
2G7C_A	Chain A, Clostridium Difficile Toxin A Fragment Bound To Agal(1,3)	487	1883	100%	2e-163	95%
ZP_05400117.1	toxin A [Clostridium difficile QCD-23m63]	478	1862	100%	4e-160	95%
ZP_06902239.1	toxin A [Clostridium difficile NAP07] >gb|EFH16631.1| toxin A [Clos	475	1853	100%	4e-159	94%
ZP_05328748.1	toxin A [Clostridium difficile QCD-63q42]	432	1763	100%	8e-143	100%
AAG18010.1	enterotoxin A [[Clostridium] difficile]	402	2297	100%	2e-128	75%
CAA73178.1	toxin A [[Clostridium] difficile]	397	2301	100%	1e-126	75%
EHJ31825.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	405	2311	100%	2e-126	75%
BAA25318.1	toxin A [[Clostridium] difficile]	384	2043	100%	7e-124	84%
ZP_04219464.1	hypothetical protein bcere0022_38930 [Bacillus cereus Rock3-44]	170	846	100%	4e-41	29%
AAO83644.1	toxin A [[Clostridium] difficile]	1639	1824	99%	0.0	99%
CAA35057.1	enterotoxin A [[Clostridium] difficile]	999	4672	99%	0.0	81%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	854	3303	99%	0.0	96%
AAO83645.1	toxin B [[Clostridium] difficile]	338	1738	99%	2e-101	41%
1814453A	toxin B	337	1728	99%	3e-101	42%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	356	1585	99%	7e-101	41%
AAO83646.1	toxin B [[Clostridium] difficile]	335	1495	99%	2e-100	41%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	346	1583	99%	1e-97	41%
CAC19891.1	toxin B [[Clostridium] difficile]	342	1402	99%	4e-96	41%
CAA43299.1	toxin B [[Clostridium] difficile]	335	1729	99%	5e-96	42%
EHJ40398.1	cell wall-binding repeat protein [Clostridium difficile 70-100-2010]	335	1492	99%	9e-94	41%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	333	1494	99%	3e-93	41%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	333	1493	99%	3e-93	41%
YP_001087135.1	tcdB gene product [Clostridium difficile 630] >ref|ZP_05349824.1|	333	1486	99%	4e-93	41%
ZP_05328744.1	toxin B [Clostridium difficile QCD-63q42]	333	1486	99%	4e-93	41%
AAG18011.1	cytotoxin B [[Clostridium] difficile]	332	1479	99%	5e-93	41%
EHJ31817.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	332	1487	99%	8e-93	41%
ADH94636.1	TcdB [[Clostridium] difficile]	331	1519	99%	2e-92	42%
ADH94635.1	TcdB [[Clostridium] difficile]	330	1517	99%	3e-92	41%
ADH94624.1	TcdB [[Clostridium] difficile]	330	1610	99%	4e-92	41%
CAA80815.1	toxin B [[Clostridium] difficile]	329	1480	99%	6e-92	41%
ADH94628.1	TcdB [[Clostridium] difficile]	329	1459	99%	8e-92	41%
ADH94623.1	TcdB [[Clostridium] difficile]	327	1452	99%	2e-91	41%
ADH94626.1	TcdB [[Clostridium] difficile]	326	1336	99%	1e-90	41%
ADH94630.1	TcdB [[Clostridium] difficile]	325	1141	99%	2e-90	41%
ADH94634.1	TcdB [[Clostridium] difficile]	324	1442	99%	3e-90	41%
ADH94633.1	TcdB [[Clostridium] difficile]	322	1444	99%	1e-89	41%
ADH94627.1	TcdB [[Clostridium] difficile]	319	1434	99%	2e-88	41%
ADH94632.1	TcdB [[Clostridium] difficile]	317	1426	99%	6e-88	41%
ADH94625.1	TcdB [[Clostridium] difficile]	316	1419	99%	1e-87	41%
ADH94631.1	TcdB [[Clostridium] difficile]	316	1420	99%	2e-87	41%
ADH94629.1	TcdB [[Clostridium] difficile]	313	1408	99%	1e-86	41%
ZP_02642051.1	cell wall binding repeat domain protein [Clostridium perfringens NCT	239	1503	99%	1e-62	29%
CAA80818.1	alpha-toxin [Clostridium novyi]	234	1467	99%	7e-62	50%
CAA88565.1	alpha-toxin [Clostridium novyi]	233	1460	99%	4e-60	50%
ZP_09815741.1	hypothetical protein LBLM1_20030 [Lactobacillus mucosae LM1] >(	184	841	99%	5e-45	27%
ZP_08160192.1	cell wall-binding repeat protein [Ruminococcus albus 8] >gb|EGC01	180	845	99%	1e-44	28%
EGU67616.1	metallo-beta-lactamase domain protein [Streptococcus mitis bv. 2	166	948	99%	2e-39	28%
ZP_04563653.1	predicted protein [Mollicutes bacterium D7] >gb|EEO33807.1| prec	151	840	99%	2e-34	29%
EGV81168.1	cell wall-binding repeat protein [Streptococcus oralis SK313]	149	750	99%	2e-34	31%
ZP_09374704.1	hypothetical protein HMPREF1021_03468 [Coprobacillus sp. 3_3_56	150	836	99%	5e-34	30%

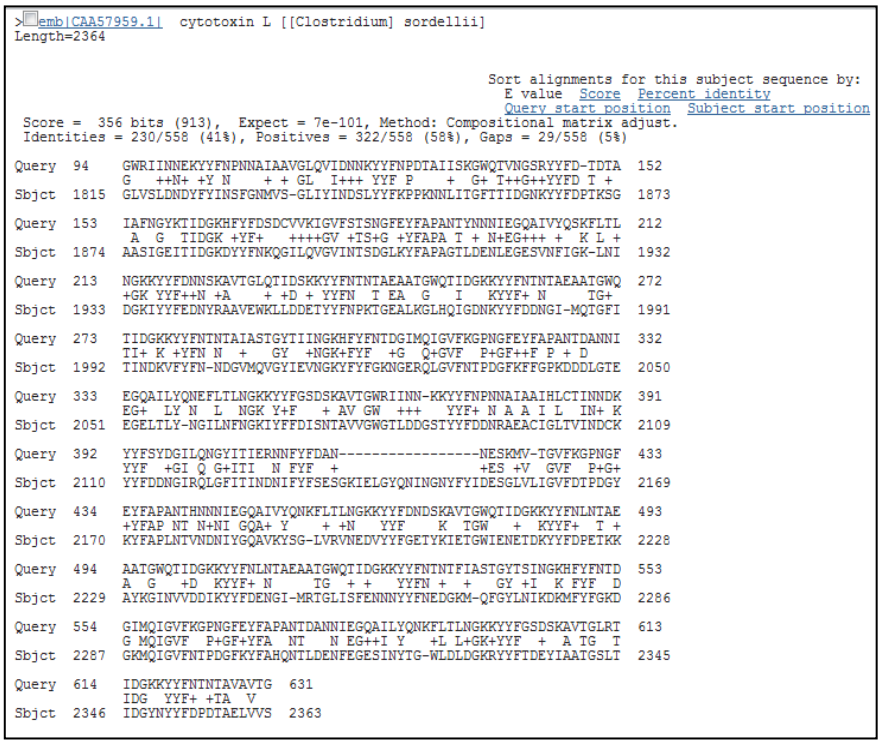
ZP_09374704.1	hypothetical protein HMPREF1021_03468 [Coprobacillus sp. 3_3_56	150	836	99%	5e-34	30%
NP_347713.1	hypothetical protein [Clostridium acetobutylicum ATCC 824] >ref|Y	328	2081	99%	2e-91	32%
ZP_03960388.1	conserved hypothetical protein [Lactobacillus vaginalis ATCC 4954	216	1053	99%	3e-56	31%
EIA18241.1	cell wall binding repeat-containing protein [Clostridium perfringens	194	1177	99%	2e-48	28%
ZP_03779155.1	hypothetical protein CLOHYLEM_06226 [Clostridium hylemonae DSM	180	548	99%	5e-44	28%
EGX29636.1	putative choline-binding protein [Streptococcus salivarius M18]	151	424	99%	1e-34	24%
ZP_04062387.1	glucosyltransferase-I [Streptococcus salivarius SK126] >gb|EEK09	151	515	99%	2e-34	27%
EHJ37067.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	588	2287	99%	0.0	99%
ZP_06144336.1	hypothetical protein RflaF_14077 [Ruminococcus flavefaciens FD-1	214	920	99%	1e-55	29%
YP_004104834.1	cell wall-binding repeat-containing protein [Ruminococcus albus 7]	191	970	99%	2e-48	28%
ZP_03960369.1	conserved hypothetical protein [Lactobacillus vaginalis ATCC 4954	182	817	99%	3e-45	55%
ZP_08065205.1	choline binding protein E [Streptococcus peroris ATCC 700780] >g	176	977	99%	1e-42	30%
ZP_06979366.1	surface protein PspA [Streptococcus pneumoniae str. Canada MDR	168	912	99%	5e-42	32%
ZP_08523504.1	metallo-beta-lactamase domain protein [Streptococcus infantis SK	172	916	99%	1e-41	37%
ZP_05853491.1	putative cell wall binding repeat-containing domain protein [Blautia	171	1144	99%	8e-41	30%
ZP_08090001.1	glucan-binding domain-containing protein [Clostridium symbiosum V	155	1011	99%	9e-37	32%
ZP_09046953.1	hypothetical protein HMPREF1020_01032 [Clostridium sp. 7_3_54FA	155	1025	99%	1e-36	30%
ZP_10038893.1	metallo-beta-lactamase domain protein [Streptococcus sp. SK140]	155	962	99%	3e-36	29%
ZP_06342308.1	mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase [Bulleid	154	982	99%	1e-35	31%
ZP_08295167.1	cell wall-binding repeat protein [Actinomyces sp. oral taxon 170 st	150	850	99%	1e-35	29%
ZP_06344526.1	hypothetical protein CLOM621_05219 [Clostridium sp. M62/1] >gb|	152	1358	99%	1e-35	31%
ZP_07458344.1	choline binding protein E [Streptococcus sp. oral taxon 071 str. 73	153	807	99%	2e-35	32%
ZP_08602448.1	hypothetical protein HMPREF0993_01825 [Lachnospiraceae bacteri	154	1086	99%	2e-35	29%
EGL91250.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK25	152	806	99%	3e-35	32%
EIC78914.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK10	150	743	99%	2e-34	27%
ZP_08163928.1	NlpC/P60 family protein [Eggerthella sp. HGA1] >gb|EGC89895.1| N	148	781	99%	5e-34	25%
2QJ6_A	Chain A, Crystal Structure Analysis Of A 14 Repeat C-Terminal Fra	642	2249	99%	0.0	100%
ZP_08849318.1	hypothetical protein HMPREF9457_01027 [Dorea formicigenerans 4	165	899	99%	6e-39	28%
ZP_08108980.1	hypothetical protein HMPREF9475_03844 [Clostridium symbiosum W	152	996	99%	1e-35	30%
ZP_05331889.1	toxin A [Clostridium difficile QCD-63q42]	711	3032	99%	0.0	99%
ZP_05399187.1	toxin A [Clostridium difficile QCD-37x79]	647	2916	99%	0.0	94%
YP_694944.1	cell wall binding repeat-containing protein [Clostridium perfringens	240	1132	99%	8e-63	28%
ZP_09123063.1	hypothetical protein SIRCR_1804 [Streptococcus cricetiHS-6] >g	162	567	99%	1e-38	30%

Alignment information for genes shown in Table 4.2:

C. difficile tcdB

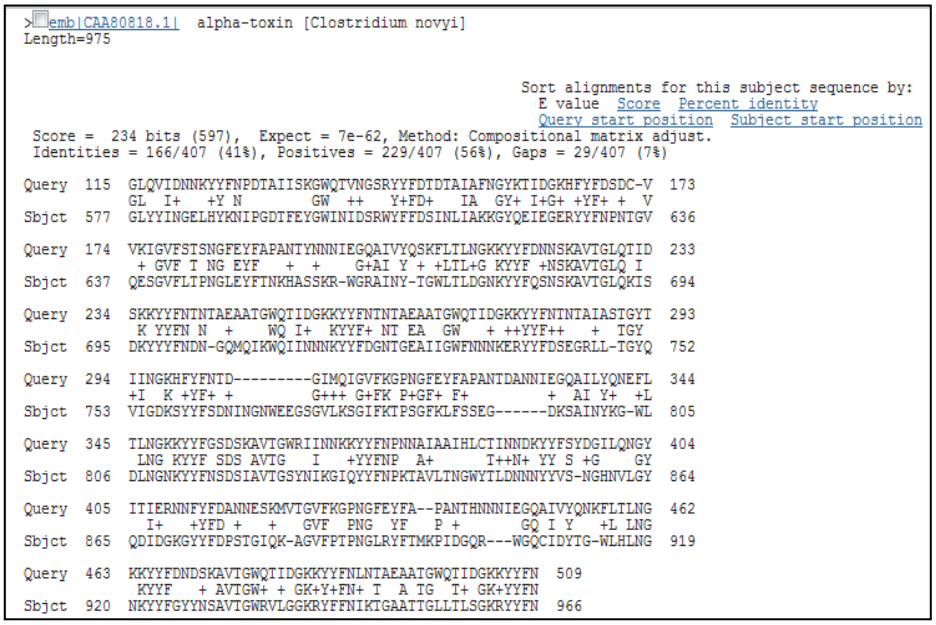
```
>prf||1814453A  toxin B
Length=576

                                                          Sort alignments for this subj
                                                            E value  Score  Percent ide
                                                          Query start position  Subje
 Score =  337 bits (863),  Expect = 3e-101, Method: Compositional matrix adjust.
 Identities = 216/520 (42%), Positives = 304/520 (58%), Gaps = 35/520 (7%)

Query  1    TGWQTINGKKYYFD-INTGAALTSYKIINGKHFYFNNDGVMQLGVFKGPDGFEYFAPANT  59
            TG+ T+    KYYF+ IN GAA      II+ K++YFN  GV+Q GVF   DGF+YFAPANT
Sbjct  64   TGFVTVGDDKYYFNPINGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPANT  123

Query  60   QNNNIEGQAIVYQSKFLTLNGKKYYFDNNSKAVTGWRIINNEKYYFNPNNAIAAVGLQVI  119
             + N+EG+AI +   K L ++    YYFD+N +     W+ ++ E +YF+P    A  GL  I
Sbjct  124  LDENLEGEAIDFTGK-LIIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQI  182

Query  120  DNNKYYFNPDTAIISKGWQTVNGSRYYFDTDTAIAFNGYKTIDGKHFYFDSDCVVKIGVF  179
             + KYYFN D  ++ KG+ ++N +++YFD D+ +   GY  IDGKHFYF  +  ++IGVF
Sbjct  183  GDYKYYFNSD-GVMQKGFVSINDNKHYFD-DSGVMKVGYTEIDGKHFYFAENGEMQIGVF  240

Query  180  STSNGFEYFAPANTYNNNIEGQAIVYQSKFLTLNGKKYYFDNNSKAVTGLQTI-DSKKYY  238
            +T +GF+YFA   N    N EG+ I Y S   L  N K YYFD++  AV G + + D  KYY
Sbjct  241  NTEDGFKYFAHHNEDLGNEEGEEISY-SGILNFNNKIYYFDDSFTAVVGWKDLEDGSKYY  299

Query  239  FNTNTAEAATGWQTIDGKKYYFNTNTAEAATGWQTIDGKKYYFNTNTAIASTGYTIINGK  298
            F+ +TAEA  G   I+  +YYFN +       G+ TI+ K +YF +++ I  +G   I+
Sbjct  300  FDEDTAEAYIGLSLINDGQYYFNDDGI-MQVGFVTINDKVFYF-SDSGIIESGVQNIDDN  357

Query  299  HFYFNTDGIMQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNEFLTLNGKKYYFGSDSK  358
            +FY + +GI+QIGVF   +G++YFAPANT  +NI GQA+ Y    + +    YYFG
Sbjct  358  YFYIDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSG-LVRVGEDVYYFGETYT  416

Query  359  AVTGWRII---NNKKYYFNPNNAIAAIHLCTINNDKYYFSYDGI----------------  399
              TGW       + KYYFNP    A    I++ KYYF   GI
Sbjct  417  IETGWIYDMENESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFN  476

Query  400  ----LQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQN  455
                +Q GYI IE   FYF    +  M  GVF  P+GF+YFA  NT + N EG++I Y
Sbjct  477  ENGEMQFGYINIEDKMFYF--GEDGVMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTG  534

Query  456  KFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAA  495
             +L L+ K+YYF ++   A TG    IDG++YYF+ +TA+
Sbjct  535  -WLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLV  573
```

C. sordellii tcsL



```
>emb|CAA57959.1|  cytotoxin L [[Clostridium] sordellii]
Length=2364

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Score =  356 bits (913),  Expect = 7e-101, Method: Compositional matrix adjust.
 Identities = 230/558 (41%), Positives = 322/558 (58%), Gaps = 29/558 (5%)

Query  94    GWRIINNEKYYFNPNNAIAAVGLQVIDNNKYYFNPDTAIISKGWQTVNGSRYYFD-TDTA  152
             G   ++N+ +Y N    + + GL  I+++ YYF P    +  G+ T++G++YYFD T +
Sbjct  1815  GLVSLDNDYFYINSFGNMVS-GLIYINDSLYYFKPPKNNLITGFTTIDGNKYYFDPTKSG  1873

Query  153   IAFNGYKTIDGKHFYFDSDCVVKIGVFSTSNGFEYFAPANTYNNNIEGQAIVYQSKFLTL  212
              A  G  TIDGK +YF+   ++++GV +TS+G +YFAPA T + N+EG+++ +  K L +
Sbjct  1874  AASIGEITIDGKDYYFNKQGILQVGVINTSDGLKYFAPAGTLDENLEGESVNFIGK-LNI  1932

Query  213   NGKKYYFDNNSKAVTGLQTIDSKKYYFNTNTAEAATGWQTIDGKKYYFNTNTAEAATGWQ  272
             +GK YYF++N +A    + +D + YYFN  T EA  G   I   KYYF+ N     TG+
Sbjct  1933  DGKIYYFEDNYRAAVEWKLLDDETYYFNPKTGEALKGLHQIGDNKYYFDDNGI-MQTGFI  1991

Query  273   TIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANNI  332
             TI+ K +YFN N  +   GY  +NGK+FYF  +G  Q+GVF  P+GF++F P + D
Sbjct  1992  TINDKVFYFN-NDGVMQVGYIEVNGKYFYFGKNGERQLGVFNIPDGFKFFGPKDDDLGTE  2050

Query  333   EGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINN-KKYYFNPNNAIAAIHLCTINNDK  391
             EG+  LY N  L  NGK Y+F   + AV GW  +++   YYF+ N A A I L  IN+ K
Sbjct  2051  EGELTLY-NGILNFNGKIYFFDISNTAVVGWGTLDDGSTYYFDDNRAEACIGLTVINDCK  2109

Query  392   YYFSYDGILQNGYITIERNNFYFDAN----------------NESKMV-TGVFKGPNGF  433
             YYF  +GI Q G+ITI  N FYF  +                  +ES +V  GVF  P+G+
Sbjct  2110  YYFDDNGIRQLGFITINDNIFYFSESGKIELGYQNINGNYFYIDESGLVLIGVFDTPDGY  2169

Query  434   EYFAPANTHNNNIEGQAIVYQNKFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAE  493
             +YFAP NT N+NI GQA+ Y    + +N   YYF     K  TGW   +  KYYF+  T +
Sbjct  2170  KYFAPLNTVNDNIYGQAVKYSG-LVRVNEDVYYFGETYKIETGWIENETDKYYFDPETKK  2228

Query  494   AATGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTD  553
             A  G   +D  KYYF+ N     TG  + +   YYFN +  +   GY +I  K FYF  D
Sbjct  2229  AYKGINVVDDIKYYFDENGI-MRTGLISFENNNYYFNEDGKM-QFGYLNIKDKMFYFGKD  2286

Query  554   GIMQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRT  613
             G MQIGVF  P+GF+YFA  NT   N EG++I Y   +L L+GK+YYF  +  A TG  T
Sbjct  2287  GKMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTG-WLDLDGKRYYFTDEYIAATGSLT  2345

Query  614   IDGKKYYFNTNTAVAVTG  631
             IDG  YYF+ +TA  V
Sbjct  2346  IDGYNYYFDPDTAELVVS  2363
```

C. novyi α toxin



```
>emb|CAA80818.1|  alpha-toxin [Clostridium novyi]
Length=975

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Score =  234 bits (597),  Expect = 7e-62, Method: Compositional matrix adjust.
 Identities = 166/407 (41%), Positives = 229/407 (56%), Gaps = 29/407 (7%)

Query  115  GLQVIDNNKYYFNPDTAIISKGWQTVNGSRYYFDTDTAIAFNGYKTIDGKHFYFDSDC-V  173
             GL  I+   +Y N          GW  ++    Y+FD+    IA  GY+ I+G+ +YF+ +  V
Sbjct  577  GLYYINGELHYKNIPGDTFEYGWINIDSRWYFFDSINLIAKKGYQEIEGERYYFNPNTGV  636

Query  174  VKIGVFSTSNGFEYFAPANTYNNNIEGQAIVYQSKFLTLNGKKYYFDNNSKAVTGLQTID  233
              + GVF T NG EYF   +  +    G+AI Y + +LTL+G KYYF +NSKAVTGLQ I
Sbjct  637  QESGVFLIPNGLEYFTNKHASSKR-WGRAINY-TGWLTLDGNKYYFQSNSKAVTGLQKIS  694

Query  234  SKKYYFNTNTAEAATGWQTIDGKKYYFNTNTAEAATGWQTIDGKKYYFNTNTAIASTGYT  293
              K YYFN N  +    WQ I+  KYYF+ NT EA  GW   + ++YYF++   +  TGY
Sbjct  695  DKYYYFNDN-GQMQIKWQIINNNKYYFDGNTGEAIIGWFNNNKERYYFDSEGRLL-TGYQ  752

Query  294  IINGKHFYFNTD---------GIMQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNEFL  344
             +I  K +YF+ +         G+++ G+FK P+GF+ F+              AI Y+  +L
Sbjct  753  VIGDKSYYFSDNINGNWEEGSGVLKSGIFKTPSGFKLFSSEG------DKSAINYKG-WL  805

Query  345  TLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKYYFSYDGILQNGY  404
              LNG KYYF SDS AVTG   I   +YYFNP  A+      T++N+ YY S +G    GY
Sbjct  806  DLNGNKYYFNSDSIAVTGSYNIKGIQYYFNPKTAVLTNGWYTLDNNNYYVS-NGHNVLGY  864

Query  405  ITIERNNFYFDANNESKMVTGVFKGPNGFEYFA--PANTHNNNIEGQAIVYQNKFLTLNG  462
               I+   +YFD +   +   GVF  PNG  YF   P +      GQ I Y   +L LNG
Sbjct  865  QDIDGKGYYFDPSTGIQK-AGVFPTPNGLRYFTMKPIDGQR---WGQCIDYTG-WLHLNG  919

Query  463  KKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFN  509
              KYYF   + AVTGW+ + GK+Y+FN+ T  A TG  T+ GK+YYFN
Sbjct  920  NKYYFGYYNSAVTGWRVLGGKRYFFNIKTGAATTGLLTLSGKRYYFN  966
```

*C. perfringens tcp*A

```
>ref|ZP_02865634.1|  TcpA [Clostridium perfringens C str. JGS1495]
 gb|EDS79391.1|  TcpA [Clostridium perfringens C str. JGS1495]
Length=1776

 Score =  464 bits (1195),  Expect = 1e-143, Method: Compositional matrix adjust.
 Identities = 247/546 (45%), Positives = 362/546 (66%), Gaps = 6/546 (1%)

Query  4    ISKEELIKLAYSIRPRENEYKTILTNLDEYNKLTTNNNENK---YLQLKKLNESIDVFMN  60
            +SKE+LI LA +  P+E EYK IL  LDEYN L  +  +N    YL+L +L++SID+++
Sbjct  1    MSKEQLIILAKNSSPKEGEYKKILELLDEYNLLNNSVEKNSIDLYLKLNELSKSIDIYLK  60

Query  61   KYKTSSRNRALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWADI  120
            KYK S RN AL  LK D+ KEVI IK++N  P+EKN+HFVW+GG +++I+++YI QW DI
Sbjct  61   KYKNSKRNNALYQLKSDLTKEVIEIKDTNLKPLEKNIHFVWVGGMINNISIDYINQWKDI  120

Query  121  NAEYNIKLWYDSEAFLVNTLKKAIVESSTTEALQLLEEEIQNPQFDNMKFYKKRMEFIYD  180
            N++Y  +WYDSEA LVN LKKAI++SS  E L  E  + +  FD+ KFY++RME I+
Sbjct  121  NSDYETIIWYDSEALLVNILKKAIIDSSNKEVLTKYESVLNDNSFDSNKFYRERMEVIFR  180

Query  181  RQKRFINYYKSQINKPTVPTIDDIIKSHLVSEYNRDETVLESYRTNSLRKINSNHGIDIR  240
            +QK F NYY +       +++D+IK +L+ +Y + +  LE Y   S       +N   DIR
Sbjct  181  KQKEFNNYYNTN--DNYTKSLNDVIKVYLIEKYLKTDEELEKYINESKEVFKANGAKDIR  238

Query  241  ANSLFTEQELLNIYSQELLNRGNLAAASDIVRLLALKNFGGVYLDVDMLPGIHSDLFKTI  300
               +  + EL +IY QELL R NLA+ASDI+R++ L   GG+YLDVD+LPGI   +FK I
Sbjct  239  EYDILDDVELKSIYEQELLMRFNLASASDIIRVIVLNKLGGIYLDVDVLPGIKKHIFKDI  298

Query  301  SRPSSIGLDRWEMIKLEAIMKYKKYINNYTSENFDKLDQQLKDNFKLIIESKSEKSEIFS  360
            ++P++I  ++W+MI+LE IMKYK+YI  YT  +F  L   L++  +  +  K+ KS+IF
Sbjct  299  NKPTNISENKWQMIQLETIMKYKQYIKGYTENSFKNLPSDLQEMLQEKVVEKNLKSDIFQ  358

Query  361  KLENLNVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLNPAIESD  420
            +L ++ +S+L+ KIAF  G + NQ LISK+ SY  NL+I Q+KNRY  +N+ L+ AIE
Sbjct  359  RLGDIFISELDTKIAFMFGKIANQVLISKKNSYSLNLIINQIKNRYNIINKCLSSAIEKG  418

Query  421  NNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSGPGAYASAYYDF  480
            +NF +T  IF   L N        F++K+  YL  G+MP+ R+T+++SGPG Y +AYYD
Sbjct  419  SNFNNTVDIFIQQL-NEFYVNEGFFVSKVMGYLGDGYMPDMRATLNISGPGIYTAAYYDL  477

Query  481  INLQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQFEKYVRDYTGG  540
            +   E ++    +   DL  F+ P+  +SQ TEQEINS W+F+Q  ++ +++K V  YT
Sbjct  478  LYFNERSLNPQILQEDLKYFEVPQALISQQTEQEINSSWTFNQVKSQIEYKKLVEKYTNK  537

Query  541  SLSEDN  546
            SLSE++
Sbjct  538  SLSEND  543
```

Streptococcus salivarius hypothetical protein

```
>gb|EGX29636.1|  putative choline-binding protein [Streptococcus salivarius M18]
Length=1089

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Score =  151 bits (381),  Expect = 1e-34, Method: Compositional matrix adjust.
 Identities = 234/992 (24%), Positives = 365/992 (37%), Gaps = 236/992 (24%)

Query  18    GAALTSYKIINGKHFYFNNDGVMQLGVFKGPDGFEYFAPANTQNNNIEGQAIVYQSKFLT  77
             G  LT +  +  + +YF  DG    G F   +G  Y+   +T          ++  + +L
Sbjct  181   GKNLITGWNNVEDREYYFQEDGKQVKGQFVEVNGKNYYYLDDHT--------GMLLVNCYLD  232

Query  78    LNGKKYYFDNNSKA---------VTGWRIINNE--KYYFNPNNAIAAVGLQVIDNNKYYF  126
              +GK Y  D N             +TG      N+  ++ +           G Q +D  + YF
Sbjct  233   KDGKHYQIDENGVVTERTKLPTNITGGHFEANDEGEWSYITEQGEKLIGFQYVDGVELYF  292

Query  127   NPDTAIISKGWQTVNGSRYYFDTDT-AIAFNGYKTIDGKHF-------------YFDSDC  172
             + D   +      TV+G  YY D +T A+  N Y+    K                YF+ D
Sbjct  293   DKDGKQLKGQEITVDGKTYYLDQNTGALLKNSYRNWSEKQIISRYKTNYIYHTSYFNRDG  352

Query  173   VVKIGVFSTSNGFEYFAPAN--------------TY----NNNIEGQAIVYQSKFLTL--  212
             +   G+  T+ GF ++   N               TY       +  ++ ++      T
Sbjct  353   IRATGLVKTAAGFIHYFDENGELLKNVAVNVGGTTYVFGERGRLARKSFIWDKVDFTFPE  412

Query  213   NGKKYYFDNNSKAVTGLQTIDSKKYYFNTNTAEAATGWQTIDGKKYYFNT----------  262
             N   YY D   AV GLQTID  + YF+ +  +A      IDGK YYF+
Sbjct  413   NVNFYYGDEKGHAVKGLQTIDGYQLYFDKDGRQAKDKIVQIDGKTYYFDKTNGRMVKNQW  472

Query  263   ----------------------NTAEAATGWQTIDGKKYYFNTNTAIASTGYTIINGKHFY  301
                                N   A TGWQ IDGK  YF      AS G   INGK++
Sbjct  473   ASVNVGGISPASKDYRSYYLGNDGAAVTGWQDIDGKHLYFTDTGIYASNGIYSINGKNYL  532

Query  302   FNTDGIMQ--IGVFKGPNG---FEYFAPANTDANNIEGQAILYQNEFLTLNGKKYYFGSD  356
             F   +++   GV   P      +Y   N D  + G+ I        ++G +Y F SD
Sbjct  533   FEKGQLVKDAYGVVDKPGAKVRLKYTYRTNADGEVLTGKQI--------IDGTEYIFASD  584

Query  357   SKAVTGWRIINNKKY--------------YFNPNNAIAAIHLCTINNDKYYFSYDGILQN  402
              + V G    + K Y               +F+ N  +  I+    I    Y    +G+L
Sbjct  585   GQVVDGVVRYDGKLYLVKDSKIEKNYFGAFFSKNEILGGINFTGI----YGTDENGVLLE  640

Query  403   GY-ITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQ--AIVYQNKFLT  459
             G   +++    YF    +S +    +K  +G  Y          +  G     I+  N  L
Sbjct  641   GVQRSLDGQLHYFQPEVKS-VDKPTWKEIDGKRYRLTKWYLPEHHAGMYTTIILTNDTLK  699

Query  460   LNGKKYYFDNDSKAVT-----GWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAE  514
             ++ K Y  DN+  A         +   D   +Y+      +  TG QTIDG + YF+ N  +
Sbjct  700   VDDKTYTIDNEGVATEFTAKNQFVRDDNWNWYYYDTDGKLLTGRQTIDGVQLYFDKNGKQ  759
```

C. perfringens cell wall binding repeat domain protein

>ref|YP_694944.1| cell wall binding repeat-containing protein [Clostridium perfri
ATCC 13124]
gb|ABG83523.1| cell wall binding repeat domain protein [Clostridium perfringens
ATCC 13124]
Length=1557

GENE ID: 4203905 CPF_0490 | cell wall binding repeat-containing protein
[Clostridium perfringens ATCC 13124] (10 or fewer PubMed links)

Sort alignments for this subj
E value Score Percent ide
Query start position Subje

Score = 240 bits (613), Expect = 8e-63, Method: Compositional matrix adjust.
Identities = 270/972 (28%), Positives = 410/972 (42%), Gaps = 207/972 (21%)

```
Query  2     GWQTINGKKYYFD---INTGAALTSYKIING--KHFYFNNDG--VMQLGVFKGPDGFEYF  54
             GWQTI GK YYF+   ++TG     Y   NG  K ++FN DG  V   G+    D  E +
Sbjct  246   GWQTIGGKTYYFENGKVSTGKKEV-YDTKNGYYKTYFFNEDGTLVTDTGIH---DYCESW  301

Query  55    APANTQNNNIEGQAIVYQSKFLTLNGKKYYFDNNSKAVTGWRIINN----EKYYFNPNNA  110
                    N +G+    ++ + T++GK YYFD  +  +T    I+N    E Y F+ +
Sbjct  302   GGKRKVYVNNKGEV---ENGWKTIDGKTYYFDQYNGMLTNVHEIDNGDKKEAYLFDNDGV  358

Query  111   I-AAVGLQVIDNNKYYFNPDTAIISKGWQTVNGSRYYFDTDTAIAFNGYKTIDGKHFYFD  169
             +    GL+ ID   YYFN D ++ S GW+T++G  YYFDT +  A    +  D  H  +
Sbjct  359   LRKGTGLKEIDGKWYYFNKDNSLAS-GWKTIDGKTYYFDTYSGRAKGSIRIYDSNHEKY-  416

Query  170   SDCVVKIGVFSTSNGFEYFAPANTYNNNIEGQAIVY-------QSKFLTLNGKKYYFDNN  222
                 K+ +F+          +TY+ N  G+  +Y        QS +  ++GK YYFD +
Sbjct  417   -----KVYLFNEDGALITEPGIHTYHENWGGERKIYVNNKGEVQSGWQIIDGKTYYFDES  471

Query  223   SKAVTGLQTIDS----KKYYFNTN-TAEAATGWQTIDGKKYYFNTNTAEAATGWQTIDGK  277
             +  VT +  I+      K Y F+ +       +GW+ I+G  YY N + +    GW+TID +
Sbjct  472   NGMVTWVHEINENDKNKSYLFDKDGVLVKGSGWKEINGNWYYLNNDNS-LLEGWKTIDDR  530

Query  278   KYYFNTNTAIASTGYTIING---KHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANNIEG  334
              YY +    + +  Y + +G    + + FN DG +  G
Sbjct  531   TYYLDKYNGMVNGVYEVNSGDKKETYLFNKDGSLVKG-----------------------  567

Query  335   QAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCT-----INN  389
                   N    +NG  YYF SD+    GW+II+ K YYFN  +    +  C       +N+
Sbjct  568   ------NGLKEVNGTWYYFNSDNSLENGWKIIDGKTYYFNKYDG--RLRGCVRVYDDLNH  619

Query  390   DKY---YFSYDGIL----------------------------QNGYITIERNNFYFDANNE  419
             +KY   +F+ DG+L                            Q+G+  I+    +FD +N
Sbjct  620   EKYKVYFFNEDGVLITEPGIHTYHETWGGERKICINNKGEVQSGWQIIDGKTYYFDEHN-  678

Query  420   SKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQNKFLTLNGKKYYFDNDSKAVTGWQT  479
```

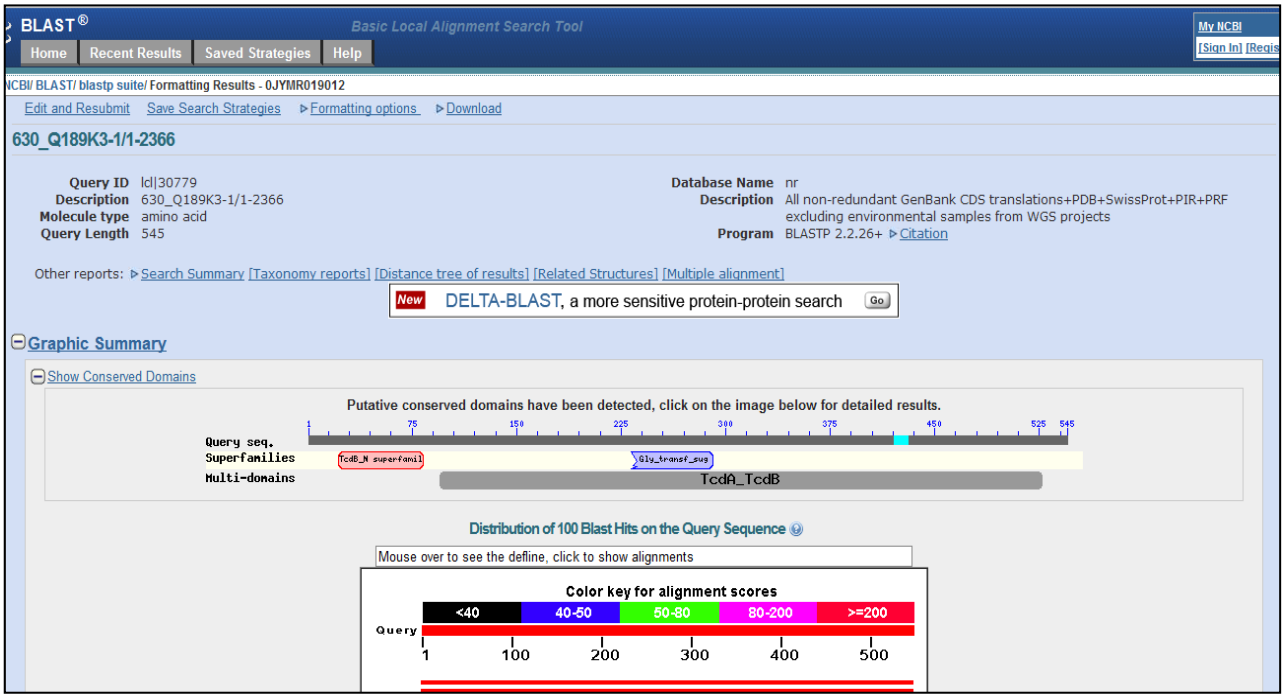
8.4.2. Toxin B protein BLAST alignments

[http://blast.ncbi.nlm.nih.gov/Blast.cgi]

8.4.2.1. Glucosyltransferase domain:

Analysis of the glucosyltransferase domain of *C. difficile* toxin B using protein BLAST revealed the following alignment:

Query coverage= 545 amino acids.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
EHJ40398.1	cell wall-binding repeat protein [Clostridium difficile 70-100-2010]	1104	1104	100%	0.0	100%
ZP_05328744.1	toxin B [Clostridium difficile QCD-63q42]	1104	1104	100%	0.0	100%
YP_001087135.1	tcdB gene product [Clostridium difficile 630] >ref|ZP_05349824.1|	1104	1104	100%	0.0	100%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	1087	1087	100%	0.0	98%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	1087	1087	100%	0.0	98%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	1075	1075	100%	0.0	97%
CAC79962.1	toxin B [[Clostridium] difficile]	881	881	100%	0.0	79%
CAC41640.1	TcdB-C34 Cluster1-2 [[Clostridium] difficile]	879	879	100%	0.0	79%
2VKD_A	Chain A, Crystal Structure Of The Catalytic Domain Of Lethal Toxir	843	843	100%	0.0	75%
EHJ31817.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	884	884	100%	0.0	79%
CAC19891.1	toxin B [[Clostridium] difficile]	884	884	100%	0.0	79%
AAG18011.1	cytotoxin B [[Clostridium] difficile]	882	882	100%	0.0	79%
CAA80815.1	toxin B [[Clostridium] difficile]	881	881	100%	0.0	79%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	838	838	100%	0.0	75%
AAC08437.1	truncated toxin A [[Clostridium] difficile]	560	560	100%	0.0	51%
CAC19893.1	truncated toxin A [[Clostridium] difficile]	557	557	100%	0.0	51%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	556	556	100%	1e-175	51%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	555	555	100%	1e-175	51%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	555	555	100%	3e-175	50%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	554	554	100%	5e-175	50%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	555	555	100%	1e-174	51%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	554	554	100%	4e-174	51%
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	554	554	100%	9e-174	51%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| T	554	554	100%	9e-174	51%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	554	554	100%	9e-174	51%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	553	553	100%	2e-173	51%

ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	554	554	100%	9e-174	51%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	553	553	100%	2e-173	51%
AFN52237.1	TcdA [[Clostridium] difficile]	552	552	100%	4e-173	50%
BAF46125.1	TpeL [Clostridium perfringens]	437	437	100%	4e-134	43%
ACF49258.1	TpeL [Clostridium perfringens A]	436	436	100%	1e-133	44%
2BVL_A	Chain A, Crystal Structure Of The Catalytic Domain Of Toxin B Fro	1095	1095	99%	0.0	99%
3SRZ_A	Chain A, Clostridium Difficile Toxin A (Tcda) Glucolsyltransferase D	556	556	99%	0.0	51%
CAA71690.1	TcdA protein [[Clostridium] difficile]	553	553	99%	0.0	50%
2BVM_A	Chain A, Crystal Structure Of The Catalytic Domain Of Toxin B Fro	1093	1093	99%	0.0	99%
4DMV_A	Chain A, Crystal Structure Of The Gt Domain Of Clostridium Difficile	555	555	99%	0.0	51%
CAC03681.1	toxin A [[Clostridium] difficile]	547	547	99%	0.0	51%
2VK9_A	Chain A, Crystal Structure Of The Catalytic Domain Of Alpha-Toxin	278	278	99%	8e-83	34%
CAA88565.1	alpha-toxin [Clostridium novyi]	277	277	99%	2e-77	34%
ZP_02865634.1	TcpA [Clostridium perfringens C str. JGS1495] >gb|EDS79391.1| T	434	434	96%	1e-132	44%
ZP_02636804.1	TpeL [Clostridium perfringens B str. ATCC 3626] >gb|EDT23010.1|	434	434	96%	1e-132	44%
ADH94629.1	TcdB [[Clostridium] difficile]	1064	1064	96%	0.0	100%
ADH94632.1	TcdB [[Clostridium] difficile]	1046	1046	96%	0.0	98%
ADH94633.1	TcdB [[Clostridium] difficile]	1045	1045	96%	0.0	98%
ADH94631.1	TcdB [[Clostridium] difficile]	1044	1044	96%	0.0	98%
ADH94627.1	TcdB [[Clostridium] difficile]	1044	1044	96%	0.0	98%
ADH94635.1	TcdB [[Clostridium] difficile]	1036	1036	96%	0.0	97%
ADH94634.1	TcdB [[Clostridium] difficile]	1036	1036	96%	0.0	97%
ADH94624.1	TcdB [[Clostridium] difficile]	1036	1036	96%	0.0	97%
ADH94636.1	TcdB [[Clostridium] difficile]	1035	1035	96%	0.0	97%
ADH94630.1	TcdB [[Clostridium] difficile]	850	850	96%	0.0	79%
ADH94626.1	TcdB [[Clostridium] difficile]	849	849	96%	0.0	79%
ADH94623.1	TcdB [[Clostridium] difficile]	848	848	96%	0.0	79%
ADH94628.1	TcdB [[Clostridium] difficile]	846	846	96%	0.0	79%
ADH94625.1	TcdB [[Clostridium] difficile]	846	846	96%	0.0	79%
YP_346342.1	hypothetical protein [Pseudomonas fluorescens Pf0-1] >gb|ABA72	88.2	88.2	94%	2e-15	22%
YP_346341.1	hypothetical protein [Pseudomonas fluorescens Pf0-1] >gb|ABA72	93.2	93.2	84%	8e-17	22%
YP_004351772.1	hypothetical protein PSEBR_a620 [Pseudomonas brassicacearum su	105	105	82%	1e-20	21%

ADH94625.1	TcdB [[Clostridium] difficile]	846	846	96%	0.0	79%
YP_346342.1	hypothetical protein [Pseudomonas fluorescens Pf0-1] >gb|ABA72	88.2	88.2	94%	2e-15	22%
YP_346341.1	hypothetical protein [Pseudomonas fluorescens Pf0-1] >gb|ABA72	93.2	93.2	84%	8e-17	22%
YP_004351772.1	hypothetical protein PSEBR_a620 [Pseudomonas brassicacearum su	105	105	82%	1e-20	21%
EIK69977.1	putative toxin protein [Pseudomonas fluorescens Q8r1-96]	103	103	82%	3e-20	21%
YP_005206075.1	cytotoxin mcf [Pseudomonas fluorescens F113] >gb|AEV60680.1|	100	100	82%	2e-19	21%
EGY29542.1	Glycosyltransferase [Candidatus Regiella insecticola R5.15]	93.2	93.2	79%	7e-17	24%
EGY28465.1	Peptidase C80 protein [Candidatus Regiella insecticola R5.15]	90.5	90.5	71%	5e-16	23%
EHV09211.1	glycosyltransferase sugar-binding region containing DXD motif fami	89.0	89.0	69%	1e-15	25%
EIO58881.1	tcdA/TcdB catalytic glycosyltransferase domain protein [Escherich	88.6	88.6	69%	1e-15	25%
ZP_03263719.1	toxin B [Escherichia coli O157:H7 str. EC4042] >gb|EDZ84682.1| t	88.6	88.6	69%	2e-15	25%
ZP_02802532.1	toxin B [Escherichia coli O157:H7 str. EC4196] >gb|EDU31090.1| t	87.8	87.8	69%	4e-15	25%
EFW63331.1	toxin B [Escherichia coli O157:H7 str. EC1212] >gb|EIP62080.1| to	87.8	87.8	69%	4e-15	25%
YP_325655.1	putative cytotoxin [Escherichia coli O157:H7 EDL933] >ref|ZP_027	87.8	87.8	69%	4e-15	25%
NP_052665.1	toxin B [Escherichia coli O157:H7 str. Sakai] >dbj|BAA31815.1| To	87.8	87.8	69%	4e-15	25%
ZP_02815165.1	toxin B [Escherichia coli O157:H7 str. EC869] >ref|ZP_05950388.1	87.8	87.8	69%	4e-15	25%
ZP_03085363.1	toxin B [Escherichia coli O157:H7 str. EC4024]	87.4	87.4	69%	4e-15	25%
EII85225.1	PF11996 family protein [Escherichia coli 3003]	75.9	75.9	69%	2e-11	25%
YP_003377848.1	toxin B [Escherichia coli O26:H-] >gb|ADB20445.1| toxin B [Esche	83.2	83.2	52%	9e-14	29%
YP_003237860.1	putative adherence factor, Efa1 homolog [Escherichia coli O26:H1	83.2	83.2	52%	1e-13	29%
YP_002756600.1	toxin B [Escherichia coli] >gb|ACL51990.1| toxin B [Escherichia co	83.2	83.2	52%	1e-13	29%
EII08397.1	PF11996 family protein [Escherichia coli 5.0959]	80.9	80.9	52%	5e-13	29%
EJE84889.1	toxin B, partial [Escherichia coli O26:H11 str. CVM10021]	85.9	85.9	49%	1e-14	30%
EHW81443.1	glycosyltransferase sugar-binding region containing DXD motif fami	85.1	85.1	49%	2e-14	30%
EHJ31823.1	hypothetical protein HMPREF1123_00959 [Clostridium difficile 050-1	275	275	47%	4e-84	52%
EGY28951.1	Cytotoxin L/B [Candidatus Regiella insecticola R5.15]	75.1	75.1	46%	4e-12	28%
ZP_05382378.1	hypothetical protein CtraD_00865 [Chlamydia trachomatis D(s)292	76.3	76.3	46%	7e-12	24%
YP_006361849.1	putative cytoadherence factor (fragment) [Chlamydia trachomatis	76.3	76.3	46%	9e-12	24%
NP_219669.1	hypothetical protein CT166 [Chlamydia trachomatis D/UW-3/CX] >	75.9	75.9	46%	1e-11	24%
ZP_05353533.1	hypothetical protein Ctra62_00850 [Chlamydia trachomatis 6276] >	75.5	75.5	46%	1e-11	24%
YP_005808436.1	hypothetical protein [Chlamydia trachomatis G/11222] >gb|ADH18	75.9	75.9	46%	1e-11	24%
YP_005815795.1	putative cytoadherence factor (fragment) [Chlamydia trachomatis	75.1	75.1	46%	2e-11	24%

Alignment information for genes shown in Table 4.2:

C. difficile tcdA

```
>emb|CAC03681.1|  toxin A [[Clostridium] difficile]
Length=553

 Score =  547 bits (1410),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 274/542 (51%), Positives = 385/542 (71%), Gaps = 2/542 (0%)

Query  1    MSLVNRKQLEKMANVRFRTQEDEYVAILDALEEYHNMSENTVVEKYLKLKDINSLTDIYI  60
            MS +++K+L K+A    R +E+EY  IL  L+EY+ ++ N    KYL+LK +N   D+++
Sbjct  1    MSSISKKELIKLA-YSVRPRENEYKTIILTNLDEYNKLTTNNNENKYLQLKKLNESIDVFM  59

Query  61   DTYKKSGRNKALKKFKEYLVTEVLELKNNNLTPVEKNLHFVWIGGQINDTAINYINQWKD  120
            + YK S RN+AL   K+ ++ EV+ +KN+N +PVEKNLHFVWIGG+++D A+ YI QW D
Sbjct  60   NKYKNSSRNRALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD  119

Query  121  VNSDYNVNVFYDSNAFLINTLKKTVVESAINDTLESFRENLNDPRFDYNKFFRKRMEIIY  180
            +N++YNV ++YDS AFL+NTLKK +VES+  + L+   E + +P+FD  KF++KRME IY
Sbjct  120  INAEYNVKLWYDSEAFLVNTLKKAIVESSTTEALQLLEEEIQNPQFDM-KFYKKRMEFIY  178

Query  181  DKQKNFINYYKAQREENPELIIDDIVKTYLSNEYSKEIDELNTYIEESLNKITQNSGNDV  240
            D+QK FINYYK+Q  +      IDDI+K++L +EY+++   L +Y   SL KI  N G D+
Sbjct  179  DRQKRFINYYKSQINKPTVPTIDDIIKSHLVSEYNRDETLLESYRTNSLRKINSNHGIDI  238

Query  241  RNFEEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGGMYLDVDMLPGIQPDLFES  300
            R    F   E  N+Y QEL+ R NLAAASDI+R+ ALK  GG+YLDVDMLPGI  DLF++
Sbjct  239  RANSLFTEQELLNIYSQELLNRGNLAAASDIVRLLALKNFGGVYLDVDMLPGIHSDLFKT  298

Query  301  IEKPSSVTVDFWEMTKLEAIMKYKEYIPEYTSEHFDMLDEEVQSSFESVLASKSDKSEIF  360
            I +PSS+ +D WEM KLEAIMKYK+YI  YTSE+FD LD++++ +F+ ++ SKS+KSEIF
Sbjct  299  IPRPSSIGLDRWEMIKLEAIMKYKKYINNYTSENFDKLDQQLKDNFKLIIESKSEKSEIF  358

Query  361  SSLGDMEASPLEVKIAFNSKGIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPAISE  420
            S L ++  S LE+KIAF    +INQ LIS + SY +NL++Q++NRY+ LN  LNPAI
Sbjct  359  SKLENLNVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLNPAIES  418

Query  421  DNDFNTITNTFIDSIMAEANADNGRFMMELGKYLRVGFFPDVKTTINLSGPEAYAAAYQD  480
            DN+F  TT  F DS+   A A+N  F+ ++  YL+VGF P+ ++TI+LSGP AYA+AY D
Sbjct  419  DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSGPGAYASAYYD  478

Query  481  LLMFKEGSMNIHLIEADLRNFEISKTNISQSTEQEMASLWSFDDARAKAQFEEYKRNYFE  540
             +  +E ++   L  +DL  F+  + N+SQ TEQE+ SLWSFD A AK Q E Y R+Y
Sbjct  479  FINLQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQIERYVRDYTG  538

Query  541  GS  542
            GS
Sbjct  539  GS  540
```

C. sordellii tcsL

```
>emb|CAA57959.1|  cytotoxin L [[Clostridium] sordellii]
Length=2364

 Score =  838 bits (2164),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 408/545 (75%), Positives = 473/545 (87%), Gaps = 0/545 (0%)

Query  1    MSLVNRKQLEKMANVRFRTQEDEYVAILDALEEYHNMSENTVVEKYLKLKDINSLTDIYI  60
            M+LVN+ QL+KM  V+FR QEDEYVAIL+ALEEYHNMSE++VVEKYLKLKDIN+LTD Y+
Sbjct  1    MNLVNKAQLQKMVYVKFRIQEDEYVAILNALEEYHNMSESSVVEKYLKLKDINNLTDNYL  60

Query  61   DTYKKSGRNKALKKFKEYLVTEVLELKNNNLTPVEKNLHFVWIGGQINDTAINYINQWKD  120
            +TYKKSGRNKALKKFKEYL  EVLELKNN+LTPVEKNLHF+WIGGQINDTAINYINQWKD
Sbjct  61   NTYKKSGRNKALKKFKEYLTMEVLELKNNSLTPVEKNLHFIWIGGQINDTAINYINQWKD  120

Query  121  VNSDYNVNVFYDSNAFLINTLKKTVVESAINDTLESFRENLNDPRFDYNKFFRKRMEIIY  180
            VNSDY V VFYDSNAFLINTLKKT+VESA N+TLESFRENLNDP FDYNKF+RKRMEIIY
Sbjct  121  VNSDYTVKVFYDSNAFLINTLKKTIVESATNNTLESFRENLNDPEFDYNKFYRKRMEIIY  180

Query  181  DKQKNFINYYKAQREENPELIIDDIVKTYLSNEYSKEIDELNTYIEESLNKITQNSGNDV  240
            DKQK+FI+YYK+Q EENPE IID+I+KTYLSNEYSK+++ LN YIEESLNKIT N+GND+
Sbjct  181  DKQKHFIDYYKSQIEENPEFIIDNIIKTYLSNEYSKDLEALNKYIEESLNKITANNGNDI  240

Query  241  RNFEEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGGMYLDVDMLPGIQPDLFES  300
            RN E+F + +   LY QELVERWNLAAASDILRIS LKE GG+YLDVD+LPGIQPDLF+S
Sbjct  241  RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQPDLFKS  300

Query  301  IEKPSSVTVDFWEMTKLEAIMKYKEYIPEYTSEHFDMLDEEVQSSFESVLASKSDKSEIF  360
            I KP S+T   WEM KLEAIMKYKEYIP YTS++FDMLDEEVQ SFES L+SKSDKSEIF
Sbjct  301  INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLDEEVQRSFESALSSKSDKSEIF  360

Query  361  SSLGDMEASPLEVKIAFNSKGIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPAISE  420
              L D++ SPLEVKIAF +  +INQ LIS+KDSYCS+L++ QI+NRYKILN++LNP+I+E
Sbjct  361  LPLDDIKVSPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKNRYKILNDNLNPSINE  420

Query  421  DNDFNTITNTFIDSIMAEANADNGRFMMELGKYLRVGFFPDVKTTINLSGPEAYAAAYQD  480
              DFNTT   F D + + +N DN  FM+++  YL+VGF PDV++TINLSGP  Y  AYQD
Sbjct  421  GTDFNTTMKIFSDKLASISNEDNMMFMIKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD  480

Query  481  LLMFKEGSMNIHLIEADLRNFEISKTNISQSTEQEMASLWSFDDARAKAQFEEYKRNYFE  540
            LLMFK+ S NIHL+E +LRNFE  KT ISQ TEQE+ SLWSF+ ARAK+QFEEYK+ YFE
Sbjct  481  LLMFKDNSTNIHLLEPELRNFEFPKTKISQLTEQEITSLWSFNQARAKSQFEEYKKGYFE  540

Query  541  GSLGE  545
            G+LGE
Sbjct  541  GALGE  545
```

C. perfringens tpeL

```
>dbj|BAF46125.1|  TpeL [Clostridium perfringens]
Length=1651

 Score =  437 bits (1123),  Expect = 4e-134, Method: Compositional matrix adjust.
 Identities = 238/548 (43%), Positives = 352/548 (64%), Gaps = 7/548 (1%)

Query  1    MSLVNRKQLEKMANVRFRTQEDEYVAILDALEEYH---NMSENTVVEKYLKLKDINSLTD  57
            M L++++QL  +A      +E EY  IL+ L+EY+   N  E   ++ YLKL +++   D
Sbjct  1    MGLMSKEQLIILAK-NSSPKEGEYKKILELLDEYNLLNNSVEKNSIDLYLKLNELSKSID  59

Query  58   IYIDTYKKSGRNKALKKFKEYLVTEVLELKNNNLTPVEKNLHFVWIGGQINDTAINYINQ  117
            IY+  YK S RN AL + K  L  EV+E+K+ NL P+EKN+HFVW+GG IN+ +I+YINQ
Sbjct  60   IYLKKYKNSKRNNALYQLKSDLTKEVIEIKDTNLKPLEKNIHFVWVGGMINNISIDYINQ  119

Query  118  WKDVNSDYNVNVFYDSNAFLINTLKKTVVESAINDTLESFRENLNDPRFDYNKFFRKRME  177
            WKD+NSDY   ++YDS A L+N LKK +++S+  + L  +   LND  FD NKF+R+RME
Sbjct  120  WKDINSDYETIIWYDSEALLVNILKKAIIDSSNKEVLTKYESVLNDNSFDSNKFYRERME  179

Query  178  IIYDKQKNFINYYKAQREENPELIIDDIVKTYLSNEYSKEIDELNTYIEESLNKITQNSG  237
            +I+ KQK F NYY         L  +D++K YL  +Y K  +EL  YI ES      N
Sbjct  180  VIFRKQKEFNNYYNTNDNYTKSL--NDVIKVYLIEKYLKTDEELEKYINESKEVFKANGA  237

Query  238  NDVRNFEEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGGMYLDVDMLPGIQPDL  297
             D+R ++   + E  ++YEQEL+ R+NLA+ASDI+R+  L ++GG+YLDVD+LPGI+  +
Sbjct  238  KDIREYDILDDVELKSIYEQELLMRFNLASASDIIRVIVLNKLGGIYLDVDVLPGIKKHI  297

Query  298  FESIEKPSSVTVDFWEMTKLEAIMKYKEYIPEYTSEHFDMLDEEVQSSFESVLASKSDKS  357
            F+ I KP++++ + W+M +LE IMKYK+YI  YT   F  L  ++Q   +  +  K+ KS
Sbjct  298  FKDINKPTNISENKWQMIQLETIMKYKQYIKGYTENSFKNLPSDLQEMLQEKVVEKNLKS  357

Query  358  EIFSSLGDMEASPLEVKIAFNSKGIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPA  417
            +IF  LGD+  S L+ KIAF    I NQ LIS K+SY  NLI+ QI+NRY I+N  L+ A
Sbjct  358  DIFQRLGDIFISELDTKIAFMFGKIANQVLISKKNSYSLNLIINQIKNRYNIINKCLSSA  417

Query  418  ISEDNDFNTTTNTFIDSIMAEANADNGRFMMELGKYLRVGFFPDVKTTINLSGPEAYAAA  477
            I + ++FN T + FI   + E   + G F+ ++  YL  G+ PD++ T+N+SGP  Y AA
Sbjct  418  IEKGSNFNNTVDIFIQQ-LNEFYVNEGFFVSKVMGYLGDGYMPDMRATLNISGPGIYTAA  476

Query  478  YQDLLMFKEGSMNIHLIEADLRNFEISKTNISQSTEQEMASLWSFDDARAKAQFEEYKRN  537
            Y DLL F E S+N  +++ DL+ FE+ +  ISQ TEQE+ S W+F+  +++ ++++
Sbjct  477  YYDLLYFNERSLNPQILQEDLKYFEVPQALISQQTEQEINSSWTFNQVKSQIEYKKLVEK  536

Query  538  YFEGSLGE  545
            Y   SL E
Sbjct  537  YTNKSLSE  544
```

C. novyi α toxin

```
>emb|CAA88565.1|  alpha-toxin [Clostridium novyi]
Length=2178

 Score =  277 bits (709),  Expect = 2e-77, Method: Compositional matrix adjust.
 Identities = 187/553 (34%), Positives = 297/553 (54%), Gaps = 18/553 (3%)

Query  3     LVNRKQLEKMANVRFRTQEDEYVAILDALEEYHNMSENTVV-EKYLKLKDINSLTDIYID  61
             L+ R+QL K+A++  + +E EY  ILDALE ++   E T V E Y KL  +N L D Y
Sbjct  2     LITREQLMKIASIPLKRKEPEYNLILDALENFNRDIEGTSVKEIYSKLSKLNELVDNYQT  61

Query  62    TYKKSGRNKALKKFKEYLVTEVLELKNNNLTP--VEKNLHFVWIGGQINDTAINYINQWK  119
              Y  SGRN AL+ F++ L +E+ EL  N+ T     KNL F+WIGG I+D ++ Y N WK
Sbjct  62    KYPSSGRNLALENFRDSLYSELRELIKNSRTSTIASKNLSFIWIGGPISDQSLEYYNMWK  121

Query  120   DVNSDYNVNVFYDSNAFLINTLKKTVVESAINDTLESFRENLNDPRFDYNKFFRKRMEII  179
              N DYN+ +FYD N+ L+NTLK  +++ +    +E  + N+ D  + +NKF+  RM++I
Sbjct  122   MFNKDYNIRLFYDKNSLLVNTLKTAIIQESSKVIIEQNQSNILDGTYGHNKFYSDRMKLI  181

Query  180   YDKQKNFINYYKAQREENPELIIDDIVKTYLSNEYSKEIDELNTYIEESLNKITQNSGND  239
             Y  ++     Y+  ++ N    +DDI+  +LSN +  +I +LN   E + NK+      D
Sbjct  182   YRYKRELKMLYENMKQNNS---VDDIIINFLSNYFKYDIGKLNNQKENNNNKMIAIGATD  238

Query  240   VRNFEEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGGMYLDVDMLPGIQPDLFE  299
             + N E     +  + Y QEL++   NLAAASDILRI+ LK+ GG+Y D+D LPG+   LF
Sbjct  239   I-NTENILTNKLKSYYYQELIQTNNLAAASDILRIAILKKYGGVYCDLDFLPGVNLSLFN  297

Query  300   SIEKPSSVTVDFWEMTKLEAIMKYKEYIPEYTSEHFDMLDEEVQSSFESVLASKSDKSEI  359
              I KP+ +  ++WE    EAI   K+ +  Y  ++ + +  E++    S      D +++
Sbjct  298   DISKPNGMDSNYWEAAIFEAIANEKKLMNNYPYKYMEQVPSEIKERILS-FVRNHDINDL  356

Query  360   FSSLGDMEASPLEV-----KIAFNSKGIINQGLISVKDSYCSNLIVKQIENRYKILNNSL  414
               LGD++ S LE+     K A   K    N  +IS  DS   N ++ Q+ENRY+ILN+ +
Sbjct  357   ILPLGDIKISQLEILLSRLKAATGKKTFSNAFIISNNDSLTLNNLISQLENRYEILNSII  416

Query  415   NPAI----SEDNDFNTTTNTFIDSIMAEANADNGRFMMELGKYLRVGFFPDVKTTINLSG  470
                  + D+  N+ +   +++     + D   F  ++  YL  GF P+V +T+  SG
Sbjct  417   QEKFKICETYDSYINSVSELVLETTPKNLSMDGSSFYQQIIGYLSSGFKPEVNSTVFFSG  476

Query  471   PEAYAAAYQDLLMFKEGSMNIHLIEADLRNFEISKTNISQSTEQEMASLWSFDDARAKAQ  530
             P  Y++A  D   F + + ++ L   +   FE S       T  E  S W      A+ +
Sbjct  477   PNIYSSATCDTYHFIKNTFDM-LSSQNQEIFEASNNLYFSKTHDEFKSSWLLRSNIAEKE  535

Query  531   FEEYKRNYFEGSL  543
             F++  + Y   +L
Sbjct  536   FQKLIKTYIGRIL  548
```

*C. perfringens tcp*A

```
>ref|ZP_02865634.1|  TcpA [Clostridium perfringens C str. JGS1495]
 gb|EDS79391.1|  TcpA [Clostridium perfringens C str. JGS1495]
Length=1776

 Score =  434 bits (1115),  Expect = 1e-132, Method: Compositional matrix adjust.
 Identities = 233/529 (44%), Positives = 342/529 (65%), Gaps = 6/529 (1%)

Query  20   QEDEYVAILDALEEYH---NMSENTVVEKYLKLKDINSLTDIYIDTYKKSGRNKALKKFK  76
            +E EY  IL+ L+EY+   N  E   ++ YLKL +++   DIY+  YK S RN AL + K
Sbjct  16   KEGEYKKILELLDEYNLLNNSVEKNSIDLYLKLNELSKSIDIYLKKYKNSKRNNALYQLK  75

Query  77   EYLVTEVLELKNNNLTPVEKNLHFVWIGGQINDTAINYINQWKDVNSDYNVNVFYDSNAF  136
              L  EV+E+K+ NL P+EKN+HFVW+GG IN+ +I+YINQWKD+NSDY   ++YDS A
Sbjct  76   SDLTKEVIEIKDTNLKPLEKNIHFVWVGGMINNISIDYINQWKDINSDYETIIWYDSEAL  135

Query  137  LINTLKKTVVESAINDTLESFRENLNDPRFDYNKFFRKRMEIIYDKQKNFINYYKAQREE  196
            L+N LKK +++S+  + L  +   LND  FD NKF+R+RME+I+ KQK F NYY
Sbjct  136  LVNILKKAIIDSSNKEVLTKYESVLNDNSFDSNKFYRERMEVIFRKQKEFNNYYNTNDNY  195

Query  197  NPELIIDDIVKTYLSNEYSKEIDELNTYIEESLNKITQNSGNDVRNFEEFKNGESFNLYE  256
               L  +D++K YL  +Y K  +EL  YI ES      N   D+R ++   + E  ++YE
Sbjct  196  TKSL--NDVIKVYLIEKYLKTDEELEKYINESKEVFKANGAKDIREYDILDDVELKSIYE  253

Query  257  QELVERWNLAAASDILRISALKEIGGMYLDVDMLPGIQPDLFESIEKPSSVTVDFWEMTK  316
            QEL+ R+NLA+ASDI+R+  L ++GG+YLDVD+LPGI+  +F+ I KP++++ + W+M +
Sbjct  254  QELLMRFNLASASDIIRVIVLNKLGGIYLDVDVLPGIKKHIFKDINKPTNISENKWQMIQ  313

Query  317  LEAIMKYKEYIPEYTSEHFDMLDEEVQSSFESVLASKSDKSEIFSSLGDMEASPLEVKIA  376
            LE IMKYK+YI  YT   F  L  ++Q   +  +  K+ KS+IF  LGD+  S L+ KIA
Sbjct  314  LETIMKYKQYIKGYTENSFKNLPSDLQEMLQEKVVEKNLKSDIFQRLGDIFISELDTKIA  373

Query  377  FNSKGIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPAISEDNDFNTTTNTFIDSIM  436
            F    I NQ LIS K+SY  NLI+ QI+NRY I+N  L+ AI + ++FN T + FI   +
Sbjct  374  FMFGKIANQVLISKKNSYSLNLIINQIKNRYNIINKCLSSAIEKGSNFNNTVDIFIQQ-L  432

Query  437  AEANADNGRFMMELGKYLRVGFFPDVKTTINLSGPEAYAAAYQDLLMFKEGSMNIHLIEA  496
             E   + G F+ ++  YL  G+ PD++ T+N+SGP  Y AAY DLL F E S+N  +++
Sbjct  433  NEFYVNEGFFVSKVMGYLGDGYMPDMRATLNISGPGIYTAAYYDLLYFNERSLNPQILQE  492

Query  497  DLRNFEISKTNISQSTEQEMASLWSFDDARAKAQFEEYKRNYFEGSLGE  545
            DL+ FE+ +  ISQ TEQE+ S W+F+  +++ ++++    Y   SL E
Sbjct  493  DLKYFEVPQALISQQTEQEINSSWTFNQVKSQIEYKKLVEKYTNKSLSE  541
```

Pseudomonas fluorescens Putative toxin A

```
>gb|EIK69977.1|  putative toxin protein [Pseudomonas fluorescens Q8r1-96]
Length=2360

 Score =  103 bits (257),  Expect = 4e-20, Method: Composition-based stats.
 Identities = 99/475 (21%), Positives = 202/475 (43%), Gaps = 42/475 (9%)

Query  92   TPVEKNLHFVWIGGQINDTAINYINQWKDV--NSDYNVNVFYDSNAFLINTLKKTVVESA  149
            T V K LHFVW+GG + +   +Y+N WK+V     Y++N++YDS+A L     + +VE+A
Sbjct  120  TEVPKTLHFVWLGGGLGNIQRDYLNVWKEVLARQGYSLNLWYDSDALLAWQTNRLIVEAA  179

Query  150  INDT-LESFRENLNDPRFDYNKFFRKRMEIIYDKQKNFINYYKAQREENPELIIDDIVKT  208
              D  L+   E +++   +    + +R  ++  +    I+          E  +D
Sbjct  180  KTDVFLQGVDERISE--VELGGLYNERTIVLKQQMHAHISAAVVGGGSADEARMD-----  232

Query  209  YLSNEYSKEIDELNTYIEESLNKITQNSGNDVRNFEEFKNGESF-NLYEQELVERWNLAA  267
             LS  Y +++  L   +E++   +    G  +R+           ++YE+E+  R N+AA
Sbjct  233  LLSRAYGQDVQVLRKQLEDNRRSVLDMDGFKLRDIATGDVSLQLQDVYEREMCLRGNMAA  292

Query  268  ASDILRISALKEIGGMYLDVDMLPGIQPDLFESIEKPSSVTVDFWEMTKLEAIMKY----  323
            ASD++R   L   GG Y DVD LP +   L        +V +  ++      +++
Sbjct  293  ASDVVRAEVLYAEGGSYTDVDHLPPLSQTL-------GAVDISGFDRNARLGVLQLLLNN  345

Query  324  -KEYIP--EYTSEHFDMLDEEVQSSFESVLASKSDKSEIFSSLGDMEASPLEVKIAFNSK  380
              E++P  + +S H+  +  E   + ++   S+   SE+F+   D  A P  ++     +
Sbjct  346  NPEWMPGRQASSSHYTHIGAEYFPALQAFAQSRPSLSEVFAQPADRLARPFALRALAMQQ  405

Query  381  GIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPAISEDNDFNTTTNT--FIDSIMAE  438
             + N  L++       N +++++   Y +++ S   A   D   +        ++ +  +
Sbjct  406  SLSNAFLMAHPGCAVLNSVIERLRANYALIDASTRLAAQRDIALSDVAGMLRLVEEVFEK  465

Query  439  ANADNGRFMME-----------LGKYLRVGFFPDVKTTINLSGPEAYAAAYQDLLMFKEG  487
             N       ME           +  +   G   + + TI L+GP A      D   + +
Sbjct  466  TNGPLTGLSMEDVPRAIALITAIATHFGDGIRFESEGTIYLTGPGAMRDGMVD---YAKA  522

Query  488  SMNIHLIEADLRNFEISKT-NISQSTEQEMASLWSFDDARAKAQFEEYKRNYFEG  541
             +   + E+  +   I+    ++ +TE+E    W  ++   +      + ++ EG
Sbjct  523  HLGAAIAESLSKEAAIAPNGTVNGATEEEQDHSWKENETDHEKWVSNEQTSWREG  577
```

Pseudomonas fluorescens mcf cytotoxin A

>ref|YP_005206075.1| cytotoxin mcf [Pseudomonas fluorescens F113]
gb|AEV60680.1| cytotoxin mcf [Pseudomonas fluorescens F113]
Length=2351

GENE ID: 11828730 PSF113_0644 | cytotoxin mcf [Pseudomonas fluorescens F113]

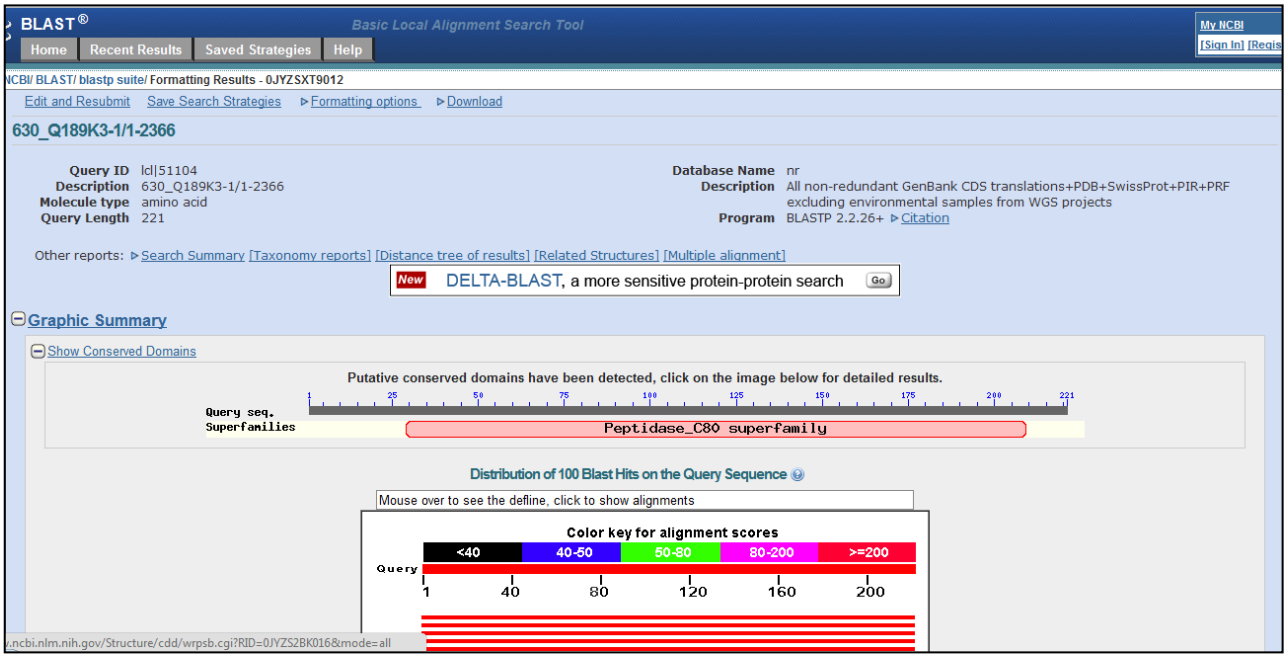
Score = 100 bits (250), Expect = 3e-19, Method: Composition-based stats.
Identities = 102/476 (21%), Positives = 203/476 (43%), Gaps = 48/476 (10%)

```
Query  94   VEKNLHFVWIGGQINDTAINYINQWKDV--NSDYNVNVFYDSNAFLINTLKKTVVESAIN  151
            V K LHFVW+GG I +   +Y++ W+ V     Y +N++YDS+A L     K +VE+A
Sbjct  122  VPKTLHFVWLGGGIGNIQRDYLHVWRQVLAQQGYTLNLWYDSDALLAWQTNKLIVEAAKA  181

Query  152  DT-LESFRENLNDPRFDYNKFFRKRMEIIYDKQKNFINYYKAQREENPELIIDDIVKTYL  210
            D  L+   E +++    +     + +R  ++  +    I+   A      E  ID      L
Sbjct  182  DVFLQGVDERISE--VELGALYNERAIVLRQQMHAHISAAVANGGSADEARID-----LL  234

Query  211  SNEYSKEIDELNTYIEESLNKITQNSGNDVRNFEEFKNGESF----NLYEQELVERWNLA  266
            S  Y +++  L   +E +    +G  +R   +   GE      ++Y++E+  R N+A
Sbjct  235  SRAYGQDVKVLRKQLEHNRRSVLDMNGFKLR---DLATGEVSLQLQDVYDREMCLRGNMA  291

Query  267  AASDILRISALKEIGGMYLDVDMLPGIQPDL----FESIEKPSSVTVDFWEMTKLEAIMK  322
            AASD++R+  L   GG Y DVD LP +   L        ++ + + V       L+ ++
Sbjct  292  AASDVVRVEVLYAEGGSYTDVDHLPPLSQTLGPVDISGFDRDARLGV-------LQLLLN  344

Query  323  YK-EYIP--EYTSEHFDMLDEEVQSSFESVLASKSDKSEIFSSLGDMEASPLEVKIAFNS  379
               E++P  + +S ++  +  E   + ++   S+   S++F    D  A P  ++
Sbjct  345  NNPEWMPGRQASSRYYAHIGAEHFPALQAFAQSRPSLSQVFGQPADRLARPFALRALAME  404

Query  380  KGIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPAISEDNDFNTTTNT--FIDSIMA  437
            + + N  L++   S   N ++++    Y +++ S   A   D   +        ++ +
Sbjct  405  QSLSNAFLMAHPGSAVLNSVIERFRANYALIDASTRLAAQRDIPLSDVAGMLRLVEEVFE  464

Query  438  EANADNGRFMME----------LGKYLRVGFFPDVKTTINLSGPEAYAAAYQDLLMFKE  486
            + N       ME          +  +   G   D + TI L+GP A      D   + +
Sbjct  465  KTNGPLTGLSMEDVPSAIALISAIATHFGDGIRFDSEGTIYLTGPGAMRDGMVD---YAK  521

Query  487  GSMNIHLIEADLRNFEISKT-NISQSTEQEMASLWSFDDARAKAQFEEYKRNYFEG  541
              ++  + E   +   I+    ++ +TE+E    W  ++   +      + ++ EG
Sbjct  522  AHLSAAVAETLRKEAAIAPNWTVNGATEEEQDHSWKENETDHEKWVSNEQTSWREG  577
```

8.4.2.2. Cysteine Protease domain:

Analysis of the cysteine protease domain of *C. difficile* toxin B revealed the following alignments:

Query coverage= 221 amino acids.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
3PA8_A	Chain A, Structure Of The C. Difficile Tcdb Cysteine Protease Dom	446	446	100%	7e-157	100%
EHJ40398.1	cell wall-binding repeat protein [Clostridium difficile 70-100-2010]	449	449	100%	3e-142	100%
ZP_05328744.1	toxin B [Clostridium difficile QCD-63q42]	449	449	100%	4e-142	100%
YP_001087135.1	tcdB gene product [Clostridium difficile 630] >ref|ZP_05349824.1|	449	449	100%	4e-142	100%
ADH94629.1	TcdB [[Clostridium] difficile]	448	448	100%	4e-142	100%
ADH94633.1	TcdB [[Clostridium] difficile]	444	444	100%	2e-140	99%
ADH94636.1	TcdB [[Clostridium] difficile]	443	443	100%	2e-140	99%
ADH94624.1	TcdB [[Clostridium] difficile]	441	441	100%	1e-139	98%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	441	441	100%	2e-139	97%
ADH94635.1	TcdB [[Clostridium] difficile]	441	441	100%	2e-139	97%
ADH94634.1	TcdB [[Clostridium] difficile]	441	441	100%	2e-139	97%
ADH94632.1	TcdB [[Clostridium] difficile]	436	436	100%	1e-137	96%
ADH94631.1	TcdB [[Clostridium] difficile]	436	436	100%	1e-137	96%
ADH94627.1	TcdB [[Clostridium] difficile]	436	436	100%	1e-137	96%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	436	436	100%	2e-137	96%
ADH94628.1	TcdB [[Clostridium] difficile]	434	434	100%	4e-137	96%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	434	434	100%	8e-137	96%
ADH94623.1	TcdB [[Clostridium] difficile]	430	430	100%	1e-135	95%
AAG18011.1	cytotoxin B [[Clostridium] difficile]	410	410	100%	1e-128	90%
EHJ31817.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	410	410	100%	1e-128	90%
ADH94625.1	TcdB [[Clostridium] difficile]	410	410	100%	2e-128	90%
ADH94630.1	TcdB [[Clostridium] difficile]	408	408	100%	6e-128	89%
ADH94626.1	TcdB [[Clostridium] difficile]	407	407	100%	1e-127	89%
CAC19891.1	toxin B [[Clostridium] difficile]	406	406	100%	3e-127	89%
CAA80815.1	toxin B [[Clostridium] difficile]	406	406	100%	4e-127	89%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	340	340	100%	4e-104	74%
3HO6_A	Chain A, Structure-Function Analysis Of Inositol Hexakisphosphate	270	270	100%	2e-87	57%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	270	270	100%	1e-79	57%

ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	270	270	100%	1e-79	57%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	270	270	100%	1e-79	57%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	270	270	100%	2e-79	57%
AFN52237.1	TcdA [[Clostridium] difficile]	270	270	100%	2e-79	57%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	270	270	100%	2e-79	57%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	270	270	100%	2e-79	57%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	269	269	100%	4e-79	57%
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	269	269	100%	6e-79	57%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	268	268	100%	6e-79	57%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	269	269	100%	6e-79	57%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| Tc	269	269	100%	6e-79	57%
ZP_02865634.1	TcpA [Clostridium perfringens C str. JGS1495] >gb|EDS79391.1| T	216	216	100%	6e-61	51%
ZP_02636804.1	TpeL [Clostridium perfringens B str. ATCC 3626] >gb|EDT23010.1|	216	216	100%	6e-61	51%
BAF46125.1	TpeL [Clostridium perfringens]	216	216	100%	7e-61	51%
ACF49258.1	TpeL [Clostridium perfringens A]	216	216	100%	9e-61	51%
CAA88565.1	alpha-toxin [Clostridium novyi]	127	127	98%	7e-30	36%
ZP_04642320.1	RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC 43969	78.6	78.6	93%	1e-13	27%
NP_930444.1	hypothetical protein plu3217 [Photorhabdus luminescens subsp. lau	45.4	45.4	92%	0.008	25%
ZP_04640347.1	Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC 439	74.7	74.7	84%	2e-12	29%
ZP_09365212.1	RTX repeat-containing cytotoxin [Vibrio ordalii ATCC 33509]	57.0	57.0	84%	1e-06	28%
ZP_00989505.1	Autotransporter adhesin [Vibrio splendidus 12B01] >gb|EAP95439.1	53.1	53.1	84%	2e-05	26%
3EEB_A	Chain A, Structure Of The V. Cholerae Rtx Cysteine Protease Doma	51.2	51.2	84%	3e-05	27%
EGQ99461.1	RTX toxin RtxA [Vibrio cholerae HE39]	52.4	52.4	84%	4e-05	27%
ZP_04961062.1	RTX toxin RtxA [Vibrio cholerae AM-19226] >gb|EDN15712.1| RTX	52.4	52.4	84%	4e-05	27%
EJH63941.1	RTX toxin RtxA [Vibrio cholerae HE-45]	52.4	52.4	84%	4e-05	27%
ZP_01982391.1	RTX toxin RtxA [Vibrio cholerae 623-39] >gb|EDL72938.1| RTX tox	52.4	52.4	84%	4e-05	27%
EJH62273.1	RTX toxin RtxA [Vibrio cholerae HE-25]	52.0	52.0	84%	4e-05	27%
ABX24512.1	hemolysin [Listonella anguillarum]	52.0	52.0	84%	5e-05	27%
YP_004191172.1	autotransporter adhesin [Vibrio vulnificus MO6-24/O] >gb|ADV889	52.0	52.0	84%	6e-05	27%
NP_762440.1	autotransporter adhesin [Vibrio vulnificus CMCP6] >gb|AAO07430.1	52.0	52.0	84%	6e-05	27%
NP_937086.1	RTX repeat-containing cytotoxin [Vibrio vulnificus YJ016] >dbj|BA	52.0	52.0	84%	6e-05	27%
EGR08431.1	RTX toxin RtxA [Vibrio cholerae HE48]	51.6	51.6	84%	7e-05	27%
EGS58021.1	RTX toxin RtxA domain protein [Vibrio cholerae HE-09]	50.8	50.8	84%	1e-04	26%
ZP_04410971.1	RTX toxin RtxA [Vibrio cholerae TM 11079-80] >gb|EEO06653.1| R	50.8	50.8	84%	1e-04	27%
ZP_04418869.1	RTX (Repeat in toxin) cytotoxin [Vibrio cholerae 12129(1)] >gb|EE	50.8	50.8	84%	1e-04	26%
EJH50736.1	RTX toxin RtxA domain protein [Vibrio cholerae HC-43B1]	50.4	50.4	84%	1e-04	26%
EGS69061.1	PGAP1-like family protein [Vibrio cholerae BJG-01]	50.4	50.4	84%	2e-04	26%
ZP_05419777.1	RTX toxins and related Ca2+-binding proteins [Vibrio cholera CIRS	48.9	48.9	84%	6e-04	27%
YP_005333215.1	RTX toxin RtxA [Vibrio cholerae IEC224] >gb|AFC58227.1| RTX tox	48.9	48.9	84%	6e-04	27%
ZP_01970588.1	RTX toxin RtxA [Vibrio cholerae NCTC 8457] >gb|EAZ74142.1| RTX	48.9	48.9	84%	6e-04	27%
ZP_01677789.1	RTX toxin RtxA [Vibrio cholerae 2740-80] >gb|EAX57825.1| RTX to	48.9	48.9	84%	6e-04	27%
AAD21057.1	RtxA protein [Vibrio cholerae]	48.9	48.9	84%	6e-04	27%
ZP_01979529.1	RTX toxin RtxA [Vibrio cholerae MZO-2] >gb|EDM53577.1| RTX tox	48.9	48.9	84%	6e-04	25%
ZP_05237921.1	RTX toxin RtxA [Vibrio cholerae MO10] >gb|EET22690.1| RTX toxin	48.9	48.9	84%	6e-04	27%
YP_002810174.1	RTX toxin RtxA [Vibrio cholerae M66-2] >ref|ZP_07008277.1| RTX	48.9	48.9	84%	6e-04	27%
NP_231094.1	RTX toxin RtxA [Vibrio cholerae O1 biovar El Tor str. N16961] >gb|	48.9	48.9	84%	6e-04	27%
ZP_01681473.1	RTX toxin RtxA [Vibrio cholerae V52] >gb|EAX61706.1| RTX toxin R	48.9	48.9	84%	6e-04	27%
ZP_01975348.1	RTX [Vibrio cholerae B33] >gb|EAZ77024.1| RTX [Vibrio cholerae B	48.9	48.9	84%	6e-04	27%
EJH31540.1	RTX toxin RtxA [Vibrio cholerae CP1041(14)]	48.9	48.9	84%	6e-04	27%
YP_005633914.1	RTX (Repeat in toxin) cytotoxin [Vibrio cholerae LMA3984-4] >gb|A	48.9	48.9	84%	6e-04	27%
ZP_01956295.1	RTX toxin RtxA [Vibrio cholerae MZO-3] >gb|EAY41476.1| RTX tox	48.5	48.5	84%	6e-04	27%
3FZY_A	Chain A, Crystal Structure Of Pre-Cleavage Form Of Cysteine Prote	47.4	47.4	84%	0.001	26%

3FZY_A	Chain A, Crystal Structure Of Pre-Cleavage Form Of Cysteine Prote	47.4	47.4	84%	0.001	26%
ZP_06028842.1	autotransporter adhesin [Vibrio cholerae INDRE 91/1] >gb|EEY491	47.8	47.8	84%	0.001	27%
ZP_04400836.1	autotransporter adhesin [Vibrio cholerae B33] >ref|ZP_04407919.1	47.4	47.4	84%	0.002	27%
ZP_04394484.1	RTX protein [Vibrio cholerae BX 330286] >gb|EEO22114.1| RTX pr	47.4	47.4	84%	0.002	27%
YP_001217005.1	RTX protein [Vibrio cholerae O395] >ref|ZP_06036949.1| RTX prot	47.4	47.4	84%	0.002	27%
EGS48622.1	PGAP1-like family protein [Vibrio cholerae HC-48A1] >gb|EGS49597	47.4	47.4	84%	0.002	26%
ZP_04405275.1	autotransporter adhesin [Vibrio cholerae TMA 21] >gb|EEO12545.1	47.4	47.4	84%	0.002	26%
ADV76432.1	RTX toxin RtxA [Vibrio vulnificus]	43.1	43.1	84%	0.046	27%
YP_004566017.1	FrpC [Vibrio anguillarum 775] >gb|AEH32975.1| FrpC [Vibrio anguill	53.9	53.9	83%	1e-05	28%
YP_003041903.1	rtx toxin RtxA [Photorhabdus asymbiotica subsp. asymbiotica ATC	52.8	52.8	83%	3e-05	27%
NP_928647.1	hypothetical protein plu1341 [Photorhabdus luminescens subsp. la	45.8	45.8	83%	0.006	25%
NP_928648.1	hypothetical protein plu1344 [Photorhabdus luminescens subsp. la	43.5	43.5	83%	0.030	26%
ADV76433.1	RTX toxin RtxA [Vibrio vulnificus]	42.7	42.7	83%	0.062	28%
YP_003041848.1	RTX toxin RtxA-like protein [Photorhabdus asymbiotica subsp. asyr	47.0	47.0	82%	0.002	24%
NP_930545.1	hypothetical protein plu3324 [Photorhabdus luminescens subsp. la	45.4	45.4	82%	0.007	27%
CBA72128.1	LRR-MCF ORF3 [Arsenophonus nasoniae]	43.5	43.5	79%	0.033	24%
AAC08437.1	truncated toxin A [[Clostridium] difficile]	159	159	68%	7e-42	50%
CAC19893.1	truncated toxin A [[Clostridium] difficile]	157	157	68%	3e-41	49%
CAA71690.1	TcdA protein [[Clostridium] difficile]	157	157	68%	3e-41	49%
ADX36388.1	putative RTX-toxin [Vibrio vulnificus]	42.7	42.7	55%	0.054	26%
EHJ31823.1	hypothetical protein HMPREF1123_00959 [Clostridium difficile 050-	79.3	79.3	33%	9e-15	46%
EHJ31824.1	hypothetical protein HMPREF1123_00960 [Clostridium difficile 050-	106	106	30%	3e-23	70%

Alignment information for genes shown in Table 4.2:

C. difficile tcdA

```
>gb|AFN52237.1|  TcdA [[Clostridium] difficile]
Length=2710

 Score =  270 bits (690),  Expect = 2e-79, Method: Compositional matrix adjust.
 Identities = 128/224 (57%), Positives = 170/224 (76%), Gaps = 3/224 (1%)

Query  1    DDNLDFSQNIVVDKEYLLE-KISS--LARSSERGYIHYIVQLQGDKISYEAACNLFAKTP  57
            D+ +DF++N  +DK YLL  KI S  +  +  + Y+HYI+QLQGD ISYEA CNLF+K P
Sbjct  545  DNEVDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNP  604

Query  58   YDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDI  117
             +S++ Q+N+ +S  +Y+ +       I E++KY+IP  + ++ K+K+TFIGHGKDEFNT
Sbjct  605  KNSIIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSE  664

Query  118  FAGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLKVKDK  177
            FA   VDSLS EI + +D  K DISPK++E+NLLGCNMFSY  NVEETYPGKLLL + DK
Sbjct  665  FARLSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDK  724

Query  178  ISELMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKEE  221
            I+  +P +++DSI + ANQYEVRINSEGR+ELL HSG+WINKEE
Sbjct  725  ITSTLPDVNKDSITIGANQYEVRINSEGRKELLAHSGKWINKEE  768


>ref|ZP_05270743.1|  toxin A [Clostridium difficile QCD-66c26]
 ref|ZP_05354979.1|  toxin A [Clostridium difficile QCD-76w55]
 ref|ZP_05383759.1|  toxin A [Clostridium difficile QCD-97b34]
 ▶6 more sequence titles
 Length=2710

 Score =  270 bits (690),  Expect = 2e-79, Method: Compositional matrix adjust.
 Identities = 128/224 (57%), Positives = 170/224 (76%), Gaps = 3/224 (1%)

Query  1    DDNLDFSQNIVVDKEYLLE-KISS--LARSSERGYIHYIVQLQGDKISYEAACNLFAKTP  57
            D+ +DF++N  +DK YLL  KI S  +  +  + Y+HYI+QLQGD ISYEA CNLF+K P
Sbjct  545  DNGVDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLFSKNP  604

Query  58   YDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDI  117
             +S++ Q+N+ +S  +Y+ +       I E++KY+IP  + ++ K+K+TFIGHGKDEFNT
Sbjct  605  KNSIIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEFNTSE  664

Query  118  FAGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLKVKDK  177
            FA   VDSLS EI + +D  K DISPK++E+NLLGCNMFSY  NVEETYPGKLLL + DK
Sbjct  665  FARLSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLSIMDK  724

Query  178  ISELMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKEE  221
            I+  +P +++DSI + ANQYEVRINSEGR+ELL HSG+WINKEE
Sbjct  725  ITSTLPDVNKDSITIGANQYEVRINSEGRKELLAHSGKWINKEE  768
```

C. sordellii tcsL

```
>emb|CAA57959.1|  cytotoxin L [[Clostridium] sordellii]
Length=2364

 Score =  340 bits (873),  Expect = 4e-104, Method: Compositional matrix adjust.
 Identities = 164/221 (74%), Positives = 190/221 (86%), Gaps = 0/221 (0%)

Query  1    DDNLDFSQNIVVDKEYLLEKISSLARSSERGYIHYIVQLQGDKISYEAACNLFAKTPYDS  60
            DDNLDF+QN V+DK+Y+ +KI S  ++  + YIHYIVQLQGDKISYEA+CNLF+K PY S
Sbjct  546  DDNLDFAQNTVLDKDYVSKKILSSMKTRNKEYIHYIVQLQGDKISYEASCNLFSKDPYSS  605

Query  61   VLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDIFAG  120
            +L+QKNIE SE AYYY   D EI+EIDKY+IP  IS++  IKLTFIGHGK EFNTD FA
Sbjct  606  ILYQKNIEGSETAYYYYVADAEIKEIDKYRIPYQISNKRNIKLTFIGHGKSEFNTDTFAN  665

Query  121  FDVDSLSTEIEAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLKVKDKISE  180
             DVDSLS+EIE  ++LAK DISPK IEINLLGCNMFSYSI+ EETYPGKLLLK+KD++SE
Sbjct  666  LDVDSLSSEIETILNLAKADISPKYIEINLLGCNMFSYSISAEETYPGKLLLKIKDRVSE  725

Query  181  LMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKEE  221
            LMPSISQDSI VSANQYEVRIN EG+RE+LDHSG+WINKEE
Sbjct  726  LMPSISQDSITVSANQYEVRINEEGKREILDHSGKWINKEE  766
```

C. perfringens tpeL

```
>ref|ZP_02636804.1|  TpeL [Clostridium perfringens B str. ATCC 3626]
 gb|EDT23010.1|  TpeL [Clostridium perfringens B str. ATCC 3626]
Length=1776

 Score =  216 bits (551),  Expect = 6e-61, Method: Compositional matrix adjust.
 Identities = 113/223 (51%), Positives = 155/223 (70%), Gaps = 5/223 (2%)

Query  1    DDNLDFSQNIVVDKEYLLEKISS--LARSSERGYIHYIVQLQGDKISYEAACNLFAKTPY  58
            +D L+F++N ++DK  LL +I+S  L    ++ Y+ YI+QLQGDK+SYEAA NLF K P
Sbjct  542  NDKLNFNENKIIDKVELLNRINSNNLINFDDKEYLRYIIQLQGDKVSYEAAINLFIKNPS  601

Query  59   DSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDIF  118
            +S+L Q   E + I+YY+N     I  I    IP I+  + KIKLTFIGHG++EFNT+ F
Sbjct  602  NSILVQ---EINNISYYFNSEYKSIDSIQFDNIPEILKGKNKIKLTFIGHGEEEFNTERF  658

Query  119  AGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLKVKDKI  178
            A   V   S +I   +D+ K + + K I+I+LLGCNMFSY+INVEETYPGKLL  V D +
Sbjct  659  ASLTVKEFSKKIYKVLDMIKSNTNVKEIQIDLLGCNMFSYNINVEETYPGKLLKVVLDYV  718

Query  179  SELMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKEE  221
             ++  +  +  I +SANQYEVRIN +G++ELL HSGEW++KEE
Sbjct  719  DKIYNADIKPEIKISANQYEVRINKDGKKELLSHSGEWLSKEE  761


>dbj|BAF46125.1|  TpeL [Clostridium perfringens]
Length=1651

 Score =  216 bits (550),  Expect = 7e-61, Method: Compositional matrix adjust.
 Identities = 113/223 (51%), Positives = 155/223 (70%), Gaps = 5/223 (2%)

Query  1    DDNLDFSQNIVVDKEYLLEKISS--LARSSERGYIHYIVQLQGDKISYEAACNLFAKTPY  58
            +D L+F++N ++DK  LL +I+S  L    ++ Y+ YI+QLQGDK+SYEAA NLF K P
Sbjct  545  NDKLNFNENKIIDKVELLNRINSNNLINFDDKEYLRYIIQLQGDKVSYEAAINLFIKNPS  604

Query  59   DSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDIF  118
            +S+L Q   E + I+YY+N     I  I    IP I+  + KIKLTFIGHG++EFNT+ F
Sbjct  605  NSILVQ---EINNISYYFNSEYKSIDSIQFDNIPEILKGKNKIKLTFIGHGEEEFNTERF  661

Query  119  AGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLKVKDKI  178
            A   V   S +I   +D+ K + + K I+I+LLGCNMFSY+INVEETYPGKLL  V D +
Sbjct  662  ASLTVKEFSKKIYKVLDMIKSNTNVKEIQIDLLGCNMFSYNINVEETYPGKLLKVVLDYV  721

Query  179  SELMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKEE  221
             ++  +  +  I +SANQYEVRIN +G++ELL HSGEW++KEE
Sbjct  722  DKIYNADIKPEIKISANQYEVRINKDGKKELLSHSGEWLSKEE  764
```

C. novyi α toxin

```
>emb|CAA88565.1|  alpha-toxin [Clostridium novyi]
Length=2178

 Score =  127 bits (318),  Expect = 7e-30, Method: Compositional matrix adjust.
 Identities = 80/225 (36%), Positives = 131/225 (58%), Gaps = 14/225 (6%)

Query  1    DDNLDFSQNIVVDKEYLLEKISSLARSS--ERGYIHYIVQLQGDKISYEAACNLFAKTPY  58
            +D L+F++   V    LL+ I  +  +   E   ++ I+Q+QGD ISYE+A N+F K P
Sbjct  551  EDGLNFNKWKRVTTSELLKVIEEVNSTKIYENYDLNMILQIQGDDISYESAVNVFGKNPN  610

Query  59   DSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDIF  118
             S+L Q  ++D    +Y+   +G +Q  +   I S  +D  KIKLT IGHG++ FN  +F
Sbjct  611  KSILIQ-GVDDFANVFYFE--NGIVQSDNINNILSRFNDIKKIKLTLIGHGENVFNPKLF  667

Query  119  AGFDVDSLSTEI-----EAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLK  173
             G  V+ L T I     +  ++     +  K ++IN+LGC MF+  +++  T+ GKL  K
Sbjct  668  GGKTVNDLYTNIIKPKLQHLLEREGVILKNKYLKINILGCYMFTPKVDINSTFVGKLFNK  727

Query  174  VKDKISELMP-SISQDSIIVSANQYEVRINSEGRRELLDHSGEWI  217
            +     +L P   S++ + +SAN+Y +RIN EG+RE+LD+ G+W+
Sbjct  728  IS---RDLQPKGFSKNQLEISANKYAIRINREGKREVLDYFGKWV  769
```

*C. perfringens tcp*A

```
>ref|ZP_02865634.1|  TcpA [Clostridium perfringens C str. JGS1495]
 gb|EDS79391.1|  TcpA [Clostridium perfringens C str. JGS1495]
Length=1776

 Score =  216 bits (551),  Expect = 6e-61, Method: Compositional matrix adjust.
 Identities = 113/223 (51%), Positives = 155/223 (70%), Gaps = 5/223 (2%)

Query  1    DDNLDFSQNIVVDKEYLLEKISS--LARSSERGYIHYIVQLQGDKISYEAACNLFAKTPY  58
            +D L+F++N ++DK  LL +I+S  L     ++ Y+ YI+QLQGDK+SYEAA NLF K P
Sbjct  542  NDKLNFNENKIIDKVELLNRINSNNLINFDDKEYLRYIIQLQGDKVSYEAAINLFIKNPS  601

Query  59   DSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDIF  118
            +S+L Q   E + I+YY+N     I  I    IP I+  + KIKLTFIGHG++EFNT+ F
Sbjct  602  NSILVQ---EINNISYYFNSEYKSIDSIQFDNIPEILKGKNKIKLTFIGHGEEEFNTERF  658

Query  119  AGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLKVKDKI  178
            A   V   S +I   +D+ K + + K I+I+LLGCNMFSY+INVEETYPGKLL  V D +
Sbjct  659  ASLTVKEFSKKIYKVLDMIKSNTNVKEIQIDLLGCNMFSYNINVEETYPGKLLKVVLDYV  718

Query  179  SELMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKEE  221
             ++  +  +  I +SANQYEVRIN +G++ELL HSGEW++KEE
Sbjct  719  DKIYNADIKPEIKISANQYEVRINKDGKKELLSHSGEWLSKEE  761


>ref|ZP_02636804.1|  TpeL [Clostridium perfringens B str. ATCC 3626]
 gb|EDT23010.1|  TpeL [Clostridium perfringens B str. ATCC 3626]
Length=1776

 Score =  216 bits (551),  Expect = 6e-61, Method: Compositional matrix adjust.
 Identities = 113/223 (51%), Positives = 155/223 (70%), Gaps = 5/223 (2%)

Query  1    DDNLDFSQNIVVDKEYLLEKISS--LARSSERGYIHYIVQLQGDKISYEAACNLFAKTPY  58
            +D L+F++N ++DK  LL +I+S  L     ++ Y+ YI+QLQGDK+SYEAA NLF K P
Sbjct  542  NDKLNFNENKIIDKVELLNRINSNNLINFDDKEYLRYIIQLQGDKVSYEAAINLFIKNPS  601

Query  59   DSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDIF  118
            +S+L Q   E + I+YY+N     I  I    IP I+  + KIKLTFIGHG++EFNT+ F
Sbjct  602  NSILVQ---EINNISYYFNSEYKSIDSIQFDNIPEILKGKNKIKLTFIGHGEEEFNTERF  658

Query  119  AGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLKVKDKI  178
            A   V   S +I   +D+ K + + K I+I+LLGCNMFSY+INVEETYPGKLL  V D +
Sbjct  659  ASLTVKEFSKKIYKVLDMIKSNTNVKEIQIDLLGCNMFSYNINVEETYPGKLLKVVLDYV  718

Query  179  SELMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKEE  221
             ++  +  +  I +SANQYEVRIN +G++ELL HSGEW++KEE
Sbjct  719  DKIYNADIKPEIKISANQYEVRINKDGKKELLSHSGEWLSKEE  761
```

Photorhabdus luminescens subsp. *laumondii* hypothetical protein

```
 GENE ID: 2801319 plu1341 | hypothetical protein
[Photorhabdus luminescens subsp. laumondii TTO1] (10 or fewer PubMed links)

 Score = 45.8 bits (107),  Expect = 0.006, Method: Composition-based stats.
 Identities = 46/186 (25%), Positives = 84/186 (45%), Gaps = 31/186 (17%)

Query  36    IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSII  95
             I+Q++ D I  +AA NL  K P  SV+ +    D++  Y+   GD           P+ +
Sbjct  2605  IIQMEDDPIVAKAAINLAGKHPDSSVVVKL---DADGKYHVIDGD-----------PAGL  2650

Query  96    SDRPKIKLTFIGHGKDEF--NTDIFAGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGC  153
             S   K++   +GHG+DE   N    +G+  D L+ +++      ++   P+ I I  +GC
Sbjct  2651  SG--KLRWQIVGHGRDESTQNNTRLSGYRADELAIKLKQFSQDFEQAGKPERISI--VGC  2706

Query  154   NMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEGRRELLDHS  213
             ++ S         Y     L  +   SE           VS  + +V +++ GR+   D +
Sbjct  2707  SLISDDKQNGFAYRFMFALDKQGIRSE-----------VSVRRSDVAVDATGRKFTRDKN  2755

Query  214   GEWINK  219
               +W+N+
Sbjct  2756  YQWVNR  2761


>ref|NP_930545.1|  hypothetical protein plu3324 [Photorhabdus luminescens subsp.
laumondii TTO1]
 emb|CAE15698.1|  unnamed protein product [Photorhabdus luminescens subsp. laumondii
TTO1]
Length=3531

 GENE ID: 2803336 plu3324 | hypothetical protein
[Photorhabdus luminescens subsp. laumondii TTO1] (10 or fewer PubMed links)

 Score = 45.4 bits (106),  Expect = 0.007, Method: Composition-based stats.
 Identities = 50/188 (27%), Positives = 89/188 (47%), Gaps = 33/188 (18%)

Query  36    IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSII  95
             I+Q++ D I+  AA NL  K P  SV+ Q +         Y     G++ ++
Sbjct  2463  ILQMEDDPIAARAAANLAGKHPDSSVVIQLDANGKYRVVY-----GDLAKLSN-------  2510

Query  96    SDRPKIKLTFIGHGKD--EFNTDIFAGFDVDSLSTEIEA---AIDLAKEDISPKSIEINL  150
               K++   +GHG+D  E N    +G+  D L+T+++     A  L K+ IS K   I+L
Sbjct  2511  ----KLRWQVVGHGRDTSEQNNIRLSGYSADELATKLKQFYQAAKLGKQAIS-KPDHISL  2565

Query  151   LGCNMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEGRRELL  210
             +GC++ S   N  + +  + ++  +        I  D   VSA    EV +++ GR+
Sbjct  2566  VGCSLISD--NKRDGFARRFITELDKQ------GIRSD---VSARSSEVAVDATGRKFTR  2614

Query  211   DHSGEWIN  218
```

Yersinia mollaretii RTX toxin

```
>ref|ZP_04642320.1|  RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC
43969]
 gb|EEQ09150.1|  RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC
43969]
Length=1998

 Score = 78.6 bits (192),  Expect = 1e-13, Method: Compositional matrix adjust.
 Identities = 58/213 (27%), Positives = 102/213 (48%), Gaps = 19/213 (9%)

Query  14   KEYLLEKISSLARSSERGY-IHYIVQLQGDKISYEAACNLFAKTPYDSVLFQ---KNIED  69
            KE+ L  +       S  GY    I+QLQGD+  +++A   FAK P  S   Q    NI D
Sbjct  533  KEHFLTLLDEGHVESSFGYEKQLIIQLQGDETCFKSAQAFFAKHPKQSEWLQLTDGNIND  592

Query  70   SEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDIFAGFDVDSLSTE  129
               + ++  D   Q  DK   P  +      +++  +GHG    ++  FAG      +
Sbjct  593  ---VFTWSTSDNHYQYTDK---PLALEKSGAVRIILVGHG----DSKQFAGMSAVQVGAT  642

Query  130  IEAAIDLAKEDISPK--SIEINLLGCNMFSYSINVEETYPGKLLLKVKDKISELMPSISQ  187
            +    D   +  + K  +I++N+ GC++F   + +E+T+PG+L + + ++       +
Sbjct  643  LGKVFDKISKQGATKFNNIKLNMAGCSLFDTYLPLEQTFPGQLAIWMNERAD--FWGLKP  700

Query  188  DSIIVSANQYEVRINSEGRRELLDHSGEWINKE  220
            + + V A +Y +R    G++E+    GEWINKE
Sbjct  701  EQLSVVAYEYPLRAVENGKKEIFV-KGEWINKE  732


>ref|ZP_04640347.1|  Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC
43969]
 gb|EEQ11141.1|  Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC
43969]
Length=1797

 Score = 74.7 bits (182),  Expect = 2e-12, Method: Compositional matrix adjust.
 Identities = 56/194 (29%), Positives = 99/194 (51%), Gaps = 16/194 (8%)

Query  34   HY----IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKY  89
            HY     I QLQGD   +EA+  LF K  Y S   Q  + D + A  +  G+    + +  Y
Sbjct  380  HYEKTLIFQLQGDDTCFEASRALFNKHRYTSEWLQ--LGDGKPAEVFTWGETYKKFV--Y  435

Query  90   KIPSIISDRPKIKLTFIGHGKDEFNTDIFAGFDVDSLSTEIEAAI-DLAKEDISPKSIEI  148
               P  +    KI++T +GHG+ E +T  F G + ++L   + +    L    +  K I +
Sbjct  436  TSPLKLDKEGKIRITLVGHGETEGDTTTFGGMNAETLKGHLSSLFARLGSSSVLIKGITL  495

Query  149  NLLGCNMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEGRRE  208
            NL GC++ +    + +T PG+L + +K + +E++   +   V+A + ++ +    G++E
Sbjct  496  NLTGCSLLNPKQPLADTLPGQLAIWLKQQ-AEIL-GLDDSNWSVNARENDLLVLENGKKE  553

Query  209  LL--DHSGEWINKE  220
            +   DH   WINKE
Sbjct  554  IRINDH---WINKE  564
```

P. asymbiotica subsp. asymbiotica RTX toxin

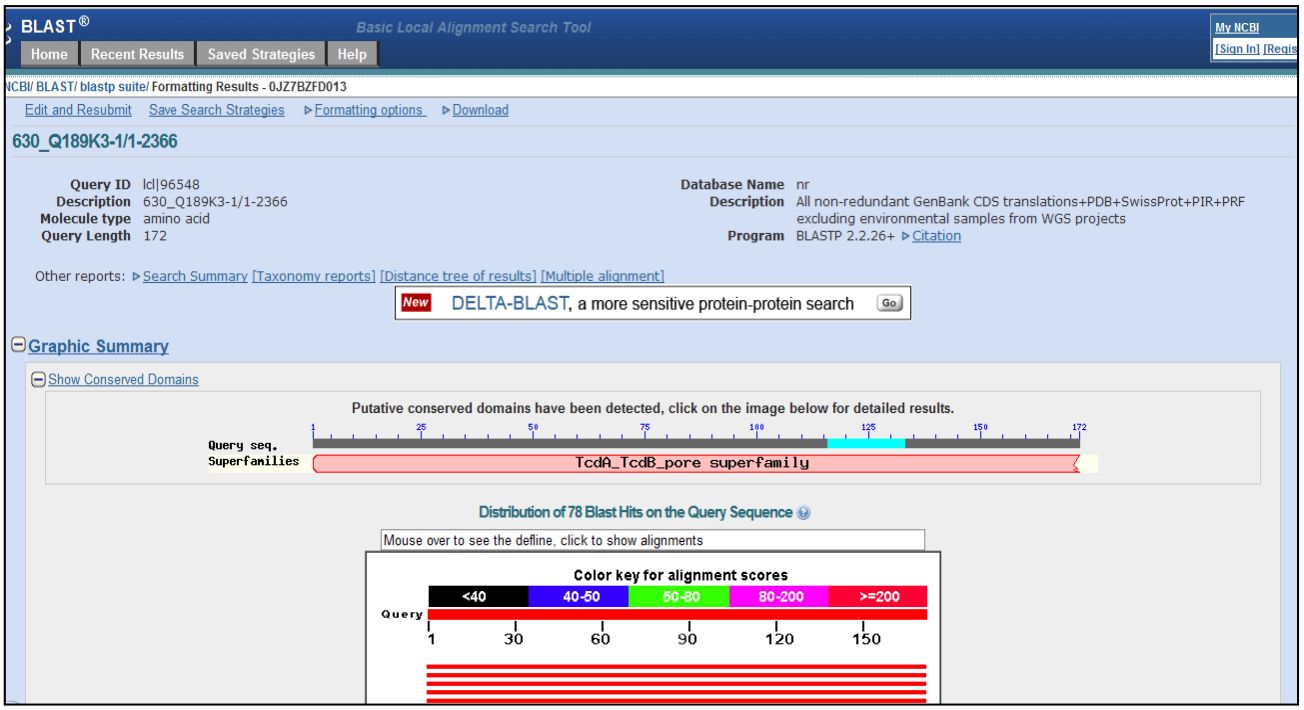
```
>ref|YP_003041903.1|  rtx toxin RtxA [Photorhabdus asymbiotica subsp. asymbiotica ATCC
43949]
 emb|CAR67497.1|  similar to rtx toxin rtxa [Photorhabdus asymbiotica subsp. asymbiotica
ATCC 43949]
 emb|CAQ85161.1|  similar to rtx toxin rtxa [Photorhabdus asymbiotica]
Length=4068

 GENE ID: 8167608 PAU_03073 | rtx toxin RtxA
[Photorhabdus asymbiotica subsp. asymbiotica ATCC 43949]
(10 or fewer PubMed links)

 Score = 52.8 bits (125),  Expect = 3e-05, Method: Composition-based stats.
 Identities = 51/186 (27%), Positives = 85/186 (46%), Gaps = 31/186 (17%)

Query  36    IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSII  95
             I+QL+ D I  +AA NL  K P  SV+ +    D++  Y+   GD           P  +
Sbjct  3001  IIQLENDPIVAKAAINLAGKHPDSSVVVKL---DADGKYHLIYGD-----------PDSL  3046

Query  96    SDRPKIKLTFIGHGKDE--FNTDIFAGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGC  153
             S   K++   +GHG+DE  +N    +G+  D L+ +++      ++   P  I I  +GC
Sbjct  3047  SG--KLRWQIVGHGRDESAYNNTRLSGYSADELAIKLKQFSQSFEQAGKPDRISI--VGC  3102

Query  154   NMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEGRRELLDHS  213
             ++ S        Y     +   DK       I  D   VSA + EV +++ GR+   D +
Sbjct  3103  SLISEDKRNGFAYR---FITALDK-----QGIRSD---VSARRSEVAVDATGRKFTRDKN  3151

Query  214   GEWINK  219
              +W+NK
Sbjct  3152  NQWVNK  3157
```

Vibrio cholerae RTX toxin

```
>gb|EGQ99461.1|  RTX toxin RtxA [Vibrio cholerae HE39]
Length=4545

 Score = 52.4 bits (124),  Expect = 4e-05, Method: Composition-based stats.
 Identities = 53/196 (27%), Positives = 87/196 (44%), Gaps = 43/196 (22%)

Query  36    IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSII  95
             IVQ++ D +  +AA NL  K P  SV+ Q     DS+  Y    GD           PS +
Sbjct  3463  IVQMENDDVVAKAAANLAGKHPESSVVVQL---DSDGNYRVVYGD-----------PSKL  3508

Query  96    SDRPKIKLTFIGHGKD--EFNTDIFAGFDVDSLSTEIEAAIDLAK--------EDISPKS  145
                 K++   +GHG+D  E N    +G+  D L      A+ LAK        E+I+ K
Sbjct  3509  DG--KLRWQLVGHGRDHSESNNTRLSGYSADEL------AVKLAKFQQSFNQAENINNKP  3560

Query  146   IEINLLGCNMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEG  205
               I+++GC++ S    + ++   G   +   D          +  + VS    E+ ++  G
Sbjct  3561  DHISIVGCSLVS---DDKQKGFGHQFINAMDA--------NGLRVDVSVRSSELTVDEAG  3609

Query  206   RRELLDHSGEWINKEE  221
             R+   D +G+W+ K E
Sbjct  3610  RKHTKDANGDWVQKAE  3625


>ref|ZP_04961062.1|  RTX toxin RtxA [Vibrio cholerae AM-19226]
 gb|EDN15712.1|  RTX toxin RtxA [Vibrio cholerae AM-19226]
Length=4558

 Score = 52.4 bits (124),  Expect = 4e-05, Method: Composition-based stats.
 Identities = 53/196 (27%), Positives = 87/196 (44%), Gaps = 43/196 (22%)

Query  36    IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSII  95
             IVQ++ D +  +AA NL  K P  SV+ Q     DS+  Y    GD           PS +
Sbjct  3476  IVQMENDDVVAKAAANLAGKHPESSVVVQL---DSDGNYRVVYGD-----------PSKL  3521

Query  96    SDRPKIKLTFIGHGKD--EFNTDIFAGFDVDSLSTEIEAAIDLAK--------EDISPKS  145
                 K++   +GHG+D  E N    +G+  D L      A+ LAK        E+I+ K
Sbjct  3522  DG--KLRWQLVGHGRDHSESNNTRLSGYSADEL------AVKLAKFQQSFNQAENINNKP  3573

Query  146   IEINLLGCNMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEG  205
               I+++GC++ S    + ++   G   +   D          +  + VS    E+ ++  G
Sbjct  3574  DHISIVGCSLVS---DDKQKGFGHQFINAMDA--------NGLRVDVSVRSSELAVDEAG  3622

Query  206   RRELLDHSGEWINKEE  221
             R+   D +G+W+ K E
Sbjct  3623  RKHTKDANGDWVQKAE  3638
```

8.4.2.3. Membrane translocation domain

Analysis of the membrane translocation domain of *C. difficile* toxin B revealed the following alignments:



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
ADH94625.1	TcdB [[Clostridium] difficile]	334	334	100%	1e-102	100%
ADH94629.1	TcdB [[Clostridium] difficile]	334	334	100%	1e-102	100%
EHJ40398.1	cell wall-binding repeat protein [Clostridium difficile 70-100-2010]	334	334	100%	1e-102	100%
YP_001087135.1	tcdB gene product [Clostridium difficile 630] >ref|ZP_05349824.1|	334	334	100%	1e-102	100%
EHJ31817.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	334	334	100%	1e-102	100%
CAA80815.1	toxin B [[Clostridium] difficile]	334	334	100%	1e-102	100%
ADH94636.1	TcdB [[Clostridium] difficile]	334	334	100%	1e-102	99%
ADH94634.1	TcdB [[Clostridium] difficile]	334	334	100%	1e-102	99%
ZP_05328744.1	toxin B [Clostridium difficile QCD-63q42]	333	333	100%	2e-102	99%
AAG18011.1	cytotoxin B [[Clostridium] difficile]	333	333	100%	2e-102	99%
ADH94628.1	TcdB [[Clostridium] difficile]	332	332	100%	6e-102	99%
ADH94633.1	TcdB [[Clostridium] difficile]	331	331	100%	1e-101	99%
ADH94623.1	TcdB [[Clostridium] difficile]	328	328	100%	2e-100	97%
ADH94630.1	TcdB [[Clostridium] difficile]	328	328	100%	2e-100	97%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	328	328	100%	2e-100	97%
ADH94635.1	TcdB [[Clostridium] difficile]	328	328	100%	2e-100	97%
ADH94624.1	TcdB [[Clostridium] difficile]	327	327	100%	4e-100	97%
ADH94632.1	TcdB [[Clostridium] difficile]	325	325	100%	1e-99	97%
ADH94631.1	TcdB [[Clostridium] difficile]	325	325	100%	1e-99	97%
ADH94627.1	TcdB [[Clostridium] difficile]	325	325	100%	1e-99	97%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	325	325	100%	1e-99	97%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	325	325	100%	2e-99	97%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	319	319	100%	3e-97	92%
ADH94626.1	TcdB [[Clostridium] difficile]	313	313	100%	2e-95	94%
CAC19891.1	toxin B [[Clostridium] difficile]	313	313	100%	2e-95	94%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	226	226	100%	7e-65	60%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	226	226	100%	7e-65	60%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	226	226	100%	9e-65	60%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	226	226	100%	9e-65	60%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| T	226	226	100%	9e-65	60%

ADH94631.1	TcdB [[Clostridium] difficile]	325	325	100%	1e-99	97%
ADH94627.1	TcdB [[Clostridium] difficile]	325	325	100%	1e-99	97%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	325	325	100%	1e-99	97%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	325	325	100%	2e-99	97%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	319	319	100%	3e-97	92%
ADH94626.1	TcdB [[Clostridium] difficile]	313	313	100%	2e-95	94%
CAC19891.1	toxin B [[Clostridium] difficile]	313	313	100%	2e-95	94%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	226	226	100%	7e-65	60%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	226	226	100%	7e-65	60%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	226	226	100%	9e-65	60%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	226	226	100%	9e-65	60%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| Tc	226	226	100%	9e-65	60%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	225	225	100%	9e-65	60%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	226	226	100%	1e-64	60%
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	226	226	100%	1e-64	60%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	225	225	100%	1e-64	60%
EHJ31824.1	hypothetical protein HMPREF1123_00960 [Clostridium difficile 050-	221	221	100%	1e-64	60%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	225	225	100%	1e-64	60%
AFN52237.1	TcdA [[Clostridium] difficile]	225	225	100%	1e-64	60%
ZP_02636804.1	TpeL [Clostridium perfringens B str. ATCC 3626] >gb|EDT23010.1|	153	153	100%	1e-39	44%
BAF46125.1	TpeL [Clostridium perfringens]	153	153	100%	1e-39	44%
ZP_02865634.1	TcpA [Clostridium perfringens C str. JGS1495] >gb|EDS79391.1| T	153	153	100%	2e-39	44%
ACF49258.1	TpeL [Clostridium perfringens A]	145	145	100%	6e-37	41%
YP_003042199.1	MCF toxin [Photorhabdus asymbiotica subsp. asymbiotica ATCC 43	48.5	48.5	100%	3e-04	24%
NP_931332.1	Mcf protein [Photorhabdus luminescens subsp. laumondii TTO1] >e	45.1	45.1	100%	0.005	24%
AAM88787.1	toxin protein [Photorhabdus luminescens]	43.1	43.1	100%	0.016	23%
YP_003468304.1	Mcf protein (fragment) [Xenorhabdus bovienii SS-2004] >emb|CBJ	50.8	50.8	98%	5e-05	24%
YP_003712501.1	hypothetical protein [Xenorhabdus nematophila ATCC 19061] >eml	50.1	50.1	98%	9e-05	26%
YP_260089.1	fitD gene product [Pseudomonas fluorescens Pf-5] >gb|AAY92255	48.5	48.5	98%	2e-04	26%
NP_930360.1	hypothetical protein plu3128 [Photorhabdus luminescens subsp. la	47.0	47.0	98%	8e-04	26%
ABY91230.1	FitD [Pseudomonas fluorescens]	45.1	45.1	98%	0.004	24%
AAR21118.1	Mcf2 [Photorhabdus luminescens]	40.4	40.4	98%	0.17	23%

AAR21118.1	Mcf2 [Photorhabdus luminescens]	40.4	40.4	98%	0.17	23%
ZP_01811976.1	calmodulin-sensitive adenylate cyclase [Vibrionales bacterium SWA	77.4	77.4	97%	2e-13	33%
CAA88565.1	alpha-toxin [Clostridium novyi]	116	116	96%	8e-27	37%
ZP_06125494.2	hypothetical protein PROVRETT_07543 [Providencia rettgeri DSM 1	42.4	42.4	96%	0.034	25%
ZP_04640347.1	Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC 439	60.8	60.8	93%	3e-08	29%
ADA82957.1	TpeL [Clostridium perfringens B]	135	135	91%	6e-34	42%
ADA82953.1	TpeL [Clostridium perfringens B]	135	135	91%	6e-34	42%
ZP_06125490.2	putative cysteine protease domain, YopT-type [Providencia rettge	39.3	39.3	88%	0.36	27%
ZP_04642320.1	RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC 4396	63.5	63.5	88%	4e-09	33%
ZP_03319418.1	hypothetical protein PROVALCAL_02362 [Providencia alcalifaciens l	50.4	50.4	88%	6e-05	29%
CBA72128.1	LRR-MCF ORF3 [Arsenophonus nasoniae]	40.4	40.4	87%	0.17	27%
ZP_10172701.1	cytotoxin FitD, partial [Pseudomonas chlororaphis O6] >gb|EIM181	45.4	45.4	77%	0.003	26%
YP_003712268.1	hypothetical protein [Xenorhabdus nematophila ATCC 19061] >eml	41.6	41.6	77%	0.068	27%
EHK88574.1	Rhs family protein [Saccharomonospora azurea SZMC 14600]	35.8	35.8	75%	4.9	25%
ZP_09835187.1	hypothetical protein SazuN_18080 [Saccharomonospora azurea NA	35.8	35.8	75%	4.9	25%
ZP_08737752.1	cytotoxin Mcf [Vibrio tubiashii ATCC 19109] >gb|EGU56431.1| cyt	43.1	43.1	73%	0.017	24%
EIF01430.1	cytotoxin FitD [Vibrio tubiashii NCIMB 1337 = ATCC 19106]	43.1	43.1	73%	0.018	24%
ZP_05884844.1	cytotoxin Mcf [Vibrio coralliilyticus ATCC BAA-450] >gb|EEX34533	53.9	53.9	73%	5e-06	31%
YP_003561334.1	putative ABC transporter substrate-binding protein [Bacillus megat	36.2	36.2	72%	3.4	29%
YP_006254759.1	heavy metal-translocating P-type ATPase [Solitalea canadensis DS	35.8	35.8	61%	5.8	29%
YP_003569985.1	zntA gene product [Salinibacter ruber M8] >emb|CBH23033.1| Cac	37.4	37.4	55%	1.5	34%

YP_006254759.1	heavy metal-translocating P-type ATPase [Solitalea canadensis DS	35.8	35.8	61%	5.8	29%
YP_003569985.1	zntA gene product [Salinibacter ruber M8] >emb|CBH23033.1| Cac	37.4	37.4	55%	1.5	34%
YP_444264.1	cadmium efflux ATPase [Salinibacter ruber DSM 13855] >gb|ABC45	37.4	37.4	55%	1.8	34%
XP_002469023.1	predicted protein [Postia placenta Mad-698-R] >gb|EED85728.1| p	37.7	37.7	52%	1.4	31%
ZP_09556685.1	hydrolase, NUDIX family [Lactobacillus kisonensis F0435] >gb|EHO	36.6	36.6	51%	1.6	33%
ZP_07743039.1	putative Flp pilus assembly protein TadC [Vibrio caribbenthicus AT	37.0	37.0	45%	1.9	35%
ZP_07109992.1	Glycosyl transferase [Oscillatoria sp. PCC 6506] >emb|CBN55142.1	36.6	36.6	44%	2.6	30%
YP_003428929.1	copper translocating P-type ATPase [Bacillus pseudofirmus OF4] >	35.4	35.4	39%	7.7	32%
YP_004732864.1	hypothetical protein WIV_gp081 [Wiseana iridescent virus] >gb|AD	35.8	35.8	35%	2.9	41%
ZP_07866892.1	conserved hypothetical protein [Capnocytophaga ochracea F0287	35.0	35.0	33%	9.3	34%

Alignment information for genes shown in Table 4.3:

C. difficile tcdA

```
>gb|AFN52237.1|  TcdA [[Clostridium] difficile]
Length=2710

 Score =  225 bits (574),  Expect = 1e-64, Method: Compositional matrix adjust.
 Identities = 104/172 (60%), Positives = 150/172 (87%), Gaps = 0/172 (0%)

Query  1     VNTLNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTAL  60
             VNTLNAAFFIQSLI+Y+S+K+ L++LS ++KVQ+YAQLFSTGLNTI D+ ++V L+S A+
Sbjct  958   VNTLNAAFFIQSLIDYSSNKDLLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNAV  1017

Query  61    DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTTAT  120
             ++TI++LPT++EG+PI++TI+DG++LGAAIKEL +  DPLL++E+EAK+G++A+N++ +
Sbjct  1018  NDTINVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLSI  1077

Query  121   TAIITSSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKHV  172
              A + S +GI +  +I L+P+AGISAGIPSLVNNEL+L DKAT VV+YF H+
Sbjct  1078  AATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNHL  1129
```

C. sordellii tcsL

```
>emb|CAA57959.1|  cytotoxin L [[Clostridium] sordellii]
Length=2364

 Score =  319 bits (817),  Expect = 3e-97, Method: Compositional matrix adjust.
 Identities = 159/172 (92%), Positives = 171/172 (99%), Gaps = 0/172 (0%)

Query  1     VNTLNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTAL  60
             VNTLN+AFFIQSLIEYN++KESLSNLSVAMKVQVYAQLFSTGLNTITDA+KVVELVSTAL
Sbjct  956   VNTLNSAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTAL  1015

Query  61    DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTTAT  120
             DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSET+DPLLRQEIEAKIGIMAVNLT A+
Sbjct  1016  DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQEIEAKIGIMAVNLTAAS  1075

Query  121   TAIITSSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKHV  172
             TAI+TS+LGIASGFSILLVPLAGISAGIPSLVNNEL+L+DKATKV+DYFKH+
Sbjct  1076  TAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILQDKATKVIDYFKHI  1127
```

C. perfringens tpeL

```
>ref|ZP_02636804.1|  TpeL [Clostridium perfringens B str. ATCC 3626]
 gb|EDT23010.1|  TpeL [Clostridium perfringens B str. ATCC 3626]
Length=1776

 Score =  153 bits (386),  Expect = 1e-39, Method: Compositional matrix adjust.
 Identities = 75/172 (44%), Positives = 121/172 (70%), Gaps = 0/172 (0%)

Query  1     VNTLNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTAL  60
             V+TLN +F IQS+I+Y +     + LS ++KVQ+Y Q+ +  L+ I DA+ +V++++ A
Sbjct  941   VSTLNTSFLIQSMIDYKAQNFDFNKLSTSVKVQIYCQITNISLSEIQDASNLVKIIAEAN  1000

Query  61    DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTTAT  120
             +  I+L+PTL+  +P+I TI+DG++L A I EL  T D LLR+E+ A+IGI++ N+T A
Sbjct  1001  EIEINLIPTLANAIPLITTIVDGINLIANIDELINTKDELLRKELAARIGIISSNMTAAI  1060

Query  121   TAIITSSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKHV  172
             ++ I         F+ LLVP+AGIS+GIP+LVNN L+L +K+ ++ +YF H+
Sbjct  1061  SSYILYFTEFGEVFNPLLVPIAGISSGIPTLVNNILILEEKSKEITEYFSHI  1112
```

C. novyi α toxin

```
>emb|CAA88565.1|  alpha-toxin [Clostridium novyi]
Length=2178

 Score =  116 bits (291),  Expect = 8e-27, Method: Compositional matrix adjust.
 Identities = 63/169 (37%), Positives = 106/169 (63%), Gaps = 3/169 (2%)

Query  4     LNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDET  63
             LN+A  +Q LI+Y    E L+N++ ++KVQ YAQ+F   +  I +A ++V ++S AL+
Sbjct  965   LNSAMLMQLLIDYKPYTEILTNMNTSLKVQAYAQIFQLSIGAIQEATEIVTIISDALNAN  1024

Query  64    IDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTTATTAI  123
              ++L  L  G  + + IIDG++L AA+ EL        R+ IEAK+G+ ++     ++++
Sbjct  1025  FNILSKLKVGSSVASVIIDGINLIAALTELKNVKTNFERKLIEAKVGMYSIGFILESSSL  1084

Query  124   ITSSLG---IASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYF  169
             I+  LG   ++    ++ VP+AGI  G+PSLVNN LVL +K  +++DYF
Sbjct  1085  ISGLLGATAVSEILGVISVPVAGILVGLPSLVNNILVLGEKYNQILDYF  1133
```

*C. perfringens tcp*A

```
>ref|ZP_02865634.1|  TcpA [Clostridium perfringens C str. JGS1495]
 gb|EDS79391.1|  TcpA [Clostridium perfringens C str. JGS1495]
Length=1776

 Score =  153 bits (386),  Expect = 2e-39, Method: Compositional matrix adjust.
 Identities = 75/172 (44%), Positives = 121/172 (70%), Gaps = 0/172 (0%)

Query  1     VNTLNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTAL  60
             V+TLN +F IQS+I+Y +     + LS ++KVQ+Y Q+ +  L+ I DA+ +V++++ A
Sbjct  941   VSTLNTSFLIQSMIDYKAQNFDFNKLSTSVKVQIYCQITNISLSEIQDASNLVKIIAEAN  1000

Query  61    DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTTAT  120
             +  I+L+PTL+  +P+I TI+DG++L A I EL  T D LLR+E+ A+IGI++ N+T A
Sbjct  1001  EIEINLIPTLANAIPLITTIVDGINLIANIDELINTKDELLRKELAARIGIISSNMTAAI  1060

Query  121   TAIITSSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKHV  172
             ++ I         F+ LLVP+AGIS+GIP+LVNN L+L +K+ ++ +YF H+
Sbjct  1061  SSYILYFTEFGEVFNPLLVPIAGISSGIPTLVNNILILEEKSKEITEYFSHI  1112
```

Photorhabdus luminescens subsp. *laumondii* hypothetical protein

```
>ref|NP_931332.1|  Mcf protein [Photorhabdus luminescens subsp. laumondii TTO1]
 emb|CAE16514.1|  Mcf protein [Photorhabdus luminescens subsp. laumondii TTO1]
Length=2997

 GENE ID: 2804155 mcf | Mcf protein
[Photorhabdus luminescens subsp. laumondii TTO1] (10 or fewer PubMed links)

 Score = 45.1 bits (105),  Expect = 0.005, Method: Composition-based stats.
 Identities = 46/195 (24%), Positives = 81/195 (42%), Gaps = 26/195 (13%)

Query  1     VNTLNAAFFIQSLIEYNSSKESLS---------NLSVAMKVQVYAQLFSTGLNTITDAAKV  52
             V+ LNA F +Q+LI++ + K             +L+ A+KV  Y  L       + D AKV
Sbjct  1710  VDGLNAGFAVQTLIQWFADKNRHDAASGVISPDLATALKVHSYLNLVQMAHGGVQDIAKV  1769

Query  53    VELVSTAL------------DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPL  100
               LV TAL            D   +L  T++EG  +   +  G  +G    EL+   + +
Sbjct  1770  TALVRTALRGEVVAAETSLKDFASNLGHTVNEGAGV---LFGGAMVGLDAYELAHAENDV  1826

Query  101   LRQEIEAKIGIMAVNLTTATTAIITSSLGIASGFSIL---LVPLAGISAGIPSLVNNELV  157
              +     ++   + +  T    +    +G ++  V L G  V L G+ G  +L
Sbjct  1827  QKAVFGTQLAFDSASFVTGAAGVGAGLVGASTAGAVLGGASVILGGLAVGFTALAQAFGA  1886

Query  158   LRDKATKVVDYFKHV  172
             + + A  V  YF  V
Sbjct  1887  VAEDAKAVGRYFDTV  1901
```

Photorhabdus luminescens subsp. *laumondii* hypothetical protein

```
 GENE ID: 2803139 plu3128 | hypothetical protein
[Photorhabdus luminescens subsp. laumondii TTO1] (10 or fewer PubMed links)

 Score = 47.0 bits (110),  Expect = 8e-04, Method: Compositional matrix adjust.
 Identities = 50/192 (26%), Positives = 83/192 (43%), Gaps = 26/192 (14%)

Query  1     VNTLNAAFFIQSLIEYNSSKE--------SLSNLSVAMKVQVYAQLFSTGLNTITDAAKV  52
             V+ LNA F IQSLI++ + K           +  +L+ A+KV  Y  L       + D  KV
Sbjct  1097  VDGLNAGFAIQSLIQWFTDKNRNDAARGIASPDLATALKVHSYLNLVQMAHGGVQDVTKV  1156

Query  53    VELVSTAL------------DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPL  100
               ELV TAL            D T  L  T++EG  +   +  G  +G    EL+   + +
Sbjct  1157  TELVRTALRGEVVAAETSLKDFTSILGHTVNEGAGV---LFGGAMVGLDAYELAHAENDV  1213

Query  101   LRQEIEAKIGIMAVNLTTATTAIITSSLGIASGFSIL---LVPLAGISAGIPSLVNNELV  157
              +     ++   + +  T T  +    +G ++  V L G  V L G+ G  +L     V
Sbjct  1214  QKAVFGTQLAFDSASFVTGIAGVGAGLVGASTAGAVLGSGGVILGGLAVGFTALAQAFGV  1273

Query  158   LRDKATKVVDYF  169
             + + A  V  YF
Sbjct  1274  VAEDAKAVGRYF  1285
```

P. asymbiotica subsp. asymbiotica RTX toxin

>ref|YP_003042199.1| MCF toxin [Photorhabdus asymbiotica subsp. asymbiotica ATCC
emb|CAR67791.1| MCF toxin [Photorhabdus asymbiotica subsp. asymbiotica ATCC 43949]
emb|CAQ85457.1| MCF toxin [Photorhabdus asymbiotica]
Length=2993

GENE ID: 8166769 PAU_03369 | MCF toxin
[Photorhabdus asymbiotica subsp. asymbiotica ATCC 43949]
(10 or fewer PubMed links)

Score = 48.5 bits (114), Expect = 3e-04, Method: Composition-based stats.
Identities = 46/195 (24%), Positives = 85/195 (44%), Gaps = 26/195 (13%)

```
Query  1     VNTLNAAFFIQSLIEYNSSKE--------SLSNLSVAMKVQVYAQLFSTGLNTIITDAAKV  52
             V+ LNA F +Q+LI++ + K         +  +L+ A+K+  Y  L      ++ D  K+
Sbjct  1746  VDGLNAGFTVQTLIQWFADKNRHDAASGVASPDLATALKIHSYLGLTQMAHGSLQDVVKI  1805

Query  53    VELVSTAL------------DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPL  100
              ELV TAL            D  + L  T++EG  +   +  GV +G    EL+   + +
Sbjct  1806  TELVRTALRGEVKAAETSFKDFALSLGHTVNEGAGV---LFGGVMVGLDAYELAHAENDV  1862

Query  101   LRQEIEAKIGIMAVNLTTATTAIITSSLGIASGFSIL---LVPLAGISAGIPSLVNNELV  157
              +      ++   + +L T    +    +G ++  ++L    V L G++ G  +L
Sbjct  1863  QKAVFGTQLAFDSASLVTGVAGVGAGLVGASTTAAVLGGAGVILGGLAVGFTALAQAFGA  1922

Query  158   LRDKATKVVDYFKHV  172
              + + A  V  YF  V
Sbjct  1923  VAEDAKAVGRYFDTV  1937
```

Yersinia mollaretii RTX toxin

>ref|ZP_04642320.1| RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC 43969]
gb|EEQ09150.1| RTX toxin and Ca2+-binding protein [Yersinia mollaretii ATCC 43969]
Length=1998

Score = 63.5 bits (153), Expect = 4e-09, Method: Compositional matrix adjust.
Identities = 58/176 (33%), Positives = 92/176 (52%), Gaps = 31/176 (18%)

```
Query  1     VNTLNAAFFIQSLIEYNSSKESLSN-LSVAMKVQVYAQLFSTGLNTIITDAAKVVELVSTA  59
             V+TLNAAF +Q+L+     ++  S+ LS  M++Q Y  L    +    DA  +  LV+ A
Sbjct  925   VHTLNAAFLLQALMGQRPQQQGSSDALSWPMQLQNYVGLIQPTIGLAEDAVHLGGLVTQA  984

Query  60    --------LDETIDLL---PTLSEGLPII-ATIIDGVSLGAAIKELSETSDPLLRQEIEA  107
                     L +T+  L    PTL   +P +   ++D  +L   I +L+ T +P+
Sbjct  985   IAGVELKPLAQTLSALHASPTLGSVMPALPGLLLDAANLTGIIAQLATTDNPI-------  1037

Query  108   KIGIMAVNLTTAT--------TAIITSSLGIASGFSILL----VPLAGISAGIPSLV  152
             +I + + NLT AT              A++TS +  A+G S +L     VPLAGI+AG+P+LV
Sbjct  1038  EIAVASTNLTMATLTTGVNVAALVTSFIPAAAGASAVLGLVAVPLAGIAAGLPALV  1093


>ref|ZP_04640347.1| Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC 43969]
gb|EEQ11141.1| Methyl-accepting chemotaxis protein [Yersinia mollaretii ATCC 43969]
Length=1797

Score = 60.8 bits (146), Expect = 3e-08, Method: Compositional matrix adjust.
Identities = 51/174 (29%), Positives = 89/174 (51%), Gaps = 19/174 (11%)

Query  2    NTLNAAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTIITDAAKVVELVSTA--  59
            +TLNAAF +Q+L+  N S    ++ LS  +++Q Y QL    L  + D + V  LV  A
Sbjct  759  HTLNAAFMLQTLMNINPSNGGINALSWPLQLQTYTQLAQNTLGLVHDVSAVANLVKLASA  818

Query  60   -----LDETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIG----  110
                 L     LL T++ G  ++  ++D  ++     +LS ++DP+   EI   I
Sbjct  819  TELKPLSAATSLLGTVAPG--VVGLLLDAANILGMSFQLSASTDPV---EINTTIANLTL  873

Query  111  ---IMAVNLTTATTAIITSSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDK  161
                ++  N+     T++  +S  ++     ++ VPLAGI+AG+P+LV N   L ++
Sbjct  874  SSLMVGTNIAALLTSLSAASAAVSGLLGMVAVPLAGIAAGLPALVGNYTTLAEQ  927
```

Xenorhabdus bovienii mcf protein

>ref|YP_003468304.1| Mcf protein (fragment) [Xenorhabdus bovienii SS-2004]
emb|CBJ81536.1| putative Mcf protein (fragment) [Xenorhabdus bovienii SS-2004]
Length=2533

GENE ID: 8832023 XBJ1_2410 | Mcf protein (fragment)
[Xenorhabdus bovienii SS-2004] (10 or fewer PubMed links)

Score = 50.8 bits (120), Expect = 5e-05, Method: Composition-based stats.
Identities = 44/187 (24%), Positives = 78/187 (42%), Gaps = 21/187 (11%)

```
Query  1     VNTLNAAFFIQSLIEYNSSKESLS--------NLSVAMKVQVYAQLFSTGLNTIITDAAKV  52
             V+ LN+AF +++LI  + +  ++  +                L    A++V  Y  L      + D AK+
Sbjct  1258  VHGLNSAFMVKTLIPWFADRKRITVASEKVPNTLETALRVHTYVNLAQMAHGMLEDGAKL  1317

Query  53    VELVSTALDETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIM  112
              L      E    +L   L  ++T   GV  G  I  ++     L    + E++    +
Sbjct  1318  ASLYRVVASEGRGVTSSLFSSLSHVST---GVGAGLNIFSVALDGIELAHAQNESQKAVY  1374

Query  113   AVNLTTATTAIITSSLGIASGF----------SILLVPLAGISAGIPSLVNNELVLRDKA  162
              L    ++T+ +GI +G          S L VPLAG+  G  +L      +  A
Sbjct  1375  GTQLAFDSAGLVTTGVGIGAGIFGAPIVAGFASSLAVPLAGLGIGFTALAEAFGRVASDA  1434

Query  163   TKVVDYF  169
               V +YF
Sbjct  1435  QAVGNYF  1441
```

Photorhabdus luminescens subsp. *laumondii* mcf2

```
>gb|AAR21118.1|  Mcf2 [Photorhabdus luminescens]
Length=2388

 Score = 40.4 bits (93),  Expect = 0.17, Method: Compositional matrix adjust.
 Identities = 45/192 (23%), Positives = 80/192 (42%), Gaps = 26/192 (14%)

Query  1     VNTLNAAFFIQSLIEYNSSKE--------SLSNLSVAMKVQVYAQLFSTGLNTITDAAKV  52
             V+ LNA F IQ+LI++ + K          +  +L+ A+KV   Y          + D   K+
Sbjct  1106  VDGLNAGFAIQTLIQWFADKNRNDAARGIASPDLATALKVHSYLNFVQMAHGGVQDVTKI  1165

Query  53    VELVSTAL------------DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPL  100
               LV TAL            D T   L  T++EG  +    +  G  +G    EL+   + +
Sbjct  1166  TALVRTALRGEVMAAETSLKDFTSILGHTVNEGAGV---LFGGAMVGLDAYELAHAENDV  1222

Query  101   LRQEIEAKIGIMAVNLTTATTAIITSSLGIASGFSIL---LVPLAGISAGIPSLVNNELV  157
              +     ++    + +  T T   +    +G ++   ++L     V L G++ G   +L
Sbjct  1223  QKAVFGTQLAFDSASFVTGTAGVGAGLMGASTAGAVLGGAGVILGGLAVGFTALAQAFGA  1282

Query  158   LRDKATKVVDYF  169
             + + A  V  YF
Sbjct  1283  VAEDAKAVGRYF  1294
```

P. fluorescens FitD cytotoxin

```
>ref|YP_260089.1|  fitD gene product [Pseudomonas protegens Pf-5]
 gb|AAY92255.1|  cytotoxin FitD [Pseudomonas protegens Pf-5]
Length=3003

 GENE ID: 3476073 fitD | cytotoxin FitD [Pseudomonas fluorescens Pf-5]
(10 or fewer PubMed links)

 Score = 48.5 bits (114),  Expect = 3e-04, Method: Composition-based stats.
 Identities = 49/191 (26%), Positives = 83/191 (43%), Gaps = 24/191 (13%)

Query  1     VNTLNAAFFIQSLIEYNSSKESLS--------NLSVAMKVQVYAQLFSTGLNTITDAAKV  52
             V+ LNA F +Q+LI++ + K             +L+ A+K+  Y L    G  T+ D AKV
Sbjct  1713  VDGLNAGFAVQTLIQWFADKNRKDAAQGVVSPDLATALKIHSYLGLAQIGHGTVQDVAKV  1772

Query  53    VELVSTAL-DETIDLLPTLSEGLPIIATIID---GVSLGAAIKELSETSDPLLRQEIEAK  108
               ELV TAL  E +     +L +     +    ++   GV  G A+  L      L   E + +
Sbjct  1773  TELVQTALRGEALAAESSLKDFASTLGHTVNEGAGVLFGGAMVGLDAYE--LAHAENDVQ  1830

Query  109   IGIMAVNLTTATTAIITSSLGIASGF------SILL----VPLAGISAGIPSLVNNELVL  158
               +    L   + + ++ + GI +G       + +L    V L G++ G   +L     +
Sbjct  1831  KAVFGTQLAFDSASFVSGAAGIGAGLIGASTTAAVLGGAGVILGGLAVGFTALAQAFGAV  1890

Query  159   RDKATKVVDYF  169
              + A  V  YF
Sbjct  1891  AEDAKAVGRYF  1901
```

P. asymbiotica subsp. asymbiotica RTX toxin

```
>ref|YP_003041903.1|  rtx toxin RtxA [Photorhabdus asymbiotica subsp. asymbiotica ATCC
43949]
 emb|CAR67497.1|  similar to rtx toxin rtxa [Photorhabdus asymbiotica subsp. asymbiotica
ATCC 43949]
 emb|CAQ85161.1|  similar to rtx toxin rtxa [Photorhabdus asymbiotica]
Length=4068

 GENE ID: 8167608 PAU_03073 | rtx toxin RtxA
[Photorhabdus asymbiotica subsp. asymbiotica ATCC 43949]
(10 or fewer PubMed links)

 Score = 52.8 bits (125),  Expect = 3e-05, Method: Composition-based stats.
 Identities = 51/186 (27%), Positives = 85/186 (46%), Gaps = 31/186 (17%)

Query  36    IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSII  95
             I+QL+ D I  +AA NL  K P  SV+ +    D++  Y+   GD           P  +
Sbjct  3001  IIQLENDPIVAKAAINLAGKHPDSSVVVKL---DADGKYHLIYGD-----------PDSL  3046

Query  96    SDRPKIKLTFIGHGKDE--FNTDIFAGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGC  153
             S   K++   +GHG+DE  +N    +G+  D L+ +++      ++   P  I I  +GC
Sbjct  3047  SG--KLRWQIVGHGRDESAYNNTRLSGYSADELAIKLKQFSQSFEQAGKPDRISI--VGC  3102

Query  154   NMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEGRRELLDHS  213
             ++ S        Y     +    DK       I  D   VSA + EV +++ GR+   D +
Sbjct  3103  SLISEDKRNGFAYR---FITALDK-----QGIRSD---VSARRSEVAVDATGRKFTRDKN  3151

Query  214   GEWINK  219
              +W+NK
Sbjct  3152  NQWVNK  3157
```

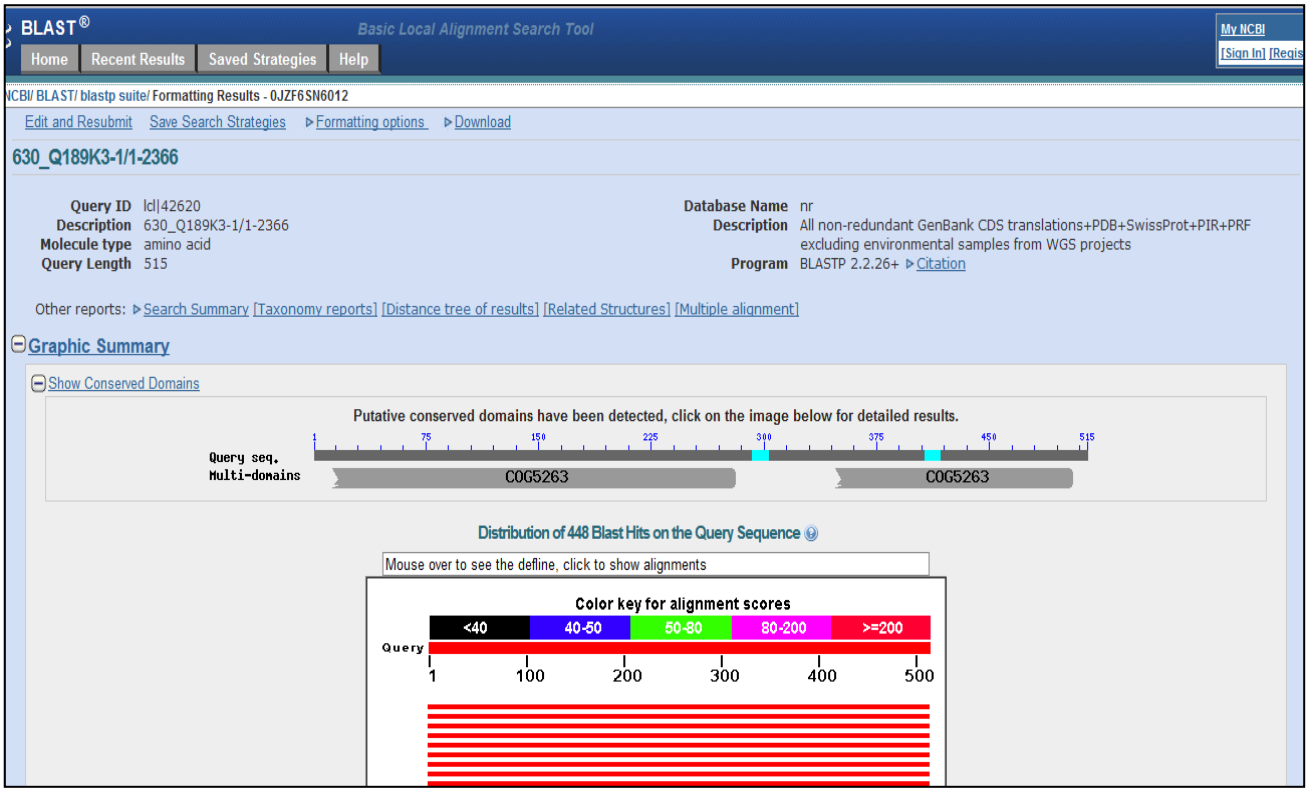
***Vibrio cholerae* RTX toxin**

```
>gb|EGQ99461.1|  RTX toxin RtxA [Vibrio cholerae HE39]
Length=4545

 Score = 52.4 bits (124),  Expect = 4e-05, Method: Composition-based stats.
 Identities = 53/196 (27%), Positives = 87/196 (44%), Gaps = 43/196 (22%)

Query  36    IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSII  95
             IVQ++ D +  +AA NL  K P  SV+ Q    DS+  Y    GD           PS +
Sbjct  3463  IVQMENDDVVAKAAANLAGKHPESSVVVQL---DSDGNYRVVYGD-----------PSKL  3508

Query  96    SDRPKIKLTFIGHGKD--EFNTDIFAGFDVDSLSTEIEAAIDLAK--------EDISPKS  145
                 K++   +GHG+D  E N    +G+  D L      A+ LAK         E+I+ K
Sbjct  3509  DG--KLRWQLVGHGRDHSESNNTRLSGYSADEL------AVKLAKFQQSFNQAENINNKP  3560

Query  146   IEINLLGCNMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEG  205
               I+++GC++ S   + ++   G   +   D          +   + VS    E+ ++  G
Sbjct  3561  DHISIVGCSLVS---DDKQKGFGHQFINAMDA--------NGLRVDVSVRSSELTVDEAG  3609

Query  206   RRELLDHSGEWINKEE  221
             R+   D +G+W+ K E
Sbjct  3610  RKHTKDANGDWVQKAE  3625


>ref|ZP_04961062.1|  RTX toxin RtxA [Vibrio cholerae AM-19226]
 gb|EDN15712.1|  RTX toxin RtxA [Vibrio cholerae AM-19226]
Length=4558

 Score = 52.4 bits (124),  Expect = 4e-05, Method: Composition-based stats.
 Identities = 53/196 (27%), Positives = 87/196 (44%), Gaps = 43/196 (22%)

Query  36    IVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSII  95
             IVQ++ D +  +AA NL  K P  SV+ Q    DS+  Y    GD           PS +
Sbjct  3476  IVQMENDDVVAKAAANLAGKHPESSVVVQL---DSDGNYRVVYGD-----------PSKL  3521

Query  96    SDRPKIKLTFIGHGKD--EFNTDIFAGFDVDSLSTEIEAAIDLAK--------EDISPKS  145
                 K++   +GHG+D  E N    +G+  D L      A+ LAK         E+I+ K
Sbjct  3522  DG--KLRWQLVGHGRDHSESNNTRLSGYSADEL------AVKLAKFQQSFNQAENINNKP  3573

Query  146   IEINLLGCNMFSYSINVEETYPGKLLLKVKDKISELMPSISQDSIIVSANQYEVRINSEG  205
               I+++GC++ S   + ++   G   +   D          +   + VS    E+ ++  G
Sbjct  3574  DHISIVGCSLVS---DDKQKGFGHQFINAMDA--------NGLRVDVSVRSSELAVDEAG  3622

Query  206   RRELLDHSGEWINKEE  221
             R+   D +G+W+ K E
Sbjct  3623  RKHTKDANGDWVQKAE  3638
```

8.4.2.4. Receptor binding domain

Analysis of the receptor binding domain of *C. difficile* toxin B revealed the following alignments:

Query Coverage = 515 amino acids.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
ZP_05328744.1	toxin B [Clostridium difficile QCD-63q42]	1023	1322	100%	0.0	100%
YP_001087135.1	tcdB gene product [Clostridium difficile 630] >ref|ZP_05349824.1|	1023	1322	100%	0.0	100%
EHJ40398.1	cell wall-binding repeat protein [Clostridium difficile 70-100-2010]	1019	1320	100%	0.0	99%
EHJ31817.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	1016	1317	100%	0.0	99%
AAG18011.1	cytotoxin B [[Clostridium] difficile]	1013	1313	100%	0.0	99%
CAA80815.1	toxin B [[Clostridium] difficile]	1010	1307	100%	0.0	99%
CAA43299.1	toxin B [[Clostridium] difficile]	1006	1393	100%	0.0	100%
ZP_06891228.1	toxin B [Clostridium difficile NAP08] >ref|ZP_06902243.1| toxin B [	998	1513	100%	0.0	95%
ZP_05400113.1	toxin B [Clostridium difficile QCD-23m63]	998	1513	100%	0.0	95%
1814453A	toxin B	989	1499	100%	0.0	100%
AAO83645.1	toxin B [[Clostridium] difficile]	989	1589	100%	0.0	99%
ZP_05270740.1	toxin B [Clostridium difficile QCD-66c26] >ref|ZP_05321143.1| toxi	940	1346	100%	0.0	89%
CAC19891.1	toxin B [[Clostridium] difficile]	928	1331	100%	0.0	87%
CAA57959.1	cytotoxin L [[Clostridium] sordellii]	776	1221	100%	0.0	74%
ZP_06979366.1	surface protein PspA [Streptococcus pneumoniae str. Canada MDR	145	550	100%	3e-35	26%
EGV01168.1	cell wall-binding repeat protein [Streptococcus oralis SK313]	132		100%	6e-30	29%
ADH94633.1	TcdB [[Clostridium] difficile]	987	1885	99%	0.0	98%
ADH94625.1	TcdB [[Clostridium] difficile]	981	1779	99%	0.0	99%
AAO83646.1	toxin B [[Clostridium] difficile]	975	1574	99%	0.0	99%
ADH94629.1	TcdB [[Clostridium] difficile]	974	1779	99%	0.0	99%
ADH94623.1	TcdB [[Clostridium] difficile]	923	1747	99%	0.0	90%
ADH94634.1	TcdB [[Clostridium] difficile]	918	1739	99%	0.0	90%
ADH94628.1	TcdB [[Clostridium] difficile]	904	1513	99%	0.0	88%
ADH94635.1	TcdB [[Clostridium] difficile]	903	1513	99%	0.0	88%
ADH94624.1	TcdB [[Clostridium] difficile]	903	1513	99%	0.0	88%
		901	1419	99%	0.0	88%

Show alignment for CAA57959.1

.nih.gov/Blast.cgi?CMD=Get&ALIGNMENTS=100&ALIGNMENT_VIE...

ADH94630.1	TcdB [[Clostridium] difficile]	901	1419	99%	0.0	88%
ADH94626.1	TcdB [[Clostridium] difficile]	898	1424	99%	0.0	88%
ADH94636.1	TcdB [[Clostridium] difficile]	887	1803	99%	0.0	87%
AAO83644.1	toxin A [[Clostridium] difficile]	353	1600	99%	1e-107	42%
ZP_05349827.1	toxin A [Clostridium difficile ATCC 43255] >gb|AAA23283.1| toxin A	353	1705	99%	4e-104	42%
YP_001087137.1	tcdA gene product [Clostridium difficile 630] >emb|CAJ67494.1| Tc	353	1706	99%	5e-104	42%
P16154.2	RecName: Full=Toxin A >emb|CAA36094.1| unnamed protein produ	349	1691	99%	2e-102	42%
ZP_05331889.1	toxin A [Clostridium difficile QCD-63q42]	236	712	99%	2e-69	42%
ZP_05399187.1	toxin A [Clostridium difficile QCD-37x79]	233	694	99%	3e-68	42%
NP_347713.1	hypothetical protein [Clostridium acetobutylicum ATCC 824] >ref|Y	245	2868	99%	8e-67	31%
ZP_03960388.1	conserved hypothetical protein [Lactobacillus vaginalis ATCC 4954	144	665	99%	1e-33	30%
YP_004104834.1	cell wall-binding repeat-containing protein [Ruminococcus albus 7]	140	721	99%	2e-32	31%
ZP_07888302.1	choline binding protein E [Streptococcus sanguinis ATCC 49296] >|	139	577	99%	3e-32	29%
EJD25793.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK1(	136	457	99%	4e-31	28%
EIC76753.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK1(	135	439	99%	8e-31	28%
EIC76035.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK6:	134	419	99%	2e-30	28%
ZP_08049975.1	choline binding protein E [Streptococcus sp. C300] >gb|EFX56765	130	435	99%	2e-29	28%
EGL91250.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK25	130	499	99%	3e-29	26%
EIC78914.1	metallo-beta-lactamase domain protein [Streptococcus oralis SK1(	130	607	99%	4e-29	29%
ZP_07458344.1	choline binding protein E [Streptococcus sp. oral taxon 071 str. 73	129	497	99%	5e-29	26%
ZP_06612389.1	choline binding protein E [Streptococcus oralis ATCC 35037] >ref|:	127	563	99%	3e-28	29%
ZP_05270743.1	toxin A [Clostridium difficile QCD-66c26] >ref|ZP_05354979.1| toxi	358	1861	99%	1e-105	43%
AFN52237.1	TcdA [[Clostridium] difficile]	347	1835	99%	1e-101	42%
BAA25318.1	toxin A [[Clostridium] difficile]	185	510	99%	2e-51	42%
ZP_04219464.1	hypothetical protein bcere0022_38930 [Bacillus cereus Rock3-44]	191	925	99%	8e-50	28%

ZP_04219464.1	hypothetical protein bcere0022_38930 [Bacillus cereus Rock3-44]	191	925	99%	8e-50	28%
ZP_06144336.1	hypothetical protein RflaF_14077 [Ruminococcus flavefaciens FD-1	172	665	99%	3e-43	31%
CAA80818.1	alpha-toxin [Clostridium novyi]	161	596	99%	4e-39	34%
ZP_08333659.1	hypothetical protein HMPREF0992_02583 [Lachnospiraceae bacteri	154	504	99%	7e-37	28%
EGU67616.1	metallo-beta-lactamase domain protein [Streptococcus mitis bv. 2	150	286	99%	2e-35	27%
ZP_06198721.1	choline binding protein E [Streptococcus sp. M143] >gb|EFA24335	142	586	99%	4e-33	29%
ZP_08849318.1	hypothetical protein HMPREF9457_01027 [Dorea formicigenerans 4	139	583	99%	7e-32	29%
ZP_09174515.1	metallo-beta-lactamase domain protein [Streptococcus sp. oral ta	138	716	99%	1e-31	31%
EGV04069.1	metallo-beta-lactamase domain protein [Streptococcus infantis SK	135	459	99%	7e-31	28%
ZP_03779155.1	hypothetical protein CLOHYLEM_06226 [Clostridium hylemonae DSM	135	722	99%	7e-31	26%
EGL98808.1	hypothetical protein NIAS840_00514 [Lactobacillus salivarius NIAS	133	504	99%	3e-30	27%
ZP_06344526.1	hypothetical protein CLOM621_05219 [Clostridium sp. M62/1] >gb|	129	751	99%	4e-29	30%
ZP_08065205.1	choline binding protein E [Streptococcus peroris ATCC 700780] >g	129	255	99%	1e-28	27%
CAA35057.1	enterotoxin A [[Clostridium] difficile]	309	1099	99%	1e-93	41%
2QJ6_A	Chain A, Crystal Structure Analysis Of A 14 Repeat C-Terminal Fra	204	559	99%	1e-57	39%
EHJ37067.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	199	676	99%	3e-56	41%
ZP_06902240.1	toxin A [Clostridium difficile NAP07] >gb|EFH16595.1| toxin A [Clos	207	674	99%	7e-54	45%
ZP_05321147.1	toxin A [Clostridium difficile CIP 107932]	152	485	99%	4e-39	46%
CAA88565.1	alpha-toxin [Clostridium novyi]	158	583	99%	1e-37	34%
EGR94324.1	metallo-beta-lactamase domain protein [Streptococcus mitis bv. 2	147	384	99%	7e-35	27%
ZP_05853491.1	putative cell wall binding repeat-containing domain protein [Blautia	147	528	99%	2e-34	29%
ZP_07694212.1	surface protein PspC [Streptococcus infantis SK1302] >gb|EFO538	127	450	99%	1e-29	30%
ADH94627.1	TcdB [[Clostridium] difficile]	962	1647	99%	0.0	95%
ADH94632.1	TcdB [[Clostridium] difficile]	961	1640	99%	0.0	95%
ADH94631.1	TcdB [[Clostridium] difficile]	960	1639	99%	0.0	95%

ADH94632.1	TcdB [[Clostridium] difficile]	961	1640	99%	0.0	95%
ADH94631.1	TcdB [[Clostridium] difficile]	960	1639	99%	0.0	95%
ZP_07405637.1	toxin A [Clostridium difficile QCD-32g58]	332	1298	99%	1e-96	42%
ZP_05321146.1	toxin A [Clostridium difficile CIP 107932]	315	820	99%	7e-91	42%
AAG18010.1	enterotoxin A [[Clostridium] difficile]	186	605	99%	6e-50	42%
CAA73178.1	toxin A [[Clostridium] difficile]	182	597	99%	2e-48	42%
EHJ31825.1	cell wall-binding repeat protein [Clostridium difficile 050-P50-2011]	186	605	99%	3e-48	42%
2G7C_A	Chain A, Clostridium Difficile Toxin A Fragment Bound To Agal(1,3)	155	499	99%	2e-40	48%
ZP_06902239.1	toxin A [Clostridium difficile NAP07] >gb|EFH16631.1| toxin A [Clos	155	593	99%	2e-40	45%
ZP_05400117.1	toxin A [Clostridium difficile QCD-23m63]	155	560	99%	2e-40	48%
EGV15112.1	cell wall-binding repeat protein [Streptococcus infantis X]	136	453	99%	2e-32	28%
ZP_08090001.1	glucan-binding domain-containing protein [Clostridium symbiosum V	139	492	99%	2e-32	32%
ZP_09046953.1	hypothetical protein HMPREF1020_01032 [Clostridium sp. 7_3_54FA	139	492	99%	2e-32	32%
ZP_08602448.1	hypothetical protein HMPREF0993_01825 [Lachnospiraceae bacteri	138	484	99%	1e-31	29%
ZP_08108980.1	hypothetical protein HMPREF9475_03844 [Clostridium symbiosum W	136	482	99%	1e-31	31%
ZP_07462919.1	choline binding protein E [Streptococcus mitis ATCC 6249] >gb|EF	136	529	99%	3e-31	28%
ZP_05328747.1	toxin A [Clostridium difficile QCD-63q42]	188	688	98%	9e-48	45%
ZP_05400116.1	toxin A [Clostridium difficile QCD-23m63]	180	665	98%	6e-45	45%
EHJ40401.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	162	613	98%	3e-39	45%
ZP_10038893.1	metallo-beta-lactamase domain protein [Streptococcus sp. SK140]	145	487	98%	4e-34	29%
EHJ36920.1	cell wall-binding repeat protein, partial [Clostridium difficile 70-100	122	439	98%	3e-29	38%
ZP_03288011.1	hypothetical protein CLONEX_00190 [Clostridium nexile DSM 1787]	159	577	98%	3e-38	29%
ZP_02234104.1	hypothetical protein DORFOR_00962 [Dorea formicigenerans ATCC	128	295	98%	2e-28	28%
BAF30978.1	collagenolytic protease [Geobacillus sp. MO-1]	132	541	98%	2e-29	36%
AAZ40138.1	toxin B [[Clostridium] difficile] >gb|AAZ40141.1| toxin B [Clostridiu	234	625	87%	1e-71	100%
AAZ40140.1	toxin B [[Clostridium] difficile]	230	627	87%	3e-70	99%
AAZ40147.1	toxin B [Clostridium difficile] >gb|AAZ40148.1| toxin B [Clostridium	229	574	87%	9e-70	98%
AAZ40139.1	toxin B [[Clostridium] difficile] >gb|AAZ40151.1| toxin B [Clostridiu	222	568	87%	5e-67	94%

Alignment information for genes shown in Table 4.3:

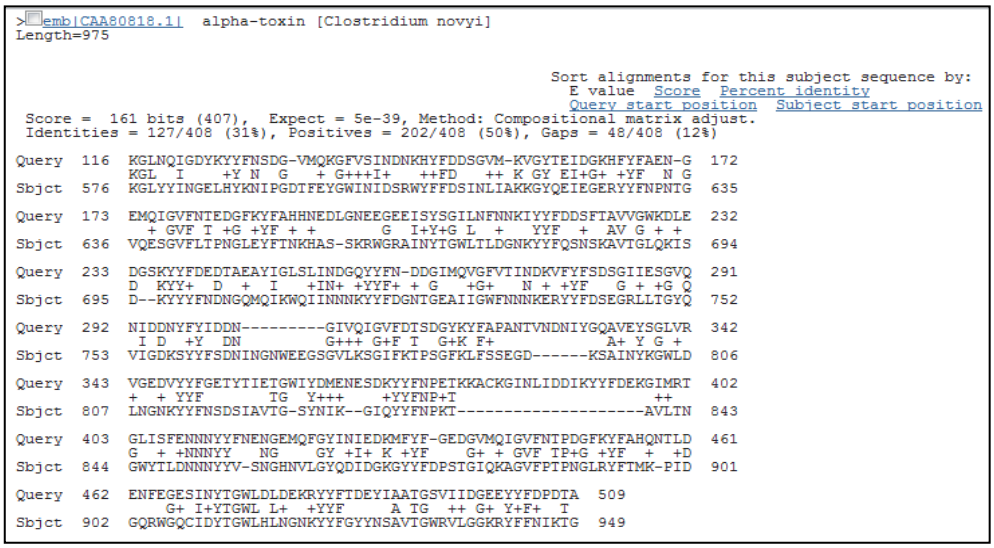
C. difficile tcdA

```
>gb|AAQ83644.1|  toxin A [[Clostridium] difficile]
Length=929

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Score =  353 bits (905),  Expect = 1e-107, Method: Compositional matrix adjust.
 Identities = 217/520 (42%), Positives = 312/520 (60%), Gaps = 31/520 (6%)

Query  1    LITGFVTVGDDKYYFNPINGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPA  60
            L+TG+ T+   KYYF+ IN GAA     II+ K++YFN  GV+Q GVF   DGF+YFAPA
Sbjct  98   LVTGWQTINGKKYYFD-INTGAALTSYKIINGKHFYFNNDGVMQLGVFKGPDGFEYFAPA  156

Query  61   NTLDENLEGEAIDFTGK-LIIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLN  119
            NT + N+EG+AI +  K L ++   YYFD++ +    W+ ++ E +YF+P+   A  GL
Sbjct  157  NTQNNNIEGQAIVYQSKFLTLNGKKYYFDNDSKAVTGWRIINNEKYYFNPDNAIAAVGLQ  216

Query  120  QIGDYKYYFNSD-GVMQKGFVSINDNKHYFD-DSGVMKVGYTEIDGKHFYFAENGEMQIG  177
             I + KYYFN D  ++ KG+ ++N +++YFD D+ +   GY  IDGKHFYF  +  ++IG
Sbjct  217  VIDNNKYYFNPDTAIISKGWQTVNGSRYYFDTDTAIAFNGYKTIDGKHFYFDSDCVVKIG  276

Query  178  VFNIEDGFKYFAHHNEDLGNEEGEEISY-SGILNFNNKIYYFDDSFTAVVGWKDLEDGSK  236
            VF+T +GF+YFA  N    N EG+ I Y S  L  N K YYFD++  AV GW+ + D  K
Sbjct  277  VFSTSNGFEYFAPANTYNNNIEGQAIVYQSKFLTLNGKKYYFDNNSKAVTGWQTI-DSKK  335

Query  237  YYFDEDTAEAYIGLSLINDGQYYFNDDGI-MQVGFVTINDKVFYF-SDSGIIESGVQNID  294
            YYF+ +TAEA   G    I+  +YYFN +        G+ TI+ K +YF +++ I  +G   I+
Sbjct  336  YYFNTNTAEAATGWQTIDGKKYYFNTNTAEAATGWQTIDGKKYYFNTNTAIASTGYTIIN  395

Query  295  DNYFYIDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSG-LVRVGEDVYYFGET  353
              +FY + +GI+QIGVF   +G++YFAPANT  +NI GQA+ Y    + +    YYFG
Sbjct  396  GKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNEFLTLNGKKYYFGSD  455

Query  354  YTIETGWIYDMENESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYY  413
              TGW      + KYYFNP    A   +  I++ KYYF  GI++ G I+ E NN+Y
Sbjct  456  SKAVTGWRII---NNKKYYFNPNNAIAAIHLCTINNDKYYFSYDGILQNGYITIERNNFY  512

Query  414  FNENGEMQFGYINIEDKMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTG  473
            F+ N E +                M  GVF  P+GF+YFA  NT + N EG++I Y
Sbjct  513  FDANNESK------------------MVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQN  554

Query  474  -WLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLV  512
             +L L+ K+YYF ++  A TG   IDG++YYF+ +TA+
Sbjct  555  KFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAA  594
```

C. sordellii tcsL

```
>emb|CAA57959.1|  cytotoxin L [[Clostridium] sordellii]
Length=2364

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Score =  776 bits (2004),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 380/515 (74%), Positives = 443/515 (86%), Gaps = 3/515 (1%)

Query  1     LITGFVTVGDDKYYFNPINGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPA  60
             LITGF T+  +KYYF+P   GAASIGE  ID K+YYFN+ G+LQ GV +T DG KYFAPA
Sbjct  1853  LITGFTTIDGNKYYFDPTKSGAASIGEITIDGKDYYFNKQGILQVGVINTSDGLKYFAPA  1912

Query  61    NTLDENLEGEAIDFTGKLIIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQ  120
              TLDENLEGE+++F GKL ID  IYYF+DNYR AVEWK LD E +YF+P+TG+A KGL+Q
Sbjct  1913  GTLDENLEGESVNFIGKLNIDGKIYYFEDNYRAAVEWKLLDDETYYFNPKTGEALKGLHQ  1972

Query  121   IGDYKYYFNSDGVMQKGFVSINDNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFN  180
             IGD KYYF+ +G+MQ GF++IND   YF++ GVM+VGY E++GK+FYF +NGE Q+GVFN
Sbjct  1973  IGDNKYYFDDNGIMQTGFITINDKVFYFNNDGVMQVGYIEVNGKYFYFGKNGERQLGVFN  2032

Query  181   TEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGSKYYFD  240
             T DGFK+F   ++DLG EEGE   Y+GILNFN KIY+FD S TAVVGW  L+DGS YYFD
Sbjct  2033  TPDGFKFFGPKDDDLGTEEGELTLYNGILNFNGKIYFFDISNTAVVGWGTLDDGSTYYFD  2092

Query  241   EDTAEAYIGLSLINDGQYYFNDDGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFYI  300
             ++ AEA IGL++IND +YYF+D+GI Q+GF+TIND +FYFS+SG IE G QNI+ NYFYI
Sbjct  2093  DNRAEACIGLTVINDCKYYFDDNGIRQLGFITINDNIFYFSESGKIELGYQNINGNYFYI  2152

Query  301   DDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGW  360
             D++G+V IGVFDT DGYKYFAP NTVNDNIYGQAV+YSGLVRV EDVYYFGETY IETGW
Sbjct  2153  DESGLVLIGVFDTPDGYKYFAPLNTVNDNIYGQAVKYSGLVRVNEDVYYFGETYKIETGW  2212

Query  361   IYDMENESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEM  420
             I   ENE+DKYYF+PETKKA KGIN++DDIKYYFDE GIMRTGLISFENNNYYFNE+G+M
Sbjct  2213  I---ENETDKYYFDPETKKAYKGINVVDDIKYYFDENGIMRTGLISFENNNYYFNEDGKM  2269

Query  421   QFGYINIEDKMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTGWLDLDEK  480
             QFGY+NI+DKMFYFG+DG MQIGVFNTPDGFKYFAHQNTLDENFEGESINYTGWLDLD K
Sbjct  2270  QFGYLNIKDKMFYFGKDGKMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTGWLDLDGK  2329

Query  481   RYYFTDEYIAATGSVIIDGEEYYFDPDTAQLVISE  515
             RYYFTDEYIAATGS+ IDG  YYFDPDTA+LV+SE
Sbjct  2330  RYYFTDEYIAATGSLTIDGYNYYFDPDTAELVVSE  2364
```

C. novyi α toxin

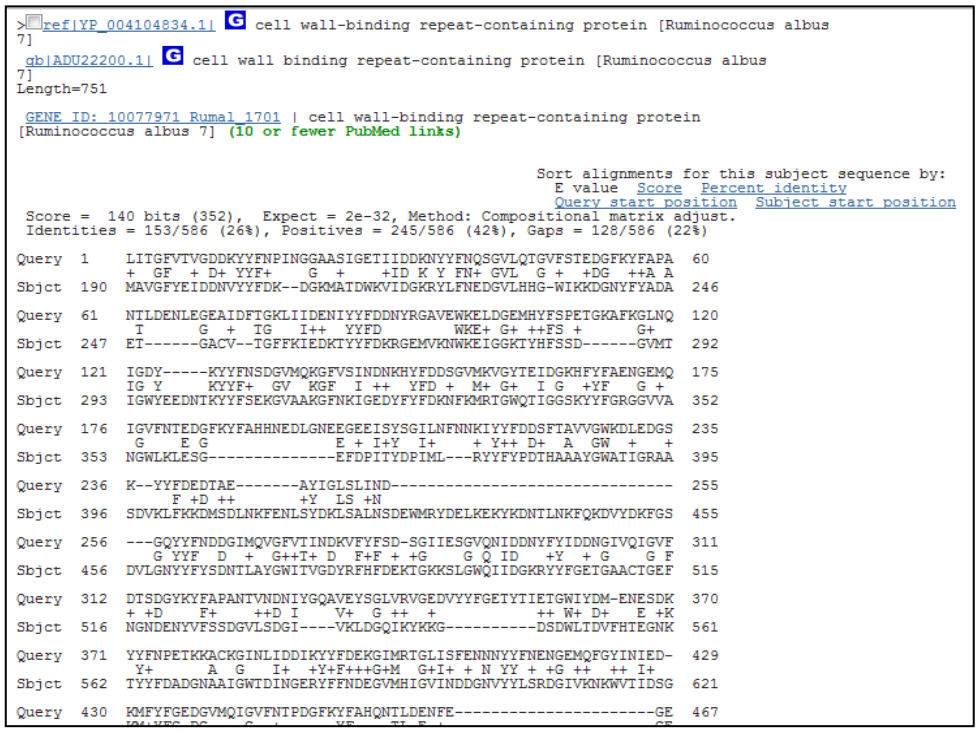


```
>emb|CAA80818.1|  alpha-toxin [Clostridium novyi]
Length=975

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Score =  161 bits (407),  Expect = 5e-39, Method: Compositional matrix adjust.
 Identities = 127/408 (31%), Positives = 202/408 (50%), Gaps = 48/408 (12%)

Query  116  KGLNQIGDYKYYFNSDG-VMQKGFVSINDNKHYFDDSGVM-KVGYTEIDGKHFYFAEN-G  172
            KGL  I    +Y N  G    + G+++I+    ++FD    ++ K GY EI+G+ +YF  N G
Sbjct  576  KGLYYINGELHYKNIPGDTFEYGWINIDSRWYFFDSINLIAKKGYQEIEGERYYFNPNTG  635

Query  173  EMQIGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAVVGWKDLE  232
             + GVF T +G +YF + +      G  I+Y+G L  +    YYF  +  AV G + +
Sbjct  636  VQESGVFLIPNGLEYFTNKHAS-SKRWGRAINYTGWLTLDGNKYYFQSNSKAVIGLQKIS  694

Query  233  DGSKYYFDEDTAEAYIGLSLINDGQYYFN-DDGIMQVGFVTINDKVFYFSDSGIIESGVQ  291
            D  KYY+  D  +  I   +IN+ +YYF+ + G   +G+    N + +YF   G + +G Q
Sbjct  695  D--KYYYFNDNGQMQIKWQIINNNKYYFDGNTGEAIIGWFNNNKERYYFDSEGRLLTGYQ  752

Query  292  NIDDNYFYIDDN---------GIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVR  342
             I D  +Y  DN          G+++ G+F T  G+K F+           A+ Y G +
Sbjct  753  VIGDKSYYFSDNINGNWEEGSGVLKSGIFKTPSGFKLFSSEGD------KSAINYKGWLD  806

Query  343  VGEDVYYFGETYTIETGWIYDMENESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRT  402
            +  + YYF        TG  Y+++     +YYFNP+T                 ++
Sbjct  807  LNGNKYYFNSDSIAVTG-SYNIK--GIQYYFNPKT--------------------AVLTN  843

Query  403  GLISFENNNYYFNENGEMQFGYINIEDKMFYF-GEDGVMQIGVFNTPDGFKYFAHQNTLD  461
            G  + +NNNYY   NG    GY +I+ K +YF    G+ + GVF TP+G +YF  +  +D
Sbjct  844  GWYTLDNNNYYV-SNGHNVLGYQDIDGKGYYFDPSTGIQKAGVFFTPNGLRYFTMK-PID  901

Query  462  ENFEGESINYTGWLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTA  509
             G+ I+YTGWL L+  +YYF     A TG  ++ G+ Y+F+  T
Sbjct  902  GQRWGQCIDYTGWLHLNGNKYYFGYYNSAVTGWRVLGGKRYFFNIKTG  949
```

Ruminococcus albus cell wall binding repeat domain protein

```
>ref|YP_004104834.1|  cell wall-binding repeat-containing protein [Ruminococcus albus
7]
 gb|ADU22200.1|  cell wall binding repeat-containing protein [Ruminococcus albus
7]
Length=751

 GENE ID: 10077971 Rumal_1701 | cell wall-binding repeat-containing protein
[Ruminococcus albus 7]  (10 or fewer PubMed links)

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Score =  140 bits (352),  Expect = 2e-32, Method: Compositional matrix adjust.
 Identities = 153/586 (26%), Positives = 245/586 (42%), Gaps = 128/586 (22%)

Query  1    LITGFVTVGDDKYYFNPINGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPA  60
            +  GF  + D+ YYF+    G  +    +ID K Y FN+ GVL  G +  +DG  ++A A
Sbjct  190  MAVGFYEIDDNVYYFDK--DGKMATDWKVIDGKRYLFNEDGVLHHG-WIKKDGNYFYADA  246

Query  61   NTLDENLEGEAIDFTGKLIIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQ  120
             T     G  +  TG   I++  YYFD        WKE+ G+ ++FS +     G+
Sbjct  247  ET------GACV--TGFFKIEDKTYYFDKRGEMVKNWKEIGGKTYHFSSD------GVMT  292

Query  121  IGDY-----KYYFNSDGVMQKGFVSINDNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQ  175
            IG Y     KYYF+  GV  KGF  I ++  YFD +  M+ G+  I G  +YF   G +
Sbjct  293  IGWYEEDNTKYYFSEKGVAAKGFNKIGEDYFYFDKNFKMRTGWQTIGGSKYYFGRGGVVA  352

Query  176  IGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGS  235
             G   E G              E + I+Y  I+    + Y++ D+  A  GW  +   +
Sbjct  353  NGWLKLESG--------------EFDPITYDPIML---RYYFYPDTHAAAYGWATIGRAA  395

Query  236  K--YYFDEDTAE-------AYIGLSLIND------------------------------  255
               F +D ++         +Y  LS +N
Sbjct  396  SDVKLFKKDMSDLNKFENLSYDKLSALNSDEWMRYDELKEKYKDNTLNKFQKDVYDKFGS  455

Query  256  ---GQYYFNDDGIMQVGFVTIINDKVFYFSD-SGIIESGVQNIDDNYFYIDDNGIVQIGVF  311
               G YYF  D  +  G++T+ D  F+F + +G    G Q ID   +Y  + G    G F
Sbjct  456  DVLGNYYFYSDNTLAYGWITVGDYRFHFDEKTGKKSLGWQIIDGKRYYFGETGAACTGEF  515

Query  312  DTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDM-ENESDK  370
            + +D    F+    ++D I    V+  G ++  +          ++ W+ D+    E +K
Sbjct  516  NGNDENYVFSSDGVLSDGI----VKLDGQIKYKKG----------DSDWLIDVFHIEGNK  561

Query  371  YYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIED-  429
             Y+      A  G    I+  +Y+F+++G+M  G+I+ + N YY + +G ++  ++ I+
Sbjct  562  TYYFDADGNAAIGWTDINGERYFFNDEGVMHIGVINDDGNVYYLSRDGIVKNKWVTIDSG  621

Query  430  KMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFE----------------------GE  467
```

S. oralis cell wall binding repeat domain protein



>ref|YP_004104834.1| cell wall-binding repeat-containing protein [Ruminococcus albus 7]
gb|ADU22200.1| cell wall binding repeat-containing protein [Ruminococcus albus 7]
Length=751

GENE ID: 10077971 Rumal_1701 | cell wall-binding repeat-containing protein [Ruminococcus albus 7] (10 or fewer PubMed links)

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Score = 140 bits (352), Expect = 2e-32, Method: Compositional matrix adjust.
Identities = 153/586 (26%), Positives = 245/586 (42%), Gaps = 128/586 (22%)

```
Query  1    LITGFVTVGDDKYYFNPINGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPA  60
            +  GF  + D+ YYF+     G  +    +ID K Y FN+ GVL  G +  +DG  ++A A
Sbjct  190  MAVGFYEIDDNVYYFDK--DGKMATDWKVIDGKRYLFNEDGVLHHG-WIKKDGNYFYADA  246

Query  61   NTLDENLEGEAIDFTGKLIIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQ  120
             I      G  +  IG    I++   YYFD         WKE+ G+ ++FS +      G+
Sbjct  247  ET------GACV--TGFFKIEDKTYYFDKRGEMVKNWKEIGGKTYHFSSD------GVMT  292

Query  121  IGDY-----KYYFNSDGVMQKGFVSINDNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQ  175
            IG Y     KYYF+  GV  KGF  I ++  YFD +  M+ G+  I G  +YF   G +
Sbjct  293  IGWYEEDNTKYYFSEKGVAAKGFNKIGEDYFYFDKNFKMRTGWQTIGGSKYYFGRGGVVA  352

Query  176  IGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGS  235
             G    E G              E + I+Y  I+    + Y++ D+  A  GW  +    +
Sbjct  353  NGWLKLESG--------------EFDPITYDPIML---RYYFYPDTHAAAYGWATIGRAA  395

Query  236  K--YYFDEDTAE-------AYIGLSLIND-------------------------------  255
              F +D ++          +Y  LS +N
Sbjct  396  SDVKLFKKDMSDLNKFENLSYDKLSALNSDEWMRYDELKEKYKDNTLNKFQKDVYDKFGS  455

Query  256  ---GQYYFNDDGIMQVGFVTIINDKVFYFSD-SGIIESGVQNIDDNYFYIDDNGIVQIGVF  311
               G YYF  D  +  G++I+ D  F+F + +G    G Q ID   +Y  + G    G F
Sbjct  456  DVLGNYYFYSDNTLAYGWITVGDYRFHFDEKTGKKSLGWQIIDGKRYYFGETGAACTGEF  515

Query  312  DTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDM-ENESDK  370
            + +D    F+    ++D I    V+  G ++  +           ++ W+ D+   E +K
Sbjct  516  NGNDENYVFSSDGVLSDGI----VKLDGQIKYKKG----------DSDWLIDVFHTEGNK  561

Query  371  YYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIED-  429
             Y+      A  G   I+  +Y+F+++G+M  G+I+ + N YY + +G ++  ++ I+
Sbjct  562  TYYFDADGNAAIGWTDINGERYFFNDEGVMHIGVINDDGNVYYLSRDGIVKNKWVTIDSG  621

Query  430  KMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFE----------------------GE  467
```